



HAL
open science

Stratégies du voyageur : analyse croisée d'entretiens semi-directifs

Amélie Martin, Sonia Adélé, Coralie Reutenauer

► **To cite this version:**

Amélie Martin, Sonia Adélé, Coralie Reutenauer. Stratégies du voyageur : analyse croisée d'entretiens semi-directifs. 13ème Journées internationales d'Analyse statistique des Données Textuelles (JADT 2016), Jun 2016, Nice, France. <hal-01360164>

HAL Id: hal-01360164

<https://hal.science/hal-01360164v1>

Submitted on 5 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Stratégies du voyageur : analyse croisée d'entretiens semi-directifs

Amélie Martin^{1,4}, Sonia Adélé^{2,3}, Coralie Reutenauer⁴

¹ INALCO + Paris – France

² IFSTTAR + Marne-la-Vallée – France

³ IRT SystemX + Palaiseau – France

⁴ SNCF Innovation & Recherche + Paris – France

Abstract

This article fosters the understanding of passenger behaviours following a research conducted within three commuter rail lines of Paris. As a starting point, semi-structured interviews with passengers outlined the behavioural key-drivers. The consolidated corpus has then been analyzed through several steps by combining two methods: a manual thematic analysis widespread in social psychology and a textometric analysis. The resulting findings are presented herein as well as insights on the joint-use of the two aforementioned approaches, both for richness of analysis and analytical protocols.

Résumé

La présente communication s'inscrit dans le cadre de travaux de recherche sur la compréhension du comportement des passagers de trois lignes de train franciliennes. Ces travaux ont débuté par une première phase exploratoire à partir d'entretiens semi-directifs afin d'identifier les déterminants des comportements des passagers. Le corpus d'entretiens a été analysé en plusieurs étapes en associant deux méthodes d'analyse : une analyse thématique manuelle traditionnellement utilisée en psychosociologie et une analyse textométrique. Nous présentons les résultats d'analyse ainsi qu'une réflexion autour de l'usage conjoint de ces deux approches, du point de vue de la richesse d'analyse et du protocole d'analyse.

Mots-clés : mobilité, transports, analyse de contenu, entretiens semi-directifs, textométrie, méthodes hybrides

1. Introduction

Cette communication restitue certains des résultats de travaux de recherche du projet multi-partenarial MIC (Modélisation, Interopérabilité, Coopération). Ce projet vise l'élaboration d'un outil de modélisation et de simulation permettant d'appréhender la mobilité multimodale (c'est-à-dire avec choix et enchaînement de plusieurs modes de transport) et le report modal en zone dense, en particulier en situation perturbée. Plus précisément, l'étude, conjointement menée par SNCF Transilien, l'IFSTTAR, SNCF Innovation & Recherche et SystemX, concerne les lignes Transilien N, U et C du sud-ouest de la région parisienne passant par la gare de Versailles Chantiers.

Pour parvenir à une modélisation réaliste, une enquête en deux phases sur la mobilité en situation normale et perturbée a été menée auprès des usagers des lignes concernées : une phase d'entretiens, dont l'analyse croisée est l'objet de cet article, puis une phase plus quantitative d'enquête par questionnaires fermés.

Les entretiens semi-directifs avaient une visée exploratoire et préparatoire des questionnaires ultérieurs. Ils permettaient aussi d'obtenir des résultats exploitables pour comprendre le comportement et les choix des usagers des transports en commun placés face à une

perturbation selon leur type (pendulaire, occasionnel ou touriste) en mettant au jour leurs opinions, jugements, attitudes, perceptions. L'analyse a été réalisée par croisement de deux approches : une analyse thématique manuelle et une analyse textométrique. Nous présentons ici la méthodologie utilisée et les principaux résultats, en insistant sur l'apport de la textométrie à l'analyse psychosociologique.

2. Etat de l'art

2.1. Mobilité : des comportements en mutation, une connaissance du voyageur à approfondir

Il est communément admis que la situation de perturbation des transports correspond à une « situation-problème » à résoudre, particulièrement intéressante à étudier. Les études sur le comportement des voyageurs en situation perturbée sont assez rares. La majorité des études réalisées concernent les automobilistes et sont plutôt expérimentales. Le conducteur, plus rarement le voyageur (Bai, 2014), est interrogé via un questionnaire sur des scénarios hypothétiques de perturbation (dit en préférences déclarées) où il doit effectuer un choix (Emmerink et al., 1996). Bien que cette technique présente de nombreux intérêts, notamment au niveau du contrôle des paramètres de la situation de choix, elle n'est pas dénuée d'inconvénients tels qu'une faible validité empirique. En effet, ce type de questionnaire ne restitue que faiblement la situation dans laquelle l'individu est amené à faire un choix (Bonsall et Palmer, 1999). L'entretien permet au contraire d'accéder à la complexité des situations rencontrées par les individus mais aussi au sens que donnent ces individus à leurs actions et aux événements qui les entourent. Cela est d'autant plus pertinent que les choix sont réalisés à partir d'une rationalité limitée (par les capacités cognitives, perceptives...) (Pronello et Camusso, 2011 ; Simon, 1955). Par ailleurs, le choix est réalisé par un individu unique dans un contexte spécifique dont les caractéristiques sont à appréhender. Différentes dimensions ont été identifiées pour leur influence sur le choix effectué par le passager des transports collectifs (Bovy et Stern, 1990 ; Fiorenzo-Catalano et al., 2003 ; Last et Manz, 2003). Elles relèvent de : l'offre de transport (les modes disponibles, leur fréquence, etc.) ; l'individu (ses caractéristiques sociodémographiques, son expérience, son niveau d'habitude, sa personnalité et notamment son attitude vis à vis du risque, ses préférences, etc.) ; le type de trajet réalisé (le moment, le caractère contraint ou non, le motif, la longueur, etc.) ; et enfin, l'information reçue, notamment concernant la perturbation (le moment de réception, le contenu, la fiabilité).

2.2. L'analyse des entretiens

2.2.1. L'analyse thématique manuelle

Dans les sciences humaines et sociales, l'analyse de contenu est définie comme un ensemble de techniques d'analyse des communications (Bardin, 1977 ; Trudel et Gilbert, 1999). Il existe une multitude de manières de la réaliser en fonction notamment du type de document à analyser, des visées du chercheur et de son ancrage théorique. Le plus souvent, l'analyse de contenu comprend trois étapes chronologiques (Wanlin, 2007) : la préanalyse (lecture flottante, repérage d'indices, découpage en unités significantes), l'exploitation (catégorisation, comptage) et l'interprétation. Sous une apparente neutralité, ce processus d'analyse de contenu est affecté continuellement par l'analyste qui oriente la forme des résultats, notamment par ses choix des regroupements d'idées. Ce codage va être construit et amélioré au cours de la lecture dans une approche constructiviste (Fallery et Rodhain, 2007). Souvent, le thème est retenu comme unité. Berelson (1952) le définit comme « une affirmation sur un sujet [...] un résumé ou une phrase condensée, sous laquelle un vaste ensemble de formulations singulières

peuvent être affectées ». Selon Unrug (1974), le thème a surtout une réalité d'ordre psychologique (Unrug, 1974). Il est une abstraction (Fallery et Rodhain, 2007). L'analyse thématique peut employer des outils assistant le codage (CAQDAS, Computer Assisted Qualitative Data Analysis Software), et plus particulièrement l'assignation de thèmes aux extraits et l'organisation de ces extraits pour l'analyse. Enfin, le chercheur utilise l'inférence pour traiter le contenu du corpus qu'il analyse. Cette inférence permet, selon Bardin (1977), le passage de la description du corpus à son interprétation.

2.2.2. L'analyse textométrique

Les approches textométriques, basées sur l'analyse statistique du lexique du corpus, reposent sur une autre dynamique interprétative qui favorise la contextualité et l'intertextualité (Pincemin, 2012) : contrairement à la démarche linéaire de préanalyse, elles favorisent une exploration délinéarisée, à travers des regroupements (cooccurrences, concordances), l'identification de saillances (indicateurs statistiques) et une navigation de l'analyste entre différents paliers d'observation. Le parcours des contenus est guidé par un fil interprétatif : on rebondit sur une idée émergente pour l'approfondir et la conforter, en jonglant entre des niveaux d'analyse locale (retour au texte, dans son déroulement linéaire ou à travers les concordanciers qui affichent mais aussi regroupent des contextes plus réduits) et globale (exploration du lexique, des indicateurs statistiques associés tels que les fréquences ou les spécificités, de la structuration des unités textuelles ou lexicales par classification, AFC...).

Même sur des volumes de données moyens, la textométrie permet une « lecture assistée » (Fallery et Rodhain, 2007) combinant découverte de pistes ou phénomènes, et retours au texte pour affiner ou confirmer ces découvertes. Ces principes d'exploration ont déjà été mis en œuvre pour analyser des entretiens ou par recoupement avec une approche psychosociologique (Poudat et Boulard, 2014).

2.2.3. Les analyses croisées

La question de l'interconnexion entre textométrie et analyses thématiques anime depuis quelques années les communautés concernées par l'analyse de contenus textuels. Même si elles semblent s'opposer, ces deux approches sont complémentaires (Glady et Leimdorfer, 2015) puisqu'elles permettent de combiner leurs avantages en limitant leurs inconvénients (par exemple, l'excès de parti pris ou le traitement décontextualisé). En effet, concernant le type d'unité d'analyse (Della Ratta-Rinaldi, 2013, cité par Glady et Leimdorfer, 2015), la textométrie se base sur l'analyse d'unités lexicales, à travers leurs occurrences et cooccurrences. Dans l'approche thématique, on partage le texte en parties définies auxquelles on attribue des étiquettes sémantiques. Concernant le moment de l'analyse statistique (Fallery et Rodhain, 2007), l'analyse statistique textométrique peut être réalisée dès le départ pour ouvrir le champ du chercheur en explorant sans a priori le corpus.

Finalement, la plupart des auteurs appellent à « un usage complémentaire de ces approches dans une démarche [...] composée de nécessaires cycles itératifs entre grilles et textes, entre codage et décodage, entre extraction et validation. » (Fallery et Rodhain, 2007, p. 14). Ainsi, l'approche textométrique permet de vérifier et d'objectiver des résultats obtenus par un autre type d'analyse, de traiter un corpus de grande taille, d'explorer des données sans a priori, de fournir des pistes de recherche (Naulin, 2013 ; Desmarais et Moscarola, 2004). Inversement, l'approche thématique permet de mettre en évidence des dimensions plus implicites non saisies par l'approche textométrique (Scholtz et Angermuller, 2013). Dans ces cas, le moment

d'usage des deux méthodologies est différent. Selon l'objectif, la textométrie est employée préalablement ou ultérieurement à l'analyse thématique (après découpage en catégories).

Dans tous les cas, l'usage conjoint de ces méthodes ouvre de façon encore plus prégnante de nombreuses questions qui doivent animer le chercheur sur les transformations qui sont appliquées aux données pour passer d'un matériel brut à une interprétation.

3. Corpus

3.1. Présentation des entretiens

Les participants ont été choisis selon plusieurs critères jugés significatifs par rapport au problème étudié : la fréquence d'usage du réseau Transilien, le motif d'usage du réseau (loisir ou contraint) et la ligne la plus utilisée pour les utilisateurs réguliers (N, U et C). Vingt-sept entretiens semi-directifs ont été menés auprès de 20 usagers réguliers, 4 utilisateurs occasionnels, 2 touristes français et 1 touriste américain francophone. L'échantillon se compose de 17 femmes et 10 hommes âgés de 19 à 68 ans (moyenne d'âge 35 ans). Il compte 21 cadres, 4 étudiants, 1 retraité et 1 agent de maîtrise.

Deux guides d'entretiens différents ont été utilisés, un premier pour les franciliens et un second pour les touristes. Tous les guides comprenaient deux parties distinctes correspondant aux déplacements en situation normale et en situation perturbée. Ces entretiens duraient en moyenne 30 min (minimum 21 min, maximum 53 min) et étaient réalisés de préférence en face à face. Ils ont été retranscrits intégralement pour analyse.

3.2. Prétraitements

Deux types de prétraitements permettent de préparer l'exploration textométrique : un découpage du corpus selon un ou plusieurs paliers textuels (phrase, paragraphe, fenêtre de mots, etc.) pour structurer l'analyse textuelle ; et un découpage du corpus en unités élémentaires qui seront nos observables. Dans notre cas d'étude, le corpus est subdivisé en entretiens. Chaque entretien est constitué des réponses concaténées des participants, sans les questions du conducteur d'entretien. Le corpus, qui compte environ 110 000 formes, a été lemmatisé et étiqueté en parties du discours par Treetagger (Schmid, 1994). Le nombre de formes par entretien varie de 1600 à 7300. Lors de nos analyses, nous avons exploité les formes ou les lemmes de façon complémentaire.

4. Méthodologie et résultats

4.1. Analyse thématique

Selon la technique décrite en 2.2.1, les entretiens ont été découpés en unités significantes¹ auxquelles a été affecté un thème et un seul. L'organisation hiérarchique de ces thèmes forme une arborescence composée d'étages thématiques (Tableau 1). L'étage le plus fin est analysé pour constituer les résultats et interprétations et cette analyse met en perspective les différents composants d'un même thème. Le logiciel Sonal (Alber, 2010) a été utilisé à cette étape.

Le corpus compte 1123 extraits. Il se répartit majoritairement entre les thèmes relatifs aux trajets en situation normale (25 %) et aux perturbations (43 %). Les corpus correspondant aux

¹ Le découpage correspond celui de la période (Rastier, 2005) : ce sont des unités sémantiquement cohérentes qui peuvent correspondre à une phrase, une demi phrase, plusieurs phrases, etc. sans respecter forcément un découpage classique de la phrase.

différents types d'usagers ont été étudiés séparément et comparés. A l'intérieur de chaque sous-thème, des codages spécifiques ont été effectués. Par exemple pour le thème "récit d'expérience", chaque récit a été coté sur deux critères : son niveau de détail (de D1, peu de détails à D3, beaucoup de détails) et le niveau de gravité de l'évènement relaté (de G1 peu grave à G3 très grave). Cette analyse montre des différences entre les perturbations rapportées. Les touristes donnent peu de détails dans leurs récits (78% classés en D1) contrairement aux réguliers (37 % classés en D1). En ce qui concerne la gravité, les réguliers sont ceux qui racontent le plus d'évènements graves (53 % classés en G3 contre 40 % environ pour les autres usagers). D'une façon générale, le discours des franciliens (utilisateurs réguliers ou occasionnels) contraste clairement avec celui des touristes. A une moindre échelle, une différence entre le discours des utilisateurs réguliers et occasionnels émerge.

Thèmes	Sous-thèmes	N	
Expérience	Historique du trajet	44	
	Connaissance / expérience / habitude	51	
	Méconnaissance / inexpérience	25	
Trajets en situation normale	Variabilité des trajets aller et retour	10	
	Vécu	72	
	Motifs des choix	Choix de mode (dont non choix) et facteurs relatifs	75
		Choix d'itinéraire (dont horaires) et facteurs relatifs	123
Information	Information de planification en situation normale	50	
	Information en situation perturbée	Habitude de consultation	68
		Fiabilité / satisfaction	66
Caractéristiques du système	Caractéristiques des TC	29	
	Caractéristiques de la clientèle	3	
Perturbations	Déterminants du choix dont les contraintes perso et pro	69	
	Alternatives perçues / description des plans A, B, C... dont télétravail	124	
	Ressentis vis-à-vis des perturbations (effort, pénibilité, impuissance, maîtrise)	67	
	Récit d'expériences	107	
	Généralités (fréquence, motif, types...)	106	
	Gestion de l'exploitant et sa perception	9	
Les autres usagers	Comportement des autres	16	
	"On-dit"	9	
Total		1123	

Tableau 1 : Arborescence thématique (N : nombre d'extraits total)

4.1.1. Utilisateurs réguliers

Les utilisateurs réguliers se rejoignent sur les caractéristiques suivantes :

- Le trajet aller est différent de celui du retour, moins contraint temporellement (surtout pour les jeunes).
- Ils se déplacent avec beaucoup de routines développées avec l'expérience aussi bien en situation normale qu'en situation perturbée. Certains usagers "experts" disposent d'un panel de solutions différentes face aux perturbations.
- Le vécu des trajets montre que la réduction du temps de trajet n'est pas le seul critère de choix puisque ce temps peut être considéré positivement, notamment si le confort permet de réaliser des activités.
- Les stratégies en cas de perturbation dépendent de la ligne utilisée, de l'offre de transport, des opportunités professionnelles (par exemple, télétravail), du sens du déplacement, de la contrainte horaire, de l'information reçue et de son moment, du motif de déplacement.

4.1.2. *Utilisateurs occasionnels*

Les occasionnels ont une connaissance assez restreinte du réseau francilien. Aussi, le recours à l'information est fréquent. Les itinéraires sont choisis surtout pour limiter le temps de trajet mais cela n'est pas une constante. Les perturbations sont des réalités un peu lointaines pour ces usagers (cf. codage des récits de perturbation ci-dessus) et leurs stratégies sont spécifiques (annulation des activités de loisir).

4.1.3. *Touristes*

Les touristes réalisent des recherches préalables soignées de leurs itinéraires. Des critères différents des autres groupes d'utilisateurs sont considérés, comme la sécurité et l'évitement des correspondances lors de voyages en groupe. Ils gardent un souvenir assez fort des perturbations rencontrées compte tenu de leur vécu pénible. Si possible, l'attente reste la solution privilégiée. L'autre solution se construit soit en suivant la foule soit en cherchant une aide humaine (personnel ou à défaut autres usagers).

4.2. *Analyse textométrique*

L'exploration textométrique du corpus a été effectuée en deux temps : la découverte et l'organisation d'idées ou de sujets évoqués dans le discours, ainsi que d'éventuels signaux faibles, afin d'identifier des paramètres, critères ou facteurs qui influent sur les pratiques des voyageurs ; et l'approfondissement des sujets préalablement identifiés lors de l'analyse thématique.

La découverte des sujets évoqués comporte deux volets :

- Le premier volet consiste à détecter des sujets saillants en parcourant le lexique. Ce parcours est effectué d'abord par fréquence décroissante. La tête de liste permet de dégager des thématiques, qui apparaîtront à travers des mots de forte fréquence et à travers des familles de mots partageant un même trait sémantique qu'on cherchera à expliciter. Le milieu ou la fin de liste sera elle utile pour confirmer ou enrichir ces thématiques (mots moins fréquents mais appartenant au même champ lexical que des mots relevés en tête de liste) ou pour dégager des signaux faibles, qui seront portés par un mot ou un ensemble de mots peu fréquents partageant un même trait sémantique (apparition plus diffuse et plus bas dans la liste). Dans un second temps, un autre parcours par ordre alphabétique est effectué : des formes lexicales de même racine se suivront et se regrouperont ainsi naturellement. Moins rigoureux qu'un regroupement par lemmes ou des regroupements morphologiques, ce procédé a le mérite d'être simple, rapide, et de permettre un meilleur contrôle de la variation morphologique.
- Le deuxième volet vise à confirmer les sujets détectés et à mieux les comprendre de deux façons. Il s'agit tout d'abord d'utiliser des concordanciers pour regrouper et lire l'information autour de mots-pivots. L'exploration du lexique donne des pistes de sujets qui demandent à être confirmées par une mise en contexte. La lecture des contextes via les concordanciers autour d'un mot ou d'un ensemble de mots-pivots relatifs à un même sujet permet de gérer l'ambiguïté lexicale des mots isolés et de distinguer les emplois, de conforter leur interprétation et d'affiner la compréhension du sujet saillant. La deuxième manière d'approfondir une idée saillante est de dégager et d'analyser les cooccurrents du ou des mots porteurs de cette idée (Firth, 1957). En pratique, les cooccurrents sont obtenus par analyse des mots statistiquement saillants dans le sous-corpus des contextes du mot-cible par rapport à l'ensemble du corpus. Dans notre cas d'étude, nous avons peu utilisé cette technique car, pour être efficace

au niveau statistique, elle nécessite de disposer d'un regroupement de contextes suffisamment volumineux.

Le deuxième temps de l'analyse textométrique (l'approfondissement de sujets préalablement identifiés lors de l'analyse thématique) a reposé sur la constitution de sous-corpus composés des segments manuellement étiquetés lors de l'analyse thématique. Ces sous-corpus ont ensuite été comparés les uns aux autres, ainsi qu'au reste du corpus, afin de faire émerger un vocabulaire qui leur est caractéristique. Le vocabulaire caractéristique est obtenu par calcul des spécificités (Lafon, 1980) des mots du sous-corpus par rapport à l'ensemble du corpus.

Le corpus d'entretiens a été exploré grâce au logiciel de textométrie TXM (Heiden et al., 2010), afin de trouver des sujets récurrents et d'autres types d'informations permettant de qualifier le parcours du voyageur. En lien avec la méthodologie décrite ci-dessus, nous avons utilisé cinq fonctionnalités du logiciel TXM : le lexique (liste de toutes les formes du corpus associées à leur fréquence), l'index (qui permet d'effectuer des requêtes ciblées sur le lexique), le concordancier (qui permet de faire apparaître une forme ou un groupe de formes en contexte), le calcul des cooccurrents, et le calcul de spécificités.

4.2.1. Découverte de sujets qualifiant les pratiques des voyageurs en situation perturbée

Pour qualifier les pratiques du voyageur, nous avons exploré successivement le lexique des noms, des adjectifs, puis des verbes selon l'étiquetage de Treetagger (Schmid, 1994). Pour chaque classe grammaticale, le lexique a été parcouru manuellement par fréquence décroissante. Les formes porteuses de traits sémantiques récurrents ont été regroupées en thématiques, dont certaines sont présentées ci-dessous. Les informations qualitatives trouvées dans chaque regroupement ont ensuite été vérifiées et approfondies par analyse des concordanciers et des cooccurrents (Lafon, 1980).

Le parcours : informations relatives aux modes de transports. Les modes de transport observés se distinguent par leur grande diversité (modes « doux » comme le vélo et modes plus traditionnels, transports en commun, transports « collaboratifs » comme le covoiturage, véhicules personnels, etc.), reflet des habitudes multimodales des voyageurs Transilien : si les modes propres à la gare de Versailles-Chantiers (lignes de trains Transilien, RER, bus) sont les plus fréquents, on retrouve également d'autres types de modes (voiture, vélo, taxi, Autolib, Vélib) moins caractéristiques de la zone focus des entretiens (Tableau 2).

Nom	F	Nom	F	Nom	F	Nom	F	Nom	F
train	504	transports	94	commun	45	N	12	D	5
bus	352	pied/pieds	62	K	38	taxi	12	véhicule	5
ligne	278	vélo	61	marche	38	tram	12	covoiturage	4
trains	147	U	52	transport	38	navette/navettes	15	T3	4
métro/métros	131	lignes	47	B	27	Vélib'	11	mode	2
voiture	118	Transilien	46	L	25	car/cars	10	moto	2

Tableau 2 : Modes de transport et leurs fréquences au sein du corpus

Le système d'information. Les formes relatives aux systèmes d'information montrent, d'une part, la diversité des supports utilisés par les utilisateurs (plans et cartes pour construire son itinéraire, GPS pour s'orienter, annonces avec dépersonnalisation – *ils, SNCF, une voix qui te parle du ciel*) et, d'autre part, la diversité des sources (internet, smartphone, humain, matériel

tel que les panneaux, hauts-parleurs, etc.). Deux formes d'accès à l'information apparaissent : passive, comme les SMS, les notifications, les annonces sonores, et active, où l'on cherche de l'aide humaine (*agent, guichet, [demander à une] personne*), regarde autour de soi pour trouver les panneaux d'affichage, consulte une application ou un site internet (*[regarder], par internet, sur vianavigo, sur le site du Transilien, les itinéraires, la RATP*), etc. En partitionnant le corpus et en effectuant un calcul de spécificités pour contraster le vocabulaire de chaque individu, on remarque également que des profils plus « experts » émergent : ce groupe d'individus cite très précisément des lieux (arrêts, villes, gares) et des noms de lignes (*bus K, bus 91, T3, Dreux-Paris, etc.*) par rapport à d'autres groupes qui restent plus évasifs.

Les déterminants du voyage. Les déplacements professionnels, désignés par des termes nombreux et fréquents (*travail, entreprise, collègue, fac...*), apparaissent comme des facteurs explicatifs majeurs de la mobilité. Mais, conformément à l'analyse de Vincent-Geslin (2012), d'autres motifs de déplacement ressortent également : le retour au domicile et surtout l'importance de la sphère familiale et de proches (*enfants, mari, amis...*) soulignent le rôle fondamental de l'ancrage familial dans les mouvements pendulaires (c'est-à-dire les déplacements quotidiens domicile-travail). D'autres facteurs qui influent sur le voyage apparaissent également, tels que le confort, la météo, le coût, l'ambiance, ou encore des contraintes physiques et matérielles (bagages, blessures, chaussures adaptées ou non, équipements SNCF, absence de permis de conduire).

Les autres. La notion de l'autre est exprimée sous quatre formes. Elle apparaît tout d'abord en termes de foule, de masse, qui correspond à un autre « subi » le plus souvent (*flux, flot, affluence, cohue, foule, fréquentation, rush, archi [plein]*). « L'autre » se présente également comme un individu auquel on s'identifie, que ce soit en termes de classe d'individus (voyageur, usager), ce qui permet d'exprimer des généralisations ou des lieux communs, ou que ce soit à travers un rapprochement (*je vois à peu près les mêmes têtes, tous les matins*).

Le ressenti du voyageur. Cette thématique s'appuie sur des adjectifs portant sur tout ou partie d'un itinéraire (un train, un mode en particulier, un horaire, etc.), et qui apportent des informations sur le vécu du voyageur, et sur ses critères de choix et de jugement. Ces adjectifs ont été sélectionnés, regroupés et analysés à partir de leurs occurrences au comparatif (*plus* ou *moins* + ADJECTIF). Nous avons groupé les adjectifs obtenus en cinq catégories de critères :

- **Parcours** : ces adjectifs se focalisent sur la temporalité de l'itinéraire, tels que *rapide* (22), *long* (25), *court* (9), *proche* (7), *fréquent* (9), *direct* (6), *lent* (1), *régulier* (1) ;
- **Facilité** : ces adjectifs qualifient le parcours en termes de complexité réelle de réalisation, tels que *simple* (16), *facile* (10), *pratique* (10), *difficile* (2), *arrangeant* (1), *complexe* (1), *contraignant* (1), *fastidieux* (1), *intéressant* (4), *optimal* (1) ;
- **Confort** : ces critères s'attachent aux conditions de voyage, avec des adjectifs comme *agréable* (8), *confortable* (5), *fatigant* (3), *sympa(thique)* (6), *reposant* (2), *convivial* (1), *glauque* (1), *propre* (2), *serré* (1), *sombre* (1), *tranquille* (1) ;
- **Aspects économiques** : *avantageux* (4), *onéreux* (2), *cher* (1), *coûteux* (1), *économique* (1) ;
- **Confiance/Fiabilité** : jugement sur la régularité d'un mode/itinéraire, comme *sûr* (2), *fiable* (3), *variable* (1), *aléatoire* (1).

Les stratégies adaptatives. Des formes (notamment les verbes) relatives aux alternatives et aux contraintes apportent un éclairage sur la manière dont les individus déterminent et

enclenchent leurs « plans B ». Ces verbes servent de pointeurs vers des informations concernant les contraintes externes subies par le voyageur (*supprimer, bloquer, arrêter*, le plus souvent au passif, renvoient à des situations de circulation perturbée), mais aussi sa liberté de choix et ses stratégies d'organisation (*planifier, anticiper, décider, préférer, privilégier, éviter*). A titre d'exemple, les concordances autour de *éviter* indiquent des stratégies pour éviter non seulement le retard, mais aussi pour réduire le facteur stress. Le contraste entre les différents groupes d'individus révèle également des indices d'une plus grande flexibilité chez les jeunes adultes (23-30 ans), avec des verbes comme *varier* (*ça peut varier, je peux prendre le train de 8 h 00 ou le train de 8 h 30*) ou *dépendre* (*ça dépend de ce que j'avais à faire la journée*).

4.2.2. Approfondir des sujets pré-identifiés

Au cours de l'analyse thématique, des extraits du corpus (unités signifiantes de type période : une phrase, plusieurs phrases ou un morceau de phrases) ont été thématiques par l'analyste. Ces extraits ont été regroupés pour former des sous-corpus qui peuvent être comparés les uns aux autres au moyen d'un calcul de spécificités afin de faire émerger un vocabulaire propre à chaque thème. Nous présentons ici trois exemples de thèmes.

Thème 1 : déterminants des choix en situation perturbée

On dégage cinq groupes de mots ou expressions : ceux qui évoquent la passivité (*attendre, de toute façon, attente*), la prise d'informations externes (*ils disent, annonce*), l'activité (*(de) trouver, solution, essaie de*), les contraintes et blocages (*ne peux pas, rendez(-vous), enfants, sachant que*) et enfin les plans ou alternatives (*si, ça dépend*).

Thème 2 : habitude de prise d'informations en situation perturbée

On observe une forte prééminence de la prise d'informations visuelles (par rapport aux annonces par exemple) avec la présence de *je regarde, je ne regarde pas, vois, application*. On note également la présence de la prise d'informations auprès des agents : *en gare, personnel, guichet, demander...*

Thème 3 : description factuelle de l'itinéraire

Les formes les plus spécifiques de cette partie laissent entrevoir une constance dans le choix des mots utilisés pour décrire un itinéraire en transports en commun. Une trame très claire se dégage : « je prends le/la [ligne ou mode] à [information temporelle/géographique] pendant [durée] » + connecteur (ex. : *ensuite*).

4.3. Complémentarité des deux méthodes

4.3.1. Au niveau des contenus

La textométrie produit des résultats concordants ou complémentaires à l'analyse thématique. Tout d'abord, certains des résultats obtenus font écho à l'analyse thématique. Il s'agit notamment de l'importance des contraintes externes qui influencent le déplacement et du rôle relatif du temps de trajet dans les choix d'itinéraire. La fréquence élevée des adjectifs relatifs au locuteur met en évidence l'importance du vécu des situations de transport dans le choix préférentiel d'une stratégie. Ensuite, certaines informations nouvelles sont identifiées. L'analyse textométrique met notamment en avant les types d'informations recherchées. Également, les noms relatifs aux stratégies adaptatives mettent en évidence une différence non perçue lors de l'analyse thématique entre le calculé, décidé de manière à optimiser son déplacement, et l'instinctif lorsque la situation est floue. Enfin, la structuration des contenus

pour un thème particulier est plus claire grâce à l'analyse textométrique. C'est le cas par exemple pour les stratégies vis à vis de l'information avec une différenciation des informations *push* (fournies à tous, stratégies passives) ou *pull* (cherchées par l'utilisateur individuellement) ou encore des informations sur soi (carte, applications, etc.) ou dans l'environnement (regarder autour de soi, panneaux, etc.).

4.3.2. Au niveau méthodologique

Les approches thématiques et textométriques convergent, se complètent, se parlent et peuvent intervenir de façon utile à différentes étapes de l'analyse. Par croisement réitéré au fil de l'analyse, elles permettent de construire une interprétation plus riche des données en compensant leurs limites respectives. Elles représentent des manières différentes de lire le discours des usagers.

Dans ce cas précis, la textométrie, moins focalisée que l'analyse thématique, a plusieurs fonctions essentielles pour l'analyse psychosociologique. D'une part, menée en amont, elle permet d'améliorer ou de préciser la catégorisation thématique et les hypothèses formulées ou encore d'identifier des pistes non perçues et de poser de nouvelles hypothèses à vérifier dans l'étape suivante (questionnaire). D'autre part, menée en aval (après le découpage en catégories thématiques), elle a un rôle de contrôle de ce découpage. Le thème est alors considéré comme une donnée de contexte. L'analyse textométrique recherche le vocabulaire spécifique de ces parties de corpus et met en évidence, le cas échéant, les ancrages lexicaux de l'analyse thématique. En plus d'une fonction de vérification, cette analyse ultérieure permet de mettre au jour des ancrages qui n'auraient pas été perçus et élargit la compréhension des catégories définies. Enfin, l'analyse textométrique permet d'approfondir les analyses thématiques en confrontant les thèmes définis aux spécificités du lexique utilisé. En effet, les mots utilisés, tout en ayant le même sens (perçu par l'analyse thématique) peuvent indiquer des différences intéressantes entre individus selon leurs caractéristiques.

Réciproquement, l'analyse thématique a permis de conforter les résultats de l'analyse textométrique et ainsi de valider la méthodologie élaborée de manière exploratoire. Elle a également permis d'aller plus loin en résolvant un des problèmes de l'analyse outillée de corpus : le découpage du corpus en unités structurantes. En effet, l'identification par l'analyste de passages thématiques permet de construire des sous-corpus qui peuvent ensuite faire l'objet de nouvelles explorations qualitatives. Enfin, l'analyse thématique est intervenue pour accélérer ou rediriger les explorations en pointant des indices ou des informations intéressantes qui ont ensuite servi d'amorces à de nouvelles explorations.

5. Conclusion et perspectives

L'analyse croisée d'entretiens relatée dans cette contribution avait un double but. D'une part, recueillir des premiers résultats sur les déterminants des choix des passagers en situation normale et perturbée, et d'autre part, préparer une seconde phase d'enquête par questionnaire auprès d'un échantillon d'utilisateurs des lignes étudiées. L'enjeu de cette phase d'enquête, en cours, est de réaliser un modèle comportemental qui pourra être intégré à un outil de simulation du système de transport multimodal.

Les résultats obtenus grâce à ces entretiens semi-directifs permettent de valider que selon le profil de l'utilisateur, ses préférences et son usage des transports mais aussi selon des éléments externes contingents tels que les contraintes familiales ou professionnelles, ou encore l'humeur ou la météo, les stratégies employées en cas de perturbation sont différentes. Tous ces éléments ont été réutilisés dans les questionnaires rédigés sous la forme de questions

relatives à ces différents déterminants des choix des usagers. Le vocabulaire employé a également pu être adapté à celui des usagers.

Cette double analyse qui a été rendue possible par la collaboration de deux équipes de recherche a également permis, outre ses objectifs initiaux, de questionner une nouvelle fois les complémentarités entre analyses textométriques et thématiques. Nous avons ainsi pu montrer plusieurs fonctions à cette double analyse conformément aux recherches menées antérieurement (Della Ratta-Rinaldi, 2013, cité par Glady & Leimdorfer, 2015 ; Desmarais & Moscarola, 2004 ; Fallery & Rodhain, 2007). Ces fonctions varient selon le moment d'usage des différentes approches.

Références

- Alber A. (2010). Voir le son : réflexions sur le traitement des entretiens enregistrés dans le logiciel Sonal. *Socio-logos*, 5.
- Bai Y. (2014). Modelling Riders' Behavioural Responses to Real-Time Information at LRT Stations in Calgary, Alberta. Master Thesis, University of Calgary.
- Bardin L. (1977). *L'analyse de contenu*. PUF.
- Berelson B. (1952). *Content analysis in Communication Research*. The Free Press.
- Bonsall P. and Palmer I. (1999). Route Choice in Response to Variable Message Signs: Factors Affecting Compliance. In Emmerink, R. and Nijkamp, P. editors, *Behavioural and Network Impacts of Driver Information Systems*. Ashgate.
- Bovy P. H. L. and Stern E. (1990). *Route Choice: Wayfinding in Transport Networks*. Kluwer.
- Della Ratta-Rinaldi F. (2013). *L'analyse textuelle en Italie : de l'exploration du texte à l'extraction de l'information*. Communication présentée aux journées d'études Usages de la lexicométrie et interprétation sociologique, Saint Quentin en Yvelines, 12 et 13 juin.
- Desmarais C. et Moscarola J. (2004). Analyse de contenu et analyse lexicale, le cas d'une étude en management public. *Lexicometrica*.
- Emmerink R. H. M., Nijkamp P., Rietveld P. and Van Ommeren J. N. (1996). Variable message signs and radio traffic information: an integrated empirical analysis of drivers' route choice behaviour. *Transportation Research Part A*, 30(2):135-153.
- Fallery B. et Rodhain F. (2007). Quatre approches pour l'analyse de données textuelles: lexicale, linguistique, cognitive, thématique. In *Actes du congrès XVIème Conférence de l'AIMS*, pages 1-16.
- Fiorenzo-Catalano S., Hoogendoorn-Lanser S., and Van Nes R.(2003). Choice-set composition modelling in multimodal travelling. In *Proc of the 10th International Conference of Travel Behaviour Research*.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In *Studies in Linguistic Analysis, Special Volume, Philological Society*, pages 1–32.
- Glady M. et Leimdorfer F. (2015). Usages de la lexicométrie et interprétation sociologique. *Bulletin de Méthodologie Sociologique*, 127:5-25.
- Heiden S., Magué J. P. et Pincemin B. (2010). TXM: Une plateforme logicielle open-source pour la textométrie-conception et développement. In *Actes des JADT 2010*, pages 1021-1032.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, 1(1):127-165.
- Last J. et Manz W. (2003). Unselected mode alternatives: what drives modal choice in long-distance passenger transport? In *Proc of the 10th International Conference of Travel Behaviour Research*.

- Naulin, S. (2013). *Les complémentarités entre la lexicométrie et l'étude "manuelle"*. Communication présentée aux journées d'études Usages de la lexicométrie et interprétation sociologique, Saint Quentin en Yvelines, 12 et 13 juin.
- Pincemin B. (2012). Sémantique interprétative et textométrie. *Texto!*, 17(3):1-21.
- Poudat C. et Boulard A. (2014). Exploration de récit de vie d'adolescents. In *Actes des JADT 2014*, pages 543–552.
- Pronello C. and Camusso C. (2011). Travellers' profiles definition using statistical multivariate analysis of attitudinal variables. *Journal of Transport Geography*, 19:1294-1308.
- Rastier, F. (2005). Mésosémantique et syntaxe. *Texto ! septembre 2005 [en ligne]*. Disponible sur : http://www.revue-texto.net/Inedits/Rastier/Rastier_Mesosemantique.html (Consulté le 16/12/2015).
- Schmid H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proc of the international conference on new methods in language processing*, pages 44-49.
- Scholz R. et Angermuller J. (2013). Au nom de Bologne? Une analyse comparative des discours politiques sur les réformes universitaires en Allemagne et en France. *Mots*, 102:21-36.
- Simon H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 59:99-118.
- Trudel P. et Gilbert W. (1999). Compléter la formation des chercheurs avec le logiciel NUD.IST. *Recherches qualitatives*, 20:87-111.
- Unrug M.-C. d' (1974). *Analyse de contenu et acte de parole, de l'énoncé à l'énonciation*. Éditions Universitaires.
- Vincent-Geslin S. (2012). Des racines et des ailes. *SociologieS*. En ligne : <https://sociologies.revues.org/3945>
- Wanlin P. (2007). L'analyse de contenu comme méthode d'analyse qualitative d'entretiens: une comparaison entre les traitements manuels et l'utilisation de logiciels. *Recherches qualitatives*, 3:243-272.

Remerciements

Ce travail a été effectué dans le cadre des recherches menées au sein de l'Institut de Recherche Technologique SystemX et a ainsi bénéficié d'une aide de l'État au titre du programme d'Investissements d'Avenir. Nous tenons à adresser également tous nos remerciements à SNCF Transilien qui a porté et appuyé ce travail.