

**Sciences humaines combinées**

ISSN : 1961-9936

: Université de Bourgogne, Université de Franche-Comté, COMUE Université  
Bourgogne Franche-Comté

6 | 2010

Actes du colloque interdoctoral 2010

## Exploration textuelle d'un discours de presse régionale du XIX<sup>e</sup> siècle : *Le Petit Comtois* (1883-1903)

01 September 2010.

**Virginie Lethier**

DOI : 10.58335/shc.191

🔗 <http://preo.u-bourgogne.fr/shc/index.php?id=191>

Virginie Lethier, « Exploration textuelle d'un discours de presse régionale du XIX<sup>e</sup> siècle : *Le Petit Comtois* (1883-1903) », *Sciences humaines combinées* [], 6 | 2010, 01 September 2010 and connection on 22 November 2024. DOI : 10.58335/shc.191. URL : <http://preo.u-bourgogne.fr/shc/index.php?id=191>

PREO

# Exploration textuelle d'un discours de presse régionale du XIX<sup>e</sup> siècle : *Le Petit Comtois* (1883-1903)

**Sciences humaines combinées**

01 September 2010.

6 | 2010

Actes du colloque interdoctoral 2010

**Virginie Lethier**

DOI : 10.58335/shc.191

 <http://preo.u-bourgogne.fr/shc/index.php?id=191>

---

Introduction

CONSTRUIRE DE GRANDS ENSEMBLES DE DONNEES TEXTUELLES DE PRESSE REGIONALE DE LA TROISIEME REPUBLIQUE

Numérique et matérialités discursives

Le projet « Base de presse comtoise de la Troisième République »

Un chantier stratégique et prototypique

Description du processus de numérisation

ANALYSER LE DISCOURS DU PETIT COMTOIS

Notre pratique d'exploration textuelle du discours du Petit Comtois

De l'analyse de la ventilation du vocabulaire en diachronie à l'analyse des stratégies de mise en scène de l'information dans le Petit Comtois

D'une vue sur la cooccurrence généralisée à une analyse réthorico-stylistique du Petit Comtois

Conclusion

---

## Introduction

- 1 Exhumer le discours d'un quotidien régional de la Troisième République, tel est l'objectif de notre thèse en sciences du langage. Exhumer dans un sens premier tout d'abord, puisqu'il s'agit bel et bien d'extraire un corpus de son support matériel initial, tombé dans un oubli presque complet, pour l'intégrer à un nouvel environnement

numérique, que nous entendons comme la condition de la valorisation et du renouvellement des lectures interdisciplinaires de cette ressource. Exhumer en un sens métaphorique ensuite, puisque nous nous proposons d'analyser un discours voué à une vie éphémère, compte tenu des conditions de production qui sont celles de la presse quotidienne.

- 2 Ces deux axes de recherche, dont le présent article vise à éclairer les enjeux, sont loin de correspondre à deux scientifiques volets cloisonnés. Dans notre perspective, l'acquisition de données textuelles numérisées et l'analyse proprement dite de ces données forment deux axes de recherche complémentaires qui s'unifient autour du programme d'une analyse textuelle du discours (Adam, Heidmann, 2005). Parce qu'elle vise à dépasser le déficit textuel de l'analyse de discours, l'analyse textuelle du discours intègre en effet à son programme celui d'une nouvelle philologie numérique, telle que J.-M. Viprey la nomme (2005) à la suite de F. Rastier (2001).
- 3 Après avoir exposé les enjeux liés à la construction de vastes ensembles de données textuelles d'intérêt historique de premier ordre pour l'analyse de discours et les sciences du texte, nous présenterons dans un second temps les principes méthodologiques présidant à l'exploration du discours du journal républicain Le Petit Comtois (1883-1903) menée dans notre thèse.

## **CONSTRUIRE DE GRANDS ENSEMBLES DE DONNEES TEXTUELLES DE PRESSE REGIONALE DE LA TROISIEME REPUBLIQUE**

### **Numérique et matérialités discursives**

- 4 Si, depuis la seconde moitié du XXe siècle, les sciences humaines portent une attention de plus en plus accrue au discours, celles-ci négligent bien souvent la dimension matérielle du discours, objet caractérisé par une double articulation sociale et textuelle. D'où un dé

ficit méthodologique important, ayant pour corolaire un manque de rigueur du processus herméneutique. L'ère du numérique apporte néanmoins des pistes prometteuses pour de nouvelles pratiques du geste de configuration des données textuelles, d'accès à la textualité et de contrôle du processus interprétatif.

- 5 En mettant à mal la représentation d'un texte existant en lui-même, hors de toute matérialité, le numérique favorise en premier lieu à une réflexion critique sur l'établissement numérique du texte, à entendre comme un garde fou crucial contre l'oubli de l'historicité du texte, la perte d'informations significatives. Préserver l'intégralité des propriétés significatives et constituantes du texte tel qu'inscrit sur son document d'origine, mais également enrichir le texte de ses lectures et renforcer sa valeur scientifique : telles sont les exigences scientifiques que se fixe le programme d'une nouvelle philologie numérique, visant à satisfaire un besoin urgent et crucial d'accès aux matérialités discursives, capitalisées dans des bases textuelles.
- 6 Parallèlement, il s'agit d'interroger les modes sophistiqués de lecture du texte tel que le permet le passage du support papier au support numérique. Ce changement de régime matériel, se caractérisant par diverses possibilités de « traitement de texte », de codage, de balisage, met à mal le « leurre de l'évidence naturelle du texte » (Adam, 2005b : 70) entretenu par le mode imprimé et nous permet d'accéder aux différents états du texte qui ne sauraient se réduire à sa seule surface graphique, pure convention. Plus, le numérique substitue à une lecture exclusivement linéaire et syntagmatique du texte, une analyse paradigmatique, tabulaire et réticulaire pour reprendre les mots de J.-M. Viprey (2005). Cette ouverture aux dimensions textuelles autres que séquentielles initie donc une approche renouvelée et privilégiée de la textualité. Le numérique bouleverse dans le même temps nos pratiques herméneutiques. Le numérique permet en effet de dépasser le carcan d'une lecture traditionnelle et intuitive des textes, dominée par la subjectivité et des jugements de savoir a priori étayés par des arguments d'autorité ; il autorise désormais une pratique de lecture contrôlée et objectivée, un traitement exhaustif et systématique des matériaux linguistiques, une approche herméneutique renouvelée sur la base d'une approche différentielle des variations du texte observées en corpus.

## **Le projet « Base de presse comtoise de la Troisième République »**

- 7 Initié en 2004 par le Centre Jacques Petit (EA EA3187) et le Laseldi (EA 02281), réunis au sein du pôle « Archives, Bases, Corpus » de la MSHE Claude-Nicolas Ledoux, le projet « Base de presse comtoise de la Troisième République » est un chantier illustrant les voies de la nouvelle philologie numérique. En effet, il s'agit d'une entreprise d'acquisition planifiée d'un vaste ensemble de données textuelles finement établies, portatives et pérennes, visant l'archivage et l'exploration d'une source figurant un creuset interdisciplinaire de premier ordre : la presse régionale de la Troisième République.
- 8 Miroir nécessairement déformant (Charaudeau, 1997) d'une société française en pleine mutation de (industrialisation, urbanisation, etc.), la presse régionale constitue en effet une source éminemment précieuse pour une meilleure connaissance d'une période fondatrice, marquée par la mise en place d'un régime parlementaire et libéral après trois grands épisodes révolutionnaires. Centrée sur la vie locale et associative, cette presse offre un accès original aux opinions individuels, la dimension réflexive des acteurs d'une société s'exerçant dans les pratiques les plus ordinaires et les plus immédiates, ainsi que dans les types de discours les plus communs. Ses colonnes densément remplies de dépêches, de chroniques, de faits-divers, offrent un éclairage primordial sur les événements politiques, sociaux et culturels, tout en offrant en palimpseste un portrait des rédacteurs et les lecteurs de cette presse. Source permettant d'observer, au jour le jour, l'évolution des discours socio-politiques antagonistes, se combattant par et pour la langue, la presse régionale s'impose enfin comme un terrain privilégié d'observation des sens et des rôles que les différentes forces en présence attribuent aux « mots ».
- 9 Parmi la vingtaine de titres de presse<sup>1</sup> destinés à constituer les fonds de la base, Le Petit Comtois (1883-1944) a été choisi pour débiter des opérations d'archivage numérique et d'analyse, en raison de son positionnement républicain et de sa publication interrompue de 1883 à 1944, qui est exceptionnelle si l'on considère que les deux tiers des quotidiens de la Troisième République ont eu une longévité inférieure à une année (Bellanger et al., 1972 : 395). L'histoire de ce quotidien,

puissant vecteur d'influence et d'éducation politique dans le Doubs de la Troisième République, débute en 1881 lorsque Jules Gros (1838-1919), avocat bisontin et Vénérable de la Loge maçonnique, futur sous-préfet et député du Doubs, décide, dans la perspective proche des élections de l'été 1881, de créer un journal visant à assurer l'élection d'A. Rambaud (1842-1905), normalien, chef de cabinet au ministère de l'Instruction publique de Jules Ferry. Le projet n'aboutira qu'en 1883, lorsque Jules Gros se sera assuré du soutien d'un groupe de notables bisontins, financiers, hommes de plume, de sciences et d'art. On citera parmi ses actionnaires le cousin de Jules Gros, Victor Delavelle, notaire et maire de Besançon, Alfred Rambaud, professeur à la Sorbonne. S'adressant au « petit peuple » de la Franche-Comté, paysans et ouvriers horlogers, à ces « nouvelles couches » chères à Gambetta, le Petit Comtois, « journal démocratique quotidien » rencontre rapidement un succès qui va perdurer jusqu'à la fin de sa parution. C'est dans ses colonnes que se façonne le paysage politique régional, par ses engagements dans chacune des élections municipales, cantonales ou législatives. C'est au fil de ses articles quotidiens, d'âpres revendications sociales et de défense farouche des valeurs républicaines contre les fureurs réactionnaires, que se créent et s'alimentent les polémiques locales. Revendiquant en 1883 l'héritage politique de Gambetta, la feuille évolue vers un radicalisme orthodoxe dès 1891. Cette ligne éditoriale radicale du quotidien perdurera jusqu'à l'arrêt de sa publication suite à un ordre de la Kommandantur le 22 mai 1944, condamnant de transparentes allusions antigermanistes.

## Un chantier stratégique et prototypique

- 10 Pour éclairer les enjeux de nos recherches en matière d'acquisition de grands ensembles de données textuelles, il convient d'insister sur la dimension prototypique de la Base « Presse comtoise », qui comptabilisera à terme ne pratique plus de 200 000 millions d'occurrences.
- 11 Mentionnons tout d'abord que de tels projets planifiés d'acquisition et de configuration de données textuelles sont actuellement peu répandus, à l'exception de quelques réalisations inaugurales, à l'instar du projet mené sous la direction de L. Frobert (ENS- UMR Triangle) sur la presse ouvrière lyonnaise du XIXe siècle, autour du titre phare de

l'Echo de la Fabrique<sup>2</sup>. Ce constat doit être éclairé à la lumière des obstacles conceptuels et techniques pesant sur de telles entreprises, s'alimentant mutuellement : les apports heuristiques de la lecture de grands ensembles de données textuelles, comparés aux moyens putatifs (humains et financiers) impliqués, sont encore très largement sous-estimés. Dans l'univers de la conservation, les campagnes d'archivage de la presse régionale du XIXe siècle, menées depuis 2007 par les bibliothèques municipales sous la tutelle de la BNF, se limitent d'ailleurs à un archivage en mode image. Au contraire de ces campagnes de numérisation motivées par des impératifs de conservation et de diffusion, notre projet vise à renforcer la valeur scientifique des données textuelles, par une numérisation en mode plein-texte, jusqu'à l'annotation fine. Corollaire de cette première exigence, il s'agit de construire des données portatives et pérennes, condition sine qua non pour la mise en œuvre d'études d'envergure et pour la confrontation des méthodes et des points de vue, entendues dans notre perspective comme autant d'exigences scientifiques. La portabilité des données s'impose de surcroît comme un impérieux besoin dans le domaine de l'analyse statistique textuelle. Afin de permettre une chaîne de traitement collaboratif et de pouvoir soumettre un même état de corpus aux fonctionnalités respectives des divers logiciels d'analyse de données textuelles, il convient en effet de disposer d'un format d'échange commun permettant de conserver les différents niveaux d'annotation du corpus. C'est tout le sens des recherches menées sur l'interopérabilité entre logiciels d'analyse textuelles dans le cadre du Réseau pour l'échange de ressources et de méthodologies en analyse de texte assistée par ordinateur ATONET, dans le cadre duquel les données de la base « presse comtoise » constituent un corpus test.

- 12 Aussi, si le projet de base de la presse comtoise a pour ambition de livrer une base textuelle de fort intérêt historique en mode texte étiqueté linguistiquement, il s'agit dans le même temps de contribuer au programme de la nouvelle archive en levant les obstacles méthodologiques et techniques pesant sur l'acquisition de grands volumes de données de qualité scientifique répondant aux normes scientifiques contemporaines.

## Description du processus de numérisation

- 13 Nous dresserons ici un rapide bilan du laboratoire d'expérimentations figurés par nos travaux relatifs à la constitution de base de presse en mode plein texte.
- 14 Le processus de numérisation a classiquement débuté par une étape d'acquisition des données en mode image respectant un cahier des charges imposant, d'une part, un traitement non-destructif des supports-papiers, et d'autre part, une exigence de résolution optimale. La qualité de résolution conditionne en effet les résultats de l'étape d'océrisation (Optical Character Recognition) lors de laquelle les signes d'un fichier image vont être transformés en mode texte sur la base du principe de reconnaissance optique des caractères.
- 15 Si l'étape d'acquisition des données image de la totalité de la collection du Petit Comtois (1883-1944) n'a posé aucune difficulté majeure, l'étape d'océrisation s'est avérée problématique. Les performances des logiciels d'océrisation actuellement disponibles sur le marché, destinés au traitement de documents récents, sont entravées par l'état médiocre du support papier initial. Un ensemble convergent de facteurs sont responsables de l'état médiocre du support-papier : la mauvaise qualité du papier utilisé, les techniques de typographie et d'impression usitées au XIXe siècle, ainsi que diverses dégradations (pliures, taches, déchirures, etc.), conséquences d'une consultation intensive et des solutions de conservation adoptés.
- 16 En dépit d'améliorations notables que nous avons opérées (1) lors d'une phase de restauration (binarisation, atténuation des courbures, etc.) du document image, destinée à atténuer les défauts du support papier, et (2) en complétant les ressources du logiciel d'océrisation par la construction de gabarits spécifiquement adaptés à la collection du Petit Comtois, nous avons observé que le taux de reconnaissance des caractères demeurait faible<sup>3</sup>. De surcroît, l'ampleur du projet « Presse Comtoise », en termes de volumes de données à traiter, est difficilement compatible avec les conditions d'ergonomie des solutions logicielles d'océrisation, conçus pour des opérations de petite envergure. En effet, le mode de correction implémenté dans ces logiciels, opéré sur la base d'un mode de lecture linéaire au moyen



d'un vérificateur orthographique et grammatical dérivé de ceux des traitements de texte, implique des opérations chronophages, fastidieuses et coûteuses ; le tout, pour une fiabilité médiocre. Nous avons estimé lors de la première campagne de numérisation du Petit Comtois que le processus de vérification des sorties OCR brut et de leur correction contrôlée, opéré à l'aide de tels outils, représentait près de 120h<sup>4</sup> par million de mots. Autant de difficultés majeures pour un projet soumis à de fortes contraintes temporelles et financières, incompatibles avec les besoins de données textuelles finement établies de l'analyse de discours ne pouvant pourtant se satisfaire, sur le plan scientifique, d'une correction automatique ayant pour corollaire une corruption des données. Aussi, sont actuellement développées au sein du pôle Archive, Base, Corpus, une interface ergonomique de correction semi-automatique des données textuelles.

- 17 Concernant l'encodage des données, nous avons opté, en collaboration avec le concepteur du logiciel Sato<sup>5</sup>, François Daoust (UQAM, Québec), pour un format pivot (Daoust & Marcoux, 2006) basé sur le XML et conforme aux recommandations TEI<sup>6</sup>. Outre le balisage instanciel, l'établissement philologique de la base, présupposant un travail collaboratif entre les différents utilisateurs de la base, reste aujourd'hui un chantier en cours.

## **ANALYSER LE DISCOURS DU PETIT COMTOIS**

### **Notre pratique d'exploration textuelle du discours du Petit Comtois**

- 18 Du premier état de la base « Presse comtoise », nous avons extrait un corpus de 630<sup>7</sup> numéros du Petit Comtois, envisagé comme un dispositif d'observation du discours du Petit Comtois dans un contexte socio-politique des plus intéressants : celui, d'une part, des « débuts » de la Troisième République et d'autre part, d'une période charnière de professionnalisation de la pratique journalistique. Regroupant plus de 26 000 articles publiés de 1883 à 1903, de forme générique et de taille extrêmement hétérogènes, ce corpus vertigineux comptabilise plus

de 5,5 millions d'occurrences, à partir desquelles nous éprouvons les méthodes de l'analyse statistique textuelle.

- 19 Dans notre thèse, nous avons qualifié notre pratique d'analyse de données textuelles assistée par informatique d'exploration textuelle du discours<sup>8</sup>, en référence à l'analyse textuelle du discours (Adam, 2005) pour marquer notre objet, le discours, ainsi que pour mettre en relief un certain parcours d'interprétation, ayant pour pierre de touche le co(n)texte. Notre pratique de lecture se construit ainsi dans un mouvement allant nécessairement du corpus au hors-corpus, au cours de laquelle le primat est donné à l'environnement linguistique immédiat, au cotexte comme réseau de textes « autres » mis en série dans le corpus.
- 20 Par le syntagme « d'exploration textuelle du discours », nous voulons enfin insister sur un certain parcours de lecture au sein de grands ensembles textuels, ayant pour objectif de recueillir des informations objectivées par le texte lui-même. Pour ce faire, notre démarche convoque les outils de la lexicométrie, « bras armé de l'analyse de discours » (Mayaffre, 2008), qui permettent de faire monter du corpus lui-même ses reliefs linguistiques, textuels, discursifs. Dans le cadre d'une démarche hypothético-déductive, l'analyse assistée par informatique convoquée pour confirmer ou infirmer des hypothèses pré-établies se révèle un puissant outil probatoire. Dans notre perspective, c'est néanmoins lorsque les méthodes de la statistique textuelle sont sollicitées pour leur pouvoir suggestif, lorsque leurs indications ne sont pas entendues comme des arguments probatoires mais comme des pistes en vue de la poursuite de l'analyse, que ces méthodes acquièrent leur pleine dimension. Science de l'écart par excellence, la statistique sert en effet une lecture des lieux de variations du texte et constitue de fait un outil profondément heuristique, comme le rappelle Jean-Pierre Massonie (1990) :

[...] la statistique devient donc l'art et la manière de poser des questions nouvelles à l'endroit qu'il convient dans le texte...Non de résoudre des problèmes, mais de les faire naître par une lecture nouvelle, en indiquant où on doit relire (Massonie, 1990 : 102).

## De l'analyse de la ventilation du vocabulaire en diachronie à l'analyse des stratégies de mise en scène de l'information dans le Petit Comtois

- 21 Pour illustrer notre démarche d'analyse à visée exploratoire et le potentiel heuristique des méthodes lexicométriques, nous nous limiterons dans le cadre de cette contribution à montrer comment des prises de vue sur le vocabulaire nous ont orienté vers une étude des stratégies de mise en scène de l'information du Petit Comtois dans notre corpus.
- 22 Dans le cadre d'une étude de l'évolution du vocabulaire en diachronie, nous avons soumis les 1000 formes les plus fréquentes du corpus aux procédures de l'Analyse Factorielle des Correspondances (Lebart & Salem, 1994). Sur la représentation AFC en mode géodésique<sup>9</sup> issue de cette procédure, on observe une opposition sur le facteur 1 des années 1883-1896 d'une part, 1897-1903 de l'autre (cf. figure 1 en annexe). Cette visualisation met également en exergue la très singulière position de l'année 1883, neutre sur l'axe 1, saillante à la fois sur l'axe 2 et sur l'axe 3.
- 23 En nous interrogeant sur les facteurs présidant au profil singulier de l'année 1883, nous avons notamment observé que l'item « court » figurait un item central du vocabulaire propre de cette année, puisque la forme réalise son seul écart-réduit significativement positif en 1883, comme on peut le constater sur l'histogramme reproduit en annexe (figure 2). Un retour au texte pointe que ce sur-emploi doit être rapporté à l'utilisation massive de la séquence « le bruit court », visant à introduire un discours rapporté dont la prise en charge est nulle, ainsi que dans l'exemple ci-dessous :

Chronique régionale, Jura, Dole, (21-08-1883)

Un empoisonnement par les champignons vient de causer une pénible émotion dans la ville de Dole. [...] Le bruit court en ville que deux autres des victimes auraient succombé aujourd'hui. Le temps me manque pour vérifier cette nouvelle. Ce qui est malheureusement

vrai, c'est que M. le docteur Gremaud a déclaré que deux des malades sont dans un état à peu près désespéré.

- 24 La prégnance de la séquence « le bruit court que X » peut dans un premier temps être entendue comme l'indice d'un réseau mal assuré de correspondants encore peu roués à la temporalité introduite par une publication quotidienne, ainsi que de la persévérance de formes de communication traditionnelle dans la construction du discours journalistique tenu par des non-professionnels. On peut ainsi légitimement parler d'une information mise en scène sur le mode de la rumeur, conformément à la définition que le Trésor de la Langue Française propose de celle-ci, à savoir une « nouvelle sans certitude qui se répand de bouche-à-oreille, bruit inquiétant qui court », d'ailleurs explicitée dans l'exemple précédent.
- 25 La fréquentation de notre corpus nous invite à dépasser cette première définition, et à insister sur la dimension idéologique de la rumeur et les stratégies socio-discursives qu'elle sert.
- 26 Mettre en scène une information comme rumeur, c'est avant tout disqualifier le discours de l'autre, comme le formule le sociologue E. Taïeb :

[...] la « rumeur » c'est aussi l'idéologie de l'autre, que les médias entendent démasquer comme telle et ne pas transformer en information légitime. Qualifier une position politique de « rumeur », c'est à la fois ne jamais lui permettre d'accéder au statut d'information, donc d'être crédible, et marginaliser la position qui a tenté de faire passer une telle information (Taïeb, 2006 : 137).

- 27 Considérons ainsi la dépêche suivante.

Calomnies de la presse officieuse (1883-08-19)

La presse ministérielle répand le bruit qu'il existe une société, au capital de 750,000 francs, ayant pour but de provoquer des pronunciamientos.

Cette société se serait acquis le concours de M. Zorilla, moyennant finances.

Il va sans dire que tout est absolument faux dans ces racontars.

- 28 Le locuteur-citant disqualifie ici explicitement les informations communiquées par « la presse ministérielle », en recourant à l'expression répandre le bruit que, assortie d'un conditionnel et d'un commentaire des plus catégoriques où le discours du locuteur-cité est relégué au statut de « racontars », substantif axiologique des plus négatifs. L'information n'est pas posée comme « bruit » parce qu'elle n'est pas vérifiée, mais parce que le quotidien la juge formellement erronée. Loin de marquer une distance avec une information incertaine, la séquence est le moyen d'une dénonciation des visées politiciennes de la source diffusant la nouvelle. Ce refus de l'information de l'autre est explicité par le cotexte de la dépêche, puisque le bulletin surmontant la rubrique des Dépêches de Nuit, dont est extraite la brève ci-dessus, accuse le gouvernement espagnol de mener une opération de désinformation, visant à dissimuler la nature et l'ampleur du mouvement insurrectionnel en cours.
- 29 Ces indications nous invitent par suite à examiner dans notre corpus les différents modes de mise en scène de l'information dans le Petit Comtois et à vérifier l'hypothèse d'une écriture journalistique, qui dans un contexte de professionnalisation de la pratique, efface progressivement les marques évidentes de sa subjectivité pour prendre la forme d'un discours « objectivisé » plus qu'objectif. L'intérêt socio-historique de cette classe d'analyse est d'autant plus grand que la diachronie de notre corpus recouvre une période charnière entre un journalisme dit « d'opinion » et un journalisme dit « d'information ».

## **D'une vue sur la cooccurrence généralisée à une analyse réthorico-stylistique du Petit Comtois**

- 30 Nous voudrions enfin présenter le potentiel heuristique d'une analyse des réseaux d'attraction et de répulsion qu'entretiennent les « mots » de notre corpus. Rappelons qu'isolé, un mot ne fait pas sens. Le sens d'une forme ne se construit qu'en co(n)texte, activé par le réseau des mots qui l'entourent, ses cooccurrents. L'étude de la co-occurrence, à l'aune du rapport fréquentiel entre deux items au sein d'une même

fenêtre contextuelle, constitue ainsi un premier palier d'accès à la sémantique. Pour étudier l'ensemble des phénomènes de co-occurrences tissées par les formes graphiques de notre corpus, nous avons construit une vaste matrice « carrée » dont les lignes et les colonnes ont des descripteurs et des contenus identiques, avec à leur intersection, le nombre de co-occurrences ligne/colonne dans un empan co(n)textuel paramétré. L'analyse factorielle des correspondances appliquée à cette matrice des 1000 vocables les plus fréquents<sup>10</sup> du corpus « Petit Comtois », offre une synthèse de la micro-distribution lexicale (figure 3).

- 31 L'analyse permet de repérer simultanément les parentés de profil co-occurentiel sous forme de vastes groupes isotropiques, représentant des classes de forte équivalence distributionnelle au sens de Z.S. Harris (1969). Celles-ci sont constituées sous l'effet conjoint d'un ensemble de contraintes linguistiques et discursives. Pour le corpus considéré, nous pouvons en l'occurrence observer quatre zones isotropiques, s'opposant paire à paire sur le facteur 1, puis se structurant plus finement sur les facteurs 2 et 3. Pour prendre connaissance des caractéristiques lexicales de ces pôles isotropiques, l'environnement Astartex permet d'obtenir des zooms centrés, fonctionnant comme des outils de navigation cliquables et orientés favorisant un mouvement d'aller-retour entre les différents états du texte.
- 32 Ainsi, si l'on zoom centré de l'isotropie observable dans le quart-ouest nord de la moitié droite du planisphère (figure 4), on observe que celle-ci renvoie au vocabulaire politique du Petit Comtois, comme en témoignent les formes renvoyant aux grands ensembles sociétaux (europe, monde) couplées avec un faisceau de désignant politiques (socialistes, réactionnaires, cléricaux, républicains) entrant en réseaux avec un ensemble de concepts tels que patrie, démocratie. Une forte dimension réthorico-stylistique se dégage ce pôle, articulée autour d'un ensemble de verbes conjugués ou à l'infinitif (crois, voulons, pense, veut, défendre) et d'adverbes (partout, toujours).
- 33 Ce sont sur les pistes d'analyse suggérées par la similarité ou la dissimilarité des profils collocatifs, appelées à être construites en objets de recherche contextualisés par l'analyste du discours et l'historien que nous voudrions insister. Le pôle isotropique présenté nous invite ainsi à questionner la rhétorique du discours politique du Petit Com-

tois. Une première entrée de cet axe de recherche pourrait être l'item vérité. L'inscription de ce celui-ci dans l'isotropie du vocabulaire politique renvoie à un trait très caractéristique de l'ethos mis en scène tant dans les professions de foi adressées aux électeurs que dans les articles d'opinion émanant de la rédaction, qui se targuent de dispenser la vérité. Cette dernière est d'ailleurs posée comme un devoir, terme qui est le plus fort co-occurent de la forme. D'un côté le martèlement de « la vérité » au gré de syntagmes péremptives et des mises en relief par des phrases présentatives. De l'autre la dénonciation incessante d'un discours mensonger visant à égarer l'opinion. Il s'agit, pour les républicains du Petit Comtois de rétablir (premier verbe co-occurent de la forme) une vérité dénaturée et de lever le voile sur le jeu des ennemis de la République. Pour ce faire, le discours politique républicain s'orne des traits rhétoriques de l'évidence, du naturel et du bon sens, ayant pour corollaire la dénonciation des faits de spoliation lexicale (Angenot, 1982). Le Petit Comtois n'a alors de cesse d'imposer le sens authentique des mots (le sien) et de « faire la loi » linguistique (Bonnafous & Tournier, 1995).

- 34 Entre autres pistes à approfondir, l'examen de l'axe de la vérité et de l'authenticité nous invite à prolonger ces remarques par une étude des désignants socio-politiques tels qu'ils circulent dans notre corpus. Prendre pour objet d'étude ces désignants, c'est s'intéresser aux formes par lesquelles la rédaction du Petit Comtois définit son républicanisme et combat les forces politiques en présence. Or, s'il est un mot dont Le Petit Comtois ne cesse en effet de dénoncer le détournement et de contester le sens, c'est bien celui de républicains, comme le suggèrent les principaux co-occurents de l'item sur l'ensemble du corpus, reflétant une conception dichotomique du terme : il est ainsi des vrais républicains (écart réduit<sup>11</sup> : +49,8) et des faux républicains (écart réduit : +10,93). Un retour au texte permet de préciser les contours de cette opposition et montre que les contextes d'emploi de ces désignants axiologisés se concentrent exclusivement sur la seconde moitié de notre corpus : la dialectique des vrais et des faux républicains ne s'impose qu'à partir de 1893, date à laquelle la ligne politique du Petit Comtois se radicalise, et où vrais devient le premier co-occurent de la forme (écart-réduit de 52,91). L'intensification de la qualification positive du républicanisme trouve, de façon générale, son origine dans la menace de l'identité républicaine par la

politique d'apaisement menée dès 1892. Le désignant vrais républicains, réservé aux anticléricaux, résonne dans ce contexte, comme une dénonciation du ralliement des catholiques à la République suite à l'encyclique papale du 20 février 1892 qui les y invite. Aux vrais républicains, sont opposés les républicains à faux nez, syntagme nominal par lequel le radical Petit Comtois réfère aux catholiques ralliés.

- 35 Autant de pistes de recherches suggérées par l'analyse, objectivées par le texte lui-même, qui montrent la valeur heuristique d'une telle méthodologie d'analyse.

## Conclusion

- 36 En guise de conclusion, nous voudrions dresser un bref bilan des deux axes de recherches autour desquels s'est articulée notre thèse de doctorat.

- 37 Nous projetions tout d'abord de contribuer à un programme de dynamisation des lectures de la presse quotidienne régionale de la Troisième République (PQR3R)<sup>12</sup> à partir des voies ouvertes par le numérique. Dans ce dessein, nous avons mis au banc d'essai un protocole de numérisation optimisant les résultats de l'océrisation et rendant compatible l'acquisition de ces données finement établies et portatives avec les conditions actuelles de la recherche. Ce protocole a constitué un essai avancé sur le plan méthodologique, qui a démontré la faisabilité d'un tel projet puisqu'à ce jour, c'est une base de 6 millions d'occurrences qui est disponible. On pourra d'ailleurs consulter en ligne un premier état de la base textuelle sur le site du Petit Comtois<sup>13</sup>.

- 38 Notre recherche se fixait ensuite pour objectif de fournir un éclairage sur le discours du Petit Comtois, analysé au travers d'un dispositif sériel d'observation de quelque 26 000 articles dont la nature et la taille étaient fort hétérogènes, publiés durant la période 1883-1903. Dans ce cadre, nous avons mis en œuvre un parcours d'analyse à visée exploratoire fondé sur l'examen de différentes prises de vues sur la matérialité du discours du Petit Comtois. Ce sont plus particulièrement sur les méthodologies de la co-occurrence généralisée présentées ci-dessus que nous voudrions insister. Articulées aux outils désormais classiques en matière d'orientation dans de vastes corpus tels que les



index de formes graphiques, les concordanciers, l'analyse factorielle des correspondances appliquée à la macro-distribution du vocabulaire, les méthodologies d'exploration de la dimension réticulaire du vocabulaire permettent de mettre en relief, dans la norme endogène du corpus, les concepts clés des discours et débats politiques les plus importants de la société française de la Troisième République tels que relayés par la presse régionale de la Troisième République. Il s'agit ainsi d'une contribution précieuse pour le développement d'une histoire linguistique des concepts telle qu'appelée par J. Guilhaumou (2000) dans le cadre d'un protocole coopératif entre les communautés des sciences du texte et l'analyse de discours.

---

ADAM, J.-M., 2005, La linguistique textuelle. Introduction à l'analyse textuelle des discours. Paris : Armand Colin.

ADAM, J.-M. & HEIDMANN, R. (Dirs.), 2005, Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité. Genève : Slatkine érudition.

ANGENOT, M., 1982, La parole pamphlétaire : contribution à la typologie des discours modernes. Paris : Payot.

BELLANGER, C., GODECHOT, J. & GUIRAL, P. (Dirs.), 1972, Histoire générale de la Presse française. Tome 3 : "De 1871 à 1940". Paris : Presses Universitaires de France.

CHARAUDEAU, P., 1997, Le discours d'information médiatique : la construction du miroir social, Paris, Nathan.

DAOUST, F., MARCOUX, Y., 2006, « Logiciels d'analyse textuelle : vers un format XML-TEI pour l'échange de corpus annotés », Actes des 8e journées internationales d'Analyse statistique des données textuelles JADT 2006, Besançon, 19-21 avril 2006, Besançon, Presses

Universitaires de Franche-Comté, p. 327-340.

GUILHAUMOU, J., 2000, « De l'histoire des concepts à l'histoire linguistique des usages conceptuels », Genèse, n°38, Paris, Belin, p. 105-118.

GUILHAUMOU, J., MALDIDIER, D., ROBIN, R., Discours et archive. Expérimentations en analyse de Discours, Liège, Mardaga, 1994.

HARRIS, Z. (1969) "Analyse du discours." in Langages, n°61. Paris : Larousse. pp. 8-45.

LEBART, L. & SALEM, A. (1994) Statistique textuelle. Paris : Dunod.

LETHIER, V. (2010) "Multidimensional Analysis based on morpho-lexical features : The example of a 19th Century regional Press Corpus along with its columns", in Belgian Journal of Linguistics, vol 23, n°1. Anvers : John Benjamins. pp. 175-190.

MASSONIE, J.-P. (1990) Analyse informatisée des textes. Besançon : Annales littéraires de l'Université de Besançon.

- MAYAFFRE, D. (2008) "L'entrelacement lexical des textes, co-occurrences et lexicométrie." *Texte et corpus*, n°3 / août 2008. pp.91-102. Disponible sur <[http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES\\_JLC07\\_mayaffre.pdf](http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_mayaffre.pdf)> (Consulté le 12/10/2009).
- PROST, A. (1974) *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*. Paris : Presses Universitaires de France.
- RASTIER, F., 2001, *Arts et sciences du texte*. Paris : Seuil.
- TOURNIER, M., BONNAFOUS, S., « Analyse du discours, lexicométrie, communication et politique », *Langages*, n°117, Paris, Larousse, 1995, p. 67-81.
- TAÏEB, E. (2006) "La 'rumeur' des journalistes." in *Diogène*, n° 213: "Rumeurs et légendes urbaines". Paris : PUF. pp. 133-152.
- VIPREY, J.-M. ,2005, « Philologie numérique et herméneutique intégrative », Adam, J.-M, Heidmann, U., (dir.) *Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité*, Genève, Slatkine, p. 51-58.
- VIPREY, J.-M. « Ergonomiser la visualisation AFC dans un environnement d'exploration textuelle : une projection 'géodésique' », *Actes des 8e journées internationales d'Analyse statistique des données textuelles JADT 2006*, Besançon, 19-21 avril 2006, Besançon, Presses Universitaires de Franche-Comté, p. 981-992.
- VIPREY, J.-M. & LETHIER, V. (2008a) "Annotation linguistique de corpus : vers l'exhaustivité par la convivialité" in HEIDEN, S & PINCEMIN, B. (éds.), *Actes des 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*. Lyon : Presses Universitaires de Lyon. pp.1151-1161.
-

# Annexe

Figure 1: Projection AFC de la ventilation du vocabulaire du Petit Comtois en diachronie

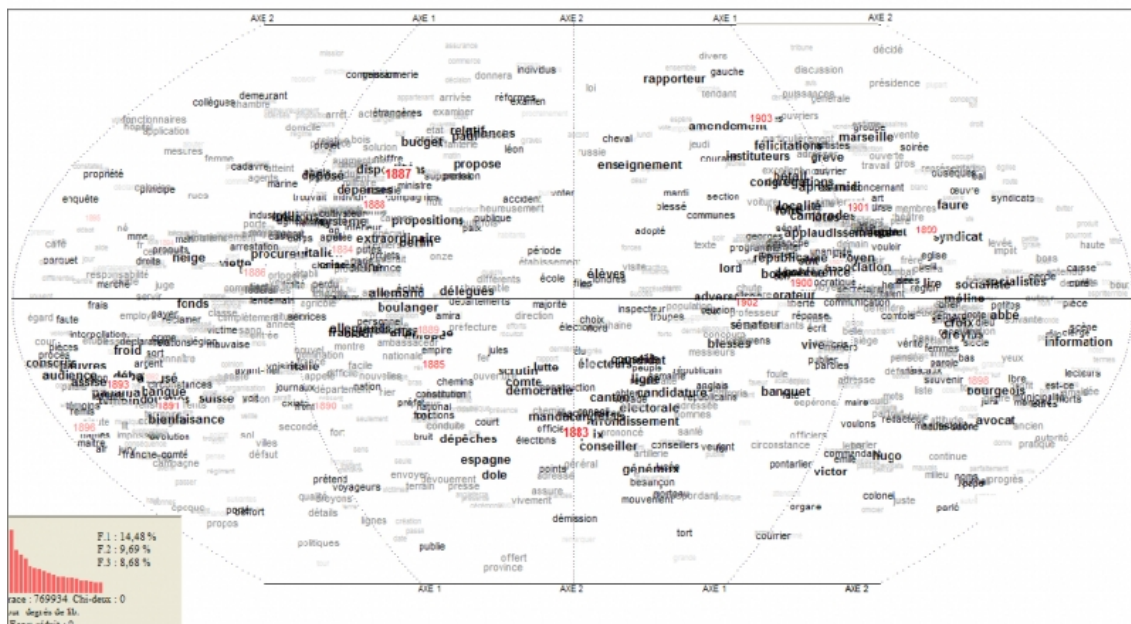


Figure 2 : Histogramme de la distribution de la forme court (en écarts-réduits).

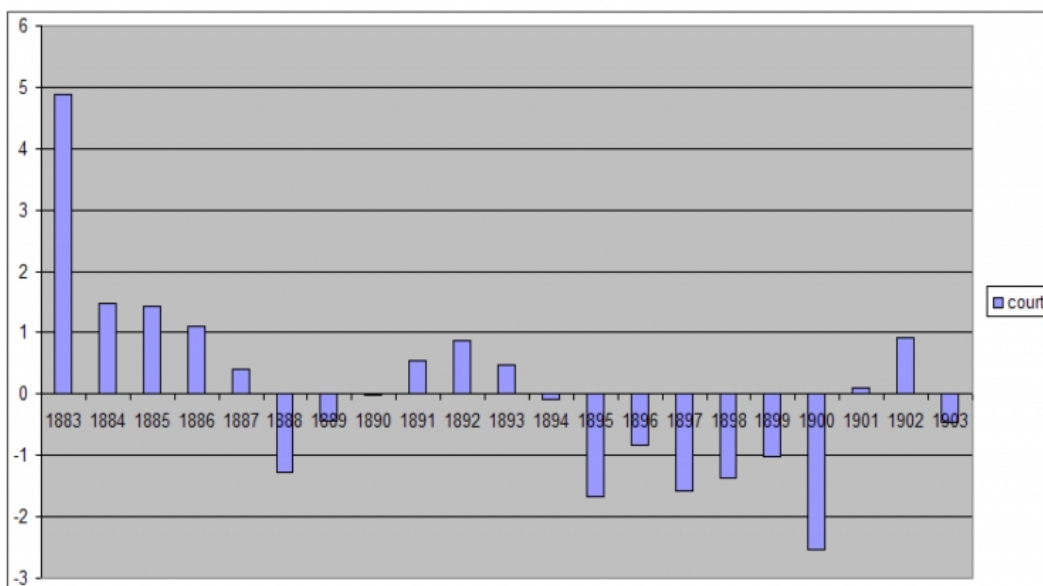
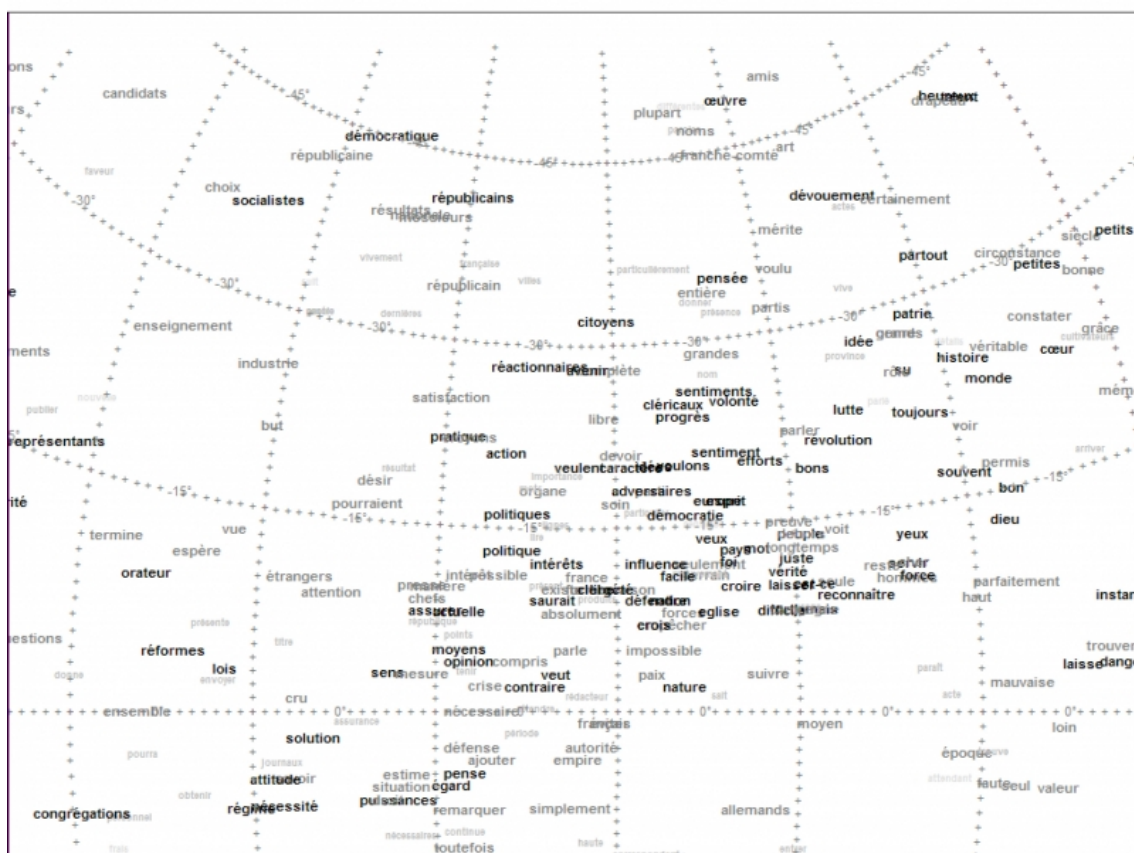




Figure 4: Zoom centré sur l'isotropie du vocabulaire politique du Petit Comtois.



1 Parmi lesquels : La Dépêche du Doubs et de Franche-Comté (1897-1933), L'Echo de la Franche-Comté (1881-1887).

2 <http://echo-fabrique.ens-lsh.fr>

3 Pour l'ensemble de la première campagne de numérisation (1883-1903), effectuée de 2005 à 2007, nous évaluons en effet le taux de reconnaissance du logiciel utilisé, en l'occurrence ABBYFineReader 8.0 à quelque 65%. Dans sa version 9.0, disponible depuis 2009, le moteur de recherche du même logiciel, appliqué à la collection des supports-papiers dont l'état s'améliore à partir des premières années du XX<sup>e</sup> siècle, s'est révélé beaucoup plus puissant, puisque le taux de reconnaissance avoisine les 75%.

4 Sur la base d'environ une heure de correction par numéro, comptant en moyenne 9000 mots.

5 <http://www.ling.uqam.ca/sato/index.html>

6 En l'occurrence, en raison du volume vertigineux des données à traiter, a été adopté un entête TEI minimal, assurant la documentation des données

et la préservation de leurs propriétés pertinentes.

7 Corpus constitué sur la base d'une procédure d'échantillonnage sur la base d'un mois par an, de sorte à mener une étude diachronique.

8 Syntagme emprunté à J.-M. Viprey et que nous faisons nôtre pleinement dans ces travaux.

9 Se référer à VIPREY (2006) pour une présentation approfondie des principes et enjeux du mode de représentation géodésique.

10 Liste de formes graphiques filtrées par un anti-dictionnaire de formes «exclues » (déterminants et pronoms, conjonctions et prépositions, ad- verbes très fréquents, nombres, verbes auxiliaires).

11 Ecart-réduit par rapport à l'équidistribution.

12 Dans le cadre des appels à projet blanc de l'ANR (édition 2009- Sciences humaines et sociales), un projet PQR3R a été déposé. Bien accueilli par les experts, ce projet n'a néanmoins pas fait l'objet d'un accord de financement.

13 <http://laseldi.univ-fcomte.fr/petit-comtois/archives.php>

---

## Français

Plongeant ses racines dans l'analyse du discours du côté de l'histoire (Guilhaumou, et al., 1974), l'analyse de données textuelles (Lebart & Salem, 1994) et l'analyse textuelle du discours (Adam & Heidmann, 2005), notre recherche entend contribuer à sa mesure au mouvement de constitution raisonnée de très grandes bases textuelles et au développement de méthodologies adaptées aux défis posés par la constitution et l'exploration de ces ensembles. Ce faisant, notre recherche, qui vise à fournir à la communauté de nouvelles archives à dimension historique, s'inscrit dans un programme de renouvellement des analyses de discours et de développement de l'histoire des concepts de la Troisième République.

---

## Virginie Lethier

Docteur en Littérature, Centre Jacques Petit - EA 3187 - UFC