

Analyse lexicométrique d'un corpus textuel avec le logiciel Alceste

Melpomeni PAPADOPOULOU – Docteure Sciences de
l'éducation et de la formation

Université de Tours – EA 7505 EES

Introduction

Objectif : étudier les effets d'une méthodologie mixte entre analyse catégorielle qualitative (Bardin, L., 2013) et analyse quantitative avec Alceste (Reinert, M., 2008) d'un corpus d'une thèse de doctorat en Sciences Humaines.

Alceste : logiciel d'aide méthodologique pour l'étude et interprétation de corpus textuels (Bart, 2011, p. 173).

Plan de la présentation :

- La méthode Alceste
- La recherche sur la FOAD avec Alceste
- Les principaux résultats et perspectives de l'utilisation d'Alceste en Sciences Humaines

Présentation d'Alceste

- Créé par Max Reinert (1990). Basé sur la théorie d'analyse statistique de Jean-Paul Benzécri (1973) and de semiose de Charles Sanders Peirce (1878).
- Objectif : représentation du corpus d'une façon graphique pour aider le chercheur à interpréter les résultats.

Etapes d'analyse par Alceste

- Analyse du vocabulaire
- Séparation du corpus
- Classification descendante hierarchique
- Résutlats

Etude sur Alceste

Problématique : quelle intervention du chercheur sur le corpus avant analyse avec Alceste?

Notre hypothèse est que l'analyse automatique avec Alceste est plus efficiente et en lien avec la question de recherche quand le chercheur :

- prépare le corpus avant d'utiliser Alceste
- divise le corpus en unités de sens en utilisant des mots-clés et le symbole \$
- combine l'analyse quantitative d'Alceste avec une analyse qualitative catégorielle (Laurence Bardin, 2013).

La méthode de recherche

Une méthode en 5 étapes. Analyse effectuée sur le corpus :

1. transcrit fidèlement
2. modifié sans enlever les questions
3. modifié, questions remplacées par leur contenu (10 mots-clés)
4. modifié, questions remplacées par les catégories de l'analyse qualitative (5 mots-clés)
5. modifié, questions remplacées les sous-catégories de l'analyse qualitative (21 mots-clés)

1. Analyse de la transcription fidèle du corpus

Corpus	Types of analysis	Classes	Selected	Donn.1	Donn.2
Asso1	Double	2	90%	Auto	Auto

Board| 1 Results of the corpus Asso1 without modifications

Modifications effectuées sur le corpus

1. Commentaires en majuscule (ex. rires --> RIRES)
2. Mots avec apostrophe (ex. 'fin --> enfin)
3. Synonymes (ex. online training, FAD, FOAD, e-learning --> distant-training)
4. Tiret du bas (ex. formation à distance --> formation_à_distance)
5. Acronymes en minuscule (ex. BAFA --> bafa)
6. Chiffres en lettres (ex. 7 --> seven or seven -->7)

2. Analyse du corpus modifié avec les questions

d37: Quel est ton rôle sur le chat ou sur le forum?

D37: Ah, (rire), à construire, à réfléchir, à approfondir je pense. Mon rôle c'est plus, j'imagine vraiment de montrer que je suis là, que je guide, que j'incite d'échanger entre eux mais encore une fois je pense que l'outil joue beaucoup quand tu vois, ça les a incités à aller à, il y a de la spontanéité, la synchrone, tu vois.

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analysed word ¹	Donn.1 ²	Donn.2
Asso1	Simple	4	78%	4		
Asso1	Double	6	66%		Auto	Auto

Board 2 Results of the corpus Asso1 including the questions

3. Questions remplacées par 10 mots-clés correspondants au contenu des questions

-*situation

Ah, (RIRE), à construire, à réfléchir, à approfondir je pense. Mon rôle c'est plus, j'imagine vraiment de montrer que je suis là, que je guide, que j'incite d'échanger entre eux mais encore une fois je pense que l'outil joue beaucoup quand tu vois, ça les a incités à aller à, il y a de la spontanéité, la synchrone, tu vois. Est-ce que toi tu as envie d'aller chercher l'info d'il y a

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analysed word	Donn.1	Donn.2
Asso1	Simple	4	83%	7		
Asso1	Simple	3	79%	4		
Asso1	Simple	3	78%	5		
Asso1	Double	4	62%		8	10
Asso1	Double	5	54%		16	18
Asso1	Double	4	43%		18	20
Asso1	Double	3	39%		12	16
Asso1	Double	3	34%		Auto	Auto
Asso1	Double	3	34%		10	12
Asso1	Double	3	32%		12	14
Asso1	Double	5	32%		14	16

Board 3 Results of the corpus Asso1 with the starred keywords corresponding to the questions

4. Questions remplacées par les catégories de l'analyse qualitative (5 mots-clés)

n° ligne	réponses	Totalité du récit	sous thèmes	thèmes	Accompagnement (a), Ingénierie pédagogique (i), Technique (t)	Interprétation
173-186	D37	<p>d37: Quel est ton rôle sur le chat ou sur le forum?</p> <p>D37: Ah, (rire), à <u>construire</u>, à <u>réfléchir</u>, à <u>approfondir</u> je pense. Mon rôle c'est plus, j'imagine vraiment de <u>montrer que je suis là</u>, que je <u>guide</u>, que j'<u>incite d'échanger entre eux</u></p>		Fonction	A-	

4. Questions remplacées par les catégories de l'analyse qualitative (5 mots-clés)

-*accomp

Ou alors en présence ils m'ont dit et après moi je suis intervenu
quoi.\$

Ah, (RIRE), à construire, à réfléchir, à approfondir je pense. Mon rôle c'est plus, j'imagine vraiment de montrer que je suis là, que je guide, que j'incite d'échanger entre eux mais encore une fois je pense que l'outil joue beaucoup quand tu vois, ça les a incités à aller à, il y a de la spontanéité, la synchrone, tu vois. Est-ce que toi tu as

4. Questions remplacées par les catégories de l'analyse qualitative (5 mots-clés)

Corpus	Type d'analyse	Classes	Selected	Fréquence minimale d'une forme analysée	Donn.1	Donn.2
Asso1	Double	2	83%		Auto	Auto
Asso1	Simple	3	81%	4		

Board 4 Results of the corpus Asso1 with the keywords corresponding to the general themes

5. Questions remplacées par les sous-catégories de l'analyse qualitative (21 mots-clés)

-*fonc_accomp

Ah, (RIRE), à construire, à réfléchir, à approfondir je pense. Mon rôle c'est plus, j'imagine vraiment de montrer que je suis là, que je guide, que j'incite d'échanger entre eux mais encore une fois je pense que l'outil joue beaucoup quand tu vois, ça les a incités à aller à, il y a de la spontanéité, la synchrone, tu vois. Est-ce que toi tu as envie d'aller chercher l'info d'il y a deux semaines. On est dans un

5. Questions remplacées par les sous-catégories de l'analyse qualitative (21 mots-clés)

Corpus	Type of analysis	Classes	Selected	Fréquence minimale d'une forme analysée	Donn.1	Donn.2
Asso1	Double	2	78%		Auto	Auto
Asso1	Simple	3	73%	4		

Board 5 Results of the Asso1 corpus with the starred keywords corresponding to the sub-themes

Interprétation des résultats

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analysed form	Don n.1	Don n.2
Assol (without modifications)	Double	2	90%		Auto	Auto
Assol (modifié)	Simple	4	85%	6		
Assol (modifié)	Simple	4	83%	7		
Assol (themes)	Double	2	83%		Auto	Auto
Assol (themes)	Simple	3	81%	4		
Assol (modifié)	Simple	3	79%	4		
Assol (modifié)	Simple	3	78%	5		
Assol (sub-themes)	Double	2	78%		Auto	Auto
Assol (without keywords)	Simple	4	78%	4		
Assol (sub-themes)	Simple	3	73%	4		
Assol (sans mots étoilés)	Double	6	66%		Auto	Auto
Assol (modifié)	Double	4	62%		8	10
Assol (modifié)	Double	5	54%		16	18
Assol (modifié)	Double	4	43%		18	20
Assol (modifié)	Double	3	39%		12	16
Assol (modifié)	Double	3	34%		Auto	Auto
Assol (modifié)	Double	3	34%		10	12
Assol (modifié)	Double	3	32%		12	14
Assol (modifié)	Double	5	32%		14	16

Board 6. Results of the Assol corpus

Interprétation des résultats

2nd corpus : four interviews in the second association

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analysed form	Donn. 1	Donn. 2
Asso2 (themes)	Simple	3	94%	4		
Asso2 (themes)	Double	2	92%		Auto	Auto
Asso2	Simple	5	86%	4		
Asso2 (themes)	Simple	4	82%	5		
Asso2	Double	4	76%		Auto	Auto

Board 7 Results of the Asso2 corpus

Interprétation des résultats

Corpus	Type of analysis	of Classes	Selected	Minimum frequency of an analysed form	Donn.1	Donn.2
Assol-Asso2	Simple	5	94%	6		
Assol-Asso2 (themes)	Simple	4	93%	5		
Assol-Asso2 (themes)	Simple	5	92%	4		
Assol-Asso2	Simple	4	83%	5		
Assol-Asso2	Simple	4	79%	4		
Assol-Asso2	Double	4	73%		Auto	Auto
Assol-Asso2	Double	4	73%		Auto	Auto
Assol-Asso2 (themes)	Double	4	73%		14	16
Assol-Asso2 (themes)	Double	4	66%		16	18
Assol-Asso2 (themes)	Double	4	62%		12	14
Assol-Asso2 (themes)	Double	4	61%		Auto	Auto
Assol-Asso2 (themes)	Double	4	61%		8	10
Assol-Asso2 (themes)	Double	4	61%		10	12

Board 8 Results of both corpus

Préconisations

- Préparer le corpus avant intégration dans Alceste
- Remplacer les questions par des mots-clés correspondant au contenu des réponses
- Marquer les interactions avec le symbole \$, qui permet la division du corpus
- Eviter une division trop importante (dans notre cas, l'utilisation des 21 mots-clés) car cela influence les résultats obtenus par Alceste
- Opter pour une méthodologie mixte (qualitative/quantitative) d'analyse du corpus

Retrouver la présentation PREZI :

<https://prezi.com/view/BWOPYICQIEpmYBlatkyo/>

Analyse lexicométrique d'un corpus textuel avec le logiciel Alceste

Intoduction

Objectif : étudier les effets d'une méthodologie mixte entre analyse catégorielle qualitative (Bardin, L., 2013) et analyse quantitative avec Alceste (Reinert, M., 2008) d'un corpus d'une thèse de doctorat en Sciences Humaines.

Alceste : logiciel d'aide méthodologique pour l'étude et interprétation de corpus textuels (Bart, 2011, p. 173).

Plan de la présentation :

- La méthode Alceste
- La recherche sur la FOAD avec Alceste
- Les principaux résultats et perspectives de l'utilisation d'Alceste en Sciences Humaines

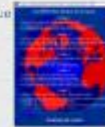
Présentation d'Alceste

Créé par Max Reinert (1990). Basé sur la théorie d'analyse statistique de Jean-Paul Benzécri (1973) and de semiose de Charles Sanders Peirce (1878).

Objectif : représentation du corpus d'une façon graphique pour aider le chercheur à interpréter les résultats.

Etapes d'analyse par Alceste

- Analyse du vocabulaire
- Séparation du corpus
- Classification descendante hiérarchique
- Résultats



Melpomeni PAPADOPOULOU
University of Tours, EA7505



Bibliographie

Bart, D. (2011). L'analyse de données textuelles avec le logiciel Alceste. *Recherches en didactiques*, 2(12), 173-184. Consulté le février 12, 2019, sur <https://www.cairn.info/revue-recherches-en-didactiques-2011-2-page-173.htm>

Clérisse, S. (2015). *Analyse de données textuelles et sciences sociales. Application et comparaison de deux outils, Calliope et Alceste*. Paris: INTD-CNAM.

Demazière, D., Brossaud, C., Trabal, P., & Meter Van, K. (2006). *Analyses textuelles en sociologie. Logiciels, méthodes, usages*. Rennes: Presses Universitaires de Rennes.

Marpsat, M. (2010). La méthode Alceste. *Sociologie*, 1(1). Consulté le avril 15, 2019, sur <https://journals-openedition-org.proxy.scd.univ-tours.fr/sociologie/312>

Masuy-Stroobant, G. (2013). Interpréter les résultats. Dans G. Masuy-Stroobant, & R. (. Costa, *Analyser les données en sciences sociales. De la préparation des données à l'analyse multivariée* (pp. 281-288). Bruxelles: P.I.E. PETER LANG.

Reinert, M. (1990). Alceste, une méthodologie d'analyse des données textuelles et une application : Aurélia, de Gérard de Nerval. *Bulletin de méthodologie sociologique*, 26(1), 24-54.

Reinert, M. (2008). Mondes lexicaux stabilisés et analyse statistique de discours. *9e Journées internationales d'Analyse statistique des Données Textuelles*, 981-993.

**MERCI POUR VOTRE ATTENTION ET
PARTICIPATION**

Contact : melpomeni.papadopoulou@univ-tours.fr