

« Si on s’abonnait au théâtre... ». Complémentarité de SATO et d’ALCESTE pour parler de la fidélité de spectateurs occasionnels

Christine Petr¹, Claire Gélinas-Chebat², Marie Lostys³

¹Maître de Conférences, IGR (IAE de Rennes), CREM UMR CNRS 6211, Rennes -France

²Professeure, UQAM – Montréal – Canada

³Directrice Communication et relations publiques, La Passerelle, Scène nationale de Saint-Brieuc, 22 041 Saint-Brieuc Cedex 2 – France

Abstract

In a telemarketing context, the concept of consumer loyalty is a very important topic accordingly to the attention it has received from enterprises and researchers. In this field, subscription is not necessarily synonymous with consumer loyalty and vice versa, not to subscribe does not mean that a consumer is not loyal. To understand more clearly why occasional consumers refrain from engaging themselves is crucial in a marketing process which specifically targets non-loyal consumers. The main purpose of this presentation is to show how SATO and ALCESTE could play a complementary role in investigating the expression of loyalty of the same consumers' discourses about subscription formulas proposed by a multidisciplinary repertoire theatre in France.

Keywords: subscription, loyalty, Computer Text Analysis, SATO, ALCESTE.

Résumé

En marketing, le concept de fidélité mérite encore aujourd’hui une attention particulière. Un consommateur qui choisit de s’abonner est-il vraiment un consommateur qui s’inscrit dans une logique d’engagement et exprime sa fidélité vis-à-vis d’une marque ou d’un établissement ? A l’inverse, un consommateur qui préfère ne pas s’engager d’une manière contractuelle et durable est-il pour autant un consommateur infidèle ? Cette recherche cherche à mettre en lumière les motifs des spectateurs occasionnels qui préfèrent ne pas s’engager puisqu’ils sont, dans une démarche marketing, les candidats à rejoindre et à fidéliser. L’objet de cette présentation est de montrer les contributions complémentaires des logiciels SATO et ALCESTE pour saisir l’expression de la fidélité à un établissement culturel exprimée dans les discours de spectateurs occasionnels d’un théâtre en France.

1. Introduction

Le concept de fidélité en marketing est une thématique qui ne cesse d’interpeller managers et académiques. Dans les stratégies de fidélisation, les entreprises élaborent des techniques de contractualisation. Or, fidéliser devrait être en principe, établir une relation avec le consommateur qui aille au-delà de la stabilité du comportement d’achat. Il devrait être question de confiance, d’engagement, etc. Le consommateur qui choisit de s’abonner est-il vraiment un consommateur qui s’inscrit dans une logique d’engagement, et qui exprime ainsi sa fidélité, vis-à-vis d’une marque ou d’un établissement ? A l’inverse, un consommateur qui préfère ne pas s’engager de manière contractuelle et durable est-il pour autant un consommateur infidèle, méritant moins d’égards car il est sensé être moins rentable (n’est-ce

pas un consommateur qui regarde plus le prix, qui est plus susceptible de zapping et moins prêt à accepter des déficiences ponctuelles, etc.).

Dans le secteur culturel, l'abonnement est-il un outil de fidélisation indice d'une fidélité véritable ? C'est dans le but de mieux comprendre ce concept de fidélité auprès de consommateurs qui ne veulent pas s'engager que des analyses de discours de consommateurs de spectacles vivants ont été réalisées.

Quarante trois individus qui fréquentent l'établissement culturel de manière occasionnelle, c'est-à-dire sans passer le cap de l'abonnement à cet établissement français, ont accepté le principe d'un entretien semi-directif. Celui-ci avait pour but de les faire parler de leur consommation de spectacles culturels, et d'exposer précisément les raisons génériques et spécifiques de leur refus de s'abonner à l'institution étudiée. Sélectionnés au cours de tous les spectacles diffusés pendant la saison artistique 2004-2005 de l'établissement, ces répondants représentent un échantillon diversifié de la population des spectateurs occasionnels de l'établissement.

Les discours ont fait l'objet d'une investigation textuelle poussée : des analyses lexicales comparatives et une analyse typologique des univers de discours. Les analyses des différences de lexique et de modalités du discours (argumentation, implication, efforts de comparaison), ont été réalisées grâce au logiciel d'analyse de données textuelles SATO (Daoust, 1996, 2004, 2007). L'analyse typologique des univers lexicaux a été réalisée par le logiciel ALCESTE (Reinert, 2002). La combinaison de ces perspectives analytiques et de ces contributions successives de SATO à ALCESTE et d'ALCESTE à SATO, a permis globalement d'affiner la compréhension des représentations de la formule commerciale de l'abonnement et à mieux saisir les motivations et les freins concernant son adoption.

2. Méthodologie

2.1. Description du corpus

Le corpus est constitué de 43 entretiens semi-dirigés individuelles menées par un ou deux animateurs auprès de 44 répondants sollicités à la sortie d'un spectacle (une entrevue a été réalisée avec un couple de répondants). Les spectateurs qui ont un abonnement régulier au théâtre n'ont pas été retenus puisque cette recherche tente de mieux comprendre les motifs évoqués pour ne pas adhérer à la formule d'abonnement.

Les animateurs ont utilisé un guide d'entretien où trois grands thèmes de discussion portant sur les formules d'abonnement étaient initiés à partir des mêmes questions. Les répondants étaient invités à commenter librement chacun de ces thèmes. Les répondants sont 33 femmes et 11 hommes. De ce groupe, 21 personnes sont encore aux études alors que 22 occupent différentes situations professionnelles. 26 répondants considèrent qu'ils attribuent un budget moyen aux activités culturelles alors que 6 n'y accordent qu'un faible budget et 11 un budget important. L'établissement de spectacles vivants qui a collaboré à cette étude appartient au réseau national des Scènes Nationales qui regroupe plus de 70 établissements en France. Institués au nom de la décentralisation culturelle, ces centres artistiques sont pluridisciplinaires. Ces organisations culturelles proposent ainsi du théâtre, de l'opéra, des spectacles chorégraphiques (danse classique, contemporaine) et des spectacles musicaux tels que de la musique classique, contemporaine et traditionnelle, des musiques du monde, du jazz, du rock, des variétés, du cabaret, etc. Enfin, le mime, le cirque et les café-théâtre et spectacles d'humour sont légitimement présents dans ces établissements.

Seule caractéristique originale par rapport aux centres culturels du secteur privé, la diffusion et le soutien à la création sont les principes fondateurs de ces établissements. Subventionnés et missionnés par le Ministère de la Culture et de la Communication pour cela, les programmations annuelles sont empreintes d'une part importante de nouveautés.

L'établissement étudié est la Scène Nationale de « La Passerelle » à Saint Briec (Côtes d'Armor, France).

2.2. Saisie et préparation des données

Chaque entrevue a été enregistrée sur support magnétophonique puis saisie sous format Word. Les transcriptions respectent les contraintes de l'oral et ont été transcrites dans leur forme la plus intégrale (une seule des deux particules de la négation, phrases incomplètes, structures syntaxiques propres à l'oral, etc). Chacune des entrevues a été codifiée de telle sorte que les interventions des animateurs et des répondants ont été distinguées. De plus, des étiquettes particulières permettent de retrouver les interventions de chacun des répondants en fonction des informations sociodémographiques correspondant, entre autres, au sexe du répondant, à son statut d'étudiant ou non, et à l'importance du budget qu'il accorde aux activités artistiques.

Les annotations éditiques sont conformes à la syntaxe de SATO : la passerelle ATONET¹ (Duchastel et al, 2005) permet de soumettre à ALCESTE les textes édités selon la syntaxe de SATO. Ce passage de SATO à ALCESTE avait déjà été réalisé avec succès et avait fait l'objet d'une présentation aux JADT 2006 (cf. Daoust et al. 2006). La stratégie de conversion des formats de données entre les logiciels qui a été présentée aux JADT 2006 (Daoust et Marcoux, 2006) est basée sur un format pivot en XML développé à partir de recommandations du *Text Encoding Initiative*. Des programmes *Perl* (passerelles) permettent de convertir les données des formats propriétaires vers le format XML-TEI et du format XML-TEI vers les formats propriétaires.

3. Résultats

3.1. Analyses lexicales

Au départ, SATO a fourni les premiers indices lexicaux appliqués à l'ensemble de notre corpus. Puis, nous avons rapidement créé un corpus secondaire, un *sous-corpus* pour ne retenir que les interventions des répondants. A cette étape de notre recherche, nous ne tentons pas de voir la pénétration des interventions des animateurs dans les réponses des interviewés mais bien les indices portant sur le concept de fidélité dans le discours des répondants. Nous avons choisi de soumettre d'abord le corpus à SATO pour mesurer le poids des interventions de chacun des intervenants et le poids des intervenants en fonction de nos variables sexe (homme/femme), statut (étudiants/non-étudiants) et budget (petit, moyen, gros).

Le corpus complet compte 89 498 occurrences (au sens de SATO²) dont 3 579 formes lexicales différentes³. Les trois premiers éléments de type « nom commun » les plus utilisés

¹ ATONET est accessible <http://www.atonet.net>

² Pour SATO, les éléments de ponctuation sont considérés comme des éléments lexicaux

³ La différence entre le nombre des occurrences et des formes lexicales tient au fait qu'une forme lexicale peut être répétée plusieurs fois.

sont dans l'ordre « abonnement » (829 fois) « spectacles » (461 fois) et « passerelle » (394 fois). La fréquence du mot « spectacle » (dans sa forme au singulier) est de 310 fois.

Les animateurs qui interviennent systématiquement dans chacune des entrevues représentent plus de 33% des interventions. Les 44 répondants ont contribué à 66% du corpus.

Lorsqu'on exclut les interventions des animateurs pour se concentrer sur les réponses des répondants seulement, ce *sous-corpus* contient 57 507 occurrences dont 3 010 formes lexicales différentes. Le premier nom commun est « abonnement » puis vient le mot « spectacles ».

Une analyse plus individuelle des interventions des répondants montre qu'un sujet en particulier, le sujet #4 contribue à lui seul à 2 849 occurrences, soit 4,95% du corpus total des répondants. Il est suivi de très près par cinq sujets, les sujets #17, 01, 19, 03 et 43. À l'opposé, six sujets s'expriment peu et donnent moins d'un pour-cent du corpus. Ce sont les sujets #32, 14, 33, 29, 13 et 25.

Lorsqu'on prend en considération le sexe des répondants, 67.18% des occurrences (soit 38 633) l'ont été par les 33 répondantes alors que 32.82% des occurrences sont attribuables aux 11 hommes. Bien que les femmes représentent 75% des répondants, seulement 68% du corpus leur est attribuable. Les hommes, soit 25% des répondants, interviennent proportionnellement plus. Sous l'angle du statut des répondants, il n'y a pas de différence notable entre les interventions des 21 étudiants et des 22 autres répondants : 49.68% des occurrences ont été fournies par les étudiants et 50.32% par les autres. Les 26 répondants qui ont un budget moyen à consacrer aux abonnements, soit 60% des répondants, représentent 66.60% du corpus. Ceux qui sont prêts à y mettre un budget important, soit 11 sujets ou 25% des répondants, ne fournissent que 12.17% des occurrences alors que 21.23% proviennent des 6 sujets, soit 14% des répondants, qui ont peu de budget à y consacrer.

3.1.1. Les caractéristiques du discours des hommes et des femmes

Pour comparer les discours des hommes et des femmes, nous avons créé des sous-textes ne comportant que le discours des hommes dans un cas et que celui des femmes dans l'autre cas. Un algorithme de distance lexicale basé sur la distance du khi2 accessible par une requête *distance* dans SATO permet de mesurer les différences des lexiques de chacun de ces groupes. La mesure évalue l'écart dans l'utilisation d'un vocabulaire donné entre les deux sous-textes. Les formes lexicales sont triées par ordre décroissant de contribution à la mesure de *distance*, ce qui permet d'identifier, par ordre d'importance, les spécificités de chaque sous-texte. L'approche est essentiellement dichotomique : nous comparons un sous-texte à un autre, via leur lexique respectif. C'est par cet indice de *distance* calculé sur la base du khi2 que SATO permet de mettre en évidence les différences les plus marquantes entre les deux lexiques. Plus précisément, SATO évalue la probabilité de retrouver chacun des mots dans un lexique spécifique d'un groupe par rapport à l'ensemble des mots du lexique de ce groupe et évalue enfin les mots contribuant le plus à discriminer les discours d'un groupe par rapport à l'autre.

Les mots qui distinguent le plus le discours des hommes de celui des femmes sont les mots suivants, dans l'ordre, utilisés plus spécifiquement par les hommes : « également », « famille », « théâtre », « frein », « presse », « saison » et « dire ». Par opposition, les points de suspension inscrits «...» dans le corpus et les mots « oui », « pouvoir », « silence », « aller » et « tarif » sont significativement plus fréquents dans le discours des femmes. Ces données sont présentées au Tableau 1 suivant. Ainsi, on constate que le mot « famille »

contribuant à une variance de 0,62 n'est présent que dans le discours des hommes. Les femmes ne parlent pas de la famille. Les mots « frein » et « presse » n'apparaissent pas non plus dans le discours des femmes. Le mot « presse » est utilisé dans le sens de média, la presse⁴. Le mot « théâtre » fréquent dans les discours des hommes et des femmes est tout de même proportionnellement plus souvent évoqué dans le discours des hommes.

Chez les femmes, le premier mot à marquer une différence par rapport au discours des hommes est le « oui ». Il apparaît 250 fois dans le discours des femmes contre 64 fois chez les hommes. Le « oui » correspond à la formule d'acquiescement, à l'ouverture phatique par excellence. À 36 reprises, le mot « pouvoir » est utilisé dans le discours des femmes contre 2 seulement chez les hommes. Ce mot est toujours associé à un verbe d'action comme aller, donner, rencontrer, participer, etc. Les femmes veulent « pouvoir » agir. Certains comportements non verbaux ont été inscrits dans les corpus, comme les pauses, les rires, etc. C'est le cas du mot « silence ». Il a été noté une seule fois chez un homme, alors qu'il a été noté 29 fois dans le discours des femmes. Ce « silence » est une marque d'hésitation. Cette hésitation est aussi marquée par les points de suspensions notés « ... » qui apparaissent également dans cette mesure de la *distance* du discours des hommes et des femmes. Le verbe « aller » utilisé 177 fois chez les femmes (contre 56 chez les hommes) est une manifestation de désir de mouvement et suggère l'importance de l'organisation de la venue (organisation de la vie familiale, garde des enfants, etc.). On le trouve en grande quantité dans des expressions comme : *pouvoir y aller, aller rejoindre, décidé d'y aller, obligé d'y aller, etc.* Un autre terme porteur de distance entre le discours des hommes et des femmes est le mot « tarif » qui apparaît 42 fois dans le discours des femmes contre 7 chez les hommes. On parle souvent du « tarif d'abonnement », du « demi-tarif » ou du « plein tarif ». Son utilisation met en évidence une préoccupation plus grande pour les femmes que les hommes sur la question du tarif.

3.1.2. Les caractéristiques du discours des étudiants et des non-étudiants

Nous avons répété la procédure de création de sous-textes en discriminant cette fois le discours des étudiants d'une part et celui des autres participants d'autre part. Puis nous avons procédé au calcul de la *distance* de SATO.

Il est possible de constater que les étudiants réfèrent beaucoup plus au « j' » lorsqu'ils s'expriment que les non-étudiants (cf. tableau 2 suivant). Ce mot est une marque significative d'implication dans le discours. Nous constatons d'ailleurs que les pronoms personnels « t' » et « nous » sont des éléments de distance entre le discours des étudiants et des non-étudiants. Les étudiants ont un discours qui est expressif bien plus qu'informatif. La recherche d'approbation semble être recherchée : le mot « vrai », toujours employé dans l'expression « c'est vrai », apparaît 108 fois dans le discours des étudiants. Il s'agit d'un élément de distance par rapport aux discours des individus qui ne sont pas étudiants.

Dans le discours des répondants qui ne sont pas étudiants, le mot « avance » est le mot qui contribue le plus à la distance entre leurs discours et celui des étudiants. Ce mot est utilisé presque exclusivement pour marquer le temps (être en avance, prévoir en avance, réserver à l'avance, etc..) et montre une certaine préoccupation au temps et à l'organisation. Les mots « disponible » (de même que « disponibilité ») d'une part, et les mots « choisir », « conditions » apparaissent dans le discours des répondants qui ne sont pas étudiants de manière plus importante que dans le discours des étudiants. Les contraintes de la vie

⁴ Une fonction *kwic* dans SATO nous a permis d'exprimer cette affirmation.

professionnelle expliquent sans doute l'émergence de ces mots dans le discours des individus qui ne sont plus étudiants. Dans cet indice de distance, les mots « cinéma » et « spectacles » sont plus utilisés par les répondants qui ne sont pas étudiants alors que le mot « concerts » l'est d'avantage par les étudiants faisant référence à des univers de pratique du spectacle légèrement différentes.

3.2. *Analyses thématiques*

ALCESTE propose une méthode qui vise à faire émerger des *mondes lexicaux* en construisant des *énoncés simples* dont l'approximation statistique correspond à des segments de texte de longueur comparable respectant les frontières de l'*unité de contexte élémentaire* (UCE), généralement la phrase. Ainsi, ALCESTE tente de faire émerger la structure du discours par le dépistage de profils de répétition dans les énoncés simples.

Les analyses avec ALCESTE n'ont porté que sur le corpus des répondants. Ce *sous-corpus* a été obtenu en créant avec SATO un texte qui éliminait du corpus initial les interventions des animateurs et ne conservait que les interventions des répondants.

ALCESTE suggère des pistes d'analyse qui méritent une attention détaillée. Ces analyses couplées avec les analyses de SATO permettent de mieux comprendre nos données. Par exemple, ALCESTE fait ressortir comme éléments caractéristiques de la classe 1, les individus 3 et 1 représentant des χ^2 respectifs de 72.77 et de 50.13. Nous savons par SATO que ces individus des locuteurs des plus bavards. Ils ont en commun d'être deux femmes, étudiantes qui ne consacrent qu'un budget moyen aux activités artistiques. Également caractéristique de cette première classe, selon ALCESTE, les étudiants se démarquent. Notre analyse comparative étudiants/ non-étudiants semble une piste intéressante pour mieux saisir le concept de fidélité. À la première section du tableau 3, nous présentons le vocabulaire caractéristique de cette classe. Les mots qui présentent les χ^2 les plus forts sont les lemmes suivants : *étudiant+*, *carte+*, *cher+ pasque*, *trouve+*, *rentabilité*.

Les éléments qui composent la classe 2 sont les individus qui ne sont pas étudiants. Un certain nombre de répondants dont les répondants 28, 38, 04 et 20, se caractérisent par un petit budget. Ces répondants, un homme et trois femmes, ne sont pas étudiants, et, sauf un, investissent moyennement dans les activités artistiques. Les mots caractéristiques de cette classe sont présentés au tableau 3, deuxième section. Nous retrouvons les mots suivants, ou leur forme lemmatisée : *avance+*, *disponibilité*, *prévoir*, *choisir*. Nous sommes confirmés dans nos analyses du discours avec SATO où ces mots sont déjà ressortis. Les femmes et en particulier les femmes avec peu de ressources financières semblent les moins intéressées par le projet d'un abonnement au théâtre.

Enfin, à partir de mots suggérés par une revue de littérature en marketing, nous avons retenu les mots *fidélité*, *confiance* et *sécurité* pour une analyse plus détaillée. Nous sommes retournés à SATO pour cette opération. Dans SATO le signe « \$ » permet de désigner une suite de caractères quelconques. Nous avons donc procédé à une requête au niveau du lexique ventilé dont les résultats sont présentés au tableau 4 selon le sexe et selon le statut.

On constate que le mot « fidèle » et les lexèmes de la même famille du corpus, soit les mots « fidélisé », « fidéliser », « fidélité » sont en presque totalité émis par les femmes. Les mots « confiance » d'une part et « sécurité », d'autre part, et les mots apparentés ne sont jamais émis dans le corpus. Nous pourrions poursuivre cette démarche avec d'autres mots ciblés pour cerner d'une manière plus pointue le concept de fidélisation que notre corpus pourrait révéler.

Cependant, plusieurs indices révélateurs de ce concept ont été soulignés et permettent d'en saisir l'essentiel.

4. Discussion et Conclusion

L'objectif de départ de cette recherche est de cerner le concept marketing de *fidélisation* de spectateurs occasionnels d'un établissement de spectacle vivant. Théoriquement, le concept de fidélisation se bâtit sur la notion de relation de confiance et d'engagement entre un consommateur et une organisation. Notre corpus constitué auprès de spectateurs occasionnels d'un théâtre en France permet de faire ressortir les raisons principales du refus de s'engager de ces spectateurs. L'univers lexical sous l'angle d'analyse des hommes et des femmes, ou des étudiants et des individus qui ne le sont plus, n'est pas le même et montre des différences intéressantes. Les résultats des analyses lexicales, permettent de constater que les répondants s'expriment proportionnellement plus que les répondantes et que ceux qui ont un budget important sont plus silencieux que ceux qui consacrent peu de budget aux activités artistiques. Les hommes justifieraient plus facilement et avec moins d'hésitation les raisons de leur refus à s'abonner. C'est l'étude du discours des individus qui ne sont plus étudiants que des préoccupations liées au temps et à l'organisation jaillissent. Dans le discours des femmes, souvent ciblées dans cette impression d'être responsables de l'organisation des activités familiales, les mots font références à des désirs d'agir et à de l'hésitation. Étonnamment dans cette étude sur l'abonnement et la notion de fidélité, jamais les mots liés à la confiance ou à la sécurité ne sont évoqués. On ne remet pas en question la confiance ou non qu'on puisse avoir avec l'établissement, ni la sécurité ou le manque de sécurité à s'abonner à long terme.

Cette présentation a pour objectif de montrer la complémentarité de SATO et d'ALCESTE dans des analyses de données lexicales. Les résultats d'analyse d'un logiciel renforcent les analyses obtenues par le second. SATO permet de mettre en évidence les différences lexicales marquantes des hommes et des femmes d'une part et des étudiants et des autres répondants d'autre part. Ces résultats sont confirmés par ALCESTE. Nous pouvons mieux comprendre nos résultats d'analyse par SATO, mais également il est possible de mieux interpréter les résultats des analyses d'ALCESTE. Ainsi, nous constatons que les mots révélateurs de la classe 2 d'ALCESTE sont exactement les mêmes mots mis en évidence par SATO dans le groupe des répondants qui ne sont pas étudiants. Dans cette classe 2 d'ALCESTE on retrouve les femmes et les individus qui investissent peu de budget dans les activités culturelles. Dans la classe 1 d'ALCESTE, les mots *cher+* et *rentabilité* apparaissent, et les étudiants sont marqueurs de cette classe. Dans SATO, les analyses du discours des étudiants montrent qu'ils tiennent un discours beaucoup plus expressif qu'informatif.

Ces analyses à la fois sur SATO et sur ALCESTE et ces mouvements de va et vient des résultats de l'un et de l'autre de ces logiciels nous paraissent extrêmement intéressants et nous permettent de saisir plus adéquatement les résultats des analyses des discours des répondants et ainsi, de mieux cerner les objectifs de recherche du professionnel ou du chercheur amené à utiliser ces logiciels.

Références

- Daoust F. (1996, 2004). *SATO 4, Manuel de référence*. Centre ATO, UQAM, Montréal.
 Daoust F. (2007). *Manuel de référence*. <http://www.ling.uqam.ca/sato/satoman-fr.html>.

Daoust F., Dobrowolski G., Dufresne M. et Gélinas-Chebat C. (2006). Analyse exploratoire d'entrevues de groupe : quand ALCESTE, DTM, LEXICO et SATO se donnent la main. In J.-M. Viprey, C. C. Condé, A. Lelu, M. Silberztein éditeurs, I, *Actes des JADT-2006* : 313-326.

Daoust F. et Marcoux Y. (2006). Logiciels d'analyse textuelle : vers un format XML-TEI pour l'échange de corpus annotés, In J.-M. Viprey, C. C. Condé, A. Lelu, M. Silberztein éditeurs, I, *Actes des JADT-2006*.

Duchastel J. et al. (2005). *ATONET, Réseau pour l'échange de ressources et de méthodologies en analyse de texte assistée par ordinateur*. <http://www.atonet.net>.

Reinert M. (2002). *Alceste, Manuel de référence*. Université de Saint-Quentin-en-Yvelines, CNRS.

Annexes

Tableau 1 : Liste des 30 premières unités contribuant le plus à la distance dans le discours des hommes et des femmes (L'astérisque marque les éléments du discours des hommes).

Fréq	tot	Hommes	Femmes	explique	cumul	
2.07	1.38	2.40	1.32	1.32	...	
2.57	1.88	2.90	1.07	2.39	.	
0.04	0.11	0.01	0.59	2.98	également	*
0.12	0.22	0.07	0.53	3.51	«	*
0.02	0.06	0.00	0.47	3.99	famille	*
0.18	0.31	0.13	0.47	4.46	théâtre	*
0.13	0.23	0.08	0.47	4.92	»	*
0.55	0.34	0.65	0.45	5.38	oui	
0.02	0.05	0.00	0.43	5.81	frein	*
6.03	6.66	5.72	0.38	6.19	,	*
0.01	0.04	0.00	0.34	6.53	presse	*
0.19	0.30	0.14	0.34	6.87	dire	*
0.05	0.11	0.03	0.34	7.21	saison	*
0.20	0.09	0.25	0.32	7.53	!	
0.01	0.04	0.00	0.30	7.83	ouaih	*
0.07	0.01	0.10	0.29	8.13	pouvoir	
0.39	0.54	0.33	0.28	8.41	pense	*
0.05	0.00	0.07	0.28	8.69	finalement	
0.65	0.83	0.57	0.27	8.96	ce	*
0.01	0.03	0.00	0.26	9.22	dvd	*
0.01	0.03	0.00	0.26	9.47	suivant	*
0.02	0.05	0.01	0.25	9.73	maximum	*
0.03	0.06	0.01	0.24	9.97	car	*
0.03	0.06	0.01	0.24	10.21	public	*
0.03	0.06	0.01	0.24	10.44	différents	*
0.05	0.01	0.07	0.23	10.67	silence	
0.01	0.04	0.00	0.23	10.90	correct	*
0.01	0.04	0.00	0.23	11.13	relativement	*
0.01	0.04	0.00	0.23	11.36	âge	*
0.04	0.00	0.06	0.22	11.59	...	

Tableau 2 : Liste des 30 premières unités contribuant le plus à la distance dans le discours des étudiants et des non-étudiants (l'astérisque marque les éléments du discours des étudiants).

Fréq	tot	Étudian	~Étudia	explique	cumul	
0.36	0.12	0.58	1.86	1.86	heu	
0.68	0.45	0.89	0.87	2.73	-	
1.86	2.21	1.53	0.76	3.49	euh *	
0.08	0.01	0.14	0.61	4.09	avance	
0.27	0.16	0.38	0.57	4.66	bah	
0.43	0.57	0.30	0.52	5.18	ben *	
1.00	1.19	0.81	0.44	5.62	j' *	
0.04	0.00	0.08	0.39	6.02	disponible	
0.05	0.01	0.09	0.35	6.36	choisir	
0.08	0.03	0.13	0.34	6.70	cinéma	
0.04	0.01	0.07	0.32	7.02	également	
0.52	0.40	0.64	0.32	7.34	spectacles	
0.03	0.00	0.05	0.29	7.63	hum	
0.53	0.41	0.63	0.28	7.91	peut	
0.29	0.38	0.22	0.27	8.18	vrai *	
0.03	0.05	0.00	0.27	8.45	concerts *	
0.53	0.42	0.63	0.26	8.71	au	
0.03	0.06	0.01	0.26	8.97	t' *	
0.21	0.28	0.14	0.26	9.23	nous *	
0.02	0.04	0.00	0.26	9.50	" *	
0.02	0.00	0.04	0.25	9.75	conditions	
0.05	0.02	0.08	0.25	10.00	place	
0.26	0.34	0.19	0.25	10.25	ouais *	
0.93	0.78	1.06	0.25	10.50	l'	
0.02	0.04	0.00	0.24	10.74	pasque *	
0.07	0.11	0.04	0.24	10.98	personnes *	
0.51	0.41	0.60	0.22	11.20	ne	
0.11	0.16	0.07	0.22	11.43	avait *	
0.04	0.07	0.02	0.22	11.65	ouai *	
0.02	0.04	0.00	0.22	11.86	p' *	
1.32	1.48	1.17	0.21	12.08	ça *	

Tableau 3 : Extrait de l'analyse des univers de discours de l'échantillon total : les vocabulaires spécifiques des 2 premières classes.

Vocabulaire spécifique de la classe 1 :

etudiant+(30), carte+(33), cher+(23), pasqu+(10), trouv+er(33), renta+ble(16), jeune+(12), voir (46), pa+yer(25), spectacle+ (61), publi+14(6), mettre.(7), interet+(8), accroch+er(4), theme+(4), justement(15), dommage+(4), formule+(23), , super(5), essa+yer(5), pens+er(16), passe(7), euro+(9), comique+(2), excessi+f(2), intermediaire+(3), plein+(6), avis(3), entretien+(3), pardon+(3), plaisir+(2), aller.(60), rembours+er(2), antigone(2), hm(2), passerelle(26), obligation+(3), tarif+(16), faire.(23), differ+ent(5), attenti+f(2), convenable+(1), splendide+(1), suggesti+f(1), vieux(1), air+(2), lycee+(1), minimum(1), paie+(1), voyage+(1), appreci+er(1), attract+ion(1), informe+(2), log+16(2), vala+ble(1), grand-chose(1), lyceens(1), p'(2), y'(6), autrement(6), completement(3), coup+(5), maximum(3), soiree+(4), attir+er(5), faite(3), bon+(1), clair+(1), fixe+(1), nouvel+(2), prenant+(1), scolaire+(1), vachement(2), culture+(1), ensemble+(2), entree+(3), facon+(3), limite+(2), amelior+er(1);

Vocabulaire spécifique de la classe 2 :

avanc+er(36), disponi+ble(23), prévoir.(20), chois+ir(20), soir+(17), spectacle+(79), pouvoir+(17), semaine+(12), horaire+(10), jour+(13), evid+ent(7), programm+er(9), reserve+(9), sortir.(10), fonction+(6), facile+(7), juin+(4), septembre+(7), billet+(8), date+(8), heure+(9), mois(23), place+(20), week-end+(4), chang+er(11), travail<(12), imprevu+(3), exactement(3), determin+er(3), liber+er(3), reserv+er(6), problem<(10), correct+(3), norma+l(5), finalement(10), forcement(18), fevrier+(2), deuxieme+(2), an+(15), attente+(2), guillemet+(2), defin+ir(3), empech+er(2), organis+er(4), refaire.(2), repondre.(4), retenir.(2), entete(2), prospectus(2), senior+(2), st(2), difficile+(6), long+(5), debut+(10), somme+(4), cherch+er(5), convenir.(7), donn+er(8), cinema<(10), decide+(2), idea+l(5), libre+(2), locati+f(2), ouvert+(2), personnel+(5), suivant+(1), carrement(2), franchement(2), generalement(5), apres-midi+(1), mardi+(1), cours(2), fin+(5), frein+(1), inconvenient+(8);

Tableau 4 :

Répartition lexicale de mots ciblés selon le sexe :

Mots ciblés	<u>Fréqtot Hommes Femmes Gramr entrée lexicale</u>				
fid\$	3	0	3	(Adjqua, *	fidèle
	1	0	1	Vparpas	fidélisé
	2	0	2	Vinf	fidéliser
	7	1	6	Nomcom	fidélité
confi\$	0 lexèmes associés				
secur\$	0 lexèmes associés				

Répartition lexicale de mots ciblés selon le statut :

Mots ciblés	<u>Fréqtot Etud. Non-étud. Gramr entrée lexicale</u>				
fid\$	3	2	1	(Adjqua, *	fidèle
	1	0	1	Vparpas	fidélisé
	2	0	2	Vinf	fidéliser
	7	6	1	Nomcom	fidélité
confi\$	0 lexèmes associés				
secur\$	0 lexèmes associés				