

ECLAIRER LE VOCABULAIRE DES QUESTIONS OUVERTES PAR LES QUESTIONS FERMÉES: LE TABLEAU LEXICAL DES QUESTIONS

par

Philippe Cibois

(Université Paris V - René Descartes, 12 rue Cujas, 75006 PARIS)

Abstract. Clarifying the Vocabulary of Open Questions by Using Closed Questions: The Lexical Table of Questions. The "Lexical Table of Questions" which cross-tabulates the vocabulary of an open question with the closed questions of a survey is a tool of data analysis which permits the clarification of each type of question by the other. An application using factorial correspondence analysis of such a table is presented concerning an opinion survey on the reform of French spelling. **Open questions. Lexical Table of Questions. Factorial Correspondence Analysis. Content Analysis.**

Résumé. Le "tableau lexical des questions" qui croise le vocabulaire d'une question ouverte avec les questions fermées d'une enquête est un instrument d'analyse des données qui permet d'éclairer l'un par l'autre les deux types de questions. On présente une application par l'analyse des correspondances d'un tel tableau à propos d'une enquête sur la réforme de l'orthographe. **Questions ouvertes. Table lexical des questions. Analyse factorielle des correspondances. Analyse de contenu.**

Quand on dépouille des enquêtes avec des questions fermées, l'analyse des données est d'une grande utilité et pour s'yramener, ceux qui utilisaient ce riche instrument qu'est laquestion ouverte devaient fermer les réponses *a posteriori*. Chacun sait que cet exercice est long et délicat même si un sociologue expérimenté peut y arriver d'une manière qui respecte les données.

Dans cet article je voudrais, en m'appuyant sur le livre de **Ludovic Lebart** et **André Salem**, *Analyse statistique des données textuelles* (Paris, Dunod, 1988), montrer comment il est possible d'aller plus loin que ces auteurs dans l'utilisation simultanée des questions fermées et des questions ouvertes. Notre but étant de rester le plus possible dans la perspective de l'analyse des données; c'est-à-dire, en laissant les traits pertinents émaner des données elles-mêmes et non d'un choix *a priori*.

Les concepts fondamentaux

On suppose que l'on traite une enquête possédant des questions à réponses fermées, chacune ayant un certain nombre de modalités de réponses possibles exclusives les unes des autres ("Masculin / Féminin" ou "Oui / Non" ou "Un peu / Beaucoup / Passionné / A la folie / Pas du tout") avec éventuellement un code de Non-réponse.

On suppose que l'on a posé une question ouverte avec des réponses de quelques lignes. On obtient donc un texte par individu, possédant un vocabulaire composé de mots (mots-outils ou non).

Rappelons que pour ce qui concerne le traitement des questions fermées, les deux outils classiques en analyse des données sont:

1) le tableau en codage logique (codage disjonctif complet) qui croise en ligne les individus de l'enquête et en colonne les modalités de réponse à toutes les questions. A l'intersection d'une ligne et d'une colonne, se trouve le nombre de fois où un individu a choisi une modalité; c'est-à-dire, une fois ou zéro fois.

2) le tableau de Burt qui est la juxtaposition de tous les tableaux croisés possibles deux-à-deux et où, en ligne et en colonne, on a toutes les modalités de l'enquête. A l'intersection d'une ligne et d'une colonne se trouve l'effectif des enquêtés qui ont choisi en même temps ces deux modalités.

Quand on dispose du vocabulaire d'une question ouverte (et en se limitant aux plus fortes fréquences), il peut sembler naturel de construire l'équivalent du tableau en codage logique où cette fois on aurait en colonne le vocabulaire rencontré dans la question ouverte et en ligne les individus de l'enquête. A l'intersection d'une ligne et d'une colonne on note "zéro" ou "un" selon que le mot en question appartient ou non au vocabulaire de l'enquêté ou, ce qui revient souvent au même, le nombre de fois où le mot en question a été employé par lui.

Ce tableau, qu'à la suite de Lebart et Salem on appellera "Tableau lexical entier" se révèle peu intéressant, comme ils l'ont d'ailleurs montré. En effet, c'est un tableau très creux (il possède beaucoup de zéros), où les facteurs sont liés à des co-occurrences. Celles-ci sont nombreuses et il faut de nombreux facteurs pour "éplucher" en quelque sorte les différentes associations. Au lieu de nous donner un facteur général, un tableau lexical complet met petit à petit en avant des co-occurrences spécifiques qui ne sont pas forcément très intéressantes.

Pour arriver à un degré de généralité suffisant, Lebart et Salem proposent de croiser le vocabulaire de la question ouverte avec une question fermée ou le pré-croisement de plusieurs questions fermées. Par exemple, ils croisent les mots les plus fréquents rencontrés dans une question ouverte (portant sur les difficultés qui peuvent faire hésiter une femme ou un couple à avoir des enfants) avec un pré-croisement de 3 niveaux d'âge avec 3 niveaux de diplôme.

Le problème posé par un tel tableau de données ne réside pas dans son traitement, qui ne pose pas de questions car les facteurs ont un caractère suffisant de généralité, mais dans sa constitution. Le choix de la question qui sert de filtre est crucial et, si l'on veut faire intervenir plusieurs questions, le pré-croisement entraîne rapidement un émiettement des données. On est donc amené à faire des choix préalables à l'analyse des données en sélectionnant *a priori* les questions que l'on va croiser avec le vocabulaire. On souhaiterait une option inverse; c'est-à-dire, faire le choix

d'une question pertinente au vu des données elles-mêmes. On envisage donc de constituer un tableau où toutes les questions fermées de l'enquête joueraient le même rôle et où ce serait l'analyse elle-même qui mettrait en relief la pertinence particulière de l'une ou de l'autre.

Ceci est possible: il suffit de s'inspirer non plus du tableau en codage logique mais du tableau de Burt et constituer ce que j'appelle le "tableau lexical des questions" qui croise le vocabulaire de la question ouverte avec l'ensemble des modalités de réponse aux questions fermées.

Pour faire comprendre ce concept de tableau lexical des questions, supposons une enquête fictive à trois questions ouvertes (Sexe codé M/F, âge codé 1/2/3, opinion codée +/=/-) et à une question ouverte où les mots émis sont symbolisés par des lettres de l'alphabet. Voici pour 4 individus les données de base:

Individu	Sexe	Age	Opinion	Vocabulaire émis
1	M	3	+	A D E F G
2	F	1	-	H E G F B
3	M	2	+	A E H C
4	F	1	=	D H E G

Le mot le plus émis a été le mot E puisque ces 4 individus l'ont pris. Ceux qui l'ont été le moins sont B et C, pris une seule fois.

Pour construire le tableau lexical des questions on met en ligne les mots par ordre de fréquence décroissante et en colonne les différentes modalités des questions fermées. A l'intersection d'une ligne et d'une colonne se trouve le nombre de fois où un mot donné a été émis en même temps qu'était possédée la modalité en question. On a ainsi les résultats suivants:

(Fréquence)	Mots	Sexe		Age			Opinion			Total
		M	F	1	2	3	+	=	-	
4	E	2	2	2	1	1	2	1	1	12
3	G	1	2	2	0	1	1	1	1	9
3	H	1	2	2	1	0	1	1	1	9
2	A	2	0	0	1	1	2	0	0	6
2	D	1	1	1	0	1	1	1	0	6
2	F	1	1	1	0	1	1	0	1	6
1	B	0	1	1	0	0	0	0	1	3
1	C	1	0	0	1	0	1	0	0	3
18	Total	9	9	9	4	5	9	4	5	54

Par exemple pour le mot H cité 3 fois, une fois il s'agit d'une homme et deux fois d'une femme; deux fois il s'agit d'âge 1, une fois d'âge 2, zéro fois d'âge 3.

française de 15 ans et plus, et nous l'avons proposé à une centaine d'étudiants en lui ajoutant une question ouverte.

On trouvera les intitulés exacts en annexe mais les questions sont centrées sur les thèmes suivants:

1) facilité ou difficulté de la langue ou de la grammaire française;

2) spécificité de l'orthographe: est-ce une matière scolaire comme une autre; un casse-tête inutile ou au contraire un charme de la langue et un élément du patrimoine;

3) à qui confier une éventuelle réforme: à l'Académie, au ministère de l'Education nationale, de la Culture, aux écrivains, aux dictionnaires, à une autre instance collective;

4) quels pourraient être les éléments d'une éventuelle réforme: une tolérance sur plusieurs façons d'écrire, une suppression des accents circonflexes, des traits d'union, des consonnes doubles, un remplacement des "ph" par des "f", ou des "x" par des "s" dans les pluriels en "oux";

5) des jugements sur la réforme: y est-on favorable, est-elle urgente pour l'apprentissage, il est impossible de la faire sans dénaturer la langue, il est possible de retoucher certaines bizarreries.

En plus de ces questions fermées qui étaient celles de l'enquête de *Lire*, on a ajouté une question ouverte qui demandait de dire en quelques phrases ce que l'on pensait du problème de la réforme de l'orthographe.

Les 96 individus interrogés ont émis 3333 mots; c'est-à-dire, des phrases moyennes de 35 mots. Le vocabulaire ainsi constitué comprend 932 mots différents.

Les 141 mots les plus fréquents émis plus de 3 fois ont été retenus et sont les suivants:

de	186	l'	131	la	104	est	80
orthographe	76	à	65	une	64	il	55
et	52	langue	52	d'	46	le	45
pas	45	en	41	des	40	les	39
un	38	qui	37	réforme	37	pour	34
que	30	on	26	n'	26	problème	25
ce	24	qu'	23	ne	23	je	22
mais	21	elle	20	apprentissage	19	plus	17
dans	16	c'	16	par	16	française	16
faire	14	s'	14	réformer	14	son	14
a	14	être	13	ou	13	sont	13
culture	12	du	12	sa	12	y	11
mots	11	difficulté	11	quelques	10	faut	10
peut	10	comme	10	notre	10	fait	10
cette	9	donc	9	au	9	trop	9
avec	9	autres	8	difficile	8	tout	8
se	8	si	8	cela	8	changer	8
sans	8	aussi	7	serait	7	pense	7
doit	7	règles	7	simplifier	7	toute	7

faciliter	7 français	7 tous	6 fautes	6
écrite	6 celle	6 me	6 partie	6
faudrait	6 certaines	6 contre	6 ceux	6
semble	6 aux	6 enseignemet	6 bien	6
évolution	6 -t-	6 pourquoi	5 patrimoine	5
beaucoup	5 changement	5 histoire	5 ont	5
gens	5 façon	5 nécessaire	5 elles	5
non	5 plutôt	5 sous	5 vraiment	5
suis	5 scolaire	5 vrai	5 absurdités	5
sens	5 apprendre	5 très	5 temps	5
peu	4 soit	4 oui	4 quoi	4
?	4 bizarrerie	4 sur	4 telle	4
pouvoir	4 charme	4 lui	4 leur	4
ils	4 ses	4 besoin	4 problèmes	4
autre	4 supprimer	4 faux	4 écrit	4
enfants	4 écriture	4 possible	4 modification	4
moins	4			

On trouve dans cette liste un mélange de mots-outils et de mots concernant le problème de la réforme de l'orthographe.

On représente dans le graphique en annexe le premier plan factoriel (1ère valeur propre égale à 0,01939 représentant 24% du Phi-deux total et 2ème valeur propre égale à 0,00867 représentant 11%). Sur ce plan on ne représente que les mots ou les modalités qui sur l'un ou l'autre des facteurs ont une contribution supérieure à la moyenne. Les mots du vocabulaire sont en minuscules et les modalités de réponse des questions fermées sont en majuscules.

Interprétation du plan factoriel

L'opposition majeure, celle du premier axe, se fait en terme d'opposition ou non à la réforme. Sur la partie droite de l'axe se trouve le vocabulaire émis par des gens qui sont très hostiles à la réforme, qui trouvent que l'orthographe est facile et est un des charmes de la langue, et qui disent non à toutes les propositions de réforme. A gauche on a les positions opposées: on est favorable à la réforme, on trouve l'orthographe difficile, qu'elle n'est pas un charme de la langue ni un patrimoine, qu'il est urgent de simplifier, que c'est un pensum, un casse-tête inutile et qui approuvent toutes les propositions faites pour réformer l'orthographe.

La deuxième opposition (de l'axe vertical), se fait entre ceux qui acceptent que des instances collectives (Académie, ministères ou autres) aient un droit de modification de l'orthographe contre ceux qui le refusent.

Les contributions au côté droit (opposé à la réforme) et inférieur (contre le droit de modification de toute instance) sont le fait des plus hostiles qui refusent même la suppression des bizarreries et absurdités. Ce vocabulaire de la droite et du bas du graphique insiste sur l'orthographe qui fait **partie** du **patrimoine** et de la **culture** légués par l'**histoire**; **pourquoi** donc en **changer** puisque cela s'est fait au cours du **temps**, c'est ce qui en **fait** le **charme** (les mots en gras sont les mots de ce vocabulaire rassemblés en une phrase en forme de type-idéal).

Quand, à partir de ce secteur de droite du graphique, on remonte vers le haut, on peut éclairer sensiblement le rapprochement à première vue paradoxal entre la modalité "assez hostile à la réforme" et l'acceptation de la modification par des instances collectives, en examinant le vocabulaire émis.

Ce qui marque cet ensemble c'est le jeu entre les différents mots: **changement, modification, simplifier, réformer**, vocabulaire qui peut sembler surprenant dans un secteur où l'on est "assez hostile" à la réforme mais qui s'explique parce que ces mots de changement sont utilisés avec la réserve de la faible importance de ce qu'il est éventuellement possible de changer. On n'accepte de ne modifier que **certaines** bizarreries ou absurdités et la finalité est de faciliter l'**enseignement** ou l'**apprentissage** pour les **enfants** quoique ce soit cet enseignement lui-même qu'il faille parfois réformer plutôt que l'orthographe.

Devant ces acceptations de certaines modifications, on s'aperçoit que le choix de "assez hostile à la réforme" se révèle être une position intermédiaire: on s'oppose à ce que l'on pense être une réforme radicale, de type phonétique, et l'on envisage des modifications prudentes qui peuvent être proposées par des instances collectives. Au lieu de laisser le soin de l'évolution à l'histoire, comme dans le groupe précédent, on accepte que l'orthographe soit considérée comme un instrument modifiable par la société, à condition que ce soit avec précaution.

Si l'on examine enfin la partie gauche du graphique, on remarque que l'on y trouve soit un vocabulaire marqué par le monde scolaire (enfants, scolaire, règles, fautes, écriture), soit un vocabulaire sans spécificité dont les mots doivent être examinés un par un. "Nécessaire" s'applique en général à la réforme, mais "pour" permet de discerner une finalité généralement altruiste: "pour les enfants d'immigrés, les enfants en difficulté scolaire, les étrangers".

Cette double perspective, celle de l'école d'une part, avec tout ce fond de difficultés douloureuses qui lui est associé (la "faute" en particulier avec son aspect moral), et la finalité "altruiste" au service des plus défavorisés, fait que le vocabulaire associé positivement à la réforme semble issu de motivations symboliques plus que techniques. On est pour la réforme parce qu'on pense qu'elle peut être une solution aux problèmes de difficultés scolaires, soit celles que l'on a ressenties soi-même, soit celles que l'on voit chez d'autres. Pour plusieurs d'ailleurs, chez ceux qui ont le plus de réponses favorables à la réforme, celle-ci est vue comme une utopie du fait de sa radicalité. Par exemple: "la réforme s'impose-t-elle vraiment?" (13 réponses favorables à la réforme sur les 15 possibles); "c'est pure utopie que de réformer, il faudrait réimprimer tous les écrits" (12 réponses favorables).

Conclusion

Ce que manifeste le plan factoriel des questions fermées et du vocabulaire de la question ouverte, c'est que la réforme de l'orthographe est perçue

comme une réforme **radicale**, bien que le thème d'une réforme de ce genre (orthographe phonétique) ne soit pas présenté par les questions fermées. Ceci s'explique pour des raisons historiques car l'arme la plus employée par les adversaires de la réforme a toujours été la mise en avant d'une orthographe phonétique qui ridiculise la réforme (car elle est illisible) et qui fait peur.

Cette réforme utopique est soit acceptée soit critiquée mais toujours sur le mode radical qui exclut une réforme modérée (on est pour, afin de lutter contre l'échec scolaire, ou on est contre, pour éviter la perte d'un patrimoine historique). Quand on commence à discuter de modifications modérées, modifications des bizarreries, on se positionne en hostilité moyenne à la réforme (perçue donc comme radicale).

De ce fait on comprend mieux pourquoi le Premier ministre dans son discours du 24 octobre 1989 installant le Conseil supérieur de la langue française éprouve le besoin de se démarquer de l'idée de réforme étant donnée la connotation radicale qui lui est associée: "Une véritable réforme, qui modifierait les principes mêmes de la graphie de notre langue, et altérerait donc son visage familier, me paraît absolument exclue. A vrai dire, personne ou presque ne la propose. (...) En revanche, en dépit des modifications effectuées depuis deux siècles, il reste encore à opérer des "rectifications utiles". Ceux qui sont favorables, comme le Premier ministre, a une réforme limitée ont donc intérêt à abandonner le terme de "réforme".

Quant à ceux qui sont contre la réforme, il leur est facile de critiquer les partisans idéologiques de la réforme (réforme "altruiste" pour les défavorisés): par exemple, l'Académie, dans sa prise de position du 16 novembre 1989, déclare que la lutte pour les défavorisés est l'argument "majeur" pour la réforme. Cette radicalisation idéologique lui permet de répondre par un argument de fait: "le nombre élevé d'enfants des familles défavorisées qui sont parvenus, depuis l'instauration de l'école primaire obligatoire, aux places les plus éminentes de l'administration, l'Université, l'armée et le Service de l'Etat, enlève à cet argument toute portée". De plus, l'Académie insiste sur l'héritage culturel et laisse le soin des ajustements au "processus naturel d'évolution": on retrouve là les thèmes de la partie droite du plan factoriel dont on voit bien comment il s'inscrit dans le débat social. On y retrouve une opposition idéologique gauche/droite, "souci des défavorisés" opposé à "souci du patrimoine national" croisée et non superposée avec une opposition blocage/modification.

=====

Annexe 2 **Implémentation logicielle**

Les données ont été traitées par un ensemble de programmes complémentaires de Tri-deux dont on utilise CODLOG et ANCORR. L'ensemble forme le logiciel "**Tri-deux**" dont une version test fonctionne. Les données doivent être présentées de la manière suivante:

1) l'enregistrement doit commencer de manière habituelle comme pour un traitement dans Tri-deux; c'est-à-dire, avec un numéro d'individu (facultatif) suivi, en format fixe, des réponses aux différentes questions fermées codées sous forme numérique. Cet ensemble forme une zone de "caractéristiques" de longueur constante pour tous les individus.

2) à la suite de cette zone de caractéristiques, on entre en format libre le texte de la question ouverte, ou une suite de mots séparés par des blancs si on emploie la technique qui consiste à donner un stimulus et à demander à l'interrogé les "mots associés" à ce stimulus. Le texte peut être disposé sur la fin de l'enregistrement des caractéristiques ainsi que sur un ou plusieurs autres enregistrements. Il n'est pas nécessaire que ce nombre soit fixe mais la fin du texte doit être notée par une étoile.

=====

