

Semaine Data SHS

*Traiter et analyser des données en sciences humaines et sociales*

MSH-LSE, Lyon, 7-11 décembre 2020.



# La textométrie une approche quantitative et qualitative des données textuelles

Bénédicte PINCEMIN

CNRS, Université de Lyon, UMR 5317 IHRIM



This work is licensed under the Creative Commons Attribution 4.0 International License.  
<http://creativecommons.org/licenses/by/4.0/>

# Plan

1. *[Intro]* Données textuelles et SHS
2. *[Pratique]* Fonctionnalités textométriques
3. *[Théorie]* Positionnement et méthodologie
4. *[Pour aller plus loin]* Ressources

# Plan

## 1. *[Intro]* Données textuelles et SHS

- Données ;  
texte numérique, pratiques et attentes :  
en sciences humaines, en sciences sociales
- Une typologie pour situer les logiciels
- La famille textométrique

## 2. *[Pratique]* Fonctionnalités textométriques

## 3. *[Théorie]* Positionnement et méthodologie

## 4. *[Pour aller plus loin]* Ressources

# Les données ?

- Des données... qui ne sont pas données (Rastier, 2008, 2011)
  - Les choisir (contenu, édition...)
  - Les trouver, les acquérir (« aspirateurs », droits -auteur, éditeur, RGPD...)
  - Les représenter numériquement (numérisation, codage...)
- Questions connexes
  - « rentabilité » :
    - capitalisation, recyclage : archivage – pérennité, organisation, description/contexte... cf. Plan de gestion de données (vendredi après-midi)
    - partage (standards et recommandations cf. XML, TEI Text encoding initiative ; droits cf. licences Creative Commons ; accessibilité... cf. FAIR : Faciles à trouver (*Findable*), Accessibles, Interopérables, Réutilisables).
  - Évaluation
    - Plan *Données de la recherche du CNRS*, Novembre 2020, p. 13 (dernière page), § Accompagner une évolution profonde de l'évaluation des scientifiques.
    - Les *Data papers*, cf. mercredi.

# Texte numérique en Sciences Humaines (1) : Éditions numériques

- Pas nécessairement pour la lecture, mais souvent attention esthétique
- Accès distant, démultiplié
- Hypertexte : navigation, documentation (web sémantique)
  - Ex. : Synopsis dynamique de *Coran 12-21* ([www.coran12-21.org](http://www.coran12-21.org))
- Philologie : représentation plurielle, dynamique, d'alternatives
  - Ex. : Facettes de la *Queste del Saint Graal* dans la *Base de français médiéval* ([bfm.ens-lyon.fr](http://bfm.ens-lyon.fr))
  - Ex. : Agenceur de seconds volumes des *Dossiers de Bouvard et Pécuchet* ([www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr))
- Outils d'analyse
  - recherche non prédéfinie (≠ index)
  - autres modes de lecture notamment pour de grands volumes (capacités cognitives – lecture non linéaire et vues de synthèse, dans le prolongement de techniques classiques comme les codes d'annotation, les fiches, les concordances...)

Coran 12-21. Traductions du X +

https://coran12-21.org/fr/sourates/s16

Rechercher

# Coran 12-21

TRADUCTIONS DU CORAN EN EUROPE, XII<sup>e</sup> - XXI<sup>e</sup>s.

PRÉSENTATION TUTORIEL CONTEXTES ACCÈS PAR SOURATES ACCÈS PAR ÉDITIONS RECHERCHER FR

S16. AL-NAHJ سورة النحل 7 ÉDITIONS

Éd. du Caire, 1924 Contexte	Hamidullah révisée, 2000 Contexte	Blachère, 1957 Contexte	Sale, 1734 Contexte	Du Ryer, 1647 Contexte	Arrivabene, 1547 Contexte	Bibliander, 1550 Contexte
<p>وَلَهُمْ عَذَابٌ يَلِيمٌ</p> <p><b>[64]</b> وَمَا أَرْسَلْنَا عَلَيْكَ الْكِتَابَ إِلَّا لِتُبَيِّنَ لَهُمُ الَّذِي اخْتَلَفُوا فِيهِ وَهُدًى وَرَحْمَةً لِّقَوْمٍ يُؤْمِنُونَ</p> <p><b>[65]</b> وَاللَّهُ أَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْبَثَ بِهِ الْأَرْضَ نَبْذًا فَجَاءَتْ بِهَا حَبًّا وَمِنْهَا لَأَنْبُوتٌ لِّقَوْمٍ يَشْكُرُونَ</p> <p><b>[66]</b> وَإِلَّا لَكُمُ فِي الْأَنْعَامِ لَعِبْرَةٌ لِّئَلَّا تُكْفِرُوا بِمَا كُفَرْتُمْ بِهِ وَمَا كُنْتُمْ بِتَالِفِينَ إِلَّا ذُرِّيًّا مِمَّنْ بَدَّلَ اللَّهُ سَبِيلَ الْكَاذِبِينَ</p> <p><b>[67]</b> وَمِن ثَمَرَاتِ الْجِبَالِ وَأَوْتَانِ النَّخْلِ الَّذِي تَجْرُدُونَ مِنْهُ سَكْرًا وَرُفًا حَسْبًا إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَعْقِلُونَ</p> <p><b>[68]</b> وَأَوْحَى رَبُّكَ إِلَى النَّحْلِ أَنِ اتَّخِذِي مِنَ الْجِبَالِ بُيُوتًا وَمِنَ الشَّجَرِ وَمِمَّا يَعْرِشُونَ</p> <p><b>[69]</b> ثُمَّ كُلِي مِن كُلِّ الثَّمَرَاتِ فَاسْلُكِي سُبُلَ رَبِّكِ ذُلُلًا يَخْرُجُ مِنْ بَطُونِهَا فَسَوَّآتُ مُخْتَلِفُ السَّوَاءِ فِيهِ سِبْغَاتُ لِّئَلَّا تُسَاقَطَ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَتَفَكَّرُونَ</p> <p><b>[70]</b> وَاللَّهُ خَلَقَكُمْ ثُمَّ تَوَفَّاكُمْ وَمِنكُمْ مَّنْ بُرِّدَ إِلَىٰ أَرْدَالِ الْأَعْمُرِ لَكِن لَّا يَلْعَلْهُمُ عَدْوُ عِلْمٍ شَتَّىٰ إِنَّ اللَّهَ عَلِيمٌ قَدِيرٌ</p> <p><b>[71]</b> وَاللَّهُ فَضَّلَ بَعْضَكُمْ عَلَىٰ بَعْضٍ فِي الرِّزْقِ فَمَا الَّذِينَ فُضِّلُوا بِرَادِّي رِزْقِهِمْ عَلَىٰ مَا مَلَكَتْ أَيْمَانُهُمْ فَهُمْ فِيهِ سَوَاءٌ أَفَبِعَدْوِهِمْ يَحْتَدُونَ</p> <p><b>[72]</b> وَاللَّهُ خَلَقَ لَكُمْ مِنْ أَنْفُسِكُمْ أَزْوَاجًا لِئَلَّا تُكْفِرُوا مِنْ أَنْفُسِكُمْ تَبَيَّنَ وَقَدْ وَرَقَكُم مِّنَ الطَّيِّبَاتِ أَفَبِالْبَاطِلِ يُؤْمِنُونَ وَيَعْبُدُونَ وَرَفَعْنَا لَكُمْ ذِكْرَنَا وَاللَّهُ عَلِيمٌ غَذِيرٌ</p> <p><b>[73]</b> وَيَعْتَدُونَ مِنَ اللَّهِ مَا لَا يَفْلَحُ لَهُمْ</p>	<p>entendent.</p> <p><b>[66]</b> Il y a certes un enseignement pour vous dans les bestiaux : Nous vous abreuvons de ce qui est dans leurs ventres. - [un produit] extrait du [mélange] des excréments</p> <p><b>[67]</b> Des fruits des palmiers et des vignes, vous retirez une boisson enivrante et un aliment excellent. Il y a vraiment là un signe pour des gens qui raisonnent.</p> <p><b>[68]</b> [Et voilà] ce que ton Seigneur révéla aux abeilles : « Prenez des demeures dans les montagnes, les arbres, et les treillages que [les hommes] font.</p> <p><b>[69]</b> Puis mangez de toute espèce de fruits, et suivez les sentiers de votre Seigneur, rendus faciles pour vous. De leur ventre, sort une liqueur, aux couleurs variées, dans laquelle il y a une guérison pour les gens. Il y a vraiment là une preuve pour des gens qui réfléchissent.</p> <p><b>[70]</b> Allah vous a créés ! Puis Il vous fera mourir. Tel parmi vous sera reconduit jusqu'à l'âge le plus vil, de sorte qu'après avoir su, il arrive à ne plus rien savoir. Allah</p>	<p>pour un peuple qui entend.</p> <p><b>[66]</b> [68] En vérité, vous avez certes un enseignement dans vos troupeaux [an'am] ! Nous vous abreuvons d'un lait pur, exquis pour les buveurs, [venant] de ce qui, dans leurs ventres, est entre un aliment digéré et du sang.</p> <p><b>[67]</b> [69] Des fruits des palmiers et des vignes, vous tirez une boisson enivrante et un aliment excellent. En vérité, en cela, est certes un signe pour un peuple qui raisonne.</p> <p><b>[68]</b> [70] Ton Seigneur a révélé aux Abeilles : « Prenez des demeures dans les montagnes, les arbres et ce qu'élevèrent les Hommes.</p> <p><b>[69]</b> [71] Mangez en outre de tous les fruits et, dociles, empruntez les chemins de votre Seigneur ! » Du ventre [des Abeilles] sort une liqueur de différents aspects où se trouve une guérison pour les Hommes. En vérité, en cela est certes un signe pour un peuple qui réfléchit.</p> <p><b>[70]</b> [72] Allah vous a créés puis Il vous rappellera [tawafā]. Il en est cependant parmi vous qui sont renvoyés jusqu'à la décrépitude,</p>	<p>and mercy unto people who believe. <b>[65]</b> GOD sendeth down water from heaven, and causeth the earth to revive after it hath been dead. Verily herein is a sign of the resurrection unto people who hearken. <b>[66]</b> Ye have also in cattle an example of instruction: we give you to drink of that which is in their bellies; a liquor between digested dregs, and blood c ; namely, pure milk d . which is swallowed with pleasure by those who drink it. <b>[67]</b> And of the fruits of palm-trees, and of grapes, ye obtain an inebriating liquor, and also good nourishment e . Verily herein is a sign unto people who understand. <b>[68]</b> Thy LORD spake by inspiration unto the bee, saying, Provide these houses f in the mountains, and in the trees, and of those materials wherewith men build hives for thee: <b>[69]</b> then eat of every kind of fruit, and walk in the beaten paths of thy LORD g . There proceedeth from their bellies a liquor of various colours h . wherein is a medicine for men a . Verily herein is a sign unto people who consider. <b>[70]</b> GOD hath created you, and he will hereafter cause you to die: and some of you shall have his life prolonged to a</p>	<p>parole, <b>[66]</b> vous avez encore une marque de sa toute-Puissance aux animaux qui vous donnent du lait pour vous nourrir, &amp; une autre marque aux fruits de la terre, <b>[67]</b> au fruit des palmiers &amp; de la vigne, dont vous tirez du vin, &amp; en recevez du profit ; Ces choses sont signes de sa toute-Puissance à ceux qui les comprennent. <b>[68]</b> Le Seigneur a inspiré à la Mouche à miel d'habiter dans les campagnes, de se loger dans les arbres, dans les ruches, <b>[69]</b> &amp; de manger de toute sorte de fruits, elle produit le miel de diverse couleur qui sert de remède aux maladies des hommes, ces choses sont signes de la toute-Puissance de Dieu à ceux qui les considèrent : <b>[70]</b> Dieu vous a créé &amp; vous fera mourir, il y a des personnes entre vous qui seront pleines d'ignominie en leur vie, afin qu'ils cognoissent que Dieu sçait tout &amp; qu'il est tout-Puissant, <b>[71]</b> il donne des biens aux uns plus qu'aux autres, les valets n'ont point de part aux facultez de leur maistre, &amp; ne leur sont pas associez, (neantmoins ils associent à Dieu un autre Dieu esgal à luy, &amp; blasphème contre sa grace.) <b>[72]</b> Dieu vous a créé hommes &amp;</p>	<p>ene habbia a venir l'ultimo di. <b>[66]</b> L'altezza è simigliante à Dio, <b>[61]</b> se Dio riguardasse à tutte le cose malfatte de gli huomini, non lascierebbe nascer in terra cosa alcuna, Ma egli piacevole lascia che la hora giunga, nella quale non si potrà fuggire, non è da ragionar di Dio <b>[62]</b> à coloro, i quali bestemmiano, non havendo essi a ricever altro che fuoco. <b>[63]</b> Giuro per me medesimo che io commandai i precetti alle genti innanzi à voi, Ma il Diavolo corrompendoli fe parer i suoi piu belli, alquale chi si raccomanda, patira gravissimi tormenti. <b>[64]</b> Non ti ho dato questo libro per altro, se non perche tu esponga gli huomini le lor contrarietà, &amp; insegni la via retta à buoni, <b>[65]</b> Dio manda di Cielo l'acqua, accioche ella faccia viva la terra, <b>[66]</b> e dalle bestie il latte con che si bea, <b>[67]</b> e da gli arbori con che si viva, e queste vi godete, le quai tutte cose son miracoli a sapienti, e non di poca importanza, <b>[68]</b> Comandammo all'api che edificassero le case loro ne monti, e ne gli arbori, <b>[69]</b> e che gustassero di tutti i fiori, accioche esse componessero il liquore che è molto salutare à gli huomini, del</p>	<p>fecit : Cui nunc se commendantes, malum grauisimum passuri sunt. <b>[64]</b> Hunc librum non alia de causa tibi commisi, nisi ut hominibus suas contrarietates exponas, et uiam rectam, atque misericordiam credentibus patefacias. <b>[65]</b> Deus de coelo dimittit imbres, ut terram mortuam uiuificent : <b>[66]</b> et de uentribus bestiarum lac purum ex egestione et sanguine decolatum ad bibendum dulce, uictuique commodum educit : <b>[67]</b> et in arboribus ac palmis et uitibus creat, quod inebriat, et optime pascit. Hincque uos assumite uobis. Quae cuncta non minima sunt miracula sapientibus et audientibus. <b>[68]</b> Apibus item iussimus, ut in montibus et arboribus sibi suas mansiones aedificarent, <b>[69]</b> et deinceps de pomis omnibus gustarent, nec quoquam pergere ad suum quaerendum commodum Dei nomine differrent, ut sic ex uteris suis syropus multicolor saluber hominibus oriretur. Quod et ita contingit summe sapientibus admirandum. <b>[70]</b> Ipse itidem Deus uos creauit, et perire faciet. Et uestri quidam ad peiorem uitam deuergerunt, usquequo a scientibus</p>

abeille

Tout surligner Respecter la casse Respecter les accents et diacritiques Mots entiers Occurrence 2 sur 2

**Synopse dynamique de Coran 12-21 (www.coran12-21.org) (1) : au survol de la souris, les versets qui correspondent dans les autres éditions se colorent.**

Éd. du Caire, 1924 Contexte	Hamidullah révisée, 2000 Contexte	Blachère, 1957 Contexte	Sale, 1734 Contexte	Du Ryer, 1647 Contexte	Arrivabene, 1547 Contexte	Bibliander, 1550 Contexte
<p>وَلَهُمْ عَذَابٌ أَلِيمٌ</p> <p><b>[64]</b> وَمَا أَنْزَلْنَا عَلَيْكَ الْكِتَابَ إِلَّا لِتُبَيِّنَ لَهُمُ الَّذِي اخْتَلَفُوا فِيهِ وَهُدًى وَرَحْمَةً لِّقَوْمٍ يُؤْمِنُونَ</p> <p><b>[65]</b> وَاللَّهُ أَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْرَجَ بِهِ الْأَرْضَ نَعْدًا مَوْتِيهَا إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَتَشَفَعُونَ</p> <p><b>[66]</b> وَإِنَّ لَكُمْ فِي الْأَنْعَامِ لَعِبْرَةً لِّتُسْئَلَكُمْ مَعَهَا فِي مَطْعُوَيْهِ مِنْ بَنٍ فَزَرْبٍ وَدَمٍ لَبَنًا خَالِصًا سَائِغًا لِلشَّارِبِينَ</p> <p><b>[67]</b> وَمِنْ ثَمَرَاتِ النَّخِيلِ وَالْأَعْنَابِ تَتَّخِذُونَ مِنْهُ سَكَرًا وَرِزْقًا حَسَنًا إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَعْقِلُونَ</p> <p><b>[68]</b> وَأَوْحَى رَبُّكَ إِلَى النَّحْلِ أَنْ اتَّخِذِي مِنَ الْجِبَالِ بُيُوتًا وَمِنَ الشَّجَرِ وَمِمَّا يَعْرِشُونَ</p> <p><b>[69]</b> ثُمَّ كُلِي مِنْ كُلِّ الثَّمَرَاتِ فَاسْلُكِي سُبُلَ رَبِّكِ ذُلًّا يَخْرُجُ مِنْ بُطُونِهَا شَرَابٌ مُخْتَلِفٌ أَلْوَانُهُ فِيهِ شِفَاءٌ لِلنَّاسِ إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَتَفَكَّرُونَ</p> <p><b>[70]</b> وَاللَّهُ خَلَقَكُمْ ثُمَّ يَتَوَقَّأَكُم مِّنكُمْ مَنْ يُرِيهِمْ إِلَى آوَادٍ أَلْفُمْ لِكَيْ لَا تَغْلِبَ عَلَيْهِمْ عِلْمٌ سَنِيًا إِنَّ اللَّهَ عَلِيمٌ قَبِيرٌ</p> <p><b>[71]</b> وَاللَّهُ فَضَّلَ بَعْضَكُمْ عَلَى بَعْضٍ فِي الرِّزْقِ فَمَا الَّذِينَ فُضِّلُوا بِرَأْسِي رَغَبِيهِمْ عَلَى مَا مَلَكَتْ أَيْمَانُهُمْ فَهُمْ فِيهِ سَوَاءٌ أَلْيَعْبَهُمُ اللَّهُ يَتَّخِذُونَ</p> <p><b>[72]</b> وَاللَّهُ خَلَقَ لَكُمْ مِنْ أَنْفُسِكُمْ أَزْوَاجًا وَعَجَلَ لَكُمْ مِنْ أَرْوَاجِكُمْ نِسِينَ وَخَفَذَةً وَرَزَقَكُمْ مِنْ الطَّيِّبَاتِ أَفِيئَاتًا لَّا يُؤْمِنُونَ وَيَعْبَتُ اللَّهُ هُمْ يَكْفُرُونَ</p>	<p>entendent.</p> <p><b>[66]</b> Il y a certes un enseignement pour vous dans les bestiaux : Nous vous abreuvons de ce qui est dans leurs ventres, - [un produit] extrait du [mélange] des excréments [intestinaux] et du sang - un lait pur, délicieux pour les buveurs.</p> <p><b>[67]</b> Des fruits des palmiers et des vignes, vous retirez une boisson enivrante et un aliment excellent. Il y a vraiment là un signe pour des gens qui raisonnent.</p> <p><b>[68]</b> [Et voilà] ce que ton Seigneur révéla aux abeilles : « Prenez des demeures dans les montagnes, les arbres, et les treillages que [les hommes] font.</p> <p><b>[69]</b> Puis mangez de toute espèce de fruits, et suivez les sentiers de votre Seigneur, rendus faciles pour vous. De leur ventre, sort une liqueur, aux couleurs variées, dans laquelle il y a une guérison pour les gens. Il y a vraiment là une preuve pour des gens qui réfléchissent.</p> <p><b>[70]</b> Allah vous a créés ! Puis Il vous fera mourir. Tel parmi vous sera reconduit jusqu'à l'âge le plus vil, de sorte qu'après avoir su, il arrive à ne plus rien savoir. Allah</p>	<p><b>[66]</b> [68] En vérité, vous avez certes un enseignement dans vos troupeaux [ 'an'âm ] ! Nous vous abreuvons d'un lait pur, exquus pour les buveurs, [venant] de ce qui, dans leurs ventres, est entre un aliment digéré et du sang.</p> <p><b>[67]</b> [69] Des fruits des palmiers et des vignes, vous tirez une boisson enivrante et un aliment excellent. En vérité, en cela, est certes un signe pour un peuple qui raisonne.</p> <p><b>[68]</b> [70] Ton Seigneur a révélé aux Abeilles : « Prenez des demeures dans les montagnes, les arbres et ce qu'élèvent les Hommes.</p> <p><b>[69]</b> [71] Mangez en outre de tous les fruits et, dociles, empruntez les chemins de votre Seigneur ! » Du ventre [des Abeilles] sort une liqueur de différents aspects où se trouve une guérison pour les Hommes. En vérité, en cela est certes un signe pour un peuple qui réfléchit.</p> <p><b>[70]</b> [72] Allah vous a créés puis Il vous rappellera [tawaffā]. Il en est cependant parmi vous qui sont renvoyés jusqu'à la décrépitude, afin qu'après avoir su ils ne sachent rien. Allah est omniscient et</p>	<p>the earth to revive after it hath been dead. Verily herein is a sign of the resurrection unto people who hearken. <b>[66]</b> Ye have also in cattle an example of instruction: we give you to drink of that which is in their bellies; a liquor between digested dregs, and blood c ; namely, pure milk d , which is swallowed with pleasure by those who drink it. <b>[67]</b> And of the fruits of palm-trees, and of grapes, ye obtain an inebriating liquor, and also good nourishment e . Verily herein is a sign unto people who understand. <b>[68]</b> Thy LORD spake by inspiration unto the bee, saying, Provide thee houses f in the mountains, and in the trees, and of those materials wherewith men build hives for thee: <b>[69]</b> then eat of every kind of fruit, and walk in the beaten paths of thy LORD g . There proceedeth from their bellies a liquor of various colours h , wherein is a medicine for men a . Verily herein is a sign unto people who consider. <b>[70]</b> GOD hath created you, and he will hereafter cause you to die: and some of you shall have his life prolonged to a decrepit age, so that he shall forget whatever he knew; for GOD is wise and powerful. <b>[71]</b> GOD causeth</p>	<p>pour conduire les vrai-croyans au droit chemin : <b>[65]</b> Dieu envoie la pluie du Ciel pour rafraichir la terre, c'est un signe evident de sa toute-Puissance à ceux qui écoutent sa parole, <b>[66]</b> vous avez encore une marque de sa toute-Puissance aux animaux qui vous donnent du lait pour vous nourrir, &amp; une autre marque aux fruits de la terre, <b>[67]</b> au fruit des palmiers &amp; de la vigne, dont vous tirez du vin, &amp; en recevez du profit ; Ces choses sont signes de sa toute-Puissance à ceux qui les comprennent. <b>[68]</b> Le Seigneur a inspiré à la Mouche à miel d'habiter dans les campagnes, de se loger dans les arbres, dans les ruches, <b>[69]</b> &amp; de manger de toute sorte de fruits, elle produit le miel de remede aux maladies des hommes, ces choses sont signes de la toute-Puissance de Dieu à ceux qui les considerent : <b>[70]</b> Dieu vous a créé &amp; vous fera mourir, il y a des personnes entre vous qui seront pleines d'ignominie en leur vie, afin qu'ils cognoissent que Dieu sçait tout &amp; qu'il est tout-Puissant, <b>[71]</b> il adorando in luogo di Dio chi non vi puo dar ne il vitto, ne il vestiro? <b>[74]</b> Non proponete esempi a chi sa ogni cosa, <b>[75]</b> ma a voi che non</p>	<p>immanzi a voi, Ma il Diavolo corrompendoli fe parer i suoi piu belli, alquale chi si raccomandanda, patira gravissimi tormenti. <b>[64]</b> ▼ Non ti ho dato questo libro per altro, se non perche tu esponga gli huomini le lor contrarietà, &amp; insegni la via retta à buoni, <b>[65]</b> Dio manda di Cielo l'acqua, accioche ella faccia viva la terra, <b>[66]</b> e dalle bestie il latte con che si bea, <b>[67]</b> e da gli arbori con che si viva, e queste vi godete, le quai tutte cose son miracoli a sapienti, e non di poca importanza, <b>[68]</b> ▼ Comandammo all'api che edificassero le case loro ne monti, e ne gli arbori, <b>[69]</b> e che gustassero di tutti i fiori, accioche esse componessero il liquore che è molto salutare a gli huomini, del che si maravigliano i buoni. <b>[70]</b> Il medesimo Dio vi ha creati, e faravi morire, e molti di voi ritorneranno à peggior vita. <b>[71]</b> ▼ Perche negate voi il ben di Dio? <b>[72]</b> Dio di voi medesimi ha prodotto le donne, la prole, i figliuoli, &amp; vi concede ogni bene, perche dunque ne credette il male, negando il ben datovi <b>[73]</b> adorando in luogo di Dio chi non vi puo dar ne il vitto, ne il vestiro? <b>[74]</b> Non proponete esempi a chi sa ogni cosa, <b>[75]</b> ma a voi che non</p>	<p>hominibus suas contrarietates exponas, et uiam rectam, atque misericordiam credentibus patefacias. <b>[65]</b> Deus de coelo dimittit imbres, ut terram mortuam uiuificent : <b>[66]</b> et de uentribus bestiarum lac purum ex egestione et sanguine decolatum ad bibendum dulce, uicuique commodum uelicit : <b>[67]</b> et in arboribus ac palmis et uitibus creat, quod inebriat, et optime pascit. Hincque uos assumite uobis. Quae cuncta non minima sunt miracula sapientibus et audientibus. <b>[68]</b> ▼ Apibus item iussimus, ut in montibus et arboribus sibi suas mansiones aedificarent, <b>[69]</b> et deinceps de pomis omnibus gustarent, nec quoquam pergere ad suum quaerendum commodum Dei nomine differrent, ut sic ex uteris suis syropus multicolor saluber hominibus oriretur. Quod et ita contingit summe sapientibus admirandum. <b>[70]</b> Ipse itidem Deus uos creauit, et perire faciet. Et uestri quidam ad peiorem uitam deuergerunt, usquequo a scientibus facti sunt inscii. <b>[71]</b> ▼ ¶ Deo huic maiorem uictus copiam, illi uero minorem tribuente, exigite uis, ditioem a suo particeps, quam ipse</p>

Synopse dynamique de Coran 12-21 ([www.coran12-21.org](http://www.coran12-21.org)) (2) : au clic-droit sur un verset, les versets qui correspondent dans les autres éditions s'ajustent au même niveau.

# Texte numérique en Sciences Humaines (1) : Éditions numériques

- Pas nécessairement pour la lecture, mais souvent attention esthétique
- Accès distant, démultiplié
- Hypertexte : navigation, documentation (web sémantique)
  - Ex. : Synopsis dynamique de *Coran 12-21* ([www.coran12-21.org](http://www.coran12-21.org))
- Philologie : représentation plurielle, dynamique, d'alternatives
  - Ex. : Facettes de la *Queste del Saint Graal* dans la *Base de français médiéval* ([bfm.ens-lyon.fr](http://bfm.ens-lyon.fr))
  - Ex. : Agenceur de seconds volumes des *Dossiers de Bouvard et Pécuchet* ([www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr))
- Outils d'analyse
  - recherche non prédéfinie (≠ index)
  - autres modes de lecture notamment pour de grands volumes (capacités cognitives – lecture non linéaire et vues de synthèse, dans le prolongement de techniques classiques comme les codes d'annotation, les fiches, les concordances...)



ités    Navigateur Web Firefox    ven. 4 déc. 10:44

Portail BFM    txm.ish-lyon.cnrs.fr/bfm/    Rechercher

Bienvenue S'inscrire Se connecter Contact fr

**CORPUS**    ACCUEIL    GRAAL    TEXT:QGRAAL\_CM

Corpus

- BFM2019
- BFMSS
- CORPTEF
- GRAAL
- PALAFRAFRO...
- PALAFRALAT...
- PALAFRAPAR

<160a>

*(Ici commence la version de la Queste del saint Graal donnée par le manuscrit K (Bibliothèque Municipale de Lyon, Palais des Arts n° 77), folios 160 recto à 224 verso. Tout le début du texte a été mutilé : la première grande lettre a été découpée, comme on le voit sur la reproduction du manuscrit, et quelques lignes du texte manquent, que nous donnons ici entre crochets, en bleu, d'après le manuscrit Z (Paris, BNF n. acq. fr. 1119, folio 138 recto, colonne a) qui est un manuscrit proche de celui que nous éditons ici.)*

**[Galaad à Kamaalot]**

§ 1

[A la veille de la Pentecoste quant li compaignon de la table ronde furent venu a Kamaalot et il orent oi le servisse et l'en voloit metre les tables a heure de]

10 nonne .<sup>1</sup> lors en[tra a cheval en la]<sup>[1]</sup> sale une mout bele damoisele, et fu venue si grant oirre que bien le pot l'en veoir, car ses chevaux en fu encore toz suanz, et ele descent et vient devant le roi si le salue, et il dist que Diex la benëie. « Sire, fet ele, por Dieu dites moi se Lancelot est ceenz. - Oil voir, fet li rois, veez le la. » Si li mostre, et ele va maintenant la ou il est, et li dist : « Lancelot je vos di de par le roi Pellés que vos avec moi venez iusqu'en cele forest. » Et il li demande a qui ele est. « Je sui, fait ele, a celui donc je vos paroil. - Et quel besoin, fet il, avez vos de moi ? - Ce verroiz vos bien (bien), fet ele. - De par Dieu, fet il, et g'irai volentiers. » Lors dist a un escuier qu'il mete la sele en son cheval, et li aport ses armes, et cil si fet tout maintenant. Et quant li rois et li autre qui ou palés estoient voient ce si lor en poise mout. Et nepoquant quant il voient qu'il ne remaindroit il l'en lessent aler. Et la reine li dist : « Que est ce Lancelot ? Nos lairez vos a cest jor qui si est hanz ? - Dame, fet la damoisele, sachiez que vos le ravroiz demain ceenz ainz hore de disner.

<160a>

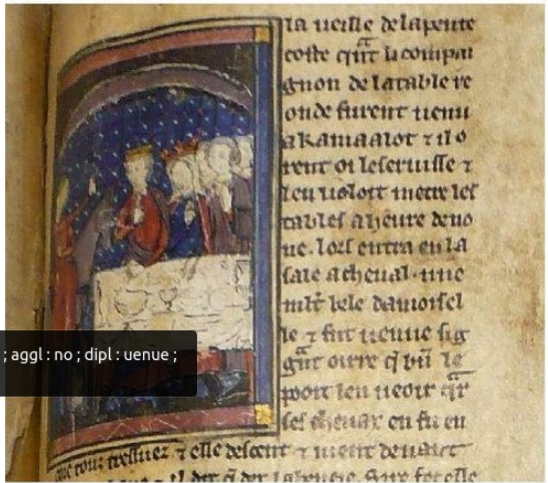
*(Ici commence la version de la Queste del saint Graal donnée par le manuscrit K (Bibliothèque Municipale de Lyon, Palais des Arts n° 77), folios 160 recto à 224 verso. Tout le début du texte a été mutilé : la première grande lettre a été découpée, comme on le voit sur la reproduction du manuscrit, et quelques lignes du texte manquent, que nous donnons ici entre crochets, en bleu, d'après le manuscrit Z (Paris, BNF n. acq. fr. 1119, folio 138 recto, colonne a) qui est un manuscrit proche de celui que nous éditons ici.)*

§ 1

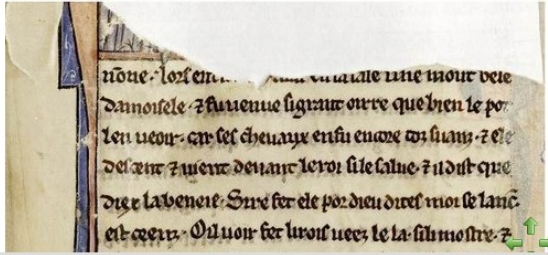
[A la ueille delapente coste qât li compaignon de latable ronde furent uenu a kamaalot ] il orent oi leferuiffe ] l'en uoloit metre lef tablef a heure de]

10 nōne .<sup>1</sup> loz en[tra acheual en la] fale une mout bele damoisele. ⁊ fuueue figant oirre que bien le pot l'en ueoir. car ses chevaux en fu encore toz suanz. ⁊ ele descent et uient devant le roi si le salue. ⁊ il dist que diex la benie. Sire, fet ele, por Dieu dites moi se Lancelot est ceenz. Oil uoient li rois, veez le la. ⁊ si li mostre, et ele ua maintenant la ou il est. ⁊ li dist lanē. ieuof di de par le roi pelles queuoef avec moi uenez iusqen cele forest. ⁊ li demande a qui ele est. ⁊ li dist que li rois li dist : « Lancelot je vos di de par le roi Pellés que vos avec moi venez iusqu'en cele forest. » Et il li demande a qui ele est. « Je sui, fait ele, a celui donc je vos paroil. - Et quel besoign, fet il, avez vos de moi ? - Ce verroiz vos bien (bien), fet ele. - De par Dieu, fet il, et g'irai volentiers. » Lors dist a un escuier qu'il mete la sele en son cheval, et li aport ses armes, et cil si fet tout maintenant. Et quant li rois et li autre qui ou palés estoient voient ce si lor en poise mout. Et nepoquant quant il voient qu'il ne remaindroit il l'en lessent aler. Et la reine li dist : « Que est ce Lancelot ? Nos lairez vos a cest jor qui si est hanz. Dame fet la damoisele, sachiez que vos lerauroiz demain ceenz ainz hore de disner.

Fragment du ms. Z (BnF, n.a.fr. 1119, col. 138a) à la place d'un fragment manquant du ms. K



Retour au manuscrit K (Lyon, BM, P.A. 77, col. 160a)



norm : venue ; pos : VERppe ; ref : qgraal\_cm, col. 160a, l. 11 ; aggl : no ; dipl : uenue ; pn : 1 ; oldid : w106\_000049 ;

page

normalisee | facsimilaire | ms\_colonne

160a / 268

GRAAL - normalisee, ...

PDF Notice

**Facettes de la Queste del Saint Graal dans la Base de français médiéval (bfm.ens-lyon.fr) : mise en correspondance fine (jusqu'au niveau du mot) de plusieurs éditions, ici normalisée, facsimilaire, et manuscrit (colonne).**

# Texte numérique en Sciences Humaines (1) : Éditions numériques

- Pas nécessairement pour la lecture, mais souvent attention esthétique
- Accès distant, démultiplié
- Hypertexte : navigation, documentation (web sémantique)
  - Ex. : Synopsis dynamique de *Coran 12-21* ([www.coran12-21.org](http://www.coran12-21.org))
- Philologie : représentation plurielle, dynamique, d'alternatives
  - Ex. : Facettes de la *Queste del Saint Graal* dans la *Base de français médiéval* ([bfm.ens-lyon.fr](http://bfm.ens-lyon.fr))
  - Ex. : Agenceur de seconds volumes des *Dossiers de Bouvard et Pécuchet* ([www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr))
- Outils d'analyse
  - recherche non prédéfinie (≠ index)
  - autres modes de lecture notamment pour de grands volumes (capacités cognitives – lecture non linéaire et vues de synthèse, dans le prolongement de techniques classiques comme les codes d'annotation, les fiches, les concordances...)

www.dossiers-flaubert.fr/jsp/public/agencement\_arbre.php?id=35

Accueil / Espace de travail / Agencements / Beautés - arborescence

Agencement: Beautés (mise le 20 jan 2015 par Stéphane Dard-Créole) Paramètres et exports

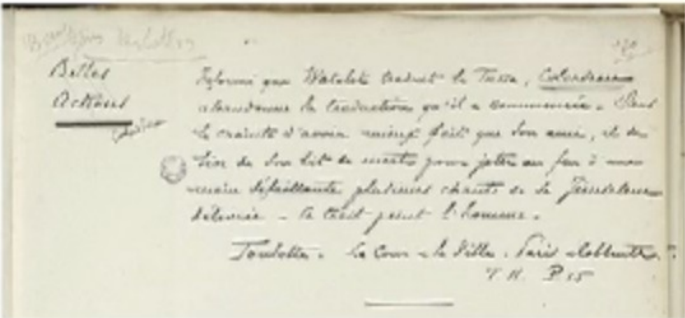
Agencement

[Tout déplier] [Tout replier] [Cacher les fragments]

- BEAUTÉS
  - BEAUTÉS DES GENS DE LETTRES
    - a) Belle actions
      - belle actions Colardeau
        - c 2021
          - Dites adans
          - Beautés
          - Gens de Lettres
          - Colardeau


Informé que Watelin traduit le Tasse, Colardeau abandonne la traduction qu'il a commencée. Dans la crainte d'avoir mieux fait que son ami, il se lève de son lit de mort pour jeter au feu d'une main défaillante plusieurs chants de sa Jérusalem dévoté. Ce trait peint l'homme.

Toulon, La Cour et la Ville, Paris et Cahiers, t. 17, p. 15

- Beautés des comédiens. Duménil
- c-2113
  - Belle actions
  - Beautés
  - Gens de Lettres

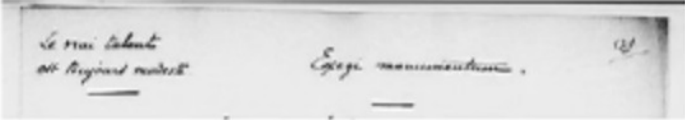
Duménil: pour déployer tous ses talents moyens il ne lui fallait pas moins de six bouteilles de vin de champagne et au dernier acte, il émit subitme. Duménil mettait à contribution les fils de l'Opéra et leur volait leurs lignes.

Biographie universelle, [art.] Duménil

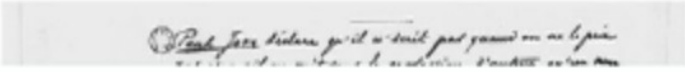

- Beautés des gens de lettres. Nicole
- c-1036
  - Le vrai talent est toujours modeste.

Il ne passait pas une crainte dans un bec, sans avoir une ceinture de sûreté pour pouvoir surager en cas de nafrage. - À Troyes, il n'osait s'enfuir quand il faisait du vent, de peur de recevoir des tuiles sur le tête.

Sainte-Beuve, Port-Royal, t. 4, page 428.


- [Le vrai talent est toujours modeste] Jovy. Dauncu
- c 1733
  - Le vrai talent est toujours modeste.

Paul Jovy déclare qu'il n'écrivait pas quand on ne le paie pas et qu'il ne voit dans la profession d'auteur qu'un moyen de faire fortune. Il distribue donc des



2007-2016 [www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr) - Stéphane Dard-Créole

**Agenceur de seconds volumes des Dossiers de Bouvard et Pécuchet** ([www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr)) : les fragments textuels des dossiers peuvent être sélectionnés et organisés pour produire des reconstitutions conjecturales du second volume, resté inachevé à la mort de Flaubert [ici extrait de la vidéo de présentation].

# Texte numérique en Sciences Humaines (1) : Éditions numériques

- Pas nécessairement pour la lecture, mais souvent attention esthétique
- Accès distant, démultiplié
- Hypertexte : navigation, documentation (web sémantique)
  - Ex. : Synopsis dynamique de *Coran 12-21* ([www.coran12-21.org](http://www.coran12-21.org))
- Philologie : représentation plurielle, dynamique, d'alternatives
  - Ex. : Facettes de la *Queste del Saint Graal* dans la *Base de français médiéval* ([bfm.ens-lyon.fr](http://bfm.ens-lyon.fr))
  - Ex. : Agenceur de seconds volumes des *Dossiers de Bouvard et Pécuchet* ([www.dossiers-flaubert.fr](http://www.dossiers-flaubert.fr))
- Outils d'analyse
  - recherche non prédéfinie (≠ index)
  - autres modes de lecture notamment pour de grands volumes (capacités cognitives – lecture non linéaire et vues de synthèse, dans le prolongement de techniques classiques comme les codes d'annotation, les fiches, les concordances...)

# Texte numérique en Sciences Humaines (2) : Humanités numériques

- Caractère innovant, valorisé
  - (financements plus que publications) ?
- Un côté sciences exactes qui peut séduire ou déplaire
  - Une juste attitude à trouver entre les deux
- Repoussoir
  - Mécaniser / automatiser ?
  - La perte du texte
- Réel apport
  - Confrontation concrète
  - Recherche de compréhension et d'intelligence des dispositifs (vs intimidation / fascination)
  - Recul critique

# Texte numérique en Sciences Humaines (2) : Humanités numériques

- Et du côté de l'analyse quantitative ?

- Que reste-t-il de la typologie proposée par Charles Muller (1987) dans son allocution inaugurale au colloque de 1986 sur les *Méthodes quantitatives et informatiques dans l'étude des textes* ? :

...il faut bien constater que, parmi les linguistes et surtout les « littéraires », c'est une minorité qui pratique ces méthodes [quantitatives et informatiques dans l'étude des textes], et que cette minorité reste très isolée, presque « marginale », un peu comme une secte hérétique, tolérée, mais suspecte. [...]

Chez certains [collègues], c'est une abstention polie, teintée d'ironie, et qui se retranche derrière l'allergie bien connue, et fort bien portée, de tout vrai littéraire pour toute mathématique ; [...] la moindre formule [...] exerce une action dissuasive [...] [et] l'abstention est sagesse.

Chez d'autres, c'est une franche hostilité qui les pousse à condamner ces techniques grossières et réductrices [...].

Il y a peut-être pire [...] : [...] l'ignorance béate et confiante, pour qui tout ce qui est calcul ou formule est vérité d'évangile.

# Texte numérique en Sciences Sociales (1)

- Articuler des données très diversifiées
  - métriques : mesures de volume, d'effectif
  - ordinales : échelles, repères (codés quanti ou quali)
  - nominales : catégories
  - textuelles : textes généralement nombreux, corpus volumineux
    - réponses à des questions ouvertes
    - textes d'information ± courts (dépêche, résumé, article)
    - entretiens (transcription ± annotée)
    - forums, tweets...

# Texte numérique en Sciences Sociales (2)

- Visualisations, synthèse
  - révéler, donner du sens – "par delà" les données ?
  - accessoirement illustrer (ex. nuages de mots)
- Quels calculs : exigence de validation
  - conditions de validité (effectifs, échantillonnage, regroupements...)
  - significativité
- Ergonomie logicielle
  - Pas forcément d'interface graphique utilisateur
  - Souvent des modules
    - habileté numérique  $\pm$  assistée  $\rightarrow$  à mettre au profit de liberté
    - mais succession d'états vs intégration complexe ?



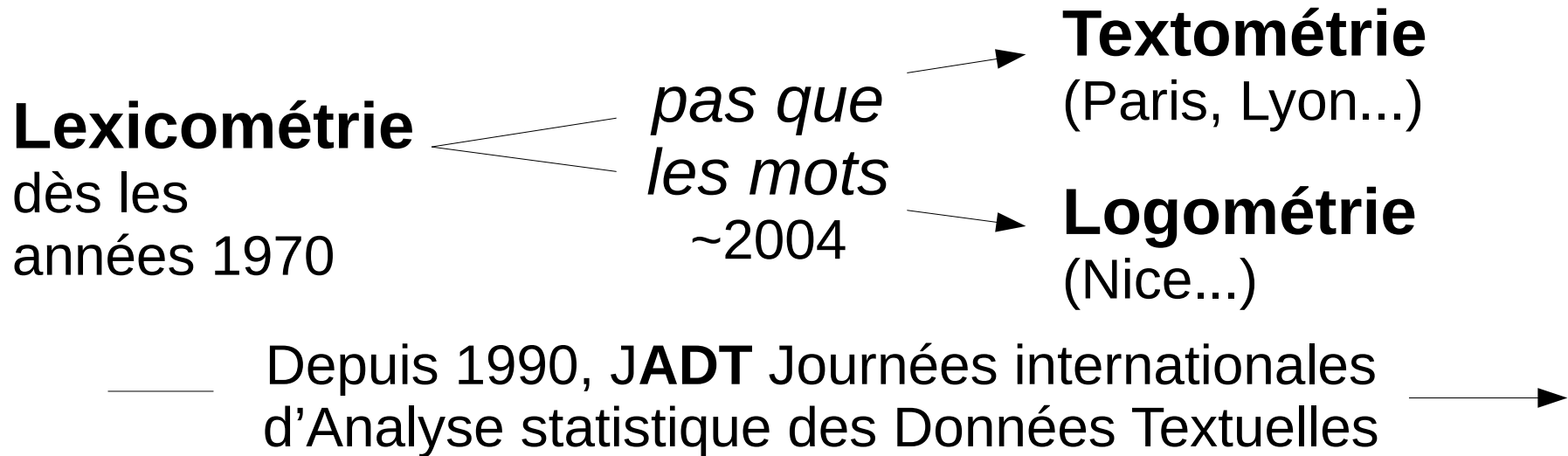
# Une typologie des approches outillées

- Derrière l'analyse de données textuelles, une grande diversité
- D'un point de vue épistémologique, liens entre approches (et outils) et communautés disciplinaires et géographiques
- Typologie « 2D » basée sur celle de Lejeune (2010, 2017)

<p><b>Textométrie,</b> Lexicométrie</p> <p><b>Concordanciers enrichis</b> avec analyse quantitative</p> <p><b>CALCULER</b> « bavard », non focalisé</p> <p><b>Textométrie</b> avec annotation ou catégories en cours d'analyse</p>	<p><b>EXPLORER</b> le matériau empirique, les données source</p> <p><b>Concordanciers KWIC</b> et moteurs de recherche linguistiques</p>
<p><b>Analyse par marqueurs</b> linguistiques ou rhétoriques ±généraux prédéfinis (dictionnaires, grammaires, indicateurs)</p> <p><b>Text Mining</b></p> <p><b>ANALYSER</b> En progressant à partir de représentations à valeur-ajoutée (catégories, annotation, visualisations, synthèses)</p>	<p><b>MONTRER</b> Qualitatif, « muet » (focalisé)</p> <p><b>Registres</b> Annotation qualitative assistée basée sur le corpus</p> <p><b>CAQDAS</b> (Computer-assisted qualitative data analysis software)</p>

**Une typologie des approches logicielles pour l'analyse de données textuelles**  
inspirée de Lejeune (2010, 2017)

# Textométrie : éléments de définition et de caractérisation



- /langue, texte/ + /mesure, statistique/ + /informatique/
- L'analyse de données : analyse factorielle (dont analyse des correspondances AFC) et classification, cf. école Benzécri (notamment années 1970-1980)
- Un noyau fonctionnel : AFC + spécificités (Lafon, MOTS 1980) + concordance/retour au texte ; mais aussi : segments répétés,...

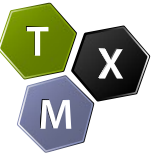
# Plan

1. *[Intro]* Données textuelles et SHS
2. ***[Pratique]* Fonctionnalités textométriques**
  - Parcours dans TXM
  - Exemples issus de projets dans différentes disciplines
3. *[Théorie]* Positionnement et méthodologie
4. *[Pour aller plus loin]* Ressources

# Parcours de fonctionnalités :

## Point de vue

- **Concrètement**, exemples de traitements réalisés par un logiciel de textométrie
- Cependant un point de vue doublement local :
  - Illustrations principalement avec le logiciel **TXM** (relativement général + dans mon laboratoire)
  - Exemples issus de collaborations de l'équipe TXM dans **différentes disciplines**
- Il s'agit donc bien d'un aperçu, pas d'un état de l'art.



<b>Projet</b>	<b>Domaine</b>	<b>Corpus</b>	<b>Biblio des exemples</b>
BFM	Linguistique	137 textes intégraux du IX <sup>e</sup> au XV <sup>e</sup> s. (4 millions de mots)  <i>Queste del Saint Graal</i> , 65+2 folios (recto-verso double colonage, 120 000 mots)	(Guillot et al. 2013 et 2015) (Guillot-Barbance et al. 2017) (Pincemin et al. 2020) (Marchello-Nizia et Lavrentiev 2019)
Pêcheurs du lac Léman	Géographie	38 entretiens transcrits (34 heures d'enregistrement, 400 000 mots)	(Le Lay et al. 2016)
ANR ANTRACT	Histoire	25 ans de diffusion hebdomadaire des <i>Actualités françaises</i> (env. 300 h de vidéo, 1,5 millions de mots)	(Carrive et al. à par.)
Équipex Matrice	Histoire, Psychologie	20 ouvrages de témoignage de la Shoah (1 million de mots) 824 réponses à une question ouverte auprès de témoins de l'attentat du 11 septembre 2001 (60 000 mots)	(Mayaffre et al. 2018)  (Dégeilh et al. 2017)
	Littérature	Recueil <i>Les Fleurs du Mal</i> , Baudelaire.  308 nouvelles russes 1900-1930 (1,4 millions de mots)	(Pincemin 2017) voir aussi (Brunet 2016), (Viprey 2002) (Lavrentiev et al. à par.)
DTH	Philosophie	4 œuvres de Hobbes (440 000 mots)	(Rebasti et Heiden 2019)

**Tableau** des principales recherches dont seront tirées les illustrations des fonctionnalités

# Édition : Accès au texte

- Pouvoir feuilleter et parcourir le corpus
- Contexte textuel
  - Mots voisins, présentation (paragrapes, position en début ou fin,...), paratexte...
- Dans TXM en particulier :
  - Édition HTML
  - Distinction possible entre ce qu'on présente et ce qui est indexé pour les recherches et les décomptes (ex. paratexte : notes, didascalies, etc.)
  - Affichage synoptique de la source (manuscrit, vidéo, etc.)

Portail BFM

portal.textometrie.org/bfm/?command=documentation&path=/GRAAL

Bienvenue S'inscrire Se connecter Contact fr

CORPUS

ACCUEIL GRAAL TEXT: QGRAAL\_CM

Corpus

- BFM2019
- BFMMSS
- CORPTEF
- GRAAL
- PALAFRAFRO...
- PALAFRALAT...
- PALAFRAPAR

mes vermeilles sanz espee et sanz escu, si dist  
 15 ausi tost com il fu enmi<sup>[4]</sup> le palais : « Pes soit a vos. »  
 Et après dist au roi la ou il le voit : « Rois Artus ge  
 t'amoign le chevalier desirré, celui qui est estrez dou  
 haut lignage le roi David, et dou parenté Joseph  
 d'Arimacie, celui par qui les merveilles de cest  
 20 païs et des estranges terres remaindront. Veez le ci. »  
 Et li rois est mout liez de ceste novele, si dist au  
 pseudome : « Sire bien soiez vos venuz se ceste paro-  
 le est veraie, et bien soit li chevaliers venuz, car se ce  
 est cil que nos atendons a achever les aventu-  
 25 res dou saint Graal onques si grant joie ne fu  
 faite d'ome come nos ferons de lui. Et qui (p. 8) que  
 il soit ou cil que vos dites ou autres. ' je voldroie  
 que bien li venist por ce qu'il est si gentis hons et  
 de si haut lignage come vos dites. - Par foi, fait  
 30 li pseudons, vos en verroiz par tens tel comence-  
 ment qui bel sera. » Et lors fet le chevalier tost desarmer  
 si remest en une cote de cendal vermeil, et il li a  
 baillié maintenant un mantel vermeil a afu-  
 blier qu'il portoit sor<sup>[5]</sup> s'espale tout de samit,  
 35 et par dedenz estoit forrez d'un blanc hermine.

§ 11

Quant il l'a vestu et apareillié si li dist : « Sivez  
 moi sire chevaliers. » Et il si fet, et il le moine tot  
 droit au siege perilleux delez quoi Lancelot se scoit,  
 et sozlieve le drap de soie que cil i avoient mis.  
 40 Si troeve les letres qui dient : ' Ci est li sieges

4 Un « petit blanc » entre 'en' et 'mi'.  
 5 Correction du copiste : 'soz' > 'sor'.

<161c>  
 Galaad sur le Siege perilleux

il soit ou cil que vos dites ou autres. je voldroie  
 que bien li venist por ce qu'il est si gentis hons et  
 de si haut lignage come vos dites. Par foi fait  
 li pseudons vos en verroiz par tens tel comence-  
 ment qui bel sera. Et lors fet le ch. tost desarmer  
 li remest en une cote de cendal vermeil et li a  
 baillié maintenant un mantel vermeil a afu-  
 blier qui portoit sor ses paules tout de samit.  
 Et par dedenz estoit forrez d'un blanc hermine  
 Quant il l'a vestu et apareillié si li dist sivez  
 moi sire chevaliers. Et il si fet et il le moine tot  
 droit au siege perilleux delez quoi Lancelot se  
 scoit et sozlieve le drap de soie que cil i avoient  
 mis. Si troeve les letres qui dient. C'est li sieges

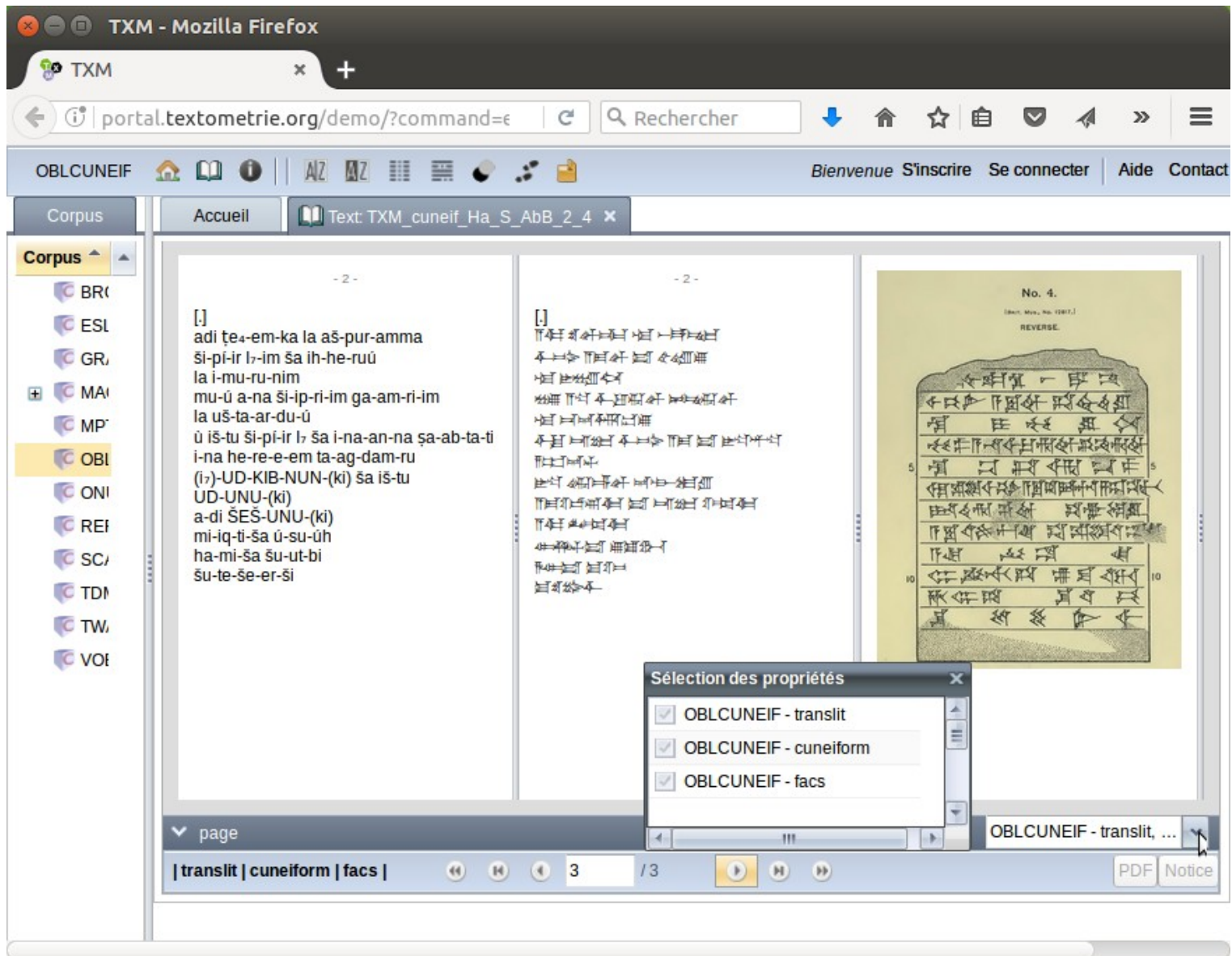
GRAAL - normalisee, ...

normalisee | ms\_colonne | 161c / 268

PDF Notice

**Edition de la Queste del Saint Graal dans TXM (Marchello-Nizia & Lavrentiev, 2019) :**  
 édition HTML du texte (avec numérotation des lignes, lettrine, etc.), apparat critique  
 (notes hypertextes, affichées mais non indexées), vue synoptique de l'image du  
 manuscrit source.





***Edition dans TXM de tablettes cunéiformes en ancien babylonien (Béranger et al, 2015).***

AFNOTICES: [frlemma="fo

Requête: [frlemma="foule

Seuils: Fmin: 1 Fmax: 999

word	Fréquence
spectateurs	232
manifestants	63
gens	23
jeunes	22
pèlerins	16
curieux	15
voyageurs	14
étudiants	13
photographes	13
invités	12
fidèles	11
visiteurs	11
face	10
journalistes	8
la	8
piétons	8
grévistés	7
Hindous	7
ouvriers	7

AFNOTICES: [frlemma="foule"][frlemma="de|du"]

Requête: [frlemma="foule"][frlemma="de|du"] Pivot: word Chercher

Clés de tri: #1 Aucun #2 Aucun #3 Aucun #4 Aucun Tri

1 - 100 / 738 Cacher paramètres

ref	Contexte gauche	Pivot	Contexte droite
1945-02-02, AFE86002962	ant la mairie en présence d'une	foule de spectateurs	- Specta
1945-03-09, AFE86003007	vec l'officiant, en arrière plan et	foule de soldats	suiwant
1945-04-13, AFE86003052	grande rue de FRANCFORT avec	foule d'Allemands	se press
1945-05-25, AFE86003109	le président HERRIOT saluant la	foule du balcon	de l'Hô
1945-06-01, AFE86003115	ABETH et MARGARET saluant la	foule du Balcon	du Pala
1945-06-15, AFE86003135	ur et Madame BENES, saluant la	foule d'une	fenêtre
1945-07-27, AFE86003194	ne tête nue, visage triste dans le	foule de St	Briec -
1945-08-03, AFE86003201	olie petite couturière (2 plans) -	Foule de soldats	(4 plans
1945-08-10, AFE86003205	ce, côté des tribunes garnies de	foule de spectateurs	- PA de
1945-08-31, AFE86003233	ierre aux journalistes (2 plans) -	Foule de journalistes	à l'intér
1945-09-21, AFE86003251	plans) - [Vue générale] de	foule de mineurs	rasseml
1945-09-28, AFE86003265	gués japonais quittent la table-	Foule de marins	améric
1945-10-05, AFE86003271	me-orient. ils sont salué par une	foule de chinois	, de Mal
1945-12-07, AFE86003360	, sortant de l'avion et saluant la	foule du haut	de la pa
1946-01-11, AFE85001264	ES du décret de nationalisation,	foules d'ouvriers	défilant
1946-01-11, AFE85001269	umentaire sur des images d'une	foule de Varsoviens	accueill
1946-01-25, AFE85001294	mbodgiens. - VG panoramique. -	Foule de jeunes	Cambo
1946-02-08, AFE85001312	et de la place Saint Pierre - La	foule des chrétiens	sur le p
1946-02-08, AFE85001312	Sedia Gestatoria " entouré de la	foule des fidèles	- PA des
1946-02-15, AFE85001319	ce ZURBIGEN fait sur un ski	Foule de spectateurs	assis su

1945 - 40

- Nature de production : Production propre
- Producteurs (Aff.) : Producteur - Les Actualités Françaises (LAF) - Paris - 1945;
- Thématique :

**TITRE PROPRE**

Visite du général de Gaulle à Boulogne Billancourt

**RÉSUMÉ**

Un long reportage sur l'île de France et la Normandie dévastées par les bombardements et les récentes inondations est précédé de brèves images de la visite du général DE GAULLE à Boulogne-Billancourt.

**SÉQUENCES**

Plusieurs plans de la visite de DE GAULLE à BOULOGNE BILLANCOURT

- Arrivée du cortège officiel devant la mairie en présence d'une **foule de spectateurs**
- Spectateurs sous la neige
- GP de DE GAULLE (de dos) prononçant un discours au balcon de l'hôtel de ville- Policiers sur le toit d'un immeuble voisin
- 2 Vues aériennes de campagnes inondées
- Plusieurs plans de péniches bloquées par les glaces
- Route enneigée et verglacée
- Roues d'un camion patinant

default 40 / 553

***Edition d'une notice documentaire de l'INA (volet de droite) : la structure de la notice issue de la base documentaire INA est visualisée, les noms des champs (Titre propre, Résumé, Séquences...) ne sont pas indexés, cependant l'information d'appartenance à un champ est bien disponible pour les interrogations. Projet ANTRACT, présentation à Blois (Carrive et al. 2018).***

AFVOIXOFFV02 [frlemma = "foule"]

Query [frlemma = "foule"]

datedediffusion	Left context	Pivot	Right context
10, 10/08/1960	du monde entier et de la	foule	des congressistes qui ser
13, 28/09/1960	nait à un dialogue avec la	foule	des envoyés de la press
16, 04/12/1957	ne Ainsi s'explique que la	foule	des étudiants en science
19, 30/11/1945	e souverain smiley par la	foule	des fidèles, à Immémoria
14, 05/10/1960	dition religieuse a réuni la	foule	des fidèles à L'origine de
12, 21/09/1960	Tapie marouane l'habituel	foule	des fidèles avec les prés
14, 21/08/1963	endant deux semaines, la	foule	des fidèles se livre à des
11, 07/02/1956	édailles d'or. euh Avec la	foule	des grands jours, monsie
17, 03/07/1963	r amenés à longchamp la	foule	des grands jours pour l'é
12, 22/05/1968	e l'elysée où se trouvait la	foule	des journalistes le chef d
18, 15/06/1960	mplètement bloqué par la	foule	des manifestants, il n'av
11, 07/10/1954	ire Madame portrait de la	foule	des parieurs du gagnant
12, 17/04/1952	a bénédiction Pascal à la	foule	des pèlerins. Hum. Pas
16, 11/05/1966	Wyszynski, on a évalué la	foule	des pèlerins à cent mille
19, 21/04/1955	bénédiction urbi et orbi à la	foule	des pèlerins réunis place
18, 24/08/1960	rd'hui L'été il ramène des	foules	des peuples des collègue
14, 21/02/1952	ce de liège au milieu de la	foule	des représentants de tou
14, 09/05/1956	portifs de Moscou. Alors la	Foule	des travailleurs a clos les

AFE86003459 - 2


DEI: Paques ; DEI: religion ; DET: catholicisme ; DEI: Pie XII ; DEI: bénédiction ; DEL: Italie ; DEL: Rome ; DEL: Vatican ;

**generique\_aff\_lig**

S 0: 0:00:03 euh

S 64: 0:00:06 à Rome à l'occasion des têtes de pas faire le pape qui n'était pas apparu en public depuis sa maladie a donné sa bénédiction urbi et orbi à la **foule** des pèlerins réunis place Saint-pierre.

AFE86003459



RESUM 00:12  Repeat Rate: 100%

***Edition d'une transcription (volet central) avec retour à la vidéo source (volet de droite) : la transcription automatique de la voix-off de la vidéo est synchronisée au mot près avec la vidéo ; une navigation hypertexte permet d'accéder depuis l'affichage du texte (en concordance ou édition) au passage concerné dans la vidéo source. Projet ANTRACT, article à paraître dans DHQ (Carrive et al., sous presse).***

# Vocabulaire

- Fonctions Lexique et surtout Index dans TXM
- Dominances (notamment à l'intérieur d'une famille de mots, d'une catégorie grammaticale ou thématique)
- Variations attestées
- Combinaisons et constructions fréquentes ou attestées, phraséologie, paradigmes « en corpus »

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [pos="S."...] >>1

Requête : [pos="S."]

Seuils : Fmin: 1 Fmax: 99999

1 -100 / 2397

word	Fréquence T=33233
cœur	110
yeux	92
âme	60
ciel	58
amour	46
œil	46
soleil	44
corps	41
air	36
soir	36
nuit	32
mer	30
beauté	29

FLEURS: [pos="J"]:lemme

Requête : [pos="J"]

Seuils : Fmin: 1 Fmax: 99999

1 -100

lemme	Fréquence
grand	72
plein	63
vieux	51
doux	43
long	42
profond	37
autre	29
cher	26
éternel	25
lourd	23
charmant	21
étrange	20
pâle	20

FLEURS: [pos="V"]:le...

Requête : [pos="V"]

Seuils : Fmin: 1 Fmax: 99999

1 -100

lemme	Fréquence
être	343
avoir	206
faire	124
voir	73
dire	55
pouvoir	55
vouloir	52
aimer	50
savoir	46
aller	43
prendre	43
venir	39
connaître	21

FLEURS: [pos="A"]

Requête : [pos="A"]

Seuils : Fmin: 1 Fmax: 99999

1 -100

word	Fréquence
pas	60
toujours	36
jamais	33
bien	21
donc	19
parfois	15
très	14
encor	13
le plus	12
loin	12
trop	11
souvent	10
à travers	9

Ouverture du tableau d'index

**Vocabulaire dominant par catégorie grammaticale** : relevés des noms communs, adjectifs qualificatifs, verbes et adverbes les plus fréquents dans les Fleurs du Mal.

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: [pos="R" ...]

Requête : [pos="R" & lemme="noir"]

Seuils : Fmin: 1 Fmax: 99999

lemme	Fréquence T=33233
noir	61
blanc	17
bleu	14
rose	13
rouge	13
brun	10
jaune	5
vermeil	5
fauve	4
blond	2
violet	2
écarlate	1
roux	1

FLEURS:[lemme="vermeil"]

Afficher paramètres

poeme_titre_ε^	Contexte gauche	Pivot	Contexte droit
BÉNÉDICTION	Retrouve l'ambrosie et le nectar	vermeil	. Il joue avec le vent, ca
ENNEMI	mon jardin bien peu de fruits	vermeils	. Voilà que j'ai touché l'
AUBE	les débauchés l'aube blanche et	vermeille	Entre en société de l'Id
PETITES VIEILLES	tombant Ensanglante le ciel de blessures	vermeilles	, Pensive, s'asseyait à l'
AME DU VIN	prison de verre et mes cires	vermeilles	, Un chant plein de lumi

FLEURS Edition - Page 11

file:///home/bpincemi/TXM/corpora/fleurs6/HTML/FLEURS/default

Ma jeunesse ne fut qu' un ténébreux orage,  
 Traversé çà et là par de brillants soleils ;  
 Le tonnerre et la pluie ont fait un tel ravage,  
 Qu' il reste en mon jardin bien peu de fruits **vermeils** .

Voilà que j' ai touché l' automne des idées,  
 Et qu' il faut employer la pelle et les râteaux  
 Pour rassembler à neuf les terres inondées,  
 Où l' eau creuse des trous grands comme des tombeaux.

Tri terminé.

**Dominances au sein d'un paradigme : emplois des adjectifs de couleur dans les Fleurs du Mal.**

TXM

Fichier Outils Affichage Aide

Corpus

FLEURS: []:pos

FLEURS: ((pos="S.")[pos="J"]|[(pos="J")[pos="S.]):pos

Requête:  Propriétés: pos

Seuils: Fmin:  Fmax:  Vmax:  Résultats par page:

t 1401 , v 6 , fmin 1 , fmax 556

pos	Fréquence T=33233
SM J	556
J SM	326
SF J	266
J SF	250
J SW	2
SW J	1

Terminé : 6 items pour 1401 occurrences.

FLEURS: [pos="J"][pos...

Requête:

Seuils: Fmin:  Fmax:

word	Fréq
longue misère	15
grands yeux	4
larges pleurs	3
Adorable sorcière	2
chaudes soirées	2
diverses beautés	2
éternel martyr	2
étranges grâces	2
grand œil	2
grand reposoir	2
grands flots	2
grand troupeau	2
innocent paradis	2

**Recherche d'une construction** : relevés des séquences nom + adjectif (antéposé ou postposé) les plus fréquentes dans les Fleurs du Mal.

Syntagme (lemmes)	F	Concordancier du syntagme « beau métier »		
beau métier	7	moi, c'est le plus	beau métier	du monde pêcheur. Alors en fait, quand
métier dur	6	est vrai que c'est un	beau métier	que c'est une denrée qui est
autre métier	4	joli métier, on a un	beau métier	. Bon, un métier de travail, faut pas rêver.
même métier	4	on a quand même un	beau métier	, on a quand même des bonnes choses.
joli métier	3	même comme un	beau métier	quoi. Ouais il y a une bonne image, et ça
métier physique	3	bah voilà c'est un	beau métier	, que quand même quelques inquiétudes
ancien métier	2	métier, qui est le plus	beau métier	du monde, pour rien au monde je ne ferais
métier artisanal	2			
métier complet	2			
métier difficile	2			
métier traditionnel	2	Concordancier du syntagme « métier dur »		
métier très individualiste	2	une image que c'est un	métier dur	parce qu'effectivement l'hiver
petit métier	2	Non que c'est un	métier dur	, pas toujours évident. C'est bien
autres métiers	1	pêcheur professionnel, c'est un	métier dur	, difficile, parce que malgré tout
beaux métiers	1	donc ça reste malgré tout un	métier dur	, un métier physique. Euh, il y a
bon métier	1	,c'est malgré tout un	métier dur	. C'est un métier de passion si on
dernier métier	1	parce que c'est un	métier dur	quand même. Et puis quand on
mêmes métiers	1			
métier ancestral	1			
métier assez différent	1			
métier bucolique	1			
métier dangereux	1			
métier extrêmement dur	1			
métier idyllique	1			
métier local	1			

**Exemple de paradigme « en corpus »** : étude des adjectifs utilisés par les pêcheurs pour qualifier leur métier : relevé complet à gauche, étude des contextes des deux expressions les plus fréquentes à droite (Le Lay et al. 2016).



# Concordance

- Examen systématique et dense des contextes locaux d'un mot ou d'une expression
- Disposition heuristique en tableau
  - Alignement et superposition des contextes + tri  
= mise en évidence visuelle de voisinages répétés

02decive - 312

God, whom nature doth dictate him to be, but distinctly the God of Abraham and Isaac, afterward being about to renew the same Covenant, by Moyses, with the whole People of Israel, **Exod. 3. v. 6.** I am saith h• the God of thy Father, the God of Abraham, the God of Isaac, and the God of Jacob. Afterward when that People, not only the freest, but also the greatest enemy to humane subjection, by reason of the fresh memory of their Aegyptian bondage, abode in the wilderness near mount Si•ai, that **anti••t** Covenant was propounded to them all to be renewed in this manner, **Exod. 19. ver. 5.** Therefore if yee will obey my voice indeed, and keep my Covenant, (to wit, that

Covenant, by Moyses, with the whole People of Israel, **Exod. 3. v. 6.** I am saith h• the God of thy Father, the God of Abraham, the God of Isaac, and the God of Jacob. Afterward when that People, not only the freest,

default | facs 312 / 407

HOBBS81LI [pos="crd"]? [word="Mat.|Acts|Gen.|Cor.|Deut.|Exod.|Sam.|Rom.|Luke|Matth.|Iohn|Joh.|Kings|Act.|Iohn|Prov.|Tim.|Heb.|Mark|Pet.|Psal.|Gal.|Tit.|Isaiah|Levit.|Artic.|Ezek.|Math.|Col.|Math|Numb.|Numbers|Zach.|Chron.|Eph.|Epist.|Isay|...]

Query [pos="crd"]? [word="Mat.|Acts|Gen.|Cor.|Deut.|Exod.|Sam.|Rom.|Luke|Matth.|Iohn|Joh.|Kings|Act.|Iohn|Prov.|Tim.|Heb.|Mark|Pet.|Psal.|Gal.|Tit.|Isaiah|Levit.|Artic.|Ezek.|Math.|Col.|Math|Numb.|Numbers|Zach.|Chron.|...]

ref	Left context	Pivot	Right context
02decive, 16.10, 268	, with the whole People of Israel,	<b>Exod. 3. v. 6.</b>	I am saith h• the God of thy Father, the
02decive, 16.10, 268	all to be renewed in this manner,	Exod. 19. ver. 5.	Therefore if yee will obey my voice indeed, and keep
02decive, 16.11, 269	the Israelites asked a King, 1.	Sam. 8. 7.	Yhey have not rejected thee, but they have rejected me
02decive, 16.11, 269	which Samuel told the Israelites, 1.	Sam. 12. 12.	Yee said unto me, nay, but a King shall
02decive, 16.11, 269	King, and that which is said,	Jer. 31. vers. 31.	I will make a new Covenant, &c. Although I
02decive, 16.12, 271	, as appears out of these words,	Exod. 31. ver. 16...	It is a perpetuall Covenant, (meaning the Sabbath)
02decive, 16.12, 271	up in the side of the Arke.	Deut. 3. ver. 26	For these retaining the faith of Abraham might be chang'd, those
02decive, 16.13, 273	wanting, he is rejectin these words,	Deut. 13. ver. 1, ...	If there arise among you a Prophet, or a dreamer
02decive, 16.13, 273	wanting, he is condemn'd by these,	Deut. 18. ver. 21...	And if thou say in thine heart, how shall we
02decive, 16.14, 274	acknowledg'd againe for the Word of God.	2 Kings 23. ver. 2.	But it is not manifest when the rest of the books
02decive, 16.14, 274	by reason of the Law cited above.	Deut. 18. ver. 21...	Whereby the Israelites were commanded not to account any man for
02decive, 16.15, 275	the people, and spake unto them.	Exod. 19. 24, 25.	It is farther manifestly, and expresly declar'd, upon occasion

201 - 300 / 707

**Concordance** : vue systématique et dense des contextes locaux, en lien avec des informations de contexte plus large (texte, chapitre, etc.).  
 Ici (Rebasti et Heiden 2019) la recherche porte sur les références bibliques dans l'œuvre de Hobbes : il est important de savoir dans quel ouvrage et à quel endroit est trouvée chaque mention. Le corpus a été construit de sorte à disposer de ces informations (métadonnées et codage de la structuration interne des textes).

## b) Le poids de la solitude

Lemme	F	Extrait du concordancier du lemme « seul »	
seul	185	d'ici et puis je suis tranquille : je suis	seul   , je vois pas pourquoi j'irais à Lugrin ou à
individualiste	19	au ventre d'aller au milieu du lac, tout	seul   . Euh, je vous parle en tant que femme.
individualisme	1	pêcheur, seul, qui va faire tout, tout	seul   , euh, alors à ce moment là, oui, il
solitaire	1	Alors d'un côté je fais tout, tout	seul   , ou je suis tout seul et puis heu,
		sur le même bateau. Après j'allais tout	seul   avec mon bateau. Voilà oui. On va parler
Syntagme (lemmes)	F	voulez progressez, faut y aller tout	seul   , se faire son expérience euh, se jeter à
être tout seul	31	me suis démener pendant sept ans tout	seul   pour avoir tous ces clients là. Donc j'ai
travailler seul	7	encore euh, je continue d'apprendre tout	seul   , quoi, en fait euh, j'ai encore euh
aller tout seul	3	quelque chose qu'on peut apprendre tout	seul   automatiquement quand on arrive sur
apprendre tout seul	3	tout seul. Si, vous apprendrez tout	seul   . Si, ben je me suis chopé la grêle, euh
être seul	3	demi, une fois qu'il aura fait ça tout	seul   , euh, il peut se lancer y a pas de souci
faire tout seul	3	voilà. Et puis après j'ai commencé tout	seul   . Alors on va prendre une journée d'été
travailler tout seul	3	bien passé, puis après j'ai continué tout	seul   . Non je m'étais jamais intéressé à,
aller seul	2	Mais là oui, on s'est débrouillé tout	seul   . Si, même, vous prenez la maison
lancer tout seul	2	vieille camionnette, je me démerde tout	seul   , voilà hin, perso. C'est pas mon
rester seul	2	c'est mathématique ! Si t'es tout	seul   , et que tu pêches, tu pêches mal
		pas, c'est ton droit, tu es tout	seul   . C'est ton boulot. T'es, t'es ton
		Ouais. Que maintenant on est tout	seul   . Ouais. Les filets. Puisque c'est pus fin
		que bon on couvre, quand on est tout	seul   on est tout seul. On est tout seul on es
		employés, mais sinon, quand on est tout	seul   on peut pas. Et vous pensez que c'est un
		ça le problème. Puis quand on est tout	seul   , quand il faut faire tout seul ça

**Exemple de concordance triée sur le contexte : étude de la façon d'employer le mot « seul » dans les entretiens avec les pêcheurs, le tri met en évidence des associations plus ou moins répétées et longues : tout seul, apprendre tout seul, quand on est tout seul (Le Lay et al. 2016).**

# Cooccurrence

- Pour un mot (ou une expression) donné(e), quels sont les mots qui apparaissent particulièrement dans son voisinage, qui semblent « attirés » ?
- Synthèse statistique des contextes locaux.

Index: "(agrair|agri..."

Query Propertie

frlemma	Frequency
agricole	81
agriculture	62
agriculteur	20
agraire	8
Agricole	2
Agraire	1
Agricoles	1

AFVOIXOFFV02: "(agrair|agric).\*\*%c (9, ..."

Query "(agrair|agric).\*\*%c

Parameters

Cooccurents properties: word Edit Thresholds: Fmin ≥

2 Cmin ≥ 2 Score ≥ 2,0

Context:  word  structure div  Use the left context

from - 9 to - 0 and from 0 to 9

Cooccurrent	Frequency	CoFreq	▲ Score	Mean distance
agriculture	62	11	17	5.3
l'	27115	112	10	3.1
pays	1312	17	8	3.1
exploitations	11	4	8	2.0
marché	201	8	7	4.2
rendement	37	5	7	5.8
paysans	54	5	6	6.8
ministre	894	12	6	3.1
accroître	36	4	5	3.8
réforme	44	4	5	.0
machines	107	5	5	.0
collective	21	3	4	5.0
Tenguiz	3	2	4	3.5
internationalement	3	2	4	6.0
industrielle	68	4	4	3.2
potentiel	27	3	4	.7

t pivot 175, v cooc 78, t cooc 0, T corpus 1568649

AFVOIXOFFV02: "paysan.\*\*%c (9, 9)"

Query "paysan.\*\*%c

Parameters

Cooccurents properties: word Edit Thresholds: Fmin ≥

2 Cmin ≥ 2 Score ≥ 2,0

Context:  word  structure div  Use the left context

from - 9 to - 0 and from 0 to 9

Cooccurrent	Frequency	CoFrequer	▲ Score	Mean distance
bretons	26	4	7	1.8
Bretagne	78	4	5	4.5
les	28383	56	5	2.7
ces	2637	12	4	3.4
manifestations	193	4	4	.0
mécontentement	11	2	4	1.5
des	20612	41	3	3.0
ouvriers	147	3	3	2.7
dirigeants	46	2	2	8.0
connaît	198	3	2	7.3
milieu	462	4	2	3.8
agricole	58	2	2	4.5
Déjà	60	2	2	4.0
agriculture	62	2	2	8.0
agitation	64	2	2	.0
dimanche	85	2	2	1.5

t pivot 83, v cooc 18, t cooc 0, T corpus 1568649

**Cooccurrence pour comparer les usages de deux familles de mots : l'hypothèse de recherche est que, dans les Actualités françaises (1945-1968), les mots « agriculteur » et « paysan » sont employés dans des contextes différents, le premier semble privilégié dans les descriptions économiques et conceptuelles, le second dans les descriptions plus politiques de manifestations avec des personnes (Carrive et al. à paraître).**

	<i>1945-1946</i>				<i>1978-1999</i>				<i>2002-2010</i>			
<i>Cooc.</i>	<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>	<i>Cooc.</i>	<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>	<i>Cooc.</i>	<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>	
le	15204	389	60	jeune	503	37	40	du	4792	73	21	
du <sup>11</sup>	4148	120	22	un	9168	95	29	un	5900	80	20	
cent	71	18	19	ce	6042	70	24	ce	3388	52	15	
deux	379	32	18	brave	29	7	12	jeune	274	16	13	
.	11027	195	12	du	6649	47	9	deux	588	15	7	
@card@ <sup>12</sup>	1362	45	10	deux	812	14	7	@card@	1457	22	6	

**Cooccurrence pour synthétiser statistiquement des contextes** : dans le corpus de récits de témoignages de la Shoah, on remarque un fort emploi du mot « homme » au sens d'« être humain » dans les témoignages rédigés au sortir des camps (1945-1946). L'observation statistique des déterminants et adjectifs immédiatement antéposés à « homme » selon les périodes d'écriture confirme la dominance de l'article défini en première période ; dans les périodes suivantes, les usages sont davantage descriptifs et narratifs : reprises anaphoriques, désignation concrète d'un individu ou d'un groupe (Mayaffre et al. 2018).

# Spécificités

- Étant donnée une division du corpus en parties (périodes, auteurs, etc.), quels mots apparaissent répartis particulièrement irrégulièrement ? → attirance mot ↔ partie
  - Pour une partie, quels mots sont sur-employés (par rapport à l'ensemble du corpus)
  - Pour un mot (ou une expression, un motif), dans quelle(s) partie(s) se trouve-t-il de façon privilégiée ?

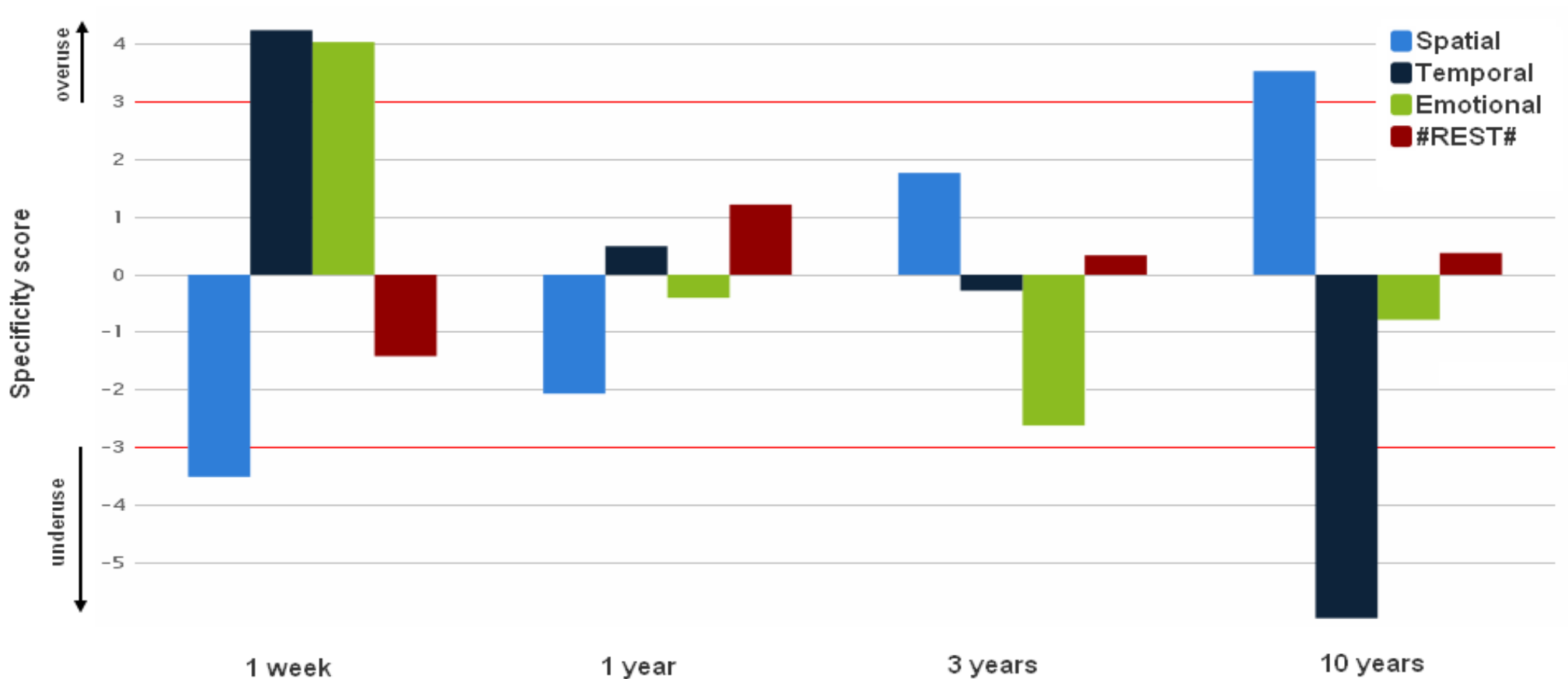
Gros établissements				Moyens établissements				Petits établissements			
Verbe	F	Fp	Sp	Verbe	F	Fp	Sp	Verbe	F	Fp	Sp
équiper	36	26	8,4	pêcher	718	463	7,0	apprendre	138	58	8,6
étudier	10	9	4,5	respecter	51	43	5,1	sauver	24	16	6,1
prévoir	23	15	4,2	concerner	19	19	5,0	espérer	29	17	5,3
surgeler	24	15	3,9	associer	22	21	4,5	lever	239	75	4,8
suivre	68	31	3,7	vendre	527	332	4,0	subir	11	9	4,7
placer	12	9	3,4	lier	31	27	3,8	travailler	416	118	4,7
HAPAX	580	180	3,0	savoir	887	539	3,7	former	53	24	4,6
confondre	5	5	3,0	parler	188	126	3,3	dépêcher	20	12	4,0
rééquilibrer	5	5	3,0	transformer	68	50	2,9	permettre	112	38	3,5
créer	33	17	3,0	payer	208	136	2,9	appréhender	7	6	3,4
diminuer	28	15	2,9	connaître	235	151	2,7	progresser	7	6	3,4
lancer	39	19	2,9	recupérer	46	35	2,6	enlever	48	20	3,3
adapter	34	17	2,8	marier	10	10	2,6	jeter	30	14	3,1
marquer	29	15	2,7	racheter	24	20	2,5	préférer	48	19	2,9
attaquer	14	9	2,6	attaquer	13	12	2,3	entourer	6	5	2,8
pousser	14	9	2,6	frayer	23	19	2,3	imprégner	4	4	2,8
développer	53	23	2,5	démarrer	29	23	2,2	essayer	153	46	2,8
embêter	22	12	2,5	perdre	62	44	2,2	préserver	11	7	2,7
donner	197	67	67	évaluer	12	11	2,2	gagner	134	41	2,7

***Spécificités des verbes employés par les pêcheurs selon la taille de leur entreprise : le calcul statistique fait apparaître des équilibres différents dans le choix des verbes (Le Lay et al. 2016).***



# Spécificités : un exemple en linguistique (Guillot et al. 2013)

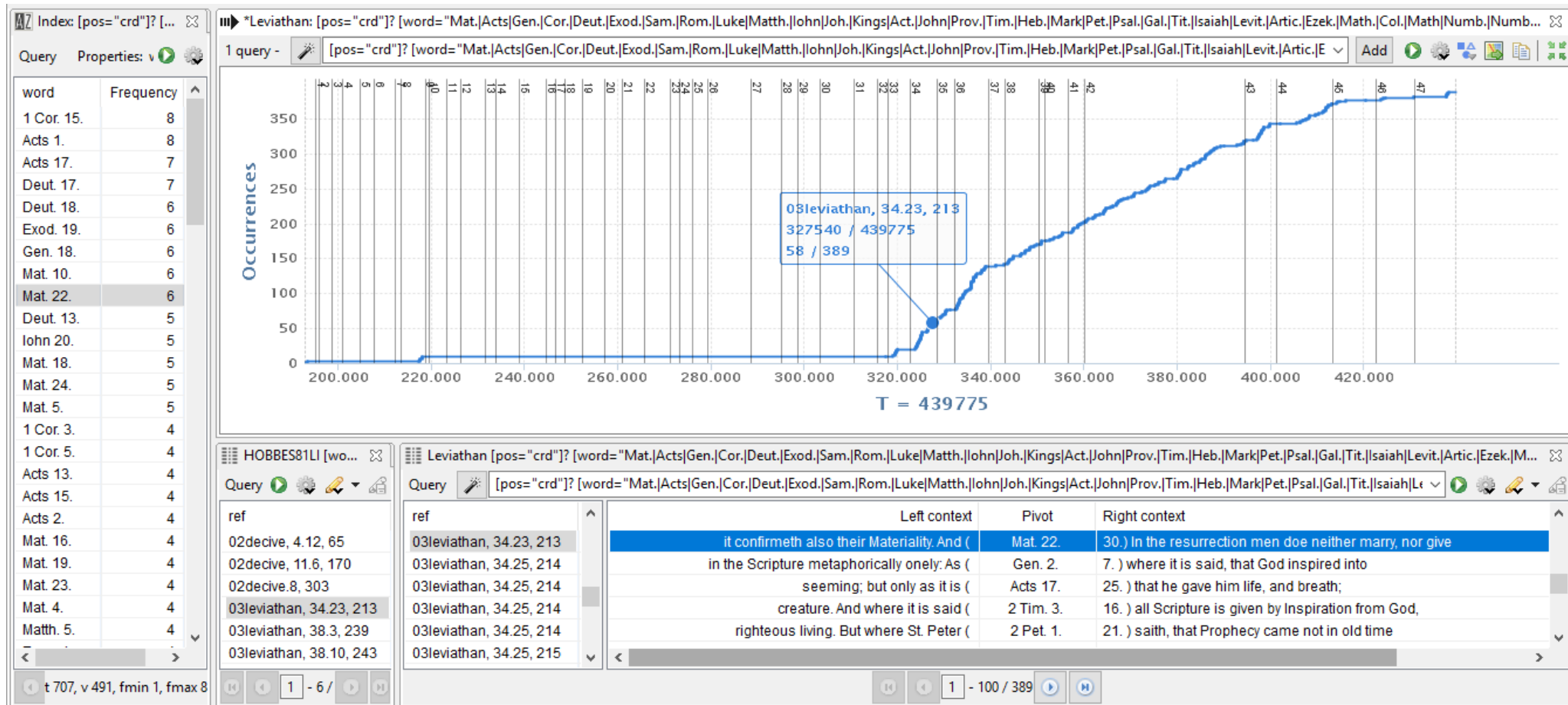
- À défaut de pouvoir observer directement la langue orale au Moyen-Âge, on cherche à la caractériser à travers l'oral représenté dans les textes (discours direct).
- On divise le corpus en deux parties : les passages au discours direct d'une part (DD), et le reste d'autre part (non DD).
- On calcule alors les spécificités des catégories grammaticales pour observer si certaines sont particulièrement sur-représentées dans l'une ou l'autre partie (on ne considère que les 9 textes pour lesquels les catégories ont été vérifiées) :
  - DD : pronoms, possessifs, adverbes de négation, verbes à l'infinitif, interrogatifs, interjections, démonstratifs...
  - Non DD : noms, déterminants, adjectifs qualificatifs, conjonctions de coordination...
- On peut ensuite analyser plus finement en étudiant des différences de répartition à des niveaux plus précis (nouveaux calculs de spécificités plus ciblés) :
  - mots de la catégorie : ex. dans les conjonctions de coordination, c'est en fait surtout le « et » qui est spécifique au non-DD)
  - sous-catégories : ex. les pronoms personnels de 3<sup>e</sup> personne sont sur-représentés dans le non-DD)
  - constructions : ex. le verbe infinitif est sur-représenté dans le DD pour les constructions « verbe modal + infinitif », tout particulièrement avec « devoir ». À l'inverse, « aller + infinitif » se révèle spécifique au non-DD.



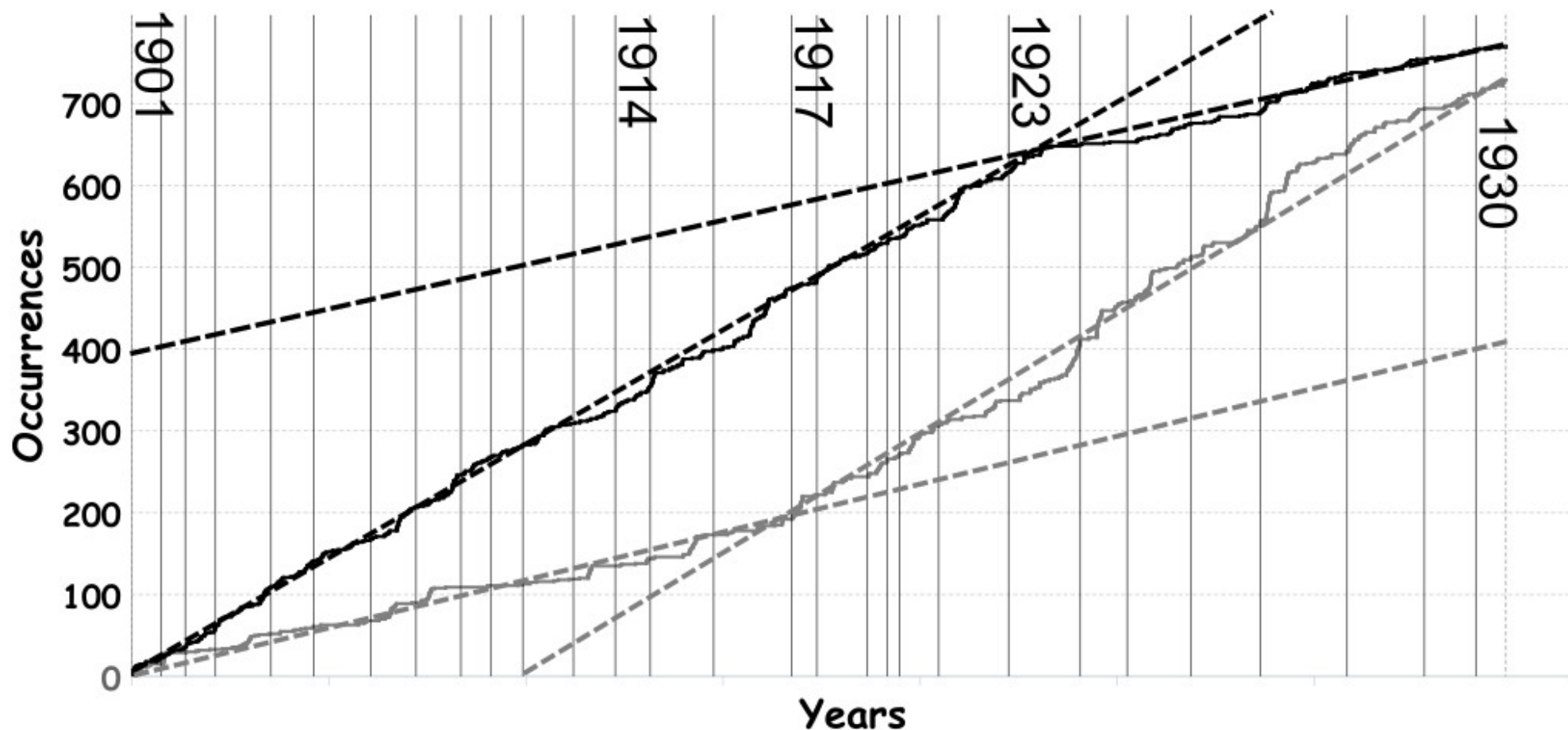
**Spécificités des types de détails mentionnés par les sujets selon la distance temporelle à l'événement** : une annotation manuelle a été effectuée pour repérer différents types de détails : temporels (indications horaires), spatiaux (noms de rues), émotionnels (sémantique lexicale). Le calcul des spécificités montre que les détails temporels et émotionnels sont sur-représentés dans les narrations juste après l'événement, alors qu'à long terme les détails spatiaux prennent davantage d'importance et les détails temporels s'estompent (Dégeilh et al. 2017).

# Progression

- Le corpus est perçu dans son déroulement, comme un continuum (plutôt que comme une succession de périodes)
- Cela peut concerner une évolution chronologique, ou le déploiement linéaire (syntagmatique) d'un texte.
- Dans TXM, la fonction prend la forme d'un graphique cumulatif : la plus ou moins forte pente représente la plus ou moins grande présence du mot (ou motif, expression).
  - D'autres types de graphiques existent dans d'autres logiciels pour répondre au même questionnement (ex. carte des sections de Lexico, topologie dans Hyperbase).



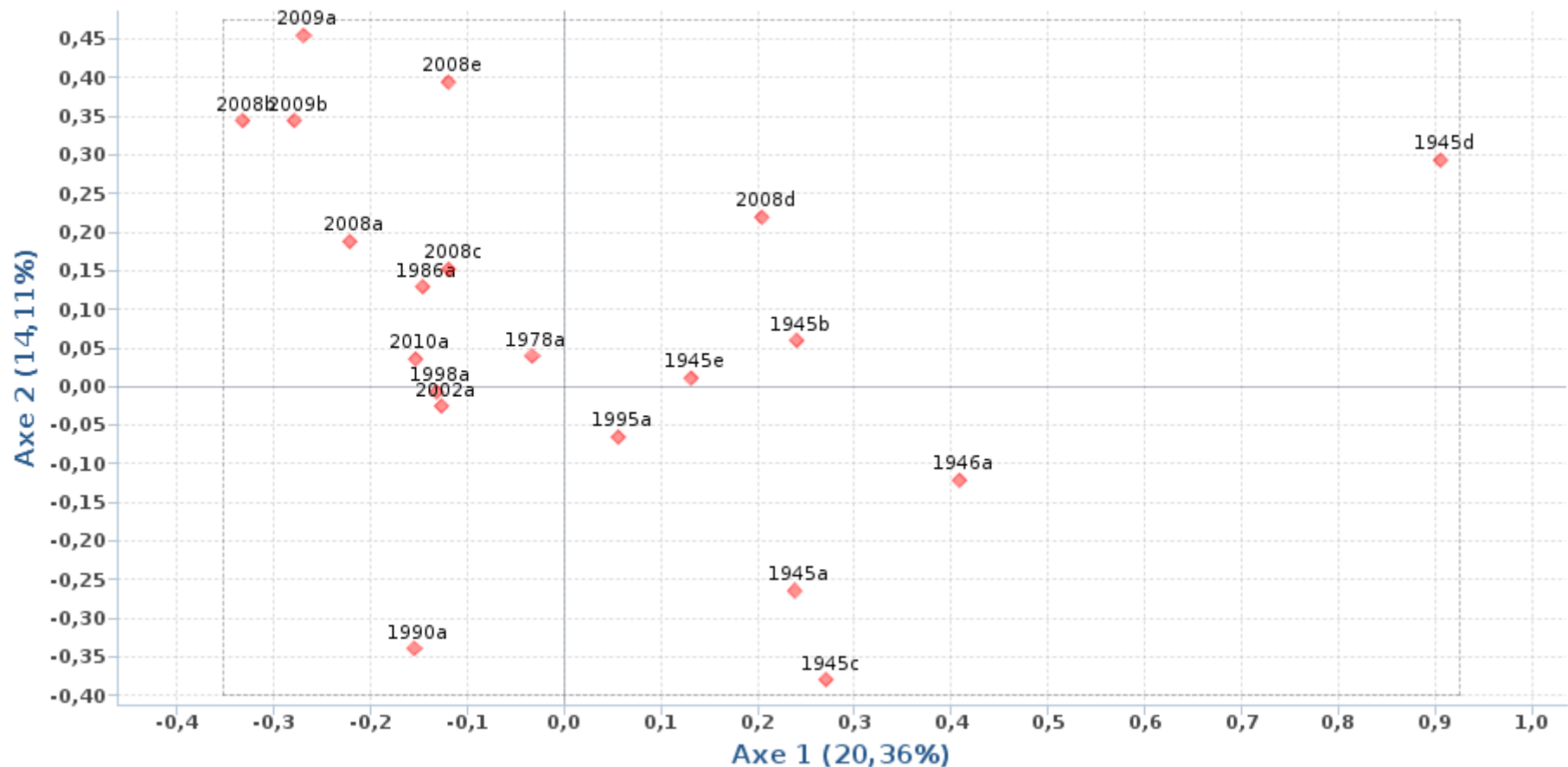
***Progression des références bibliques dans le Léviathan de Hobbes : les références bibliques apparaissent quasi-absentes de la première moitié de l'œuvre (courbe « plate »), et deviennent subitement beaucoup plus présentes dans les chapitres 34 à 44 (Rebasti et Heiden 2019).***



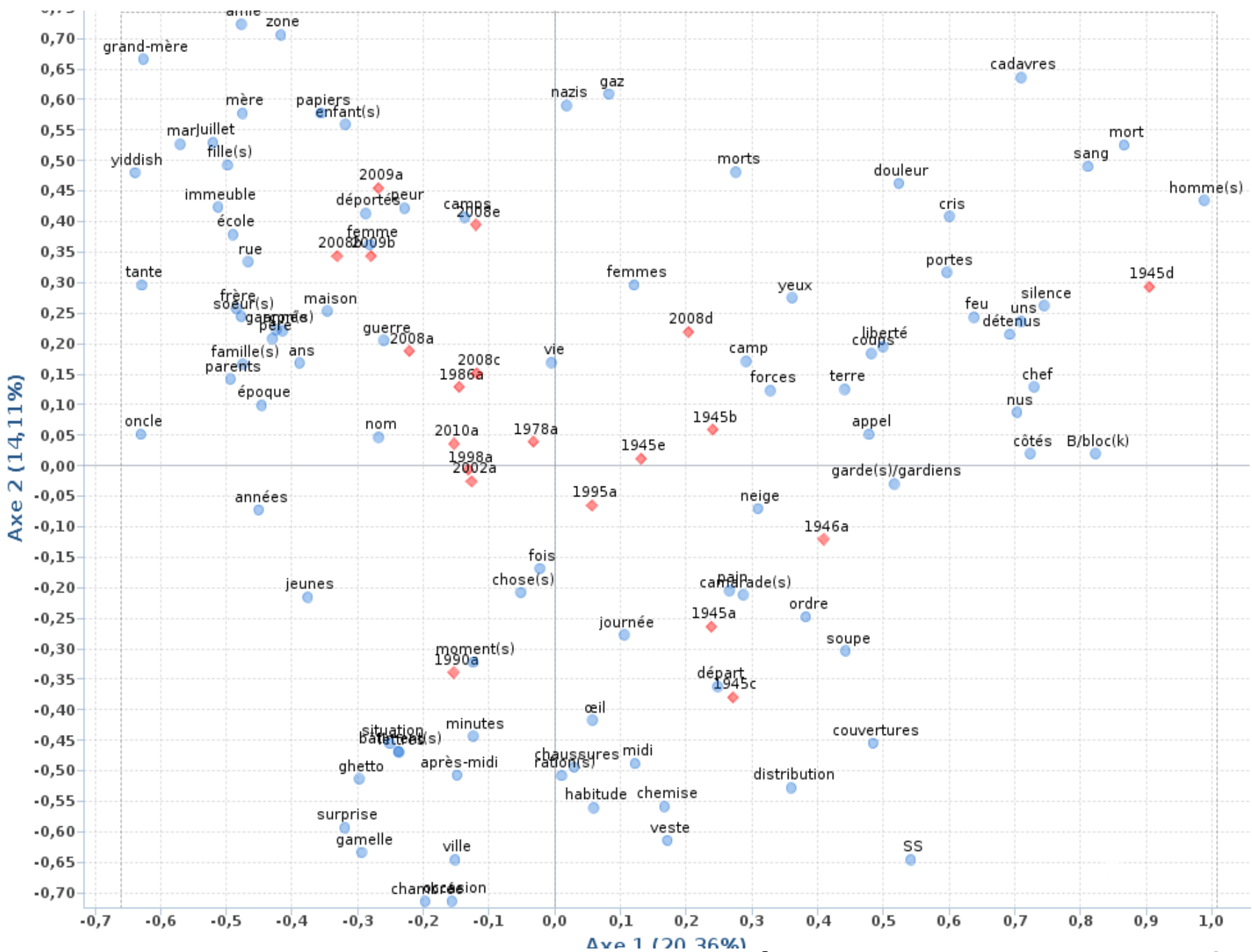
*Progression des lemmes dusha (âme, courbe noire) et tovarisch (camarade, courbe grise) dans un corpus de nouvelles russes : dans ses analyses d'ensemble, le corpus révèle des évolutions lexicales fortes autour de la période de la Révolution de 1917. On observe ici des changements marqués de fréquence moyenne d'emploi pour deux mots aux profils opposés (Lavrentiev et al. à paraître).*

# Analyse factorielle des correspondances (AFC)

- Synthèse géométrique des contrastes au sein du corpus.
- La visualisation privilégiée est un plan 2D,
  - qui concentre (mathématiquement) le maximum d'information sur les contrastes,
  - mais qui ne représente pas tout.
- La lecture de la visualisation 2D doit donc être contrôlée à l'aide d'autres informations, telles que :
  - Spécificités calculées sur la même table,
  - Contributions des points aux dimensions représentées, déformation liée à la projection,
  - Stabilité de la représentation (ellipses de confiance).



**Analyse factorielle des correspondances des 20 témoignages de la Shoah** : pour le calcul les témoignages sont représentés ici par leur fréquence d'emploi des 400 noms communs les plus fréquents. Il ressort que, dès au niveau de ce vocabulaire, les témoignages rédigés à la sortie des camps se démarquent de ceux rédigés ultérieurement (Mayaffre et al. 2018).

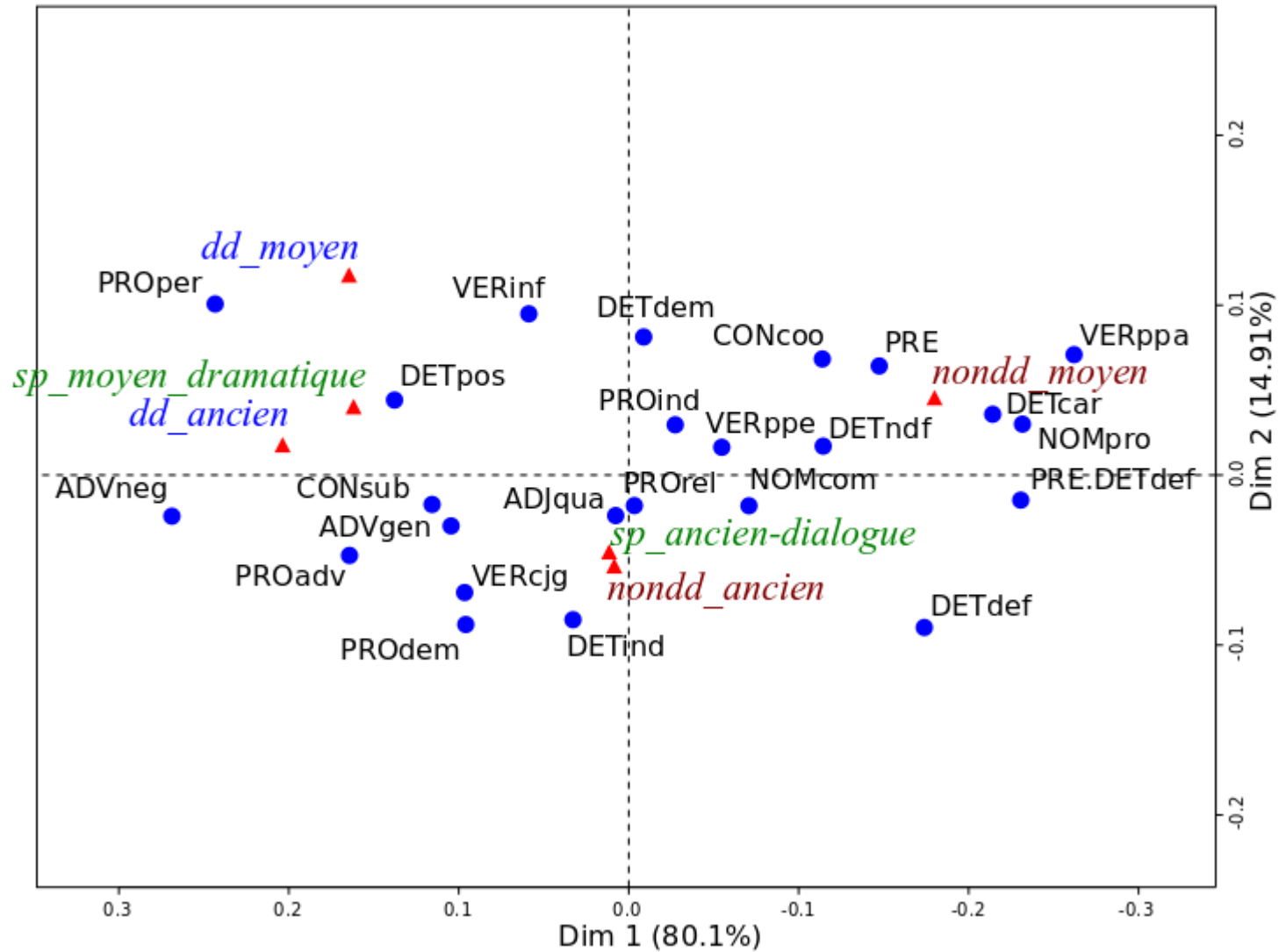


**Analyse factorielle des correspondances des 20 témoignages de la Shoah : même analyse, en affichant les 97 noms structurant le plan (contribution > 0,5 % ou  $\cos^2$  au plan > 0,5). Le vocabulaire sensoriel de la douleur et du quotidien des camps contraste avec un vocabulaire familial ou plus abstrait (Mayaffre et al. 2018).**

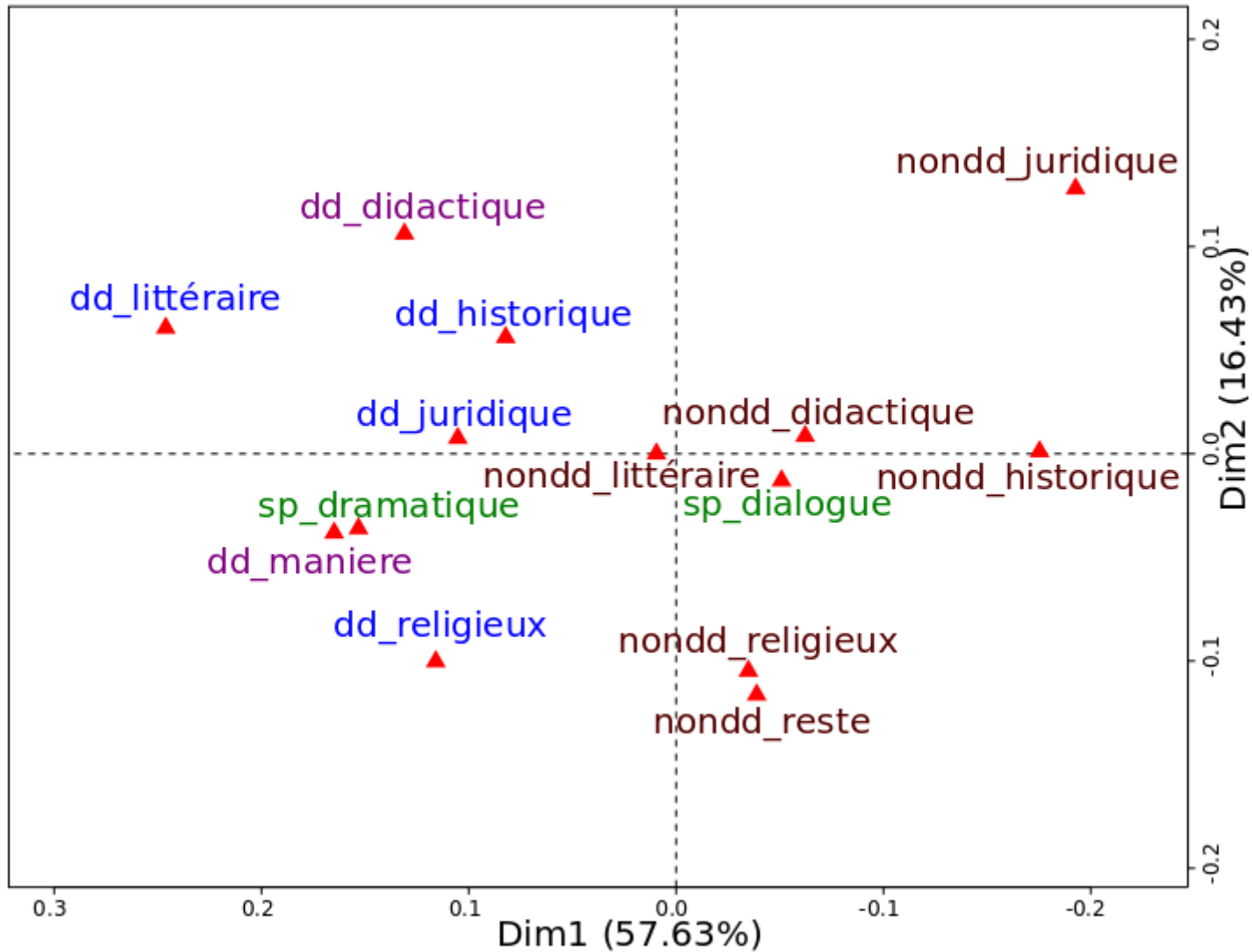


# AFC en linguistique : exemple de l'oral représenté

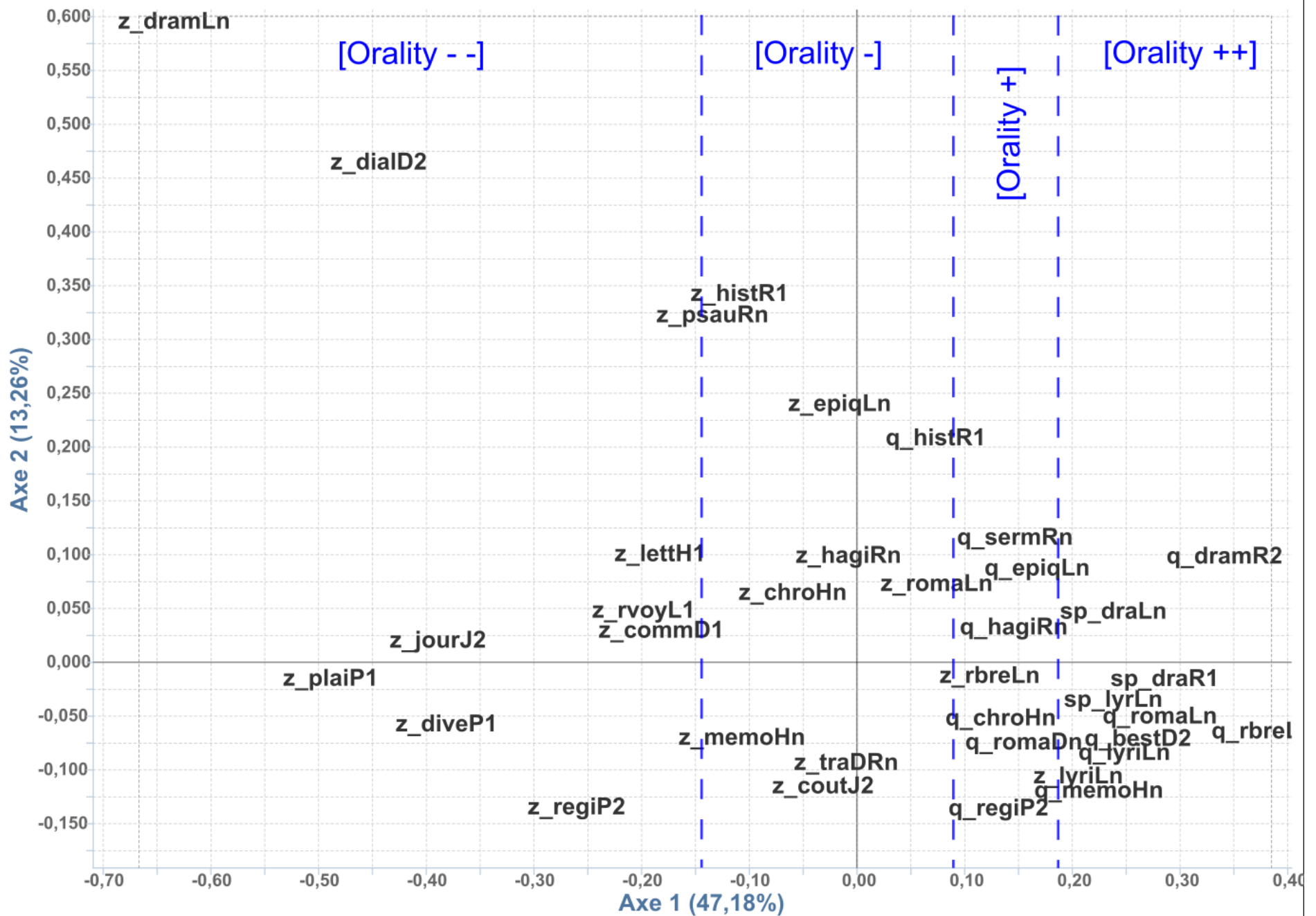
- À défaut de pouvoir observer directement la langue orale au Moyen-Âge, on cherche à la caractériser à travers l'oral représenté dans les textes (discours direct).
- On fragmente le corpus en unités de discours (périodes, domaines, genres, ou textes), en séparant aussi les passages au discours direct d'une part, et le reste d'autre part.
- Les unités de discours sont représentées par leur composition en catégories grammaticales.
- On cherche alors à observer quelle est la principale dimension de contraste entre les unités de discours : est-ce que, au niveau linguistique des catégories grammaticales, les contrastes sont plus forts entre discours direct et non discours direct, ou selon d'autres facteurs (chronologie, type de texte, etc.)
- Étude développée dans plusieurs publications successives :
  - (Guillot et al. 2015) : 2 périodes, 5 domaines, 20 textes vérifiés
  - (Guillot-Barbance et al. 2017) : 40 genres – point de vue plus linguistique
  - (Pincemin et al. 2020) : 40 genres, distinction vers/prose au sein des genres, 137 textes – point de vue plus statistique



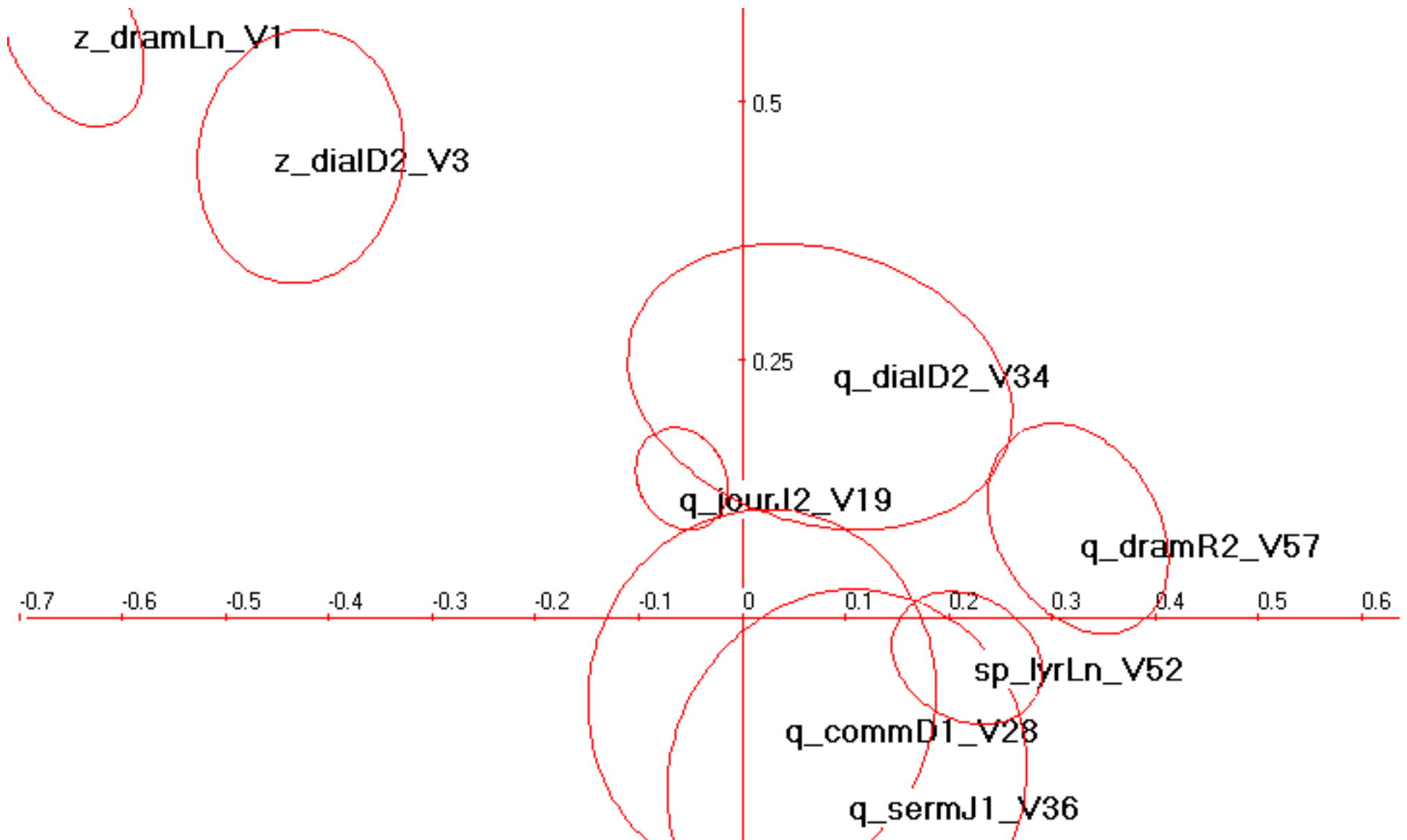
**Analyse factorielle des correspondances du contraste entre discours direct et non discours direct au niveau des périodes : les périodes opposent Ancien et Moyen français (Guillot et al. 2015).**



***Analyse factorielle des correspondances du contraste entre discours direct et non discours direct au niveau des domaines (Guillot et al. 2015)***

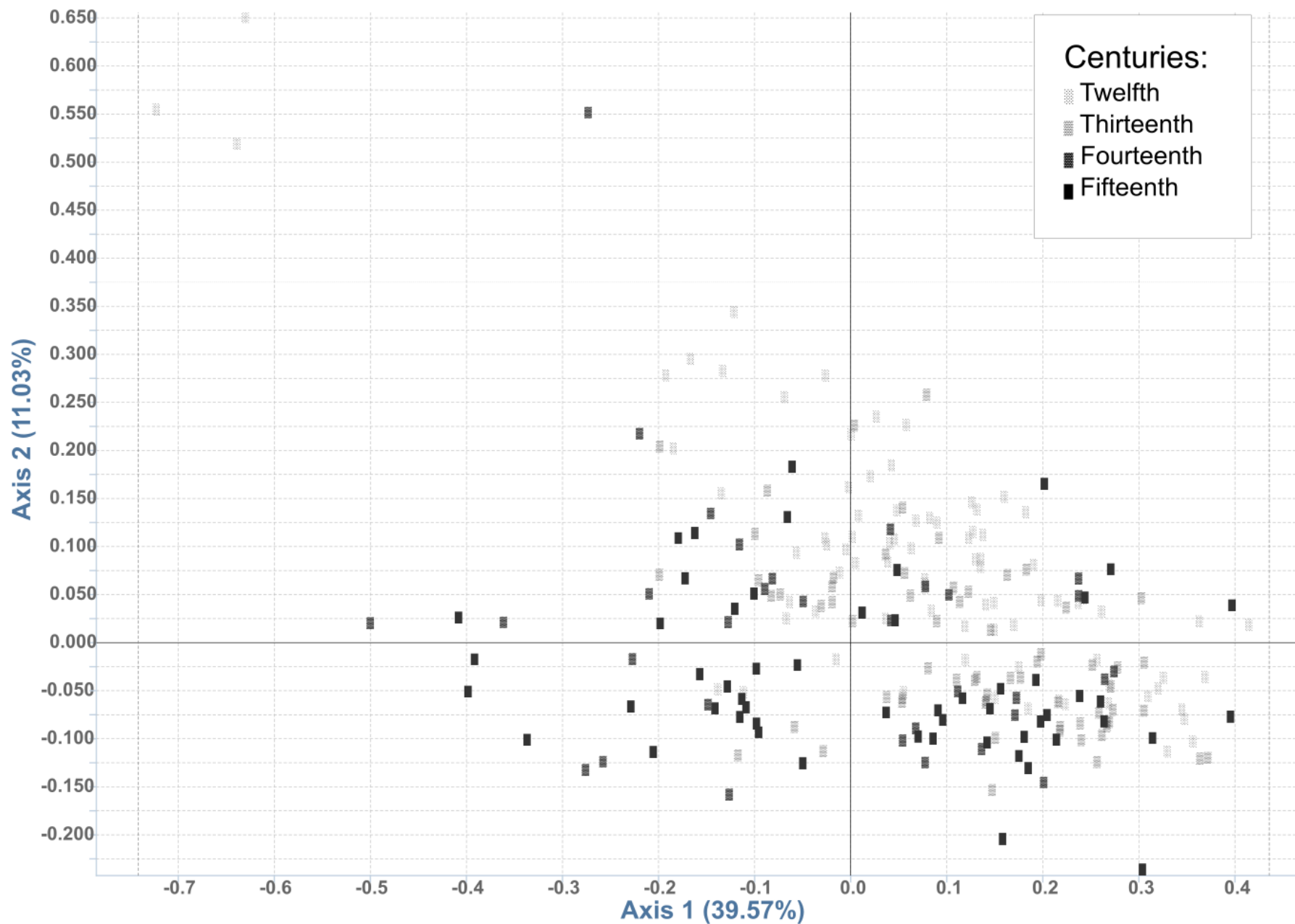


**Analyse factorielle des correspondances du contraste entre discours direct et non discours direct au niveau des genres : le préfixe z\_ représente le non DD. Les points peu contributeurs (< 2 %) et mal représentés ( $\cos^2$  au plan < 0,3) ont été filtrés pour la lisibilité du plan mais l'ensemble des positions des points a été examiné (Pincemin et al. 2020).**



**Analyse factorielle des correspondances du contraste entre discours direct et non discours direct au niveau des genres : affichage des plus grandes ellipses de confiance, montrant que seules 5 unités de discours (dont 2 qui n'ont pu être représentées) n'ont pas une position interprétable (sur 59 au total).**  
*Rq. Le tableau de données construit par TXM a été analysé par DtmVic pour le calcul des ellipses de confiance (Pincemin et al. 2020).*



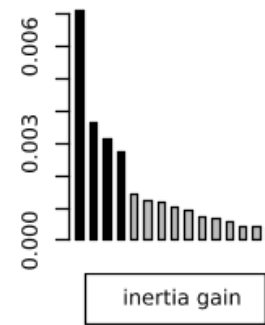
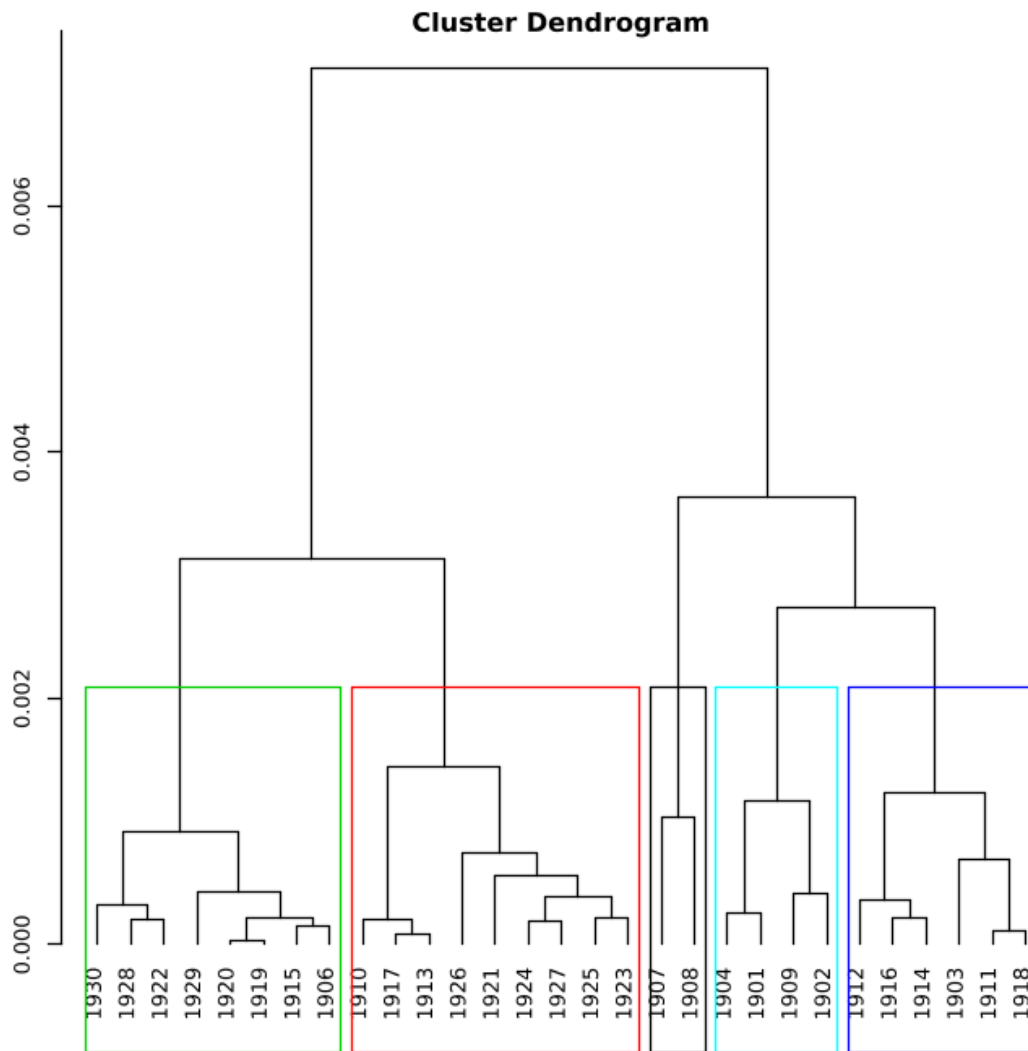


**Analyse factorielle des correspondances du contraste entre discours direct et non discours direct au niveau des textes, colorés selon le siècle : l'examen des coordonnées sur le 1<sup>er</sup> axe ne montre que 4 inversions DD/nonDD au sein d'un même texte, dont 2 quasi nulles, sur 75 textes concernés. La chronologie apparaît portée par l'axe 2 (Pincemin 2020).<sup>54</sup>**

# Classification

- Dans TXM : classification ascendante hiérarchique (en aval d'une AFC).
  - Convient pour des parties pas trop nombreuses (lisibilité).
- Pour des regroupements de mots, voir par exemple la méthode Reinert implémentée dans IRaMuTeQ
  - Des passerelles existent pour qu'un même corpus puisse être analysé par les deux logiciels (TXM et IRaMuTeQ).

## Hierarchical clustering



**Classification ascendante hiérarchique des années du corpus de nouvelles russes :** l'AFC du corpus sur les 200 mots (lemmes) les plus fréquents (à droite) avait montré un effet chronologique, la CAH (sur les coordonnées des 5 premières dimensions de l'AFC) vient confirmer une rupture majeure autour de 1917 (Lavrentiev et al. à paraître).

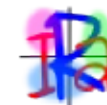




wagon	travail	froid	coup	pain	rue	enfant	adèle	vie	juif
porte	kommand	neige	oeil	soupe	pari	mère	simon	situation	pays
bâtiment	block	fatigue	tête	ration	hôtel	soeur	répondre	mémoire	guerre
salle	crématoire	sommeil	regarder	manger	rafle	frère	revoir	sentiment	allemagne
mètre	camp	jambe	crier	pomme	argent	famille	filles	condition	politique
étage	four	douleur	main	morceau	appartemen	père	embrasser	survie	france
monter	gaz	soif	bras	margarine	juillet	école	jeune	avenir	nazi
lit	ss	soif	frapper	lait	carte	parent	siwek	vivre	europe
cour	kapo	allonger	visage	distribution	police	oncle	question	réalité	gouveremen
train	chantier	réveiller	poing	gamelle	ville	ainé	inquiéter	récit	communiste
descendre	appel	glacial	jeter	distribuer	habiter	fi	demander	monde	mondial
barbelé	travailler	force	hurler	viande	lyon	ainsi	aimer	espoir	population
escalier	détenu	fièvre	cri	sucré	faux	tante	questionner	physique	hitler
bois	chef	corps	rire	nourriture	identité	naître	remercier	survivre	allemand
châlit	équipe	endormir	lever	cigarette	arrondissement	régina	jacob	circonstance	hitlérien
couloir	creuser	plaie	sang	beurre	visa	slatka	véronique	événement	résistant
immense	orchestre	faim	regard	confiture	quartier	âge	visite	détail	pétain
camion	vorarbeiter	dormir	voir	litre	métier	régine	saluer	comprendre	antisémite
fenêtre	rang	mort	homme	tranche	rome	garçon	rachel	conscience	citoyen
mur	commando	geler	taper	boire	franc	mari	rencontrer	existence	loi
traverser	sélection	boue	ouvrir	thé	tailleur	grand_mère	rassurer	respect	résistance
route	chambre	vent	bâton	saucisson	patron	lén	hélène	histoire	propagande
planche	daw	nu	baisser	boîte	lettre	année	heureux	témoin	hongrie
long	ordre	soleil	revolver	gâteau	hiv	filles	rendre	crime	populaire
arriver	kommandos	cauchemar	cravache	chocolat	drancy	cousin	sourire	contraire	régime
pièce	huta	tomber	tourner	chaud	vél	nancy	ravir	agir	journal
heure	rassemblement	typhus	tourner	tabac	belleville	classe	adresser	question	français
fer	baraque	réchauffer	genou	fromage	boulevard	manier	amoureux	oublier	statut
	hirkonass	mourir	apercevoir	assiette	papier	habiter	poser	esprit	déporté
		nuît	tirer			âné		poser	antisémitisme

**Regroupement de mots par la méthode Reinert : une classification descendante hiérarchique est appliquée sur des « unités de contexte » (par exemple des fragments d'une quarantaine de mots tenant compte de la ponctuation). Les groupes de mots sont les mots caractéristiques de ces classes (chi-2). Calcul réalisé avec IRaMuTeQ.**

Exemple ici sur le corpus des témoignages de la Shoah (Mayaffre et al. 2018).



# D'autres fonctionnalités dans d'autres logiciels

- *Pour la diachronie* : il existe des variantes plus spécialisées des calculs proposées par Salem (1991) qu'on trouvera dans d'autres logiciels :
  - Lexico : <http://www.lexi-co.com/>
  - Trameur : <http://www.tal.univ-paris3.fr/trameur/>
- *Pour d'autres visualisations "spatiales"*
  - en réseau : analyse arborée d'Hyperbase, analyse des similitudes d'IRaMuTeQ, arbre de longueur minimale dans DtmVic ;
  - DtmVic : cartes de Kohonen
    - <http://hyperbase.unice.fr/>
    - <http://ancilla.unice.fr/hyperbase/>
    - <http://iramuteq.org/>
    - <http://www.dtmvic.com/>

# Plan

1. *[Intro]* Données textuelles et SHS
2. *[Pratique]* Fonctionnalités textométriques
3. ***[Théorie]* Positionnement et méthodologie**
  - Texte et langue
  - Sens et interprétation construite
  - Choix des calculs
  - Critiques
4. *[Pour aller plus loin]* Ressources

# Le rapport à la langue et aux textes

- Robustesse : prendre les données comme elles sont.
  - Mais connaître ses données, construire son corpus : enjeu d'interprétabilité (Pincemin 2012).
- L'attention au texte, à l'expression, à la textualité
  - Source plutôt que ressource (Valette 2016) ? cf. le retour au texte
  - *Des textes plutôt que du texte* (Péry-Woodley 1995) : plutôt des textes intégraux, avec des métadonnées, pouvant être structurés (paragrapes, sections, tours de parole, etc.)
- Langue : pas de supériorité de la lemmatisation
- L'attention à la formulation
  - ex. de la question ouverte « Quelles sont les raisons qui, selon vous, peuvent faire hésiter une femme ou un couple à avoir un enfant ? » : manque d'argent ≠ raisons financières (Lebart & Salem 1994)

# Sens et interprétation

- Sémantique contextuelle
  - On « fait parler » le corpus par rapport à lui-même
  - Une approche ontologique (type Web Sémantique) est possible mais pas première
    - Plus favorable pour découvrir de l'inattendu plutôt que retrouver ce qu'on a mis ?
- Place de l'interprétation, pluralité des lectures
  - Il n'y a pas qu'une seule analyse d'un corpus
    - Pas automatique, avec une entrée → une sortie
    - L'évaluation de la qualité de l'analyse ne peut se faire par « banc d'essai », mais se fonde sur le caractère clair, argumenté, reproductible.
  - Les calculs fournissent un résultat, pas une réponse

# Le rapport au logiciel

- Pas automatisé de l'analyse, mais outil
  - Ordinateur : mémoire et vitesse de calcul
  - Chercheur : conduit l'analyse, contrôle, interprète
- Exigences
  - Connaître / lire son corpus (relire (différemment) vs s'épargner la lecture)
  - Composer son corpus – préparation numérique (progressif)
  - Trouver / évaluer les points d'entrée
  - Comprendre (vs produire) :
    - le traitement, les résultats : qu'est-ce que cela dit et comment on y arrive, sur quelles bases, quel point de vue ;
    - règles de lecture (« herméneutique des sorties logicielles », Rastier 2011), ex. pente vs altitude dans diagrammes de Progression, proximité des points dans un plan factoriel, etc.

# Conception des traitements et choix des calculs

- Concordance → pas n'importe quel KWIC (*keyword in context*)
  - Pour avoir les effets heuristiques de rapprochement des contextes : tri (non trivial), superposition verticale, contexte sur une ligne
  - Références de localisation globale
- Spécificités → centrales par rapport à des fréquences relatives ou d'autres mesures statistiques (chi-2, log-likelihood, etc.)
  - Calcul exact, non paramétrique – pas de seuil de fréquence pour son applicabilité
  - Mais prise de distance, ou souplesse, par rapport à la notion de significativité
    - On ne modélise pas la langue : on mesure l'écart à l'aléatoire sachant qu'elle n'est pas aléatoire
    - l'indice sert plus à hiérarchiser qu'à valider
- AFC → mieux adaptée aux tableaux de contingence (que ACP, MDS...)

# Limites et critiques

- International : des calculs originaux, pas toujours bien identifiés
  - AFC moins connue qu'ACP
  - Spécificités non (re)connues comme un test exact de Fisher
- L'indigence des « mots » sur lesquels se base l'analyse ?
  - Polysémie, synonymie, implicite, néologies...
  - Mots découpés automatiquement ?
  - Plutôt : des unités de travail au croisement desquelles pourront apparaître les unités faisant sens – les unités d'analyse sont construites
- voir aussi (Pincemin 2020).



# Plan

1. *[Intro]* Données textuelles et SHS
2. *[Pratique]* Fonctionnalités textométriques
3. *[Théorie]* Positionnement et méthodologie
4. ***[Pour aller plus loin]* Ressources**
  - Logiciel TXM
  - Bibliographie

# TXM : genèse et orientations fondatrices

- Une première génération de logiciels précurseurs (années 1980-1990) et très complets
  - Lexico, Hyperbase, Spad-T, Weblex, Astartex...
- Projet ANR Textométrie (2007-2010) : Fédération des recherches et développements en textométrie autour de la création d'une plateforme logicielle ouverte
  - Collaboratif, open-source, mutualisation de développements
  - Prise en compte des corpus structurés et enrichis (XML TEI, TAL,...)

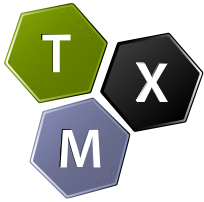


**Textométrie**  
Fédération des recherches et développements  
en textométrie autour de la création  
d'une plateforme logicielle ouverte

– accessible (affichés)  
– modifiables & écrits  
– enrichies (enrichissements)  
– normalisées  
– partageables / publiables

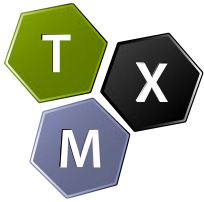
– accessible (affichés)  
– modifiables & écrits  
– enrichies (enrichissements)  
– normalisées  
– partageables / publiables

– accessible (affichés)  
– modifiables & écrits  
– enrichies (enrichissements)  
– normalisées  
– partageables / publiables



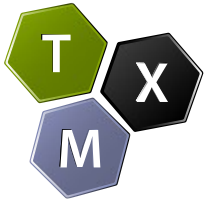
# TXM : le traitement des corpus structurés et enrichis

- Possibilités actuelles
  - Représenter les structurations internes au texte (et les métadonnées textuelles), notamment avec XML TEI
  - Enrichissements apportés par des outils de Traitement automatique des langues
    - En particulier, la lemmatisation et l'identification des catégories grammaticales
- Philologie numérique
  - Concilier Édition & Base de données & Moteur de recherche
- Multiplicité des imports
  - prendre le texte tel qu'il est (vs dépouiller ou équiper)



# TXM : open-source

- Non boîte noire
- Modularité
  - composants experts : R, CQP ; Textometry
  - articulation à d'autres logiciels
    - exports possibles dans des formats appropriés (csv, svg)
    - transfert privilégié au niveau de la table lexicale, ex. :
      - Construire une table mots x textes dans TXM (en profitant du retour au texte, du moteur de recherche CQP, etc.) et l'analyser avec R (pour d'autres calculs statistiques), ou avec DtmVic (pour les ellipses de confiance)
      - Entrer dans TXM une table de contingence entre deux séries de phénomènes (ex. types x positions) et obtenir avec TXM des analyses des correspondances et des spécificités.
  - Scriptage : pouvoir personnaliser/compléter le logiciel sans entrer dans le coeur du code
- Gratuité
  - citer le logiciel
  - communauté
- Partage des données // ouverture du code (Guillot, Heiden, Lavrentiev 2017)



# TXM : d'autres caractéristiques

- Interface utilisateur graphique
  - Outils des scénarios, en particulier le retour au texte
- Multiplateforme et en ligne
  - Windows, MacOS, Linux
  - Portail à installer sur un serveur pour publier ou partager des corpus en ligne
- Point d'entrée : le site <http://textometrie.org>
  - Téléchargement, documentation du logiciel et de la méthode
  - Wiki et liste utilisateurs
  - Bibliothèque de fichiers : macros, corpus
  - Wiki développeurs

# Bibliographie sur la méthode (1)

- Ouvrages

- Lebart Ludovic, Salem André (1994) - *Statistique Textuelle*. Paris, Dunod.
- Lebart Ludovic, Salem André et Berry Lisette (1998). *Exploring Textual Data*. Dordrecht, Kluwer Academic.
- Lebart Ludovic, Pincemin Bénédicte, Poudat Céline (2019) – *Analyse des données textuelles*. Québec, Presses de l'Université du Québec.
- Née Emilie (dir.) (2017) – *Méthodes et outils informatiques pour l'analyse des discours*, Rennes, Presses universitaires de Rennes.
- Poudat Céline, Landragin Frédéric (2017) – *Explorer un corpus textuel. Méthodes – pratiques – outils*. Louvain-la-Neuve, De Boeck.
- Bernard Michel, Bohet Baptiste (2017) – *Littérométrie. Outils numériques pour l'analyse des textes littéraires*. Paris, Presses de la Sorbonne nouvelle.
- Fénelon Jean-Pierre (1981) - *Qu'est-ce-que l'analyse des données ? Exposé accessible aux non-mathématiciens*, Lefonen, Paris. Réédité en 1999 par SEISAM, Paris.
- Brunet Étienne : les trois tomes des *Écrits choisis* chez Champion : *Comptes d'auteurs. Études statistiques, de Rabelais à Gracq* (2009) ; *Ce qui compte. Méthodes statistiques* (2011) ; *Tous comptes faits. Questions linguistiques* (2016).

# Bibliographie sur la méthode (2)

- Autres documents

- Mayaffre Damon (2007) – « Analyses logométriques et rhétoriques des discours », in Stéphane Olivési (dir.), *Introduction à la recherche en SIC*, Grenoble, Presses universitaires de Grenoble, 2007, pp. 153-180.
- Mayaffre Damon (2010) – *Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique*, Mémoire d'Habilitation à diriger des recherches en Histoire, Université Nice Sophia Antipolis.  
<https://tel.archives-ouvertes.fr/tel-00655380>
- Revue *MOTS* : les numéros des années 1980 comprennent de nombreux articles de référence sur les modèles statistiques mobilisés. Disponible en ligne : <https://www.persee.fr/collection/mots>
- Revue *Corpus* : <https://journals.openedition.org/corpus/>
- Colloque JADT : édition courante : <http://jadt.org>, Actes : <http://lexicometrica.univ-paris3.fr/jadt/index.htm>

# Bibliographie des exemples présentés (1)

- Béranger Marine, Heiden Serge, Lavrentiev Alexei (2015) - « Reengineering Akkadian Tablets with TEI and TXM for Linguistic Analysis », *Text Encoding Initiative Conference and Members' Meeting 2015*, Oct. 28-31, Lyon.
- Carrive Jean, Beloued Abdelkrim, Goetschel Pascale, Heiden Serge, Laurent Antoine, Lisena Pasquale, Mazuet Franck, Meignier Sylvain, Pincemin Bénédicte, Poels Géraldine, Troncy Raphaël (à paraître) - « Transdisciplinary Analysis of a Corpus of French Newsreels: The ANTRACT Project », *Digital Humanities Quarterly*, Special Issue on AudioVisual DH: Challenges and Possibilities.
- Carrive Jean, Goetschel Pascale, Mazuet Franck (2018) - « Favoriser la recherche sur l'image : le projet ANTRACT », Table-ronde aux 22e *Rendez-vous de l'Histoire* « La puissance des images », Blois, 9-13 octobre 2018.
- Dégeilh Fanny, Hirst William, Heiden Serge, Pincemin Bénédicte, Decorde Matthieu, Meksin Robert, Eustache Francis, Peschanski Denis (2017) - « Temporal evolution of memory for the attack of September 11, 2001: Lexical analysis of flashbulb memories through 10-year follow-up », *CPA National Convention*, Toronto (Ontario), 8-10 juin 2017.
- Guillot Céline, Heiden Serge, Lavrentiev Alexei, Pincemin Bénédicte (2015) - « L'oral représenté dans un corpus de français médiéval (9e-15e) : approche contrastive et outillée de la variation diasystémique », in Kirsten Jeppesen Kragh et Jan Lindschouw (éds), *Les variations diasystémiques et leurs interdépendances dans les langues romanes - Actes du Colloque DIA II à Copenhague (19-21 nov. 2012)*, Strasbourg : Éditions de linguistique et de philologie, coll. Travaux de Linguistique Romane / Sociolinguistique, dialectologie, variation, ISBN 978-2-37276-002-7, p. 15-28 <https://halshs.archives-ouvertes.fr/halshs-00760647v2>
- Guillot Céline, Lavrentiev Alexei, Pincemin Bénédicte, Heiden Serge (2013) - « Le discours direct au Moyen Âge : vers une définition et une méthodologie d'analyse », in Dominique Lagorgette, Pierre Larrivée (dir.), *Représentations du sens linguistique 5* (Sélection de contributions présentées à la cinquième édition de la conférence internationale Représentations du sens linguistique, Chambéry, 25-27 mai 2011), Chambéry : Éditions de l'université de Savoie, Laboratoire langages, littératures, sociétés, coll. langages, n°14, ISBN 978-2-919732-15-9, p. 17-41. <http://halshs.archives-ouvertes.fr/halshs-00820262>
- Guillot-Barbance Céline, Pincemin Bénédicte, Lavrentiev Alexei (2017) - « Représentation de l'oral en français médiéval et genres textuels », *Langages*, 208 (4/2017), p. 53-68. <https://halshs.archives-ouvertes.fr/halshs-01495132>



# Bibliographie des exemples présentés (2)

- Lavrentiev Alexei M., Sherstinova Tatiana Yu., Chepovskiy Andrey M., Pincemin Bénédicte (à paraître) - « Using TXM Platform for Research on Language Changes over Time: the Dynamics of vocabulary and Punctuation in Russian Literary Texts », *Tomsk State University Journal of Philology*.
- Le Lay Yves, Heiden Serge, Merchez Luc, Pincemin Bénédicte (2016) - « Retour de pêche. Le métier de pêcheur à travers le discours des professionnels français du lac Léman », in Émeline Comby, Yannick Mosset et Stéphanie de Carrara (éds.), *Corpus de textes : composer, mesurer, interpréter*, Lyon : ENS éditions, collection « Sociétés, Espaces, Temps », ISBN 978-2-84788-827-0, p. 117-133.  
<https://halshs.archives-ouvertes.fr/halshs-01423605>
- Marchello-Nizia Christiane et Lavrentiev Alexei (éds) (2019) - *Queste del saint Graal*, Lyon, ENS de Lyon, 2019. Publié en ligne par la Base de français médiéval, [http://catalog.bfm-corpus.org/qgraal\\_cm](http://catalog.bfm-corpus.org/qgraal_cm). Dernière révision le 2018-11-30.
- Mayaffre Damon, Pincemin Bénédicte, Heiden Serge, Weyl Philippe (2018) - « L'évolution de la mémoire de la Shoah au prisme de la statistique textuelle », in Denis Peschanski & Brigitte Sion (dir.), *La vérité du témoin. Mémoire et mémorialisation*, vol. 2, collection Mémoire(s), Hermann Éditeurs, Paris & Institut National de l'Audiovisuel, Bry-sur-Marne, chapitre VI, p. 93-124.  
<https://halshs.archives-ouvertes.fr/hal-01890536v1>
- Pincemin Bénédicte, Lavrentiev Alexei, Guillot-Barbance Céline (2020) - « Using the First Axis of a Correspondence Analysis as an Analytic Tool. Application to Establish and Define an Orality Gradient for Genres of Medieval French Texts », in Domenica Fioredistella Iezzi, Damon Mayaffre, Michelangelo Misuraca (eds), *Text Analytics. Advances and Challenges*, Studies in Classification, Data Analysis, and Knowledge Organization, Heidelberg : Springer, p. 127-143.  
[https://doi.org/10.1007/978-3-030-52680-1\\_11](https://doi.org/10.1007/978-3-030-52680-1_11), <https://halshs.archives-ouvertes.fr/halshs-03070182>
- Rebasti Francesca, Heiden Serge (2019) – « The Problem of Hobbes and the Bible: A Textometric Approach », *Digital Humanities Conference (DH 2019)*, Utrecht University, July 9-12,  
<https://dev.clariah.nl/files/dh2019/boa/0887.html>

# Autres références citées (1)

(dans la présentation ou pendant les questions)

- Geffroy Annie, Lafon Pierre (1982) - « L'insécurité dans les grands ensembles. Aperçu critique sur "Le vocabulaire français de 1789 à nos jours" d'Étienne Brunet », *MOTS*, 5, 129-141. [https://www.persee.fr/doc/mots\\_0243-6450\\_1982\\_num\\_5\\_1\\_1079](https://www.persee.fr/doc/mots_0243-6450_1982_num_5_1_1079)
- Guillot Céline, Heiden Serge, Lavrentiev Alexei (2017) - « Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique », *Diachroniques. Revue de Linguistique française diachronique*, 7, 168-184. <https://halshs.archives-ouvertes.fr/halshs-01809581>
- Lafon Pierre (1980) - « Sur la variabilité de la fréquence des formes dans un corpus », *MOTS*, 1, 127-165. [https://www.persee.fr/doc/mots\\_0243-6450\\_1980\\_num\\_1\\_1\\_1008](https://www.persee.fr/doc/mots_0243-6450_1980_num_1_1_1008)
- Lejeune Christophe (2010) - « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches Qualitatives*, 9, p. 15-32. <http://orbi.ulg.ac.be/handle/2268/61098>
- Lejeune Christophe (2017) - « Analyser les contenus, les discours ou les vécus ? À chaque méthode ses logiciels ! », in Marie Santiago-Delefosse & Maria del Rio Carral (eds.), *Les méthodes qualitatives en psychologie et sciences humaines de la santé*, Malakoff, Dunod, chapitre 9, p. 203-224. <https://orbi.uliege.be/handle/2268/206219>
- Muller Charles (1987) - « Introduction », in Étienne Brunet (dir.), *Méthodes quantitatives et informatiques dans l'étude des textes*, Genève, Slatkine, et Paris, Champion, p. 9-12.
- Péry-Woodley Marie-Paule (1995) – « Quels corpus pour quels traitements automatiques ? », *Traitement automatique des langues*, 36 (1-2), p. 213-232.

# Autres références citées (2)

(dans la présentation ou pendant les questions)

- Pincemin Bénédicte (2012) - « Hétérogénéité des corpus et textométrie », in Julien Longhi et Nathalie Garric (éds), *Langages*, 187, Paris : Armand Colin, 13-26.
- Pincemin Bénédicte (2018) – *Sept logiciels de textométrie*. Document de travail, juillet 2018, 11 pages.  
<https://halshs.archives-ouvertes.fr/halshs-01843695>
- Pincemin Bénédicte (2020) - « La textométrie en question », in Véronique Magri (dir.), *Linguistique et traitements quantitatifs, Le français moderne*, 88 (1), 26-43.  
<https://halshs.archives-ouvertes.fr/halshs-02902088>
- Pincemin Bénédicte, Heiden Serge, Decorde Matthieu (2020) - « Textometry on Audiovisual Corpora: Experiments with TXM software », in *Proceedings of 15th International Conference on Statistical Analysis of Textual Data (JADT 2020)*, Université de Toulouse 3 Paul Sabatier, juin 2020.  
<https://halshs.archives-ouvertes.fr/halshs-02779055>
- Rastier François (2008) - « Que cachent les "données textuelles" ? », in Serge Heiden & Bénédicte Pincemin (éds), *Actes des 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, Lyon, 12-14 mars 2008, Lyon : Presses Universitaires de Lyon, vol. I, 13-26.  
<http://lexicometrica.univ-paris3.fr/jadt/jadt2008/pdf/rastier.pdf>
- Rastier François (2011) – *La mesure et le grain*. Paris, Champion.
- Salem André (1991) – « Les séries textuelles chronologiques », *Histoire & Mesure*, 6 (1-2), 149-175.  
[http://www.persee.fr/doc/hism\\_0982-1783\\_1991\\_num\\_6\\_1\\_1390](http://www.persee.fr/doc/hism_0982-1783_1991_num_6_1_1390)
- Valette Mathieu (2016) – « Analyse statistique des données textuelles et traitement automatique des langues. Une étude comparée », in Damon Mayaffre, Céline Poudat, Laurent Vanni, Véronique Magri, Peter Follette (éds), *Statistical Analysis of Textual Data. JADT 2016. Proceedings of 13th International Conference on Statistical Analysis of Textual Data*, vol. II, 697-706.  
<http://lexicometrica.univ-paris3.fr/jadt/jadt2016/01-ACTES/84134/84134.pdf>
- Viprey Jean-Marie (2002) – *Analyses textuelles et hypertextuelles des Fleurs du mal*. Paris, Champion.