

Les correspondances de Machiavel et Guichardin au prisme de la textométrie.

Enjeux et perspectives

Marc LASSON (Ingénieur d'études) - Corinne MANCHIO (PRCE Université Paris 8) - Hélène MIESSE (MCF Université de Liège)

TEXTOMÉTRIE ET CORPUS ÉPISTOLAIRES DU XVI^e S.

La Florence du début du *Cinquecento*, soumise à des changements de régimes successifs en lien avec les conflits qui secouent la Péninsule italique, est traditionnellement considérée comme un laboratoire où s'élaborent de nouvelles façons de faire et de dire la politique.

De multiples études ont montré l'existence de termes et concepts politiques clés chez différents auteurs florentins de cette époque, mais il manque toutefois une étude systématique, comparatiste et historicisée du lexique politique florentin.

Les corpus épistolaires se composent de documents attestant un usage vivant de la langue, précisément situables dans l'espace et le temps et déclinés en grande quantité. Moyennant une préparation adéquate des textes, ils se prêtent aisément à la formalisation mathématique, ce qui permet de construire un réseau de données éclairantes et de disposer de relevés systématiques fiables.

CRITÈRES D'HOMOGENÉITÉ

- Nature des corpus : lettres abondantes (2214 et 3975 missives)
- Origine géographique/linguistique des scripteurs : Florence
- Inscription dans un moment historique : les guerres d'Italie à l'origine d'une expérience républicaine inédite
- Statut des auteurs, aussi rédacteurs
- Écriture diplomatique

CRITÈRES D'HÉTÉROGÉNÉITÉ

- Fonction et autonomie des auteurs (Machiavel est secrétaire, donc subordonné aux autorités qu'il sert, son écriture est parfois « chorale »)
- Type de lettres (codification des écrits de chancellerie)
- Chronologie (fin d'activité en 1512 vs 1540)

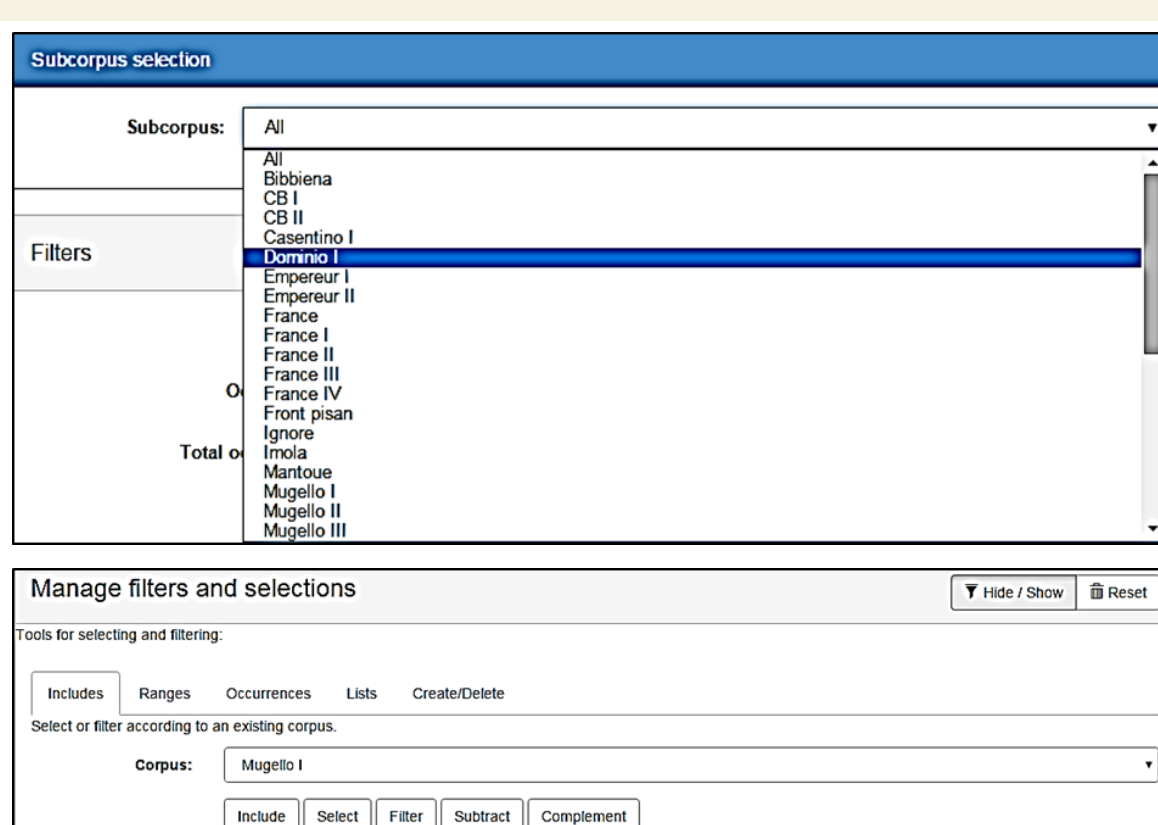
MACHIAVEL ANALYSE TEXTUELLE OUTIL

MACHIAVO est un logiciel créé pour l'étude de la correspondance officielle de Machiavel. Il est commandé par une interface Web qui confère à l'utilisateur un accès facile aux ressources du programme avec son navigateur. Les analyses et les pré-calculs statistiques sont faits par un back-end mis en œuvre avec le langage de programmation Python dont le rôle est d'initialiser la base de données. Le Framework open-source Django sert à générer du contenu HTML. Le langage de programmation JavaScript est sollicité pour l'affichage des données et les résultats côté utilisateur.

DIFFÉRENTS NIVEAUX D'APPREHENSION

- ⇒ Index des lettres
- ⇒ Index des mots, pour le relevé des graphies, des hapax, etc. (25007 entrées chez Machiavel)
- ⇒ Index des familles de mots (6399 chez Machiavel)
- ⇒ Intervalle ou « période » déterminée
- ⇒ Type de lettre

RÉSULTAT : partition en sous-corpus et affinage des filtres d'interrogation



MACHIAVO ET MOCACCINO

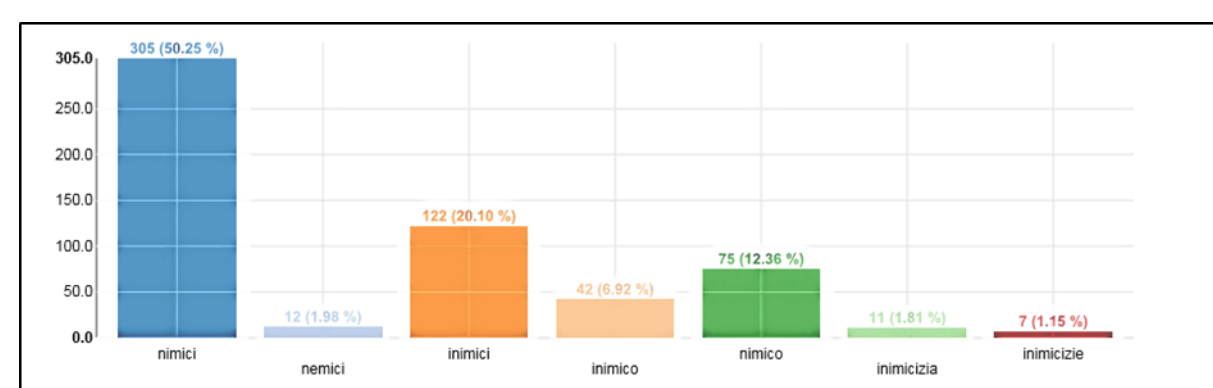
Nos recherches se concentrent sur la confrontation entre MACHIAVEL et GUICHARDIN à partir de sous-corpus numériquement, chronologiquement et fonctionnellement comparables et sur des termes clés (*nemico*, *guerra*, *Francia*) : 220 lettres de légation de Machiavel (1499-1511) et 288 lettres de Guichardin lorsqu'il est ambassadeur puis gouverneur (1511-1513 et 1521-1522).

VISUALISATION DES DONNÉES

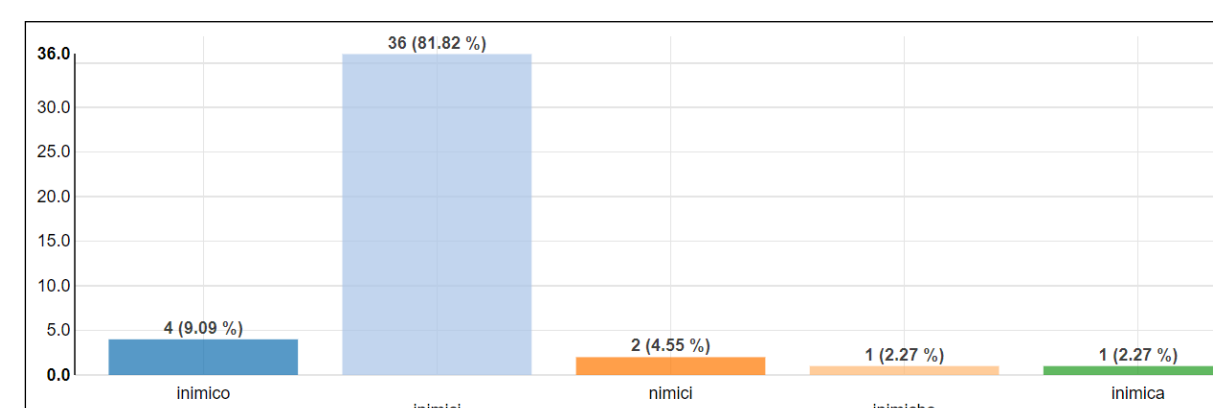
- ⇒ concordanciers (occurrences)
- ⇒ structures arborescentes (collocations)
- ⇒ histogrammes interactifs
- ⇒ histogrammes de distribution

INSTABILITÉ DE LA LANGUE : NEMICO

Chez Machiavel, *nimici* = 50,25% - *inimici* = 20,10%



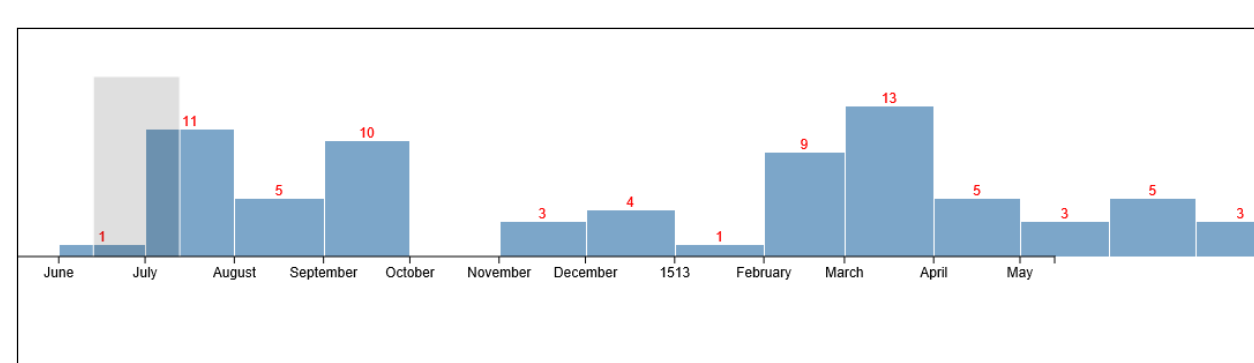
Chez Guichardin, *nimici* = 4,55% - *inimici* = 81,82%



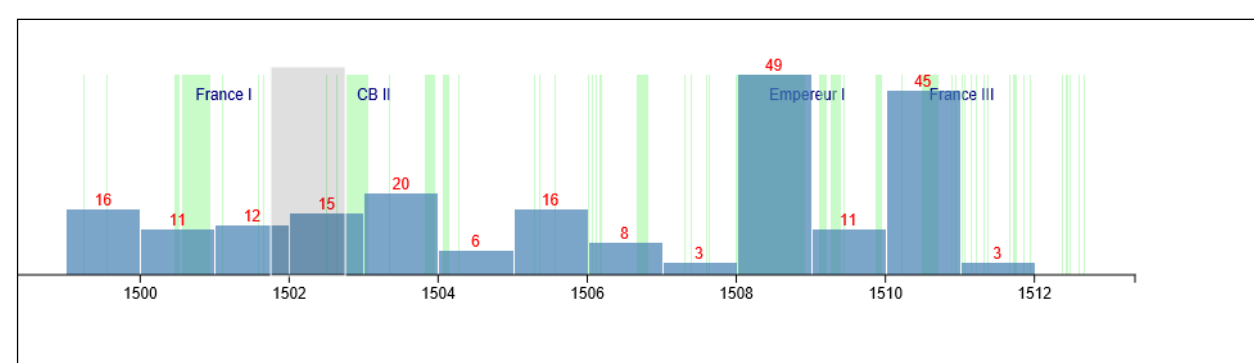
LES TEMPS DES USAGES

L'histogramme semi-inerte avec en abscisse le temps et en ordonnée le nombre de lettres dans lesquelles apparaît le mot interrogé. On constate ainsi comment certains termes sont surreprésentés pendant un temps et disparaissent ensuite (soit au profit d'autres expressions, soit car ils ne sont plus opérants pour décrire la réalité et les enjeux du temps présent). Il est également possible de mettre en relation la longueur ou la concision des lettres avec les données historiques dont nous disposons.

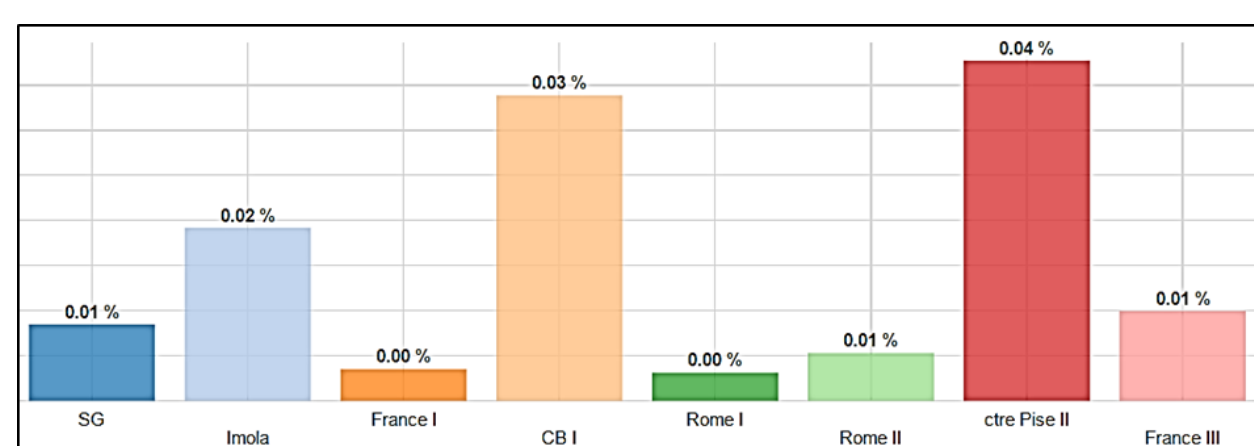
RÉPARTITION DES OCC. DE GUERRA CHEZ GUICHARDIN



RÉPARTITION DES OCC. DE GUERRA CHEZ MACHIAVEL



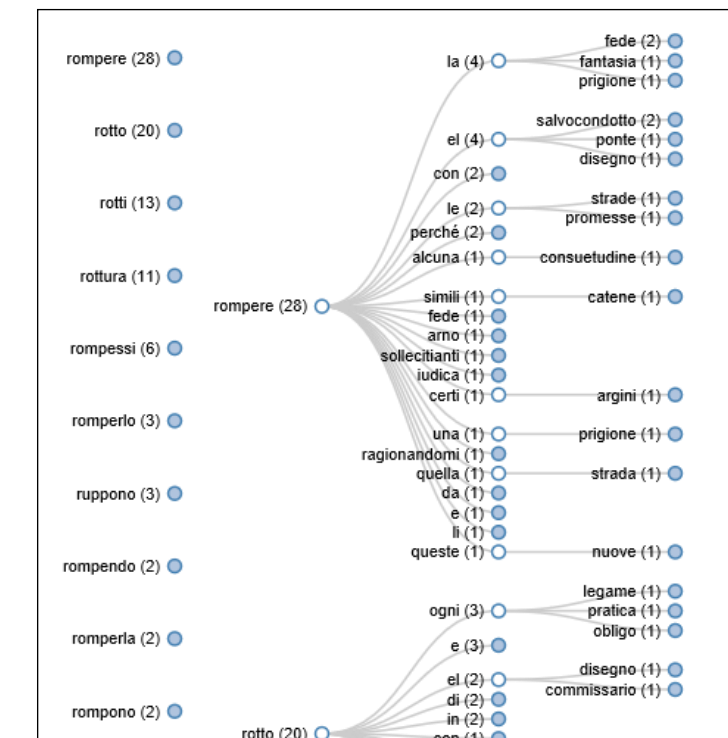
LES LIEUX DES USAGES : LA MALATTIA



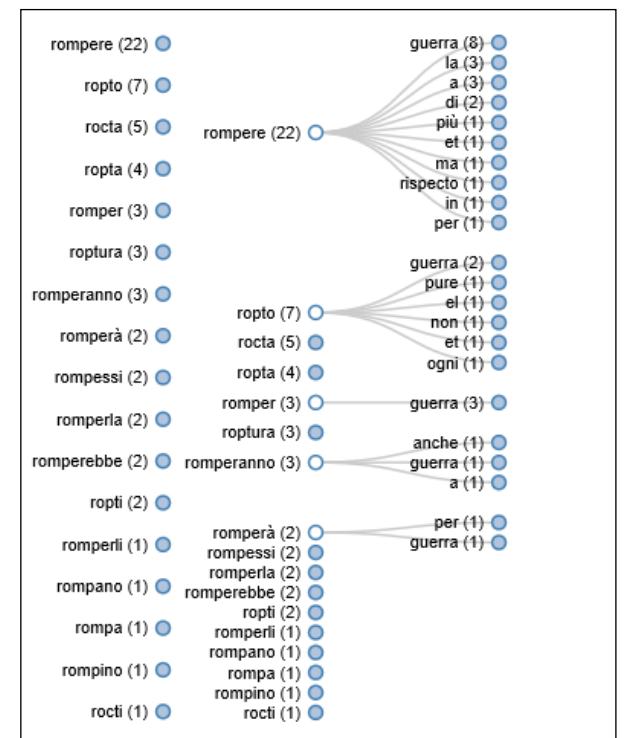
Les termes *malattia* (substantif), *malato* et *ammalato/i* (adjectifs) représentent 48 occurrences sur l'ensemble des lettres de Machiavel. Ils sont essentiellement employés quand il est *a palazzo* ou lorsqu'il se trouve dans les campements autour de Pise, et aussi de façon significative, lorsqu'il se trouve en légation en France et auprès de César Borgia.

FORMULES FIGÉES, POLYSÉMIE : ROMPERE (LA) GUERRA

Machiavel



Guichardin



Les structures arborescentes permettent de visualiser un nombre très important de résultats en une seule image ; elles apportent aussi des éléments clés pour dégager et cataloguer les différents usages et la polysémie des termes.

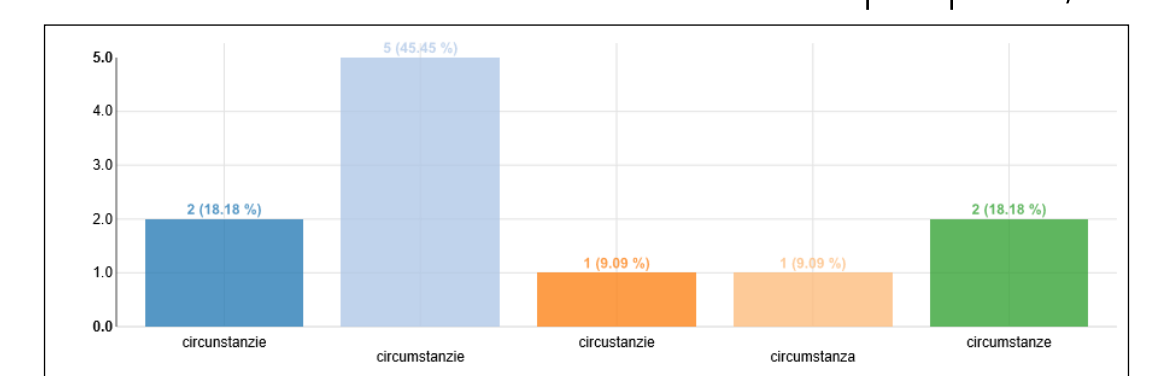
ENJEUX DE L'ÉTUDE COMPARATIVE

Ils touchent à la fois à la linguistique (photographier un état de langue ancien, anormé, pour lequel les dictionnaires manquent) et au discours politique renaissant (identifier les termes, et parfois même les structures discursives profondes, autour desquels se focalise la réflexion politique).

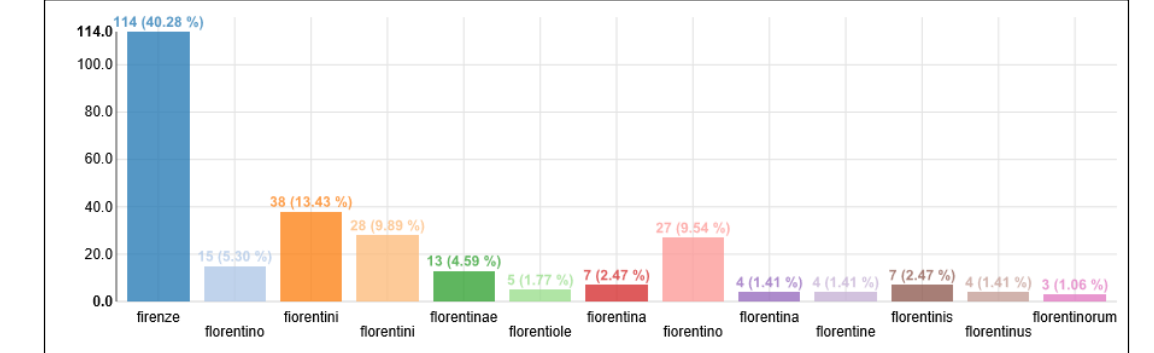
Il ne s'agit pas de penser la base de données comme une lime qui mettrait tous les textes termes au même niveau, en gommant les différences, les nuances ou les préférences, mais de la concevoir de façon dynamique comme un espace multidimensionnel au sein duquel l'utilisateur aurait la possibilité d'appréhender aisément l'ensemble des caractéristiques de chaque mot du corpus, ses relations aux autres et son poids dans les structures discursives, en les mettant en lien avec les auteurs, leurs fonctions et la chronologie. Ceci permettra de dégager un socle lexical commun aux auteurs mais aussi de distinguer d'éventuelles zones de fracture, dans les usages (fréquences, distribution) comme dans les emplois (polysémies, associations, etc.).

Ces objectifs déterminent autant de défis techniques pour l'avenir :

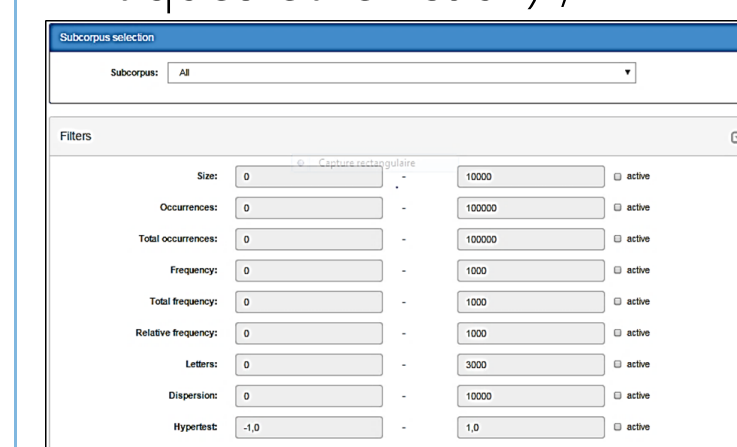
- ⇒ Rationalisation du regroupement en familles et/ou lemmatisation, création d'un dictionnaire de référence pour les noms communs et les noms propres ;



Histogrammes des flexions de *circostanze* et de *Firenze*



- ⇒ Multiplication des sous-corpus ;
- ⇒ Affinage des outils de prise en compte des spécificités matérielles et langagières des corpus ;
- ⇒ Développement des outils de mesure statistiques (poids des mots et relations entre les champs sémantiques et lexicaux) ;



- Fréquence en 1/10000 dans un sous-corpus
- Indice de dispersion
- Indice de spécificité

⇒ Publication de l'outil.