

[Saadi Lahlou](#)

Réponse à une question ouverte: incidence du mode de questionnement

Book section

Original citation:

Beaudouin, Valérie and Lahlou, Saadi and Yvon, François (1993) *Réponse à une question ouverte: incidence du mode de questionnement*. In: École Nationale Supérieure de Télécommunications, (corp. ed.) Actes des secondes journées internationales d'analyse statistique de données textuelles : [Montpellier, 21 et 22 octobre 1993]. École Nationale Supérieure de Télécommunications, Paris, France, pp. 133-145.

© 1993 [ENST](#)

This version available at: <http://eprints.lse.ac.uk/32919/>
Available in LSE Research Online: May 2011

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's submitted version of the book section. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

(1993) - BEAUDOUIN, Valérie, LAHLOU, Saadi, YVON, François.
Analyse d'une question ouverte : incidence du mode de
questionnement. *Actes des secondes journées internationales d'analyse
statistique des données textuelles*. Montpellier, 21-22 octobre 1993.
Paris, ENST. pp. 133-145.

RÉPONSE À UNE QUESTION OUVERTE : INCIDENCE DU MODE DE QUESTIONNEMENT

Valérie BEAUDOUIN

CREDOC, Paris, France

Saadi LAHLOU

E D F - D E R / GRETS, Clamart, France

François YVON

E N S T, Paris, France

SUMMARY

Two methods of data collection (mailing and telephone interviews) were applied to the same open-ended question : "What is an ideal breakfast ?" on equivalent samples (N=1000). Lexicometric analysis and descending classification demonstrate important lexical, syntactic and semantic differences in the corpuses.

Telephone answers are shorter, lexically and syntactically poorer ; they focus on objective product descriptions (breakfast ingredients) with a prototypic approach. In contrast, written answers, longer and richer, are more global, conceptual and idiosyncratic, they also include contextual aspects (time, place, attendees...) and personal judgments.

Cette contribution expose des recherches sur l'incidence du mode de questionnement sur les réponses à une question ouverte.

Ce travail a pu être réalisé grâce à une collaboration étroite entre la SECODIP et le CREDOC en vue de faire progresser les méthodes d'enquête. On compare ici les réponses obtenues à une même question ouverte dans deux protocoles : auto-administré postal ("AUTO") et enquête téléphonique ("TELE"). Après un premier test des questions effectué par la SECODIP auprès de 80 membres de son panel de consommateurs, le corpus AUTO a été recueilli par la SECODIP auprès de 1001 panélistes en septembre 1992 (Beaudouin et al. 1993).

Le corpus TELE a été obtenu dans le cadre de l'enquête CREDOC sur la consommation auprès d'un échantillon de 1005 personnes en novembre 1992 (Lahlou, Collierie de Borely & Beaudouin, 1993).

Les différentes techniques de questionnement sont plus ou moins bien adaptées au recueil de questions ouvertes pour l'analyse lexicale (Yvon, 1990). Actuellement les modes d'interrogation les plus répandus sont :

- l'entretien retranscrit, en face à face ou par téléphone : l'enquêteur note ou saisit sur clavier les réponses de l'enquêté.

- le questionnaire auto-administré : l'enquêté se charge lui-même d'écrire la réponse sur papier, Minitel ou micro-ordinateur.

1. Le petit déjeuner idéal : une même question, deux modes d'enquête (téléphone, auto-administré)

Nous comparons ici deux modes de questionnement : l'enquête téléphonique et le questionnaire auto-administré. La même question a été posée aux deux échantillons décrits ci-dessus :

Un petit déjeuner idéal, à quoi ça vous fait penser ?

Si la formulation est parfaitement semblable dans les deux protocoles, trois différences majeures les distinguent :

- le mode d'interrogation (téléphone, auto-administré) ;

- le contexte de la question. Dans TELE, les questions précédentes ne portaient pas sur le petit déjeuner mais sur d'autres aspects de la consommation alimentaire. En revanche, l'ensemble du questionnaire AUTO était centré sur le petit déjeuner.

<p>1. Un petit déjeuner idéal, à quoi ça vous fait penser ?</p> <p>2. Décrivez votre petit déjeuner habituel</p> <ul style="list-style-type: none">- en semaine (ambiance, lieu, heure, produits consommés)- en week-end, en vacances (ambiance, lieu, heure, produits consommés)
--

Figure 1
Questionnaire SECODIP

- l'échantillon. Celui de TELE est représentatif de la population française. Celui de AUTO est un échantillon de *ménagères panélistes* représentatif des *ménages* français.

Examinons d'abord les sources de variations en dehors du mode de questionnement, c'est-à-dire l'incidence possible des différences de contexte et d'échantillon.

D'une part, il était à craindre que l'enquête de AUTO, après lecture de l'intégralité du questionnaire, soit conduit à faire une description du petit déjeuner idéal reprenant les rubriques imposées pour la description du petit déjeuner habituel (ambiance, lieu, heure, produits). Or, ce n'est pas ce qui se produit dans la majorité des cas.

D'autre part, dans la pré-enquête auto-administrée, où seule figurait la Question 1, on obtenait des réponses du même type que celles de l'enquête finale. Les questions sur la description fine des produits, heure et lieu ont précisément été rajoutées pour obtenir systématiquement des précisions concrètes qui manquaient souvent.

En ce qui concerne les échantillons, les panélistes sont exclusivement des femmes issues de ménages représentatifs de la population française, alors que l'échantillon de l'enquête téléphonique est composé d'hommes et de femmes. L'influence du sexe sera étudiée plus loin. Les panélistes ont l'habitude de remplir, chaque semaine, un questionnaire décrivant leurs achats alimentaires. On peut penser que ceci sélectionne des ménagères assez appliquées et motivées. Cela devrait aussi les orienter vers des descriptions plus fines (en terme de marque, de conditionnement) des produits consommés. Ce dernier point ne s'est pas vérifié, il est plus difficile de se prononcer sur l'influence de la motivation, nous y reviendrons.

2. Comparaison de la richesse lexicale

2.1. Réponses écrites plus longues et plus variées que les réponses orales

Le nombre moyen de mots par réponse est nettement plus élevé par écrit (19) que par téléphone (15). Il en va de même pour la taille du vocabulaire (AUTO : 2 370 mots distincts ; TELE : 874).

Les courbes de Zipf permettent de localiser les fréquences d'occurrences pour lesquelles les écarts sont les plus importants.

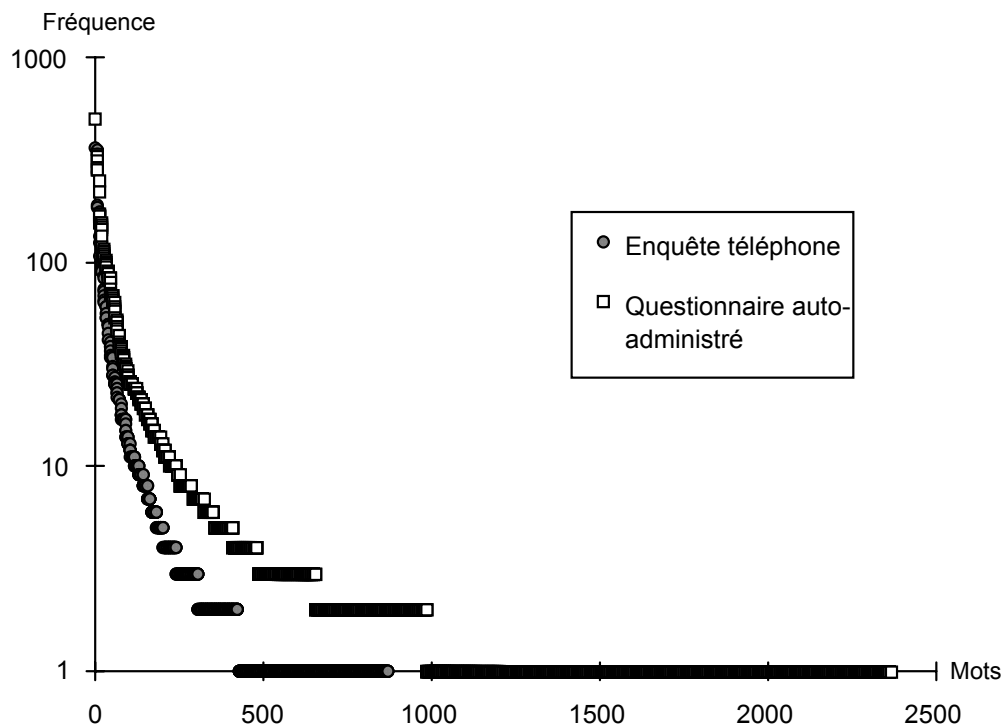


Figure 2

Courbes de Zipf sur les deux corpus de réponses

L'écart est principalement dû aux mots ayant un très faible nombre d'occurrences (deux fois plus nombreux dans AUTO). Pour les fréquences de mots élevées, on a à peu près autant de mots en auto-administré que par téléphone.

Les réponses du corpus AUTO sont plus longues et utilisent un vocabulaire beaucoup plus varié. C'est la première différence majeure entre les deux corpus.

On peut avancer les explications suivantes :

- le temps. Par écrit, les personnes interrogées ont le temps de construire leur représentation mentale de l'objet et de choisir les mots les

mieux adaptés à l'énonciation de cette représentation. Il en découle logiquement des réponses plus longues et plus riches.

Notons au passage que le temps est sans doute un facteur déterminant pour l'expression de la part la plus personnelle de la représentation mentale. Placé dans une situation d'urgence par l'appel téléphonique, l'enquêté aura tendance à donner une réponse courte qui a toutes les chances de reproduire des stéréotypes (culturels, publicitaires ...) ; phénomène déjà observé sur d'autres corpus (Beaudouin, Hébel et Lahlou, 1992).

- la motivation. Il semble que l'enquêté interrogé par téléphone s'investisse peu dans sa réponse. Interrompu dans ses activités quotidiennes par l'appel téléphonique, il n'a souvent qu'un souci, celui de se débarrasser au plus vite de l'enquêteur. Notons que les relances insistantes atténuent l'influence de ce phénomène sur les corpus recueillis.

2.2. Réduire la taille du vocabulaire

Pour comparer les profils lexicaux des réponses avec une bonne robustesse statistique, il est souvent nécessaire de diminuer le nombre de mots entrant dans l'analyse. Différentes techniques (Benzécri et al., 1981 ; Lebart et Salem, 1989) permettent d'organiser cette réduction du vocabulaire. La plus commune consiste à éliminer les mots-outils, qui sont à la fois très fréquents et peu pertinents. Ensuite, on peut faire disparaître les mots trop rares, en se fixant un seuil en-deça duquel les mots sont éliminés. Enfin, une dernière méthode de réduction consiste à lemmatiser les réponses. Cette réduction du vocabulaire n'est pas neutre : elle a tendance à réduire les singularités.

En utilisant conjointement ces méthodes, et en se fixant un seuil de quatre occurrences pour les mots trop rares, nous avons obtenu les résultats suivants sur nos deux corpus :

	Téléphone		Auto-administré	
	Avant réduction	Après réduction et reconnaissance des mots-outils	Avant réduction	Après réduction et reconnaissance des mots-outils
Nombre de MOTS DISTINCTS	874	232	2 370	453
dont nombre d'hapax	440	0	979	0
Nombre de MOTS PLEINS		165		317
Nombre de MOTS OUTILS		67		136
Nombre moyen de MOTS PAR RÉPONSE	15,4	13,4	18,7	13,4

Figure 3

Comparaison de la richesse lexicale sur les deux corpus et effet de la réduction

Après traitement, on constate d'une part que la taille du vocabulaire a été divisée par quatre dans TELE, par cinq dans AUTO. Plus la taille du vocabulaire initiale est grande, plus les effets relatifs de la réduction sont importants. On a vu sur les deux corpus que l'écart maximal au niveau de la répartition du vocabulaire se situait dans les faibles fréquences. En éliminant les mots peu employés, on a tendance à réduire mécaniquement l'écart absolu entre les deux corpus.

L'élimination des mots-outils a tendance à être un peu plus défavorable aux textes écrits. En effet, la proportion des mots-outils est plus importante en auto-administré : 43 % contre 41 %. Ceci est assez prévisible : dans la rédaction écrite, le recours à l'appareillage syntaxique est beaucoup plus naturel que dans le langage parlé, où la coordination ou la juxtaposition se substituent facilement à la subordination. Nous reviendrons sur ce point lors de l'étude des mots les plus fréquents.

En fin de course, le vocabulaire analysé ne représente plus que 13 % du vocabulaire initial dans AUTO, et 19 % dans TELE. Les nombres moyens de mots par réponse ont été égalisés (13,4).

En résumé : les réponses écrites sont plus longues, puisent dans un vocabulaire plus vaste et ont davantage recours aux outils syntaxiques que les réponses orales. Ceci est en partie dû au respect des règles d'expression du discours (respect de la syntaxe) qui pèsent plus dans la rédaction que dans l'expression orale. La réduction de la taille du vocabulaire indispensable à l'analyse lexicale a tendance à réduire les différences entre l'écrit et l'oral.

3. Les différences de contenu

3.1. Les mots les plus fréquents

L'examen des mots les plus fréquents, - cités par plus d'une personne sur dix - donne un aperçu des différences de contenu dans les réponses.

Auto-administré				Téléphone			
336	<i>journée</i>	107	jus de fruit	413	beurre	134	chocolat
281	café	102	grillé	361	café	118	grillé
251	pain	100	chocolat	344	confiture	106	thé
237	bon	99	frais	343	pain	83	bon
202	beurre	97	<i>repas</i>	278	jus d'orange		
195	confiture	94	jus d'orange	255	tartine		
186	<i>famille</i>	93	<i>commenc+</i>	197	croissant		
174	lait	93	<i>détend+</i>	191	lait		
161	croissant	90	<i>vacances</i>	171	café au lait		
156	bien	89	<i>calme</i>	166	céréale		
156	<i>temps</i>	89	fruit	139	oeuf		
155	<i>moment</i>	86	<i>table</i>	135	fruit		
115	<i>faire</i>			135	jus de fruit		

Figure 4

Les mots les plus fréquents (en auto-administré et par téléphone)

Légende

Les termes en “police standard” correspondent aux produits.

Les termes en “gras” aux caractérisants hédoniques.

Les termes en “italique” aux termes caractérisant le contexte, ...

Sur les deux corpus, ce sont les produits composant le petit déjeuner qui sont les plus souvent cités (description pragmatique, en termes d'objets-du-monde). L'ensemble café, pain, beurre, confiture vient en tête sur les deux sous-échantillons - c'est le prototype du petit déjeuner le plus répandu. Notons que par téléphone, les mots les plus fréquents ont tendance à être encore plus souvent répétés que par écrit.

3.2. Structure syntaxique

La syntaxe est plus complexe dans AUTO, comme en témoigne la liste des mots-outils les plus fréquents.

Auto-administré				Téléphone	
507	et	117	c'est	184	avec
331	ou	116	se	178	et
314	en	113	que	99	ou
289	pour	106	me	86	en
283	avec	106	on		
148	dans	104	qui		
143	je	98	ne		

Figure 5

Les mots-outils les plus fréquents (en auto-administré et par téléphone)

Les conjonctions de coordination et de subordination ont tendance à disparaître dans TELE (sans doute davantage du fait de la retranscription).

Le *je* est presque absent de TELE tandis que dans AUTO, 14 % des enquêtés l'emploient. Par téléphone, la présence de l'énonciateur est tellement évidente qu'elle n'a pas besoin d'être signalée. Par écrit, le respect de la syntaxe pousse à employer le *je* plus fréquemment.

Tout se passe comme si, en passant de l'écrit à l'oral, on passait du syntagme nominal avec subordonnées au syntagme nominal simple. La structure phrastique avec subordonnées cède la place à une juxtaposition de syntagmes nominaux.

Réponses écrites :

* Aux périodes de vacances car c'est là qu'on l'apprécie le plus. On a le temps. Idéal veut dire avec le temps, le soleil dehors, un pain très frais et chaud, et encore mieux à deux.

* À quelque chose de bien préparé, d'attendu, c'est un moment privilégié qui prépare la journée, que l'on ne devrait jamais manquer car c'est très important et cela évite les coups de pompe de la matinée.

Réponses orales :

* Jus d'orange, com flakes, beurre, pain, laitage, fruit, c'est tout.

* Une orange pressée ou un pamplrousse pour la vitamine C, du lait avec du cacao, une tartine de beurre ou de confiture et un yaourt.

L'analyse comparée de l'effet de l'écholalie (phénomène de répétition qui se propage de la question à la réponse) montre aussi les différences dans le traitement de la syntaxe entre l'oral et l'écrit : ainsi *petit-déjeuner* apparaît avec des fonctions syntaxiques différentes. Dans AUTO (263 occurrences), il figure comme embrayeur de réponse (*le petit-déjeuner idéal c'est...*) ; dans TELE (54 occurrences), il est support d'une caractérisation (*petit-déjeuner anglais...*).

3.3. Contenu sémantique

Le contexte social, géographique, l'ambiance et de place du petit déjeuner dans la journée n'apparaissent de façon significative que dans AUTO. Dans TELE, le petit déjeuner est un ensemble d'ingrédients, mis sur la table, préparés et/ou consommés. La description est principalement objective, succincte et pragmatique : il s'agit d'une *focalisation locale* :

Café au lait, tartines de pain frais avec beurre et confiture de fraises

Un jus d'orange, des oeufs brouillés avec bacon ou jambon, fromage, toasts, confiture, thé

A l'inverse, dans AUTO, la focalisation est *globale* ; les descriptions sont enrichies de considérations physiologiques, sociales et morales. L'esprit ne se contente plus d'énumérer les ingrédients situés sur la table mais de voir aussi ce qu'il y a autour dans l'espace (famille, musique, radio...) et dans le temps (point de départ de la journée...). La description est alors multimodale, elle inclut aussi des commentaires et des jugements personnels. Par exemple :

Aux vacances, c'est un des rares moments où l'on prend son temps pour le petit déjeuner, détendu, au calme, instants délicieux

x sur une terrasse, dans le jardin, où l'on savoure ce que l'on a envie de manger, où l'on prend le temps d'écouter. A partager avec ceux que l'on aime.

Il nous semble que ces deux types de réponses correspondent à deux modes cognitifs différents, et déjà repérés dans deux enquêtes (Beaudouin, Hébel et Lahlou, 1992) : l'approche prototypique (Rosch, 1976 ; De Fornel, 1990) et l'approche conceptuelle (Beaudouin et Lahlou, 1992).

Cette dernière n'a pas forcément une valeur supérieure. La simplicité a une efficacité pragmatique et communicative évidente. La présence immédiate de l'enquêteur joue un rôle, l'enquêté cherchant à lui faire comprendre le plus rapidement possible son point de vue. La communication sur les objets (produits) est plus immédiate que la communication sur les concepts.

3.4. Typologie des réponses

Pour confirmer ces résultats, examinons une analyse simultanée des deux corpus, par classification descendante hiérarchique, à l'aide d'Alceste (Reinert 1983, 1987, 1990, 1992).

L'analyse met en évidence six classes qui se répartissent en deux grandes surclasses : les énumérations de *produits* d'une part et les évocations du *contexte* (spatial, temporel et social) d'autre part.

Figure 6

Typologie des réponses à la question sur le petit déjeuner idéal

Dès le premier noeud les réponses AUTO se distinguent des réponses TELE. La classe “contexte” (31 % des réponses) est constituée à 93 % par des réponses de AUTO : ces réponses ont trait au cadre et à la convivialité. La classe “produit” (69 % des réponses, correspondant à des énumérations de produits) est composée aux trois quarts de réponses provenant de TELE.

	“produits”	“contexte”
TELE	72 %	7 %
AUTO	28 %	93 %
Total	100	100

Figure 7

Contenu selon le questionnement

Les réponses qui prennent en compte le contexte et l'entourage sont pratiquement absentes de TELE (5 % seulement). En revanche (cf. Figure 8), AUTO fournit à la fois des réponses en terme de noms de produits (38 %) et des réponses évoquant le contexte (62 %). La "représentation sociale" du petit déjeuner qui transparaît dans AUTO est donc beaucoup plus variée que celle de TELE. Ces différences ne sont pas dues à la structure par sexe des échantillons, comme le montre à l'évidence le tableau suivant :

	TELE			AUTO	Ensemble
	Homme	Femme	Total	Femme	
“produits”	94 %	97 %	95 %	38 %	67 %
“contexte”	6 %	3 %	5 %	62 %	33 %
Total	100	100	100	100	100

Figure 8

Répartition des réponses suivant le sexe et le mode d'enquête

La différence de profil ne peut donc s'expliquer par la différence de structure de l'échantillon (absence d'hommes dans le panel), mais bien par l'incidence du mode de questionnement. En bref, TELE favorise des réponses descriptives centrées sur l'objet tandis que AUTO autorise un élargissement du champ des associations.

Conclusion

L'interrogation téléphonique (urgence, présence d'un interlocuteur) induit des réponses plus courtes, un vocabulaire pauvre ; l'absence de structure syntaxique, des descriptions (noms de produits) visant d'abord l'efficacité pragmatique.

A l'inverse, les réponses écrites sont plus longues, riches en vocabulaire, hapaxiques, syntaxiquement bien formées, plus conceptuelles, et font appel à un champ de représentation large, englobant des aspects sociaux.

Il semble bien que le temps soit une condition nécessaire à l'élaboration d'une réponse plus conceptuelle, moins descriptive et sans doute plus personnelle.

Le mode de questionnement a donc une forte incidence sur les réponses ouvertes. Si l'on veut accéder aux représentations immédiates et spontanées, l'enquête téléphonique paraîtra bien adaptée. L'auto-administré est préférable pour obtenir une expression variée et complexe des représentations mentales.

Références

- Beaudouin V., Lahlou S., *Analyse lexicale : outils d'exploration des représentations*. Cahier de recherche n°48, CREDOC, Paris, 1993.
- Beaudouin V., Hébel P., Lahlou S., Le Bihan H., "Enquête sur la qualité perçue", *Comprendre et évaluer la qualité*. Cahier de recherche n°39, CREDOC, Paris, 1992.
- Beaudouin V., (CREDOC), Gilles M.-O., Lahlou S. (CREDOC), Marsal F., Peyre L., Wattrelot S., Yvon F., (ENST), *Analyse automatique du discours : le petit déjeuner*. Rapport de groupe de travail de 3e année à l'ENSAE, juin 1993.
- Benzécri J-P.(éd.) *Pratique de l'analyse des données, Tome 3 : Linguistique et lexicologie*. CNRS-Dunod, Paris, 1981.
- Fornel M. de., "Sémantique du prototype et analyse de conversation", *Cahiers de linguistique française*, n° 11, Genève, 1990.
- Lahlou S., Collerie de Borely A., Beaudouin V., *Où en est la consommation aujourd'hui ? : une enquête sur le consommateur français des années 90*. Cahier de recherche n°46, CREDOC, Paris, 1993.
- Lebart L., Salem A.- *Analyse statistique des données textuelles*. Préface de Christian Baudelot.- Dunod, Paris, 1988.
- Reinert M., "Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte", *Les cahiers de l'analyse des données*, Vol. VIII, n° 2, p 187-198, 1983.
- Reinert M., "Classification descendante hiérarchique et analyse lexicale par contexte : application au corpus des poésies d'Arthur Rimbaud", *Bulletin de Méthodologie Sociologique*, n°13, 1987.
- Reinert M., *La méthodologie ALCESTE et l'analyse d'un corpus de 304 récits de cauchemars d'enfants*, Convention internationale "Ricerca qualitativa e computer nelle scienze sociali", 1992.
- Rosch E. R., Mervis C. B., Gray W., Johnson D. N., Boyes-Braem P., "Basic Objects in Natural Categories", *Cognitive Psychology*, 8, p 382-439, 1976.
- Yvon F., *L'analyse lexicale appliquée à des données d'enquête : état des lieux*. Cahier de recherche n°5, CREDOC, Paris, 1990.