

Quantitative analysis of a textual corpus in Human Sciences: The case of the Alceste french software

Introduction

Objective : study the effects of the combination of a qualitative analysis of the content (Bardin, L., 2013) and a quantitative analysis of the corpus with Alceste (Reinert, M., 2008) in a case study on Human Sciences.

Alceste : is used as a methodological assistance for the study and interpretation of textual corpus (Bart, 2011, p. 173).

Plan of the presentation :

- Presentation of the Alceste method
- Presentation of the case study on E-learning using Alceste
- Presentation of the main results and perspectives of the use of Alceste in Human Sciences

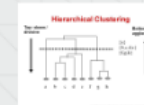
Presentation of Alceste

Created by Max Reinert (1990). Based on the theory of data analysis of Jean-Paul Benzécri (1973) and the semiosis of Charles Sanders Peirce (1878).

Objective : representation of the corpus in a graphic way to help the analyst interpret the results.

Steps of the Alceste's analysis

- Analysis of the vocabulary
- Separation of the corpus
- Hierarchical downward classification
- Results



Quantitative analysis of a textual corpus in Human Sciences: The case of the Alceste french software

Introduction

Objective : study the effects of the combination of a qualitative analysis of the content (Bardin, L., 2013) and a quantitative analysis of the corpus with Alceste (Reinert, M., 2008) in a case study on Human Sciences.

Alceste : is used as a methodological assistance for the study and interpretation of textual corpus (Bart, 2011, p. 173).

Plan of the presentation :

- Presentation of the Alceste method
- Presentation of the case study on E-learning using Alceste
- Presentation of the main results and perspectives of the use of Alceste in Human Sciences

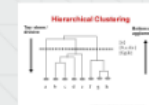
Presentation of Alceste

Created by Max Reinert (1990). Based on the theory of data analysis of Jean-Paul Benzécri (1973) and the semiosis of Charles Sanders Peirce (1878).

Objective : representation of the corpus in a graphic way to help the analyst interpret the results.

Steps of the Alceste's analysis

- Analysis of the vocabulary
- Separation of the corpus
- Hierarchical downward classification
- Results



Introduction

Objective : study the effects of the combination of a qualitative analysis of the content (Bardin, L., 2013) and a quantitative analysis of the corpus with Alceste (Reinert, M., 2008) in a case study on Human Sciences.

Alceste : is used as a methodological assistance for the study and interpretation of textual corpus (Bart, 2011, p. 173).

Plan of the presentation :

- Presentation of the Alceste method
- Presentation of the case study on E-learning using Alceste
- Presentation of the main results and perspectives of the use of Alceste in Human Sciences

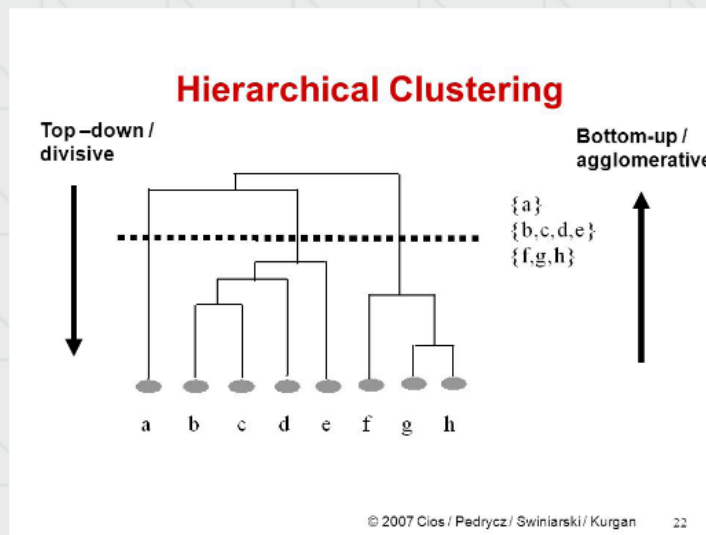
Presentation of Alceste

Created by Max Reinert (1990). Based on the theory of data analysis of Jean-Paul Benzécri (1973) and the semiosis of Charles Sanders Peirce (1878).

Objective : representation of the corpus in a graphic way to help the analyst interpret the results.

Steps of the Alceste's analysis

- Analysis of the vocabulary
- Separation of the corpus
- Hierarchical downward classification
- Results



Alceste - Animacoop-(4).bt (140 Ko) Code 121 - Lundi 22 Avril 2019 à 14h27

Les différentes étapes de l'analyse

Etape 1

- Lecture du corpus et **extraction des formes**
- Catégorisation, lemmatisation et comptage des formes
- Constitution du **dictionnaire des formes réduites**

Etape 2

- Définition et segmentation des unités textuelles du corpus
- Construction des **tableaux de données** destinés à la classification
- **Première Classification Descendante Hiérarchique**
- **Deuxième Classification Descendante Hiérarchique**

Etape 3

- Croisement des classifications, définition des classes retenues
- Présences et absences des formes
- **Analyse Factorielle des Correspondances**

Etape 4

- Sélection des unités textuelles par classe
- Segments répétés des classes et du corpus
- **Classification Ascendante Hiérarchique du corpus**
- **Classification Ascendante Hiérarchique par classe**

Etape 5

- Réseaux de proximité de formes, réseaux neuronaux
- Cartographies du corpus en unités textuelles, analyse temporelle
- Nuages de mots du corpus et par classe
- Courbes d'accroissement du vocabulaire du corpus et par individu
- Classement des individus et des variables
- Création des rapports détaillé et de synthèse

Analyse en cours

Les différentes étapes de l'analyse

Etape 1

- Lecture du corpus et extraction des formes
- Catégorisation, lemmatisation et comptage des formes
- Constitution du dictionnaire des formes réduites

Etape 2

- Définition et segmentation des unités textuelles du corpus
- Construction des tableaux de données destinés à la classification
- Première Classification Descendante Hiérarchique
- Deuxième Classification Descendante Hiérarchique

Etape 3

- Croisement des classifications, définition des classes retenues
- Présences et absences des formes
- Analyse Factorielle des Correspondances

Etape 4

- Sélection des unités textuelles par classe
- Segments répétés des classes et du corpus
- Classification Ascendante Hiérarchique du corpus
- Classification Ascendante Hiérarchique par classe

Etape 5

- Réseaux de proximité de formes, réseaux neuronaux
- Cartographies du corpus en unités textuelles, analyse temporelle
- Nuages de mots du corpus et par classe
- Courbes d'accroissement du vocabulaire du corpus et par individu
- Classement des individus et des variables
- Création des rapports détaillé et de synthèse

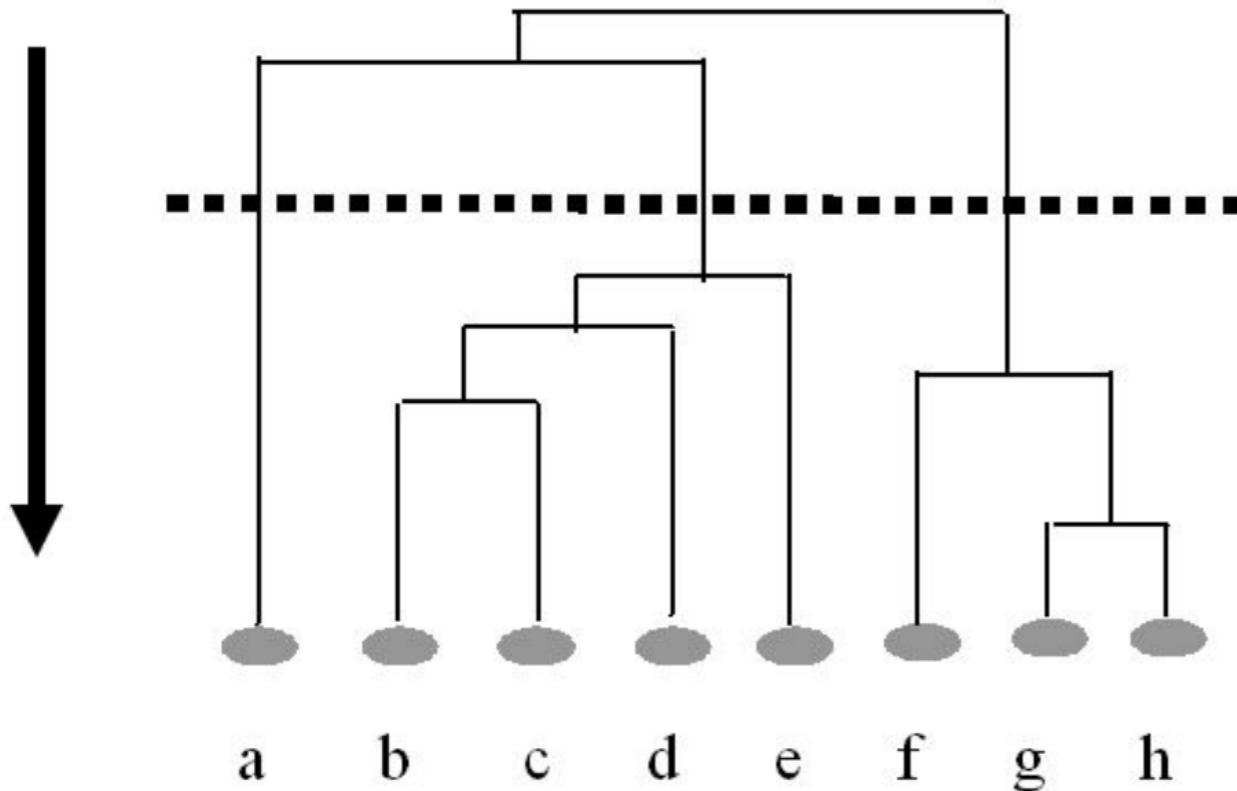
Analyse en cours

on

Hierarchical Clustering

Top-down /
divisive

Bottom-up /
agglomerative



{a}
{b, c, d, e}
{f, g, h}

A case study with Alceste

Research problem : how can the researcher prepare the corpus for an analysis by Alceste, without intervening on the results.

Our hypotheses are that the automatic analysis by Alceste is more efficient and relevant to the research problem when the researcher :

- prepares the corpus before to use Alceste
- divides the corpus in separate context units by using starred keywords and \$
- combines the quantitative analysis method of Alceste with a qualitative method of content analysis (Laurence Bardin, 2013).

The research method

A method in 6 steps :

1. first analysis of the corpus as accurately transcribed
2. modifications made on the corpus
3. second analysis of the modified corpus by including the questions of the interviewer
4. interactions replaced by 10 keywords corresponding to the contents of the questions
5. combination of a qualitative and quantitative method of data analysis
6. use of 21 keywords corresponding to the sub-themes of our qualitative analysis

1. First analysis of the corpus as accurately transcribed

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

2. Modifications made on the corpus

1. Comments in capital letters (ex. he thinks --> HE THINKS)
2. Words with apostrophe (ex. 'fin --> enfin)
3. Synonyms (ex. online training, FAD, FOAD, e-learning --> distant-training)
4. Underscore (ex. formation à distance --> formation_à_distance)
5. Acronyms in diminutive (ex. BAFA --> bafa)
6. Numbers in figures (ex. 7 --> seven or seven -->7)

3. Second analysis of the modified corpus by including the questions of the interviewer

d37: what is your role on the chat or forum?

D37: ah, (laughter), to build, to think, to deepen I think. My role it is more, I really imagine to show that I am there, that I guide, that I encourage them on exchanging between them but once again I think that the tool plays an important role when you see that it encouraged them to go online, there is spontaneity, because things happen at the same time in distance, you see?

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

Including the questions of the interviewer

37: what is your role on the chat or forum?

037: ah, (laughter), to build, to think, to deepen I think. My role it is more, I really imagine to show that I am there, that I guide, that I encourage them in exchanging between them but once again I think that the tool plays an important role when you see that it encouraged them to go online, there is spontaneity, because things happen at the same time in distance, you see

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

4. Questions were replaced by keywords of its content

-*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

Case	Step #	Case	Step #	Information	Case #	Step #
Case 1	Step 1	Case 1	Step 1	Information	Case 1	Step 1
Case 1	Step 2	Case 1	Step 2	Information	Case 1	Step 2
Case 1	Step 3	Case 1	Step 3	Information	Case 1	Step 3
Case 1	Step 4	Case 1	Step 4	Information	Case 1	Step 4
Case 1	Step 5	Case 1	Step 5	Information	Case 1	Step 5
Case 1	Step 6	Case 1	Step 6	Information	Case 1	Step 6
Case 1	Step 7	Case 1	Step 7	Information	Case 1	Step 7
Case 1	Step 8	Case 1	Step 8	Information	Case 1	Step 8
Case 1	Step 9	Case 1	Step 9	Information	Case 1	Step 9
Case 1	Step 10	Case 1	Step 10	Information	Case 1	Step 10
Case 1	Step 11	Case 1	Step 11	Information	Case 1	Step 11
Case 1	Step 12	Case 1	Step 12	Information	Case 1	Step 12
Case 1	Step 13	Case 1	Step 13	Information	Case 1	Step 13
Case 1	Step 14	Case 1	Step 14	Information	Case 1	Step 14
Case 1	Step 15	Case 1	Step 15	Information	Case 1	Step 15

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

5. combination of a qualitative and quantitative method of data analysis

Titre de l'étude	un thème	Date	Accompagnement (si applicable, si possible en français)	Statistique
101 - Quel est le rôle de l'enseignant ?				
102 - À quel moment de la journée l'enseignant est-il le plus actif ?				

--"recoup
 Ou alors en présence ils m'ont dit et après moi je suis intervenu
 tout ?
 Ah, (RBE), à construire, à réfléchir, à approfondir le sens. Mon
 rôle c'est plus, j'imaginais vraiment de montrer que je suis là, que je
 guide, que j'aide à 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 synchronie, tu vois. Est-ce que toi tu as

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

n° ligne	réponses	Totalité du récit	sous thèmes	thèmes	Accompagnement (a), Ingénierie pédagogique (i), Technique (t)	Interprétation
173-186	D37	d37: Quel est ton rôle sur le chat ou sur le forum? 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>		Fonction	A-	

-*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

Qualitative method of data analysis

écrit	sous thème	thèmes	Accompagnement (A), Ingénierie pédagogique (I), Technique (T)	Interprétation
en rôle sur le chat ou sur le forum?		Function	A-	
à construire, à réfléchir, à approfondir je pense. Mon rôle c'est plus, est de montrer que je suis là, que je guide, que j'aide d'échanger entre eux				

```
.*accomp  
Du alors en présence ils m'ont dit et après moi je suis intervenu  
quoi.5  
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
```

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

6. Use of 21 keywords corresponding to the sub-themes of our qualitative analysis

-*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

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

t plus, j'imagine vraiment de montrer que je suis là,
e j'incite d'échanger entre eux mais encore une fois
il joue beaucoup quand tu vois, ça les a incités à al
la spontanéité, la synchrone, tu vois. Est-ce que to
llier chercher l'info d'il y a deux semaines. On est d

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

Interpretation of the results

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analyzed form	Donn. 1	Donn. 2
Asso1 (without modifications)	Double	2	90%		Auto	Auto
Asso1 (modifié)	Simple	4	85%	6		
Asso1 (modifié)	Simple	4	83%	7		
Asso1 (themes)	Double	2	83%		Auto	Auto
Asso1 (themes)	Simple	3	81%	4		
Asso1 (modifié)	Simple	3	79%	4		
Asso1 (modifié)	Simple	3	78%	5		
Asso1 (sub-themes)	Double	2	78%		Auto	Auto
Asso1 (without keywords)	Simple	4	78%	4		
Asso1 (sub-themes)	Simple	3	73%	4		
Asso1 (sans mots étoilés)	Double	6	66%		Auto	Auto
Asso1 (modifié)	Double	4	62%		8	10
Asso1 (modifié)	Double	5	54%		16	18
Asso1 (modifié)	Double	4	43%		18	20
Asso1 (modifié)	Double	3	39%		12	16
Asso1 (modifié)	Double	3	34%		Auto	Auto
Asso1 (modifié)	Double	3	34%		10	12
Asso1 (modifié)	Double	3	32%		12	14
Asso1 (modifié)	Double	5	32%		14	16

Board 4 Results of the Asso1 corpus

2nd corpus : four interviews in the second association

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analyzed 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

Corpus	Type of analysis	Classes	Selected	Minimum frequency of an analyzed form	Donn. 1	Donn. 2
Asso1-Asso2	Simple	5	94%	6		
Asso1-Asso2 (themes)	Simple	4	93%	5		
Asso1-Asso2 (themes)	Simple	5	92%	4		
Asso1-Asso2	Simple	4	83%	5		
Asso1-Asso2	Simple	4	79%	4		
Asso1-Asso2	Double	4	75%		Auto	Auto
Asso1-Asso2	Double	4	73%		Auto	Auto
Asso1-Asso2 (themes)	Double	4	73%		14	16
Asso1-Asso2 (themes)	Double	4	66%		16	18
Asso1-Asso2 (themes)	Double	4	62%		12	14
Asso1-Asso2 (themes)	Double	4	61%		Auto	Auto
Asso1-Asso2 (themes)	Double	4	61%		8	10
Asso1-Asso2 (themes)	Double	4	61%		10	12

Board 8 Results of both corpus

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

2nd corp

Corpus

Asso2
(themes)

Asso2
(themes)

Asso2

Asso2
(themes)

Asso2

Board 7

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

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

A case study with Alceste

Research problem : how can the researcher prepare the corpus for an analysis by Alceste, without intervening on the results.

Our hypotheses are that the automatic analysis by Alceste is more efficient and relevant to the research problem when the researcher :

- prepares the corpus before to use Alceste
- divides the corpus in separate context units by using starred keywords and \$
- combines the quantitative analysis method of Alceste with a qualitative method of content analysis (Laurence Bardin, 2013).

Research perspectives

- Prepare the corpus before its integration for analysis on Alceste.
- Replace the text of the questions by keywords corresponding to the contents of the answers.
- Mark the interactions with the symbol \$, which allows the division of the corpus.
- Avoid a division too precise (in our case, the use of 21 starred keywords) because it can intervene with the results obtained by the software.
- Opt for a combined method of qualitative and quantitative corpus analysis

Quantitative analysis of a textual corpus in Human Sciences: The case of the Alceste french software

Introduction

Objective : study the effects of the combination of a qualitative analysis of the content (Bardin, L., 2013) and a quantitative analysis of the corpus with Alceste (Reinert, M., 2008) in a case study on Human Sciences.

Alceste : is used as a methodological assistance for the study and interpretation of textual corpus (Bart, 2011, p. 173).

Plan of the presentation :

- Presentation of the Alceste method
- Presentation of the case study on E-learning using Alceste
- Presentation of the main results and perspectives of the use of Alceste in Human Sciences

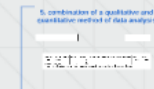
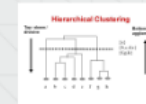
Presentation of Alceste

Created by Max Reinert (1990). Based on the theory of data analysis of Jean-Paul Benzécri (1973) and the semiosis of Charles Sanders Peirce (1878).

Objective : representation of the corpus in a graphic way to help the analyst interpret the results.

Steps of the Alceste's analysis

- Analysis of the vocabulary
- Separation of the corpus
- Hierarchical downward classification
- Results



Melpomeni PAPAPOULOU
University of Tours, EA7505