

# EpistETmeth n°1

05.12.22

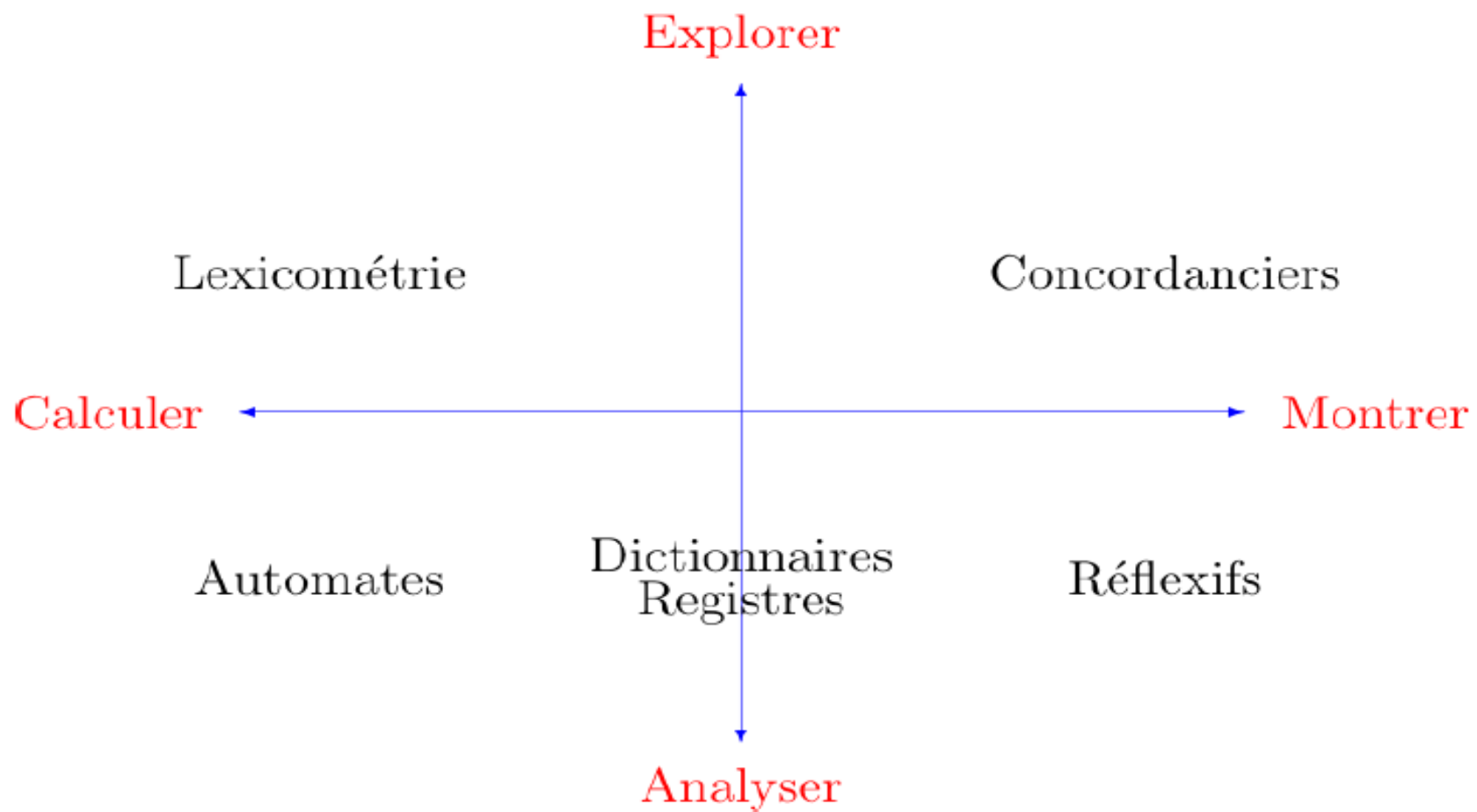
16h30 à 18h

**Analyse quantitative de données qualitatives**

**Agathe Dirani, CREAD, Université Rennes 2**

# Quelques points à aborder

- Diversité de finalités
- Atout et limites des logiciels
- Les logiciels au CREAD



*Figure 1. Familles de fonctionnalités*

# Lexicométrie

- Description des procédés :
  - l'exploration statistique du lexique d'un corpus de textes. Les outils statistiques classiques, comme les tableaux de fréquences, les tris croisés voire l'analyse factorielle des correspondances, permettent (notamment) de déterminer la fréquences des mots d'un corpus, leur répartition dans différents textes ou leurs éventuelles sur- (ou sous-) représentation
  - Recours à la lemmatisation
  - Dispositifs classiques de visualisation statistique (histogrammes, boîtes à moustaches, dendrogrammes)
- Objectifs :
  - descriptifs et quantitatifs
- Ex. de logiciels :
  - Lexico, Spad-T, Sato, Hyperbase et le projet d'unification Textométrie
- Atouts et limites
  - l'hypothèse que les phénomènes de récurrence ont une pertinence
- Disciplines les plus concernées / exemples de travaux
  - Analyse de la propagande dans la presse
  - En sciences politiques, Corinne Gobin, étude des acceptions du travail dans les institutions européennes.
  - En sociologie, Jules Duchastel, pour rendre compte des événements de crise.

# Concordanciers

- Description des procédés :
  - Réagencement du texte
  - Les segments du texte reprenant une même expression sont alignés les uns en-dessous des autres, ce qui permet de les replacer dans leur contexte
  - seuls les éléments explicitement (et délibérément) recherchés par l'utilisateur apparaissent dans les concordances.
- Objectifs :
  - essentiellement exploratoire et qualitatif
- Ex. de logiciels :
  - Lexico et Hyperbase, Tropes
- Atouts et limites
  - Contextualisation des mots, permet de travailler sur des expressions
- Disciplines les plus concernées / exemples de travaux
  - Littérature et linguistique

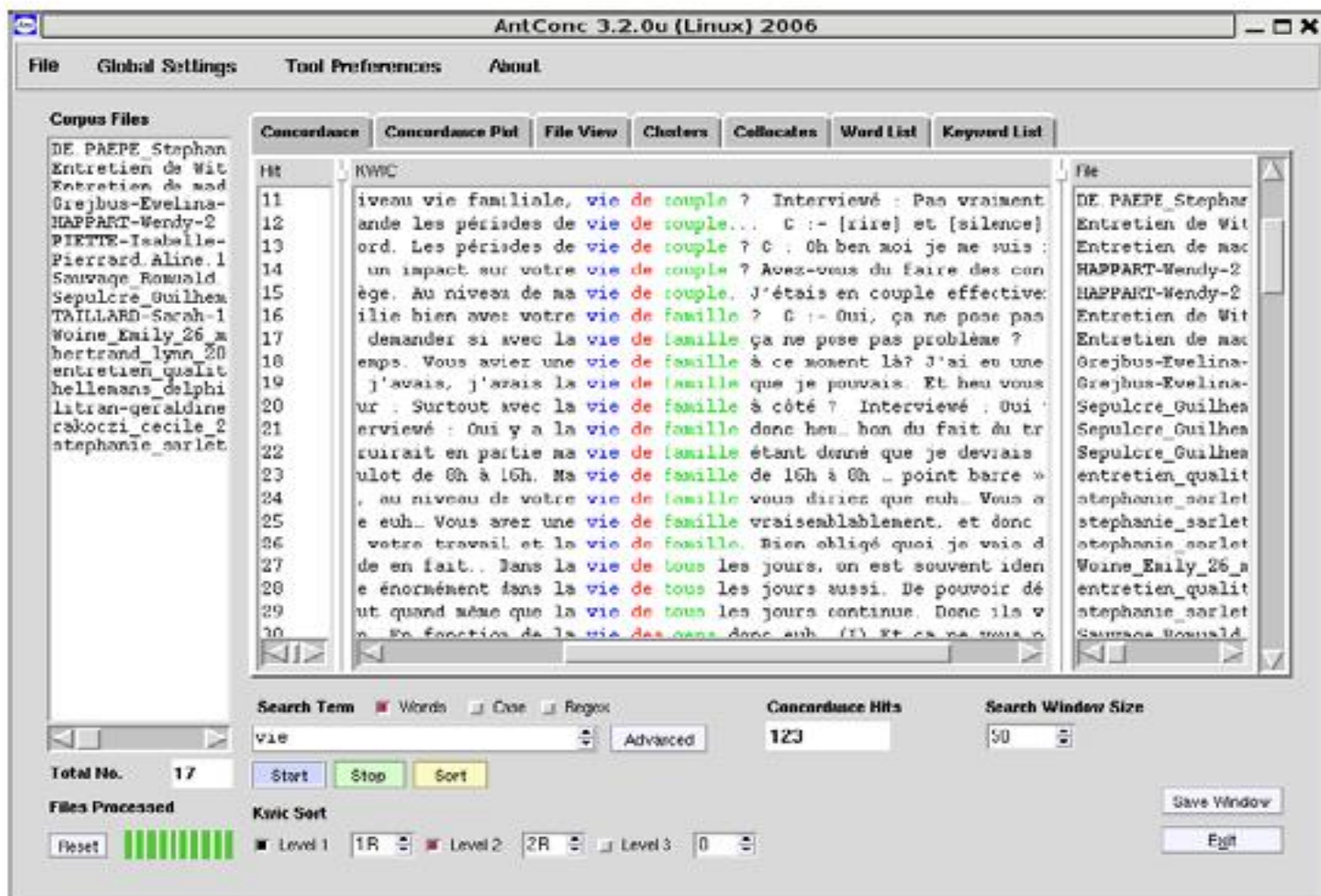


Figure 3. Concordancier (copie d'écran de AntConc)

# Automates

- Description des procédés :
  - Assignation automatique d'étiquettes à des portions de texte
  - Analyse de co-occurrence → segmentation automatique des corpus de texte
  - Statistique benzecriste
  - Pari d'économie de temps
  - Représentation des agrégats calculés sous forme d'arborescences ou de réseaux
  - Pas d'étiquetage nécessaire
  - Interprétation des angles et non de la distance entre les mots
- Objectifs :
  - les mêmes que la lexicométrie (descriptif et quantitatif) + valorisation des vertus synthétiques de la représentation graphique
- Ex. de logiciels :
  - Leximappe, Candide et Réseau-Lu (à partir du calcul des cooccurrences, Spad-T ou Alceste (sur base de l'analyse des correspondances)
- Atouts et limites
  - Interprétation non ancrée et possiblement fausse si non associée à une connaissance fine du corpus
  - Rôle avant tout heuristique (soutien à la formulation d'hypothèses) mais non probatoire (test des hypothèses)

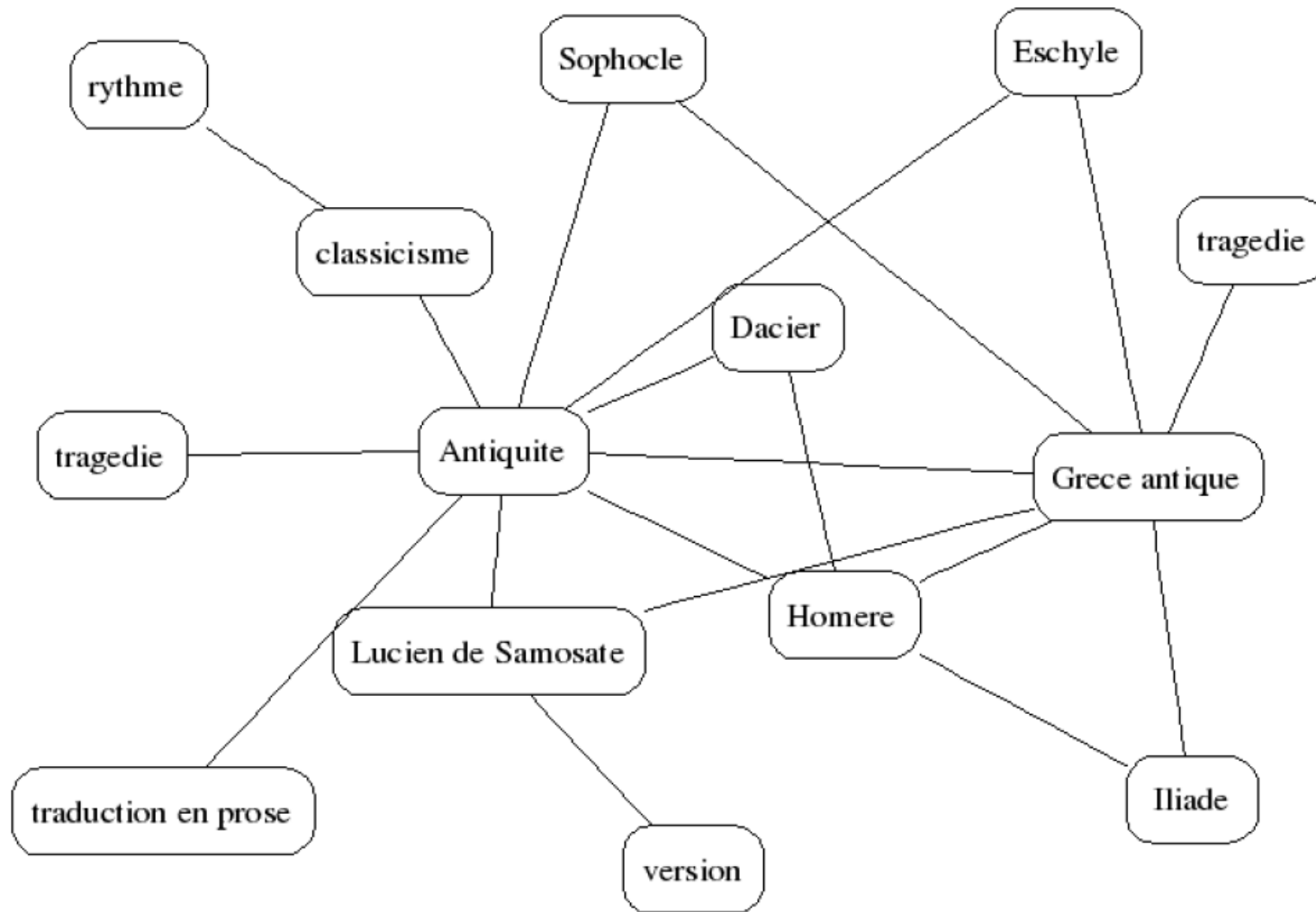


Figure 4. Automate (d'après Candide)

Source : Lejeune, 2010



# Réflexifs

- Description des procédés :
  - Lecture exhaustive et répétée des entretiens
  - Annotation des segments (rappelle le surlignage)
  - Le support informatique assiste l'explicitation des codages.
- Objectifs :
  - méthode par théorisation ancrée (ou enracinée), la « Grounded Theory Methodology » (Barney Glaser et Anselm Strauss, 2010)
  - Analyse résolument qualitative
  - Usage plus quantitatif lorsque qu'il s'agit par exemple de croiser l'analyse selon les caractéristiques des enquêt.é.e.s (selon le genre par ex.)
- Ex. de logiciels :
  - l'acronyme CAQDAS pour « Computer Assisted Qualitative Data Analysis Software », des outils comme NVivo, ATLAS.ti, Weft QDA ou TamsAnalyser reposent essentiellement sur le travail intellectuel du chercheur
- Atouts et limites
  - Tendance à l'obsession du codage au détriment de l'analyse
  - Mais également un gage de scientificité
- Disciplines les plus concernées / exemples de travaux
  - de la gestion aux sciences de l'éducation en passant par la sociologie

# *Dictionnaires et registres*

- Description des procédés :
  - Le chercheur définit les catégories d'analyse (liste d'indicateurs), alors que l'attribution de ces catégories à des passages de textes est confiée à différents codeurs.
  - pari qu'il existe une signification stable des catégories d'analyse, sur laquelle on peut s'appuyer.
- Objectifs :
  - Troisième voie entre automates et outils réflexifs
- Ex. de logiciels :
  - Tropes et Prospéro, Cassandre
- Atouts et limites
  - Limite la subjectivité du codage
- Disciplines les plus concernées / exemples de travaux
  - Utilisation des synonymes, des registres argumentatifs ou des logiques d'actions se rencontrent plutôt dans les sciences de la culture (comme la sociologie, l'histoire ou l'anthropologie).
  - Utilisation des catégories grammaticales : sciences du langage
  - Psychologie, ethnométhodologie

# Atouts et limites des logiciels

- Représentations négatives et méconnaissance
  - « l'assimilation de l'ordinateur à un artéfact technique produit par l'industrie et procédant de l'application systématique de la logique mathématique » (Lejeune, 2010)
- Des atouts
  - une facilitation de l'exploration du matériau empirique
  - une explicitation (et, partant, une reproductibilité) des procédures analytiques
  - une ouverture sur l'analyse à plusieurs ou un archivage de l'activité scientifique.
  - se donner une contrainte de travail participant à garantir la rigueur de sa production.
  - dimension structurante et fédératrice de la discussion entre pairs : de Merton (1953) à Latour (1995)

# Atouts et limites des logiciels

- Des limites
  - Un adjutant mais pas une garant du sérieux d'une recherche
  - la vigilance épistémologique et la rigueur méthodologique restent de mise
  - Discuter la validité du codage
  - Risque de produire une recherche très descriptive
  - Des outils complémentaires d'une connaissance intime du corpus
- C'est en combinant les fonctionnalités de l'analyse textuelle et les logiciels que l'on peut espérer seconder efficacement une recherche

# Éléments de conclusion

- « Quelle que soit la stratégie pour laquelle on opte, j'espère avoir montré qu'il n'est pas de technique qui garantisse la scientificité ou l'originalité d'une recherche. » (Lejeune, 2007, p.210)
- « Qu'elle soit ou non informatisée, la méthode nécessite de la rigueur, et, dans tous les cas de figure, la qualité de l'interprétation revient toujours au scientifique. En sociologie qualitative, l'atout des logiciels d'analyse réside en définitive dans la facilitation offerte pour échanger et discuter entre chercheurs. Aussi, plutôt que de brandir l'outil, tel un bouclier face à la critique, il importe d'ouvrir les boîtes noires et de partager les expériences. » (Lejeune, 2007, p. 210)

# Références citées

- Glaser, B.G., & Strauss, A.L. (2010). *La découverte de la théorie ancrée : stratégies pour la recherche qualitative*. Paris : Armand Colin.
- Lejeune, C. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. Dans C. Brossaud, & B. Reber (Éds), *Humanités numériques. Nouvelles technologies cognitives et concepts des sciences sociales* (pp. 197-214). Paris : Hermès.
- Lejeune, C. (2010). Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative. *Recherches qualitatives*. 9. pp. 15-32.
- Latour, B. (1995). *La science en action*. Paris : Gallimard.
- Merton, R.K. (1953). *Éléments de méthode sociologique*. Paris : Plon.