

L'individu dans le(s) groupe(s) : focus group et partitionnement du corpus

Daniel Pélissier

Université Toulouse 1 Capitole - daniel2.pelissier@ut-capitole.fr

Abstract

Lexicometric analyzes of the focus groups depend in particular on the choice of partitioning of the corpus by researcher. After having proposed a typology of possible partitioning, we present the results of an experiment of one of these approaches on a corpus of ten focus groups. These analyzes highlight some contributions and limitations of lexicometry compared to conversational analysis.

Résumé

Les analyses lexicométriques des focus groups dépendent notamment des choix de partitionnement du corpus par le chercheur. Après avoir proposé une typologie des partitionnements possibles, nous présentons les résultats d'une expérimentation d'une de ces approches sur un corpus de dix focus groups. Ces analyses mettent en évidence certains apports et limites de la lexicométrie par rapport à l'analyse conversationnelle.

Keywords: Focus groups, partitioning, individual, group.

Mots clefs : Focus groups, partitionnement, individu, groupe.

1. Introduction

La lexicométrie a étudié d'abord des discours écrits (articles de journaux, discours politiques, etc.) et des réponses à des questions ouvertes (Lebart et Salem, 1988) puis s'est intéressée aux conversations orales retranscrites (Rouré et Reinert, 1993; Bonneau et Dister, 2010). L'analyse de ces dernières est en effet plus délicate en raison de textes en général plus courts, de syntaxes particulières. Les focus groups appartiennent à cette famille de données en posant le problème particulier du nombre important de participants. Selon certains auteurs, ce type de données est difficile à analyser avec des logiciels de lexicométrie (Duchesne et Haegel, 2014).

Pourtant, l'analyse lexicométrique a été utilisée dans plusieurs études (Guerrero et al., 2009; Grésillon et al., 2012; Hulin, 2013; Bengough et al., 2015; Brangier et al., 2015) et des articles méthodologiques ont analysé l'efficacité des traitements lexicométriques (Dransfield et al., 2004; Peyrat-

Guillard et al., 2014).

Ainsi, la possibilité de traiter les focus groups par la lexicométrie est établie. Cependant, les apports spécifiques d'une approche quantitative sont à préciser dans un domaine dominé par les approches qualitatives dont l'analyse conversationnelle. Par exemple, le lien entre focus groups et représentations sociales est mis en avant (Jovchelovitch, 2004) et la classification descendante hiérarchique (CDH) de Reinert (1983) forme des mondes lexicaux (Ratinaud et Marchand, 2015) dont la nature est proche des représentations sociales. Nous insisterons, dans cet article, sur la place de l'individu dans le(s) groupe(s), problématique que la lexicométrie permet d'approcher par un jeu de variables adapté. Mais cette analyse suppose de préparer le corpus avec des méthodes spécifiques.

Nous présenterons ainsi une typologie des méthodes de préparation d'un corpus de focus groups en complétant les analyses de Peyrat-Guillard et al. (2014) et en mettant en exergue celles centrées sur l'individu. Puis, nous analyserons les résultats de l'expérimentation d'une de ces méthodes en montrant en quoi elle permet une compréhension des discours de l'individu dans le(s) groupe(s).

2. Typologie des partitionnements d'un corpus de focus groups

Avant de commencer le traitement lexicométrique de focus groups, le corpus exige une préparation spécifique. En effet, certaines décisions de partitionnement détermineront notamment les méthodes lexicométriques employables et les analyses possibles.

Les textes des modérateurs sont souvent supprimés du focus groups (Guerrero et al., 2009 ; Peyrat-Guillard et al., 2014) car ses interventions, dans le cadre d'un focus group servent à fluidifier les échanges sans les orienter. Cependant, il peut être conseillé de comparer les résultats avec ou sans les interventions du modérateur (Peyrat-Guillard et al., 2014).

La deuxième question porte sur la partition du corpus issu du focus group. Plusieurs méthodes existent. Une première possibilité est d'analyser le focus group comme une entité sans prendre en compte les échanges entre les individus. Soit chaque focus group constitue un texte sans distinction d'individu (Dransfield et al., 2004) ; l'argument avancé par les utilisateurs de cette méthode est de faciliter les analyses statistiques mais cela n'est pas une évidence, le nombre de segments étant stable. Soit le focus group est partitionné en thèmes à partir d'une analyse de contenu (Bengough et al., 2015) ; cette approche permet de comparer par exemple les résultats d'une analyse thématique avec celle proposée au chercheur par la lexicométrie. La deuxième famille de partition est celle qui souhaite conserver les échanges du focus group. Soit la partition peut être centrée sur les individus, dite

decrowded (Peyrat-Guillard et al., 2014) ; les textes des interventions de chaque individu sont alors rassemblés (Guerrero et al., 2009). Soit chaque intervention est considérée comme un texte, approche dite *crowded* (Peyrat-Guillard et al., 2014).

Chacune de ces méthodes a des avantages et des inconvénients. Nous ne pensons pas qu'une partition soit à privilégier mais que la décision dépend des analyses envisagées par le chercheur selon sa problématique. Dans cet article, nous nous centrerons sur la deuxième famille qui permet d'étudier l'individu dans le(s) groupe(s) et pas seulement les thèmes abordés.

3. Résultats de l'expérimentation du partitionnement par locuteur

Nous avons pu expérimenter ces méthodes de partition d'un corpus de focus groups à partir d'une recherche que nous avons menée auprès de jeunes diplômés de l'enseignement supérieur (niveaux bac+3 et bac+5). Les discussions des focus groups concernaient la communication numérique de recrutement des banques et ces jeunes diplômés échangeaient sur les dispositifs utilisés par les entreprises pour recruter. Nous avons animé puis retranscrit 10 focus groups de 6 à 7 personnes soit 67 locuteurs au total.

3.1. Préparation du corpus et partitionnement

Une fois les textes préparés (anonymisation, intégration des noms propres (BNP, Facebook, etc.) au dictionnaire, adaptation du dictionnaire selon les spécificités du discours, etc.), nous avons décidé de supprimer les interventions du chercheur car elles restaient neutres par rapport aux discours des jeunes diplômés que nous souhaitons analyser.

Nous avons alors créé une partition par tours de parole selon ce principe : (variables entre crochets)

[Groupe1, Ingénieurs , NUM1, 18ans, masc]: il y a des choses marquantes, il y a un site web où on n'a pas beaucoup d'informations et un autre site où il y a beaucoup d'informations.

[Groupe1, Ingénieurs , NUM2, 20ans, masc]: je suis d'accord avec toi.

En effet, nous souhaitons repérer des discours individuels dans les focus groups et pouvoir associer des variables de profil à un locuteur.

Les variables utilisées (tableau 1) ont été déterminées selon nos hypothèses de recherche et leur accessibilité puis ont été associées par un script automatique à chaque intervention de locuteur.

Tableau 1. Variables du focus groups associées aux locuteurs.

Num	Code variable	Valeur	Source	Description
1	num	1, 2, 3, etc.		Numéro de chaque intervenant
2	formation	3IL : école		Désignation du

Num	Code variable	Valeur	Source	Description
		d'ingénieur LPB : licence professionnelle banque		groupe
3	groupe	1, 2, 3, etc. 10 groupes au total		Numéro du groupe
4	sexe	M, F		
5	participation	TA, PA, A TA : très actif A : actif PA : pas actif	Statistiques SONAL selon le nombre d'interventions	Indicateur quantitatif de la participation de chaque intervenant
6	initial	STS, IUT	Données organisme de formation	Formation initiale des intervenants

Le corpus se présentait ainsi de cette façon pour être utilisé dans Iramuteq (Ratinaud, 2009) :

**** *num_44 *formation_LPB *groupe_1 *sexe_M *participation_ A
*initial_STS moi je veux bien commencer. Quand je suis allé sur le site de la
SG, ... Les caractéristiques du corpus obtenu et traité à l'aide du logiciel
Iramuteq sont alors les suivantes : 1876 textes allant d'une seule forme (Oui
par exemple) pour les plus courts à 126 formes ou 280 occurrences pour le
plus long, 40404 occurrences et 2094 formes au total, 21,54 occurrences par
texte en moyenne, les hapax représentent 41,26% des formes. Chaque texte
correspond alors à une intervention d'un locuteur dans un focus group.

3.2. Choix méthodologiques

Si la CDH de Reinert est la plus souvent citée dans la littérature (Duchesne et al., 2010 ; Gresillon et al., 2012; Hulin, 2013; Peyrat-Guillard et al., 2014; Brangier et al., 2015; Freitas et Luis, 2015, etc.) d'autres techniques sont impliquées comme l'analyse factorielle (Dransfield et al., 2004; Guerrero et al., 2009) ou plus rarement l'analyse de similitude (Bengough et al., 2015). Notre choix de la classification de Reinert est lié à nos hypothèses de recherche qui associent les discours de ces jeunes diplômés aux représentations sociales. Or, la CDH de Reinert (1983) favorise le repérage de représentations sociales (Ratinaud et Marchand, 2015). Nous avons effectué plusieurs CDH simples sur segments de texte en faisant varier le nombre de classes demandées, le nombre minimum de segments par classe. Nous avons choisi de retenir les formes dont la fréquence est supérieure à 3 (soit 687 formes dans ce cas) pour centrer le traitement sur les formes les plus présentes. Au terme de ces simulations, nous avons retenu une CDH qui présente 15 classes avec un taux de segments classés de 83,63%.

3.3. Exemple d'utilisation de variables, groupes et degré de participation

Chaque intervention ayant été associée à des variables de contexte, la méthode choisie permet de vérifier le lien existant entre les groupes et chaque classe repérée. Ainsi, pour ce corpus de focus groups, la classe 1 ($\text{Chi}^2=20,82$, recherche d'emploi) et la classe 12 ($\text{Chi}^2=16,76$, articles de journaux) sont associées aux étudiants de 3IL. La classe 7 ($\text{Chi}^2=32,17$, Dupuy) et la classe 13 ($\text{Chi}^2=11,44$, avantages et valeurs) sont plutôt liées au groupe des licences banques (fig. 1).

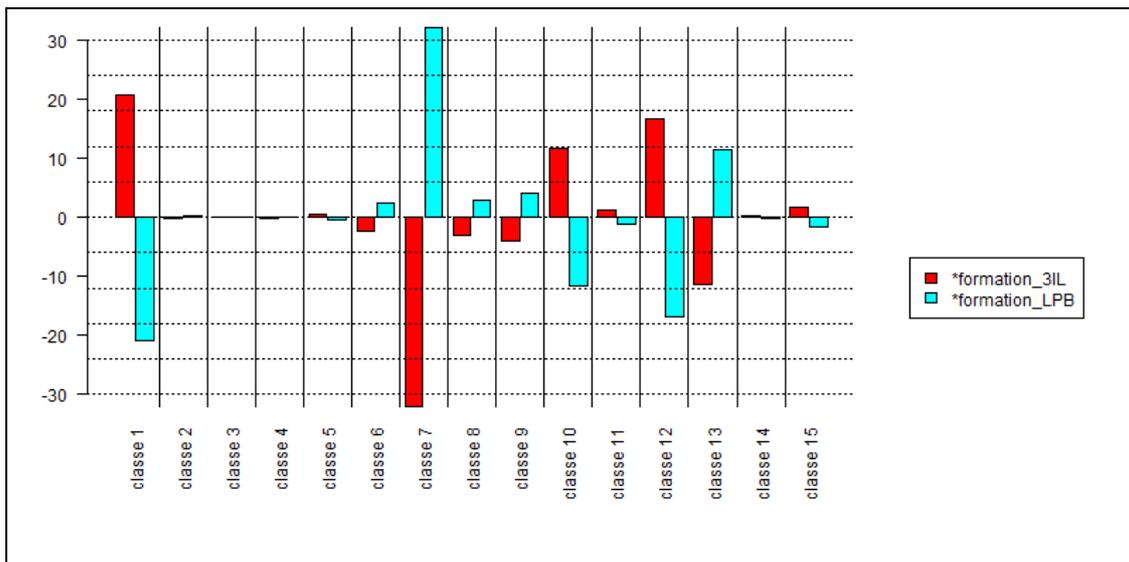


Figure 1. Chi^2 par classe pour la variable 'formation'.

De même, la variable sur la participation (tableau 1 et fig 2.) a permis d'associer certaines classes avec cette caractéristique. Les résultats de la CDH permettent ainsi de poser une hypothèse sur le degré de consensus entourant une représentation sociale.

eff. s.t.	eff. total	pourcentage	chi2	Type	forme
13	61	21.31	4.19		*participation_PA

Figure 2. Association de la classe 8 avec la variable participation.

En effet, la classe 8 sur la taille de l'organisation est associée aux locuteurs qui ont peu participé globalement (Variable PA (Peu Actif), $\text{Chi}^2=4,19$; fig. 2) comme pour la classe 3 (mobilité). Les discussions sur la recherche d'emploi (classe 1), la banque Dupuy (classe 7) ou les classements des sites internet et témoignages sont dominées par les locuteurs les plus actifs (Variable TA (Très actif) : $\text{Chi}^2=5,69$ pour la classe 1 et Variable A (Actif) : $\text{Chi}^2=7,51$ pour la classe 7). Elles peuvent être perçues comme plus conflictuelles ou engagées. Les échanges sur la taille ont ainsi laissé plus de places aux locuteurs peu

actifs avec des discussions plus consensuelles moins conflictuelles que pour des représentations moins stabilisées. Cette hypothèse renvoie alors à la structure possible de cette représentation sociale construite autour d'un noyau central stable qui exigerait des études complémentaires pour être confirmée.

3.4. Repérage de discours individuels par l'analyse factorielle de correspondance (AFC)

Le partitionnement effectué permet aussi de repérer des individus dont les discours sont différents (fig. 3) grâce à une AFC réalisée à la suite d'une CDH de Reinert. Dans ce cas, deux individus se détachent principalement : 17 et 37. Le retour au texte permet de confirmer ce repérage. L'autre intérêt est aussi de souligner des regroupements d'individus différents de leur rattachement à un focus groups. L'AFC, en mettant en évidence des ensembles de locuteurs, propose une approche qui dépasse la frontière de chaque focus groups pour proposer une analyse de l'individu dans les groupes.

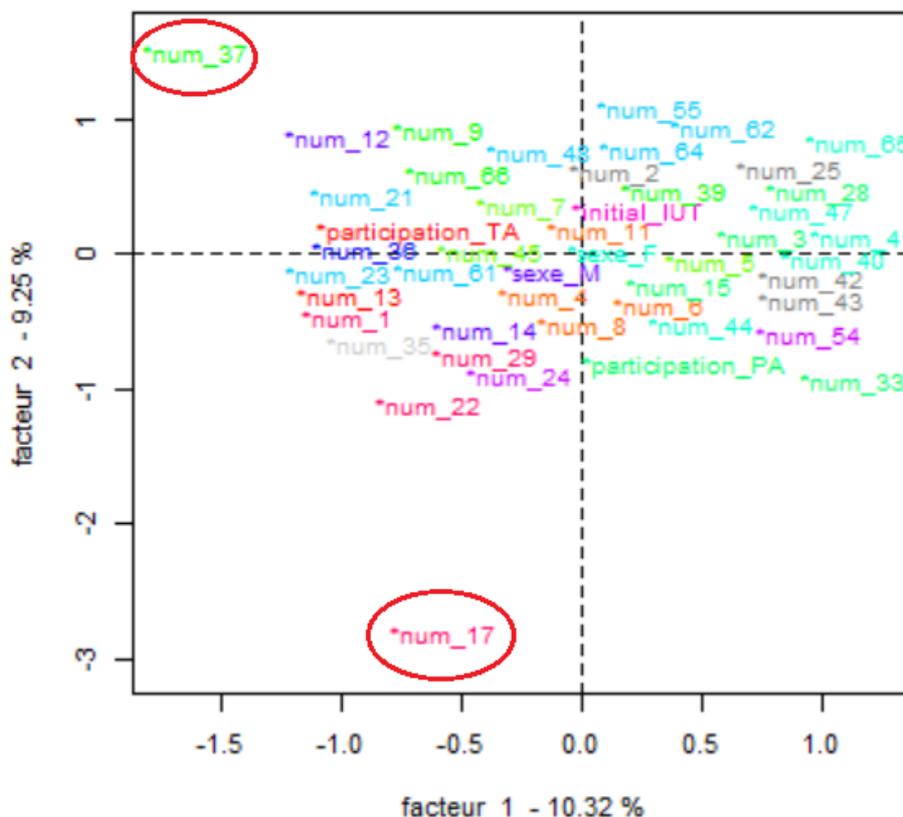


Figure 3. AFC à partir de la CDH présentant les variables (F1/F2, 19,57 % de l'inertie).

4. Conclusion

Les méthodes lexicométriques utilisées pour analyser des focus groups

dépendent notamment de la partition du corpus effectuée en amont. Dans notre recherche, l'association de variables à chaque intervention de locuteur a permis de repérer des sous-groupes d'individus à l'intérieur des focus groups, des discours d'individus isolés ou des sous-groupes associés à plusieurs focus groups qui n'apparaissaient pas de façon évidente pendant les échanges. Cette approche a cependant certaines limites. D'abord, la procédure automatisée d'association des variables utilisée dans cette expérimentation ne permet pas de repérer l'évolution des thèmes pendant la discussion, une variable repérant les tours de paroles aurait alors été nécessaire. Ensuite, le repérage des individus s'est fait sur une AFC qui explique une faible part de la variance (19,57 %) et les causes de la singularité des discours est ainsi difficile à associer à la CDH. Enfin, d'autres méthodes auraient pu être investies (analyse des antiprofiles, spécificités, similitudes, etc.).

Sans remplacer l'analyse conversationnelle qui apporte des nuances spécifiques, certaines méthodes lexicométriques peuvent ainsi permettre de comprendre le corpus différemment et compléter la compréhension de ce type de données riches et profondes en dépassant notamment la frontière de chaque focus groups et faciliter une approche transversale du sens.

Remerciements : merci à Pascal Marchand, Pierre Ratinaud et Lucie Loubère pour leur initiation à la lexicométrie et à Iramuteq.

References

- Bengough, T., Bovet E., Bécherraz C., Schlegel S., Burnand B., et Pidoux, V. (2015). Swiss family physicians' perceptions and attitudes towards knowledge translation practices. *BMC Family Practice*, décembre: 1–12.
- Bonneau, J., and Dister, A. (2010). Logométrie et modélisation des interactions discursives, l'exemple des entretiens semi-directifs. *Journées internationales d'Analyse statistique de Données Textuelles*, pp. 253–264.
- Brangier, E., Barcenilla, J., Bornet, C., Roussel, B., Vivian, R., and Bost, A. (2015). Prospective ergonomics in the ideation of hydrogen energy usages. In *Proceedings 19th Triennial Congress of the IEA*. Melbourne, pp. 1–2.
- Dransfield, E., Morrot, G., Martin, J.-F., and Ngapo, T.-M. (2004). The application of a text clustering statistical analysis to aid the interpretation of focus group interviews. *Food Quality and Preference*, 15(4): 477–488.
- Duchesne, S., and Haegel, F. (2014). *L'entretien collectif*. Armand Colin. Paris.
- Duschesne, S., Florence Haegel, Elizabeth FRAZER, Virginie Van Ingelgom, and Guillaume Garcia, André-Paul Frogner. (2010). Europe between integration and globalisation social differences and national frames in the analysis of focus groups conducted in France, francophone Belgium and the United Kingdom. *Politique Européenne*, 30(1): 67–105.

- Freitas, E. A. M., and Luis, M. A. V. (2015). Perception of students about alcohol consumption and illicit drugs. *Acta Paul Enferm.*, 28(5): 408–414.
- Gresillon, E., and Marianne Cohen, Julien Lefour, Lydie Goeldner et Laurent Simon. (2012). Les trames vertes et bleues habitantes : un cheminement entre pratiques et représentations. L'exemple de la ville de Paris (France). *Développement Durable et Territoires*, 3: 2-17.
- Guerrero, L., Guàrdia, M., and Xicola, J. (2009). Consumer-driven definition of traditional food products and innovation in traditional foods. A qualitative cross-cultural study. *Appetite*, 52(2): 345–354.
- Hulin, T. (2013). Enseigner l'activité « écriture collaborative ». *Tic&société*, 7(1): 89–116.
- Jovchelovitch, S. (2004). Contextualiser les focus groups : comprendre les groupes et les cultures dans la recherche sur les représentations. *Bulletin de Psychologie*, 57(3): 245–261.
- Lebart, L., and Salem, A. (1988). *Analyse statistique des données textuelles*. Dunod. Paris.
- Peyrat-Guillard, D., Lancelot Miltgen, C., et Welcomer, S. (2014). Analysing conversational data with computer-aided content analysis: The importance of data partitioning. *Journées internationales d'Analyse statistique des Données Textuelles*, pp. 519–530.
- Pélessier, D. (2016), Pourquoi et comment utiliser la lexicométrie pour l'analyse de focus groups ?, *Présence numérique des organisations*, 11/07/2016.
- Ratinaud, P. (2009). Iramuteq. Lerass.
- Ratinaud, P., and Marchand, P. (2015). Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014). *Mots. Les Langages du Politique*, 108(2): 57–77.
- Reinert, M. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Les Cahiers de L'analyse Des Données*, 8(2): 187–198.
- Rouré, H., and Reinert, M. (1993). Analyse d'un entretien à l'aide d'une méthode d'analyse lexicale. *Journées internationales d'Analyse statistique de Données Textuelles*. ENST, Paris, pp. 418-42