



Bulletin de méthodologie sociologique

Bulletin of sociological methodology

82 | 2004

April

Méthode et stratégie d'analyse des questions ouvertes du panel électoral français

Mathieu Brugidou, Nadine Mandran, Michèle Moine et Annie-Claude Salomon



Édition électronique

URL : <http://journals.openedition.org/bms/1068>

ISSN : 2070-2779

Éditeur

Association internationale de méthodologie sociologique

Édition imprimée

Date de publication : 30 avril 2004

Pagination : 5-26

ISSN : 0759-1063

Référence électronique

Mathieu Brugidou, Nadine Mandran, Michèle Moine et Annie-Claude Salomon, « Méthode et stratégie d'analyse des questions ouvertes du panel électoral français », *Bulletin de méthodologie sociologique* [En ligne], 82 | 2004, mis en ligne le 09 juillet 2008, consulté le 30 avril 2019. URL : <http://journals.openedition.org/bms/1068>

Ce document a été généré automatiquement le 30 avril 2019.

© BMS

*Méthode et stratégie d'analyse des questions ouvertes du panel électoral français**

Mathieu Brugidou, Nadine Mandran, Michèle Moine et Annie-Claude Salomon

- 1 Cet article se propose de décrire un dispositif d'enquête électoral qui comporte de nombreuses questions ouvertes, puis de détailler les stratégies d'analyses retenues et enfin, à partir de quelques exemples de traitement, de montrer les avancées méthodologiques – et leurs limites – des approches mêlant des dimensions quantitative et qualitative.
- 2 Lors de la séquence électorale du printemps 2002, un dispositif d'enquête comprenant un panel a permis de mesurer les opinions et les comportements politiques des français en trois vagues d'enquêtes¹. La première partie de la communication décrira le dispositif, les types de questions ouvertes posées et reviendra sur la question de la qualité des « données ».
- 3 Les réponses à une partie de ces questions ont été intégralement retranscrites et ont été analysées par le logiciel d'analyse textuelle Alceste. Le choix de ces questions ainsi que l'approche en terme d'analyse des données textuelles seront explicités dans un deuxième temps de l'intervention. Les stratégies d'analyse à partir d'une question ouverte portant sur la participation au premier tour de l'élection présidentielle seront commentés à deux niveaux :
 - au niveau linguistique, en montrant en quoi, la méthode – au delà de l'analyse thématique – permet de faire apparaître des discours,
 - au niveau de la sociologie politique, en explicitant certaines des hypothèses guidant les stratégies d'analyse.
- 4 Il est par ailleurs possible de croiser les classes produites par Alceste avec les questions fermées et de revenir à une approche « explicative » plus classique. Dans une troisième partie, deux de ces protocoles seront exposés : l'un partant directement de la

classification des discours - à propos d'une question ouverte sur les principaux problèmes environnementaux - l'autre impliquant une analyse de contenu préalable des lexiques - à propos de questions portant sur les raisons de la défaite de la gauche ou de la victoire de la droite.

- 5 Ces approches posent toutefois une série de questions épineuses tant du point de vue technique qu'épistémologique que nous ne pourrions qu'indiquer en guise de conclusion.

Présentation des questions ouvertes

Le dispositif

- 6 A l'occasion des élections présidentielle et législatives de 2002, le Centre d'étude de la vie politique française (CEVIPOF) de Sciences Po Paris, le Centre d'informatisation des données socio-politiques (CIDSP) de Sciences Po Grenoble, le Centre d'études et de connaissances sur l'opinion publique (CECOP) et l'Université de Montréal (Etudes électorales) ont uni leurs efforts, avec l'aide du Ministère de l'Intérieur et de la Fondation nationale des sciences politiques (FNSP), pour mettre en place une grande enquête électorale réalisée tout au long du processus électoral, d'avril à juin 2002.
- 7 L'enchaînement des deux scrutins a présenté une occasion favorable pour conduire une enquête par panel permettant de suivre les trajets électoraux individuels de 1417 électeurs, de la veille de l'élection présidentielle aux élections législatives, en passant par les deux tours de l'élection présidentielle. Trois enquêtes ont donc été réalisées selon le calendrier suivant : avant le premier tour de l'élection présidentielle (vague 1), après le deuxième tour de l'élection présidentielle (vague 2), après le deuxième tour des élections législatives (vague 3).

Les questions ouvertes

- 8 Une des originalités de ce dispositif repose sur la relative importance accordée aux questions ouvertes, assez rares habituellement dans les enquêtes électorales, pour des raisons de coût d'administration notamment.
- 9 Quatre types de questions figurent dans le PEF2002, qui peuvent être identifiés² comme suit :
 - les questions ouvertes recodées par l'enquêteur : informations factuelles, signalétiques, dont le recodage peut être revu, à partir du verbatim recueilli, selon d'autres hypothèses ;
 - les postes « autre » : reliquat d'une question fermée, en général d'effectif faible ; cette modalité permet de s'assurer que l'univers des réponses possibles était bien balisé par les items proposés ;
 - les questions ouvertes demandant une « simple citation », sans développement ni argumentation : le recodage manuel fin est nécessaire et s'effectue d'après les catégories d'analyse de la science politique et les hypothèses formulées ;
 - les questions ouvertes demandant une argumentation plus ou moins développée sous forme de « justification ». Les réponses sont plus complexes et ne peuvent pas être réduites à des items simples ; elles relèvent d'une analyse de discours. Les stratégies d'analyse relèvent, comme la catégorie précédente, des décisions du chercheur.

- 10 Le Panel Electoral Français 2002 comporte neuf³ questions ouvertes de ces deux derniers types. Celles-ci peuvent faire l'objet de traitements spécifiques selon les choix méthodologiques opérés avant et après la réalisation du terrain.

La qualité des données recueillies

- 11 Avant d'aborder la qualité des données elles-mêmes, il faut souligner l'importance capitale de la qualité de leur collecte. Le professionnalisme de l'enquêteur est pour beaucoup dans la richesse du verbatim obtenu, que ce soit pour les relances ou pour la saisie des réponses. Les interrelations enquêteur-enquêté sont à ce sujet déterminantes.
- 12 Après expertise des fichiers de données, différents problèmes ont été recensés. Certains tiennent au format d'extraction des questions, à la qualité de la saisie des réponses, aux types de questions et donc aux types de traitements envisageables (post-codage réalisé par l'enquêteur, post-codage manuel fin par l'équipe PEF2002, traitement exploratoire par analyse des données textuelles). Un travail important de préparation des fichiers et des données s'est révélé nécessaire avant toute exploitation des données.

Constitution des fichiers, formats d'extraction

- 13 Les fichiers bruts de l'enquête électorale 2002 n'ont pas tous été construits de la même façon selon la vague d'enquête. La configuration idéale consiste à obtenir un fichier par question, et une ligne par répondant, avec l'identifiant de l'enquêté et le verbatim recueilli. Le format d'extraction des verbatims doit être suffisamment important pour s'assurer que l'intégralité des réponses est recueillie.
- 14 Ainsi, pour les questions ouvertes *post-codées* par l'enquêteur Sofres (professions, diplôme, catégorie sociale subjective), la saisie littérale des réponses est fournie sur 240 caractères, ce qui s'avère suffisant pour ce type de questions appelant des réponses courtes. Il en est de même pour les réponses « autres » aux questions d'intention de vote, de proximité partisane, de sentiment français/européen, parti pour lequel on ne voterait en aucun cas.
- 15 En revanche, pour les questions ouvertes *non post-codées*, l'intégralité des verbatims des vagues 2 et 3 nécessite un champ de 900 à 1200 caractères. En vague 1, les verbatims tiennent sur 240 caractères, sans doute en raison de l'administration en face à face, où l'enquêteur saisit la réponse de l'enquêté sous son regard, voire sa dictée.

Préparation des données pour exploitation

- 16 Avant tout traitement, les données recueillies doivent être nettoyées et homogénéisées, la qualité de la collecte et de la saisie pouvant être variable. Le nettoyage diffère selon les traitements envisagés, post-codage ou analyse textuelle.
- 17 Les questions susceptibles d'être post-codées manuellement peuvent être dispensées de cette phase, au moins dans un premier temps : trois questions de la vague 1 du PEF2002 sont dans ce cas : le problème le plus important en France, les mesures figurant dans le programme de Chirac et celui de Jospin. En revanche, les questions ouvertes visées par l'analyse des données textuelles doivent être obligatoirement nettoyées : corrections orthographiques, saisie phonétique, homogénéisation des abréviations, des formats de caractères (maj/min, accentuation), erreurs de saisie...

Les différents types de traitement

- 18 Chaque type de question ouverte requiert un type de traitement particulier. Lors de l'élaboration d'un questionnaire, il importe donc d'avoir présent à l'esprit non seulement la problématique envisagée mais aussi le mode d'exploitation des réponses envisagé. Par conséquent le choix de la formulation d'une question est important.
- 19 Les questions ouvertes du Panel électoral français 2002 offrent la possibilité de pratiquer différents traitements. Les questions recueillant des informations factuelles (profession, diplôme, classe sociale subjective), ainsi que les reliquats de questions fermées (poste « autre ») peuvent être fait l'objet d'un post-codage systématique. Il s'effectue sur la base de consignes précises à l'enquêteur qui code simultanément au recueil ou en fin de questionnaire. Il peut également être réalisé par le chercheur sur la base de ses propres hypothèses.
- 20 Un post-codage manuel, plus fin, permet de traiter les questions appelant des réponses courtes, des citations simples (« quel est le problème le plus important en France ? »). A partir de l'analyse détaillée d'un échantillon de réponses, une grille de codage est élaborée qui décline les thèmes rencontrés en sous-thèmes. A rechercher l'homogénéité à l'intérieur des thèmes, il importe toutefois de veiller à catégoriser sans trop réduire l'information. Cette grille de codage est ensuite appliquée à l'ensemble des réponses.
- 21 Les questions ouvertes, à multiples relances, qui visent à obtenir des réponses les plus complètes possibles, argumentées, relèvent de l'analyse de discours (« pour quelles raisons avez-vous voté pour Jacques Chirac au 2^{ème} tour de l'élection présidentielle ? » ou « Concernant l'environnement, quel est à votre avis le problème le plus important actuellement ? »).

L'analyse textuelle des questions ouvertes : stratégies d'analyse

- 22 Pour notre propos, il était important de disposer de techniques d'analyse des questions ouvertes qui, permettent d'aller au-delà de la simple analyse de contenu manuelle. Le recours à l'analyse textuelle pour ce type de corpus s'avère particulièrement productif (Lebart et Salem,1993). Celle-ci permet notamment, à travers l'analyse de récurrences formelles dans les réponses, d'identifier différentes isotopies. La méthode Alceste⁴ utilisée ici privilégie une approche statistique qui identifie dans un corpus de texte donné (les réponses à la question ouverte) des sous-ensembles homogènes de verbatims sur la base de leur profil lexical.⁵ Les classes de réponses obtenues peuvent être caractérisées par les variables sociologiques disponibles dans l'enquête qu'elles soient socio-démographiques ou d'attitudes. L'analyse textuelle présente ainsi l'avantage de dissocier assez clairement deux moments de l'analyse : d'abord la formalisation des données et leur « traitement » puis l'étude sociologique de ces résultats qu'il faut au préalable construire et argumenter.
⁶ Au-delà des commentaires statistiques, on cherche en effet à livrer une analyse linguistique et sociologique des textes. L'interprétation des classes identifiées par la méthodologie Alceste s'appuie, sur une caractérisation « externe » de la classe (opposition et proximité par rapport aux autres classes telles qu'elles sont par exemple visualisées par le graphique factoriel) et, sur des critères « internes » (mots et réponses caractéristiques de la classe). Par ailleurs, les variables « illustratives » (variables linguistiques : catégorisation des mots-outils etc. et variables socio-démographiques

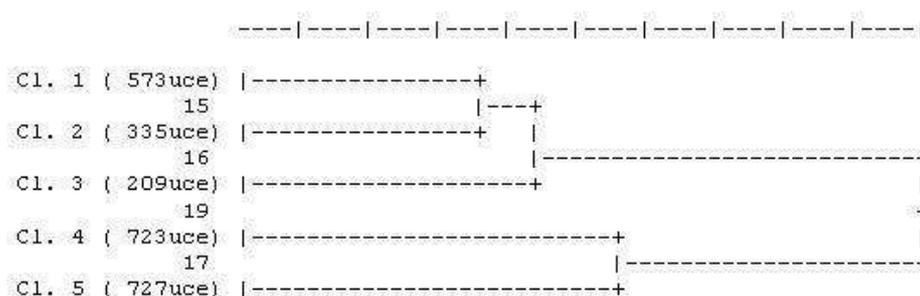
décrivant les répondants) nous permettent d'identifier certaines caractéristiques des classes de réponses. L'analyse propose ainsi un « parcours interprétatif » qui s'appuie sur les trois types d'indicateurs mobilisés.

- 23 Un premier exemple portant sur les raisons de la participation au scrutin du premier tour de l'élection Présidentielle permet de rendre compte de ce travail interprétatif.⁷

Identifier des classes de réponses : l'exemple d'une question ouverte sur la participation

- 24 Le traitement par la méthode Alceste permet de classer 66% des individus en 5 classes.
- 25 La classification descendante hiérarchique effectuée sur les réponses⁸ distingue d'abord l'univers des raisons formelles du vote de l'univers des raisons liées au contenu du vote (premier embranchement sur l'arbre de la classification, cf. graphique 1).

Graphique 1 : Dendrogramme des classes



- 26 Du point de vue *des raisons formelles du vote*, la classification distingue, dans un deuxième temps, ceux qui disent que le vote est un **droit** (*acquis de haute lutte dans le passé, des gens sont morts pour ça, on évoque aussi la lutte des femmes pour obtenir le droit de vote etc.*) de ceux qui parlent du vote comme d'un **devoir** (*un acte civique, un rôle voire une habitude que tout citoyen français se doit de remplir*). Il s'agit respectivement des classes 4 et 5 (environ 30% des individus classés à chaque fois - sachant qu'ils sont environ 3000 sur 4000 à être classés).
- 27 Du côté des raisons liées au contenu du vote, la classification distingue, d'un côté, ceux qui évoquent un **changement** (*voter pour changer les choses, changer de président, il s'agit de faire avancer, bouger les choses, de les améliorer etc.* environ 8% des individus classés) et, d'un autre côté, la classe regroupant ceux qui veulent exprimer leur **opinion** (*avis, qui veulent se faire entendre, leur voix, donner leur point de vue, voire contester et exprimer leur mécontentement*). Cette branche se subdivise à son tour en deux classes très proches : l'une évoque le **premier tour** et ses aspects stratégiques (et l'on retrouve *mécontentement, choix candidat etc.* 22% des classés) et l'autre est centrée sur les aspects expressifs du vote (*exprimer son avis, son opinion etc.*) (13% des classés).
- 28 Alceste permet de caractériser les classes par une série de variables illustratives. En première analyse, il semble que l'on puisse résumer les choses de la manière suivante. Une analyse de correspondance réalisée à partir d'un tableau croisant les classes Alceste et les mots oppose sur le premier axe "raisons formelles du vote" aux "raisons liées au contenu" (et ici il y a davantage de « ni gauche ni droite » vs « gauche » ou « droite ») et sur l'axe vertical des individus plutôt à gauche (« **Droit** » et « **Opinion** ») vs des individus

plus à droite (« **Devoir** » et « **Changement** »). Dans la classe « **Changement** », on trouve plus de sympathisants du FN qu'ailleurs (mais aussi des sympathisants du PC) et des "ni gauche, ni droite". Les sympathisants du RPR sont sur-représentés dans la classe « **Devoir** », cependant que l'on trouve beaucoup de personnes proches de l'extrême gauche (mais aussi du PS) dans la classe « **Opinion** ». Les sympathisants du PS sont sur-représentés dans la classe « **Droit** ».

Figure 1 : Analyse de correspondance réalisée à partir d'un tableau croisant les classes Alceste, des mots opposés et des orientations gauche/droite

<p>le-vote==Devoir</p> <p>longueur moyenne des réponses = 4,56</p> <p>mots</p> <p>+ de substantifs, verbes modaux = devoir, pas-personnalisation (voix narrative)</p> <p>Objet-de-valeur</p> <p>- un devoir, une responsabilité, un rôle, une habitude</p> <p>Type-d'épreuves, de-compétences</p> <p>- remplir, accomplir, faire un acte (civique)</p> <p>Type-d'actants (Agent)</p> <p>- citoyens, français, électeur</p> <p>Segments répétés les plus fréquents : c'est un devoir, il faut</p>	<p>le-vote==Changement</p> <p>longueur moyenne des réponses = 10,43</p> <p>mots</p> <p>+ de verbes, verbes modaux = vouloir</p> <p>Objet-de-valeur</p> <p>- un espoir, (promesse)</p> <p>Type-d'épreuves, de-compétences</p> <p>- changer, avancer, bouger, améliorer</p> <p>- représenter, être, choisir</p> <p>Type-d'actants (Bénéficiaire)</p> <p>- La France</p> <p>Segments répétés les plus fréquents : pour espérer de changer, faire changer les choses, que ça change</p>
<p>le-vote==Droit</p> <p>longueur moyenne des réponses = 9,90</p> <p>mots</p> <p>+ de substantifs, verbes modaux = falloir, pas-personnalisation (voix narrative)</p> <p>Objet-de-valeur</p> <p>- un droit, un principe, un privilège, une liberté, une occasion</p> <p>Type-d'épreuves, de-compétences</p> <p>- (primaire) utiliser, servir, profiter, respecter</p> <p>- (secondaire) acquiescer, obtenir, battre</p> <p>Type-d'actants (Agent)</p> <p>- gens, femmes, morte</p> <p>Segments répétés les plus fréquents : droit de vote, il faut, c'est un droit et un devoir</p>	<p>le-vote==Premier tour</p> <p>longueur moyenne des réponses = 13,75</p> <p>mots</p> <p>(beaucoup de mots-outils, marqueurs d'argumentation, verbes modaux = pouvoir, avoir, vouloir) personnalisation du discours (narrateur=je) + Opinion</p> <p>longueur moyenne des réponses = 9,69</p> <p>mots</p> <p>+ de verbes, personnalisation du discours (narrateur=vois)</p> <p>Objet-de-valeur</p> <p>- un souhait (voix, opinion, avis, point de vue)</p> <p>Type-d'épreuves, de-compétences</p> <p>- exprimer, donner (son opinion), contester, se faire entendre</p> <p>Type-d'actants (Bénéficiaire)</p> <p>- le pays</p> <p>Segments répétés les plus fréquents : pour donner mon avis, pour donner mon opinion (Opinion)</p> <p>premier tour, c'est important, mes idées, mon mécontentement (Premier tour)</p>

- 29 En termes socio-démographiques, les revenus modestes et les niveaux d'études faibles sont sur-représentés dans la classe « **Changement** ». Les plus âgés et les catholiques pratiquants sont nombreux dans la classe **Devoir** cependant que les niveaux d'études supérieurs et les revenus les plus élevés sont sur-représentés dans les classes « **Droit** » et « **Opinion/Premier tour** » avec davantage de jeunes dans le groupe « **Opinion/Premier tour** ».

La construction d'une typologie de discours (interprétation linguistique)

- 30 Ces classes de réponses peuvent être analysées du point de vue de leur contenu et des thèmes qu'elles abordent mais il est aussi possible de montrer qu'elles manifestent l'existence de discours différents sur la participation électorale. A partir des spécificités lexicales (mots-pleins et mots-outils sur-représentés au sens du chi-2) et des verbatims caractéristiques de ces classes, il est possible de proposer une typologie de « micro-récits » sur la participation, présentée de manière synthétique dans le tableau 1.

- 31 A droite, le vote est plutôt décrit comme un acte (*acte, avancer, bouger, améliorer* etc.) alors qu'à gauche il s'agit plutôt d'une prise de *parole* (*mot, voix, opinion, avis* etc.). Bref, le vote est un acte de langage avec ses deux dimensions.
- 32 Dans les réponses, le vote est décrit comme le récit d'une action impliquant différents participants ou rôles (« actants »).
- 33 Les énoncés de la classe **Devoir** comportent, comme la classe Droit, une sur-représentation de substantifs, ne sont pas ou peu personnalisés. C'est souvent une « voix narrative » impersonnelle qui s'exprime : « le vote est un devoir », « le vote est un droit ».
- 34 *L'objet de l'action* (ou « objet de valeur ») dans la classe Devoir est le vote défini comme un *devoir*. On trouve toutes les variations sémantiques autour de cette idée : du devoir moral impérieux décrit comme une *obligation* à une norme sociale plus ou moins contraignante : une *responsabilité*, un *rôle* de *citoyen* voire au frontière de la classe, « faire comme les autres », jusqu'à *l'habitude*.
- 35 *Le type de compétence requis*, ou d'épreuve à passer, est essentiellement l'accomplissement d'une action : *accomplir, remplir* (un devoir), *faire un acte civique*. La simplicité de l'action relatée explique souvent le côté très lapidaire des réponses.
- 36 *Le type d'actant* mis en scène est principalement celui de l'agent, le responsable de l'action : les personnes interrogées se définissent comme *citoyen français, électeur*.
- 37 Les énoncés de la classe **Droit**, plus longs en moyenne que ceux de la classe précédente, proposent un récit de l'action un peu différent :
- 38 *L'objet de l'action* est le vote défini cette fois comme un *droit*. Là aussi, on recense toute une série de variations sémantiques, de quasi-synonymes qui oscillent autour d'un noyau central : un droit peut être ainsi décrit comme un *principe* (ce qui le rapproche du Devoir) un *privilege*, une *liberté*, voire comme une *occasion*.
- 39 *Le type de compétence requis* implique un rôle plus ou moins actif selon que l'on décrit l'action « primaire » - il s'agit d'*utiliser*, de *respecter*, de *se servir* ou de *profiter* d'un droit - ou « secondaire » - dans le passé, ce droit a été *acquis, obtenu* de haute *lutte* par des *gens* qui se sont *battus*. L'action est décrite « en puissance » et non pas « en acte » comme pour le devoir : elle est soit au passé, soit minimale, quasiment sur le mode passif, puisqu'il s'agit de *respecter* ce que d'autres ont acquis (« *des gens se sont battus pour que les femmes aient le droit de vote, donc j'en use* »).
- 40 *Le type d'actant* mis en scène est, pour les deux classes développant des raisons formelles, principalement celui de l'agent : ce sont cette fois les *gens*, figure plus ordinaire et moins institutionnelle que celle du citoyen, ou des *femmes* dont les luttes passées sont rappelées ainsi que la figure plus historique de ceux qui sont *morts*.
- 41 Les énoncés de la classe **Opinion/premier tour**⁹ se caractérisent par des énoncés assez longs, voir très longs (presque 14 mots pour la classe premier tour qui s'attarde souvent sur des considérations plus ou moins stratégiques).¹⁰ A la différence des classes qui déclinent les raisons formelles de voter, celles liées au contenu sont marquées par un suremploi de verbes et une personnalisation plus importante du discours.
- 42 *L'objet de l'action* est moins longuement décrit que dans les classes précédentes. Le vote est défini essentiellement comme un *souhait*, une *opinion*, un *avis* voire un *point de vue*. Ce qui est mis ici en valeur c'est l'expression d'une subjectivité ou d'une individualité.
- 43 *Le type de compétence requis* implique un sujet de l'action - qui se confond avec le narrateur - qui sache *s'exprimer, donner son point de vue* mais aussi *se faire entendre* et pour

certaines *contester*. On oscille ici entre la description d'une action centrée sur un individu qui se définit par sa capacité à exprimer *son point de vue* et un individu qui *conteste* un état de chose. La première action semble davantage correspondre à la classe Opinion, la seconde à la classe Premier tour.

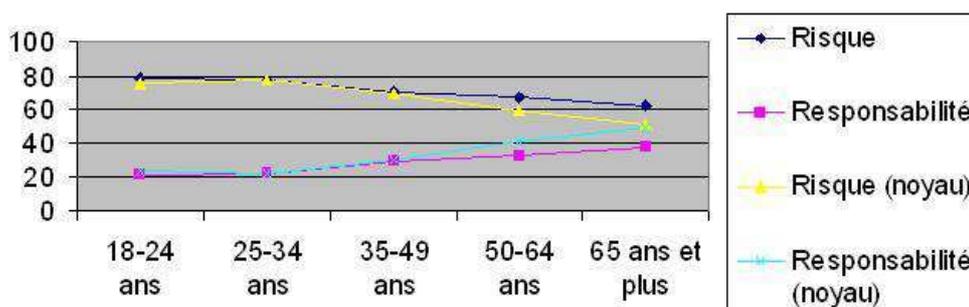
- 44 *Le type d'actant mis en scène, outre le narrateur sujet de l'action, est le « bénéficiaire » de l'action, c'est-à-dire ici le pays, équivalent structural des gens à qui il s'oppose sur l'axe individuel vs collectif mais avec qui pourtant il partage une dimension commune : une représentation « ordinaire » et non solennelle du politique (vs la France, les citoyens français).*
- 45 Les énoncés de la classe **Changement** comme ceux de la classe précédente sont marqués par un suremploi des verbes et un usage privilégié d'un des verbes modaux, *vouloir*. Le discours est toutefois moins personnalisé.
- 46 La description de l'action est moins centrée sur *l'objet de l'action* que les classes Devoir et Droit : le vote y est succinctement glosé comme un *espoir*.
- 47 *Le type de compétence requis* définit le vote essentiellement comme une action. Mais, à la différence de la classe Droit qui décrivait formellement cet acte (ou disons son aspect procédural), ce type de discours décrit l'action en fonction de sa fin et de son objet : il s'agit de *changer, de faire bouger, d'améliorer (les choses)*, en un mot *d'avancer*. Pour cela, il faut *choisir, élire* le (bon) *président*. Comme la classe Opinion/premier tour, *le cours des choses* est dénoncé mais ici on ne contente pas de contester en faisant entendre sa voix, on *change* les choses ou du moins, on *essaye* et on *espère*. Nous sommes ici aux antipodes de l'action racontée dans la classe Droit : les discours s'opposent terme à terme sous le rapport des types de raisons (contenu vs formelles) comme du point de vue des types d'actes (en acte vs en puissance). L'horizon temporel constitue une troisième dimension transversale qui achève de saturer l'opposition entre ces discours : ici, le narrateur se projette dans le passé, là, dans le futur.
- 48 La France est désignée comme le principal bénéficiaire de l'action.

Retour aux individus et aux variables explicatives : les déterminants sociaux et politiques des répondants

- 49 Bien sûr, ces caractérisations sont idéal-typiques et très « qualitative », mêmes si elles sont construites à partir d'une procédure quantitative qui identifie des spécificités lexicales et des réponses caractéristiques des classes. La typologie de discours obtenue peut-être analysée sous la forme d'une variable fermées à *n* modalités.
- 50 Ainsi par exemple, l'analyse de la question ouverte identifiant les principaux problèmes environnementaux, posée dans la vague post-présidentielle fait-elle apparaître deux discours portant, l'un, sur les risques et l'autre, sur les responsabilités. Ces deux discours, bien que non exclusifs logiquement, se différencient assez bien selon des critères socio-démographiques ou politiques.
- 51 Il se distribuent notamment selon une logique d'âge qui apparaît dans le graphique 2. Le discours sur la responsabilité est plus répandu à mesure que l'âge croît – l'accroissement étant assez régulier à partir de 34 ans. Ce phénomène s'observe mieux encore si l'on ne prend en compte que les individus au centre des classes¹¹ : les 18-24 ans sont 76% à développer une réponse sur le thème des risques cependant que les plus de 65 ans ne sont

que 51% dans ce cas. Ils abordent beaucoup plus que leurs cadets le thème de la responsabilité.

Graphique 2 : Risques et responsabilité selon l'âge



- 52 Le retour aux individus de l'enquête à partir des variables créées par Alceste présente toutefois des difficultés liées aux spécificités de la méthode.¹² Il est parfois préférable d'un point de vue statistique d'opter pour une analyse lexicale et thématique comme le montre l'exemple suivant sur la victoire de la droite et la défaite de la gauche.

Analyse lexicale et analyse thématique

- 53 Lors de la troisième vague, l'une des deux questions ouvertes suivantes était posée aléatoirement à chaque enquête¹³ :

Selon vous, pourquoi la gauche a-t-elle perdu ?

Et, selon vous, pourquoi la droite a-t-elle gagné ?

- 54 Pour analyser ces deux questions, un double objectif a été défini : le premier est d'isoler les différents discours produits par les enquêtés, le second est de caractériser le répondant par l'ensemble des thèmes qu'il a évoqué. La première étape consiste à analyser les corpus de chacune des questions, puis à dégager les champs lexicaux et leurs associations. Dans une deuxième étape, une liste de thèmes est construite à partir de ces champs lexicaux. La comparaison des distributions des thèmes abordés par les enquêtés permet de mettre en évidence des différences attendues selon la formulation de la question. Mais au-delà de ce constat, les mêmes thématiques sont présentes dans les deux corpus, la différence se fait sur leur fréquence d'apparition. Enfin, les croisements des thèmes abordés par les locuteurs avec leurs caractéristiques sociodémographiques conduisent à la conclusion que certaines relations sont indépendantes de la question posée.

Analyse textuelle, définition des champs lexicaux

- 55 L'analyse lexicale conduit à une typologie d'un certain nombre de discours tenus par les répondants sollicités à s'exprimer sur les résultats des élections de 2002. Les résultats de l'analyse lexicométrique mettent en exergue des thèmes plus spécifiques de la victoire de la droite : le refus de la cohabitation, la peur de l'extrême droite et la volonté de changement, et d'autres spécifiques de la défaite de la gauche : des promesses non tenues, un vote contestataire et l'abstention.

- 56 Le premier axe de chacune des deux analyses factorielles des correspondances du tableau croisant le vocabulaire et les classes d'énoncés (Figure 3), peut être considéré comme un axe temporel. Dans le cas des réponses au questionnaire sur les causes de la défaite de la gauche, les énoncés faisant référence à l'ancienne campagne s'opposent à ceux émettant un avis sur les faits du printemps 2002. Pour les réponses relatives à la victoire de la droite, des commentaires sur la situation politique à la veille des élections sont opposés à d'autres concernant les élections. L'intervalle de temps semble cependant plus étendu pour la défaite de la gauche que pour la victoire de la droite. Dans les deux cas, le deuxième axe de ces plans factoriels oppose des propos relatifs à une évaluation des actions de la gauche aux conséquences que ces actions ont eues sur les résultats du scrutin.

Figure 3 : Deux analyses factorielles des correspondances du tableau croisant le vocabulaire et les classes d'énoncés

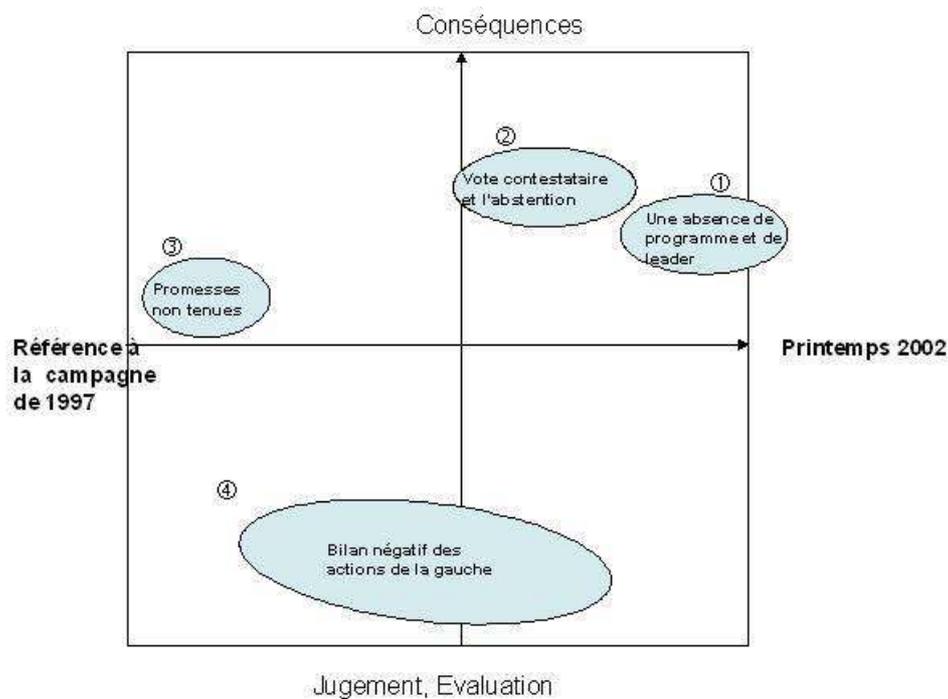
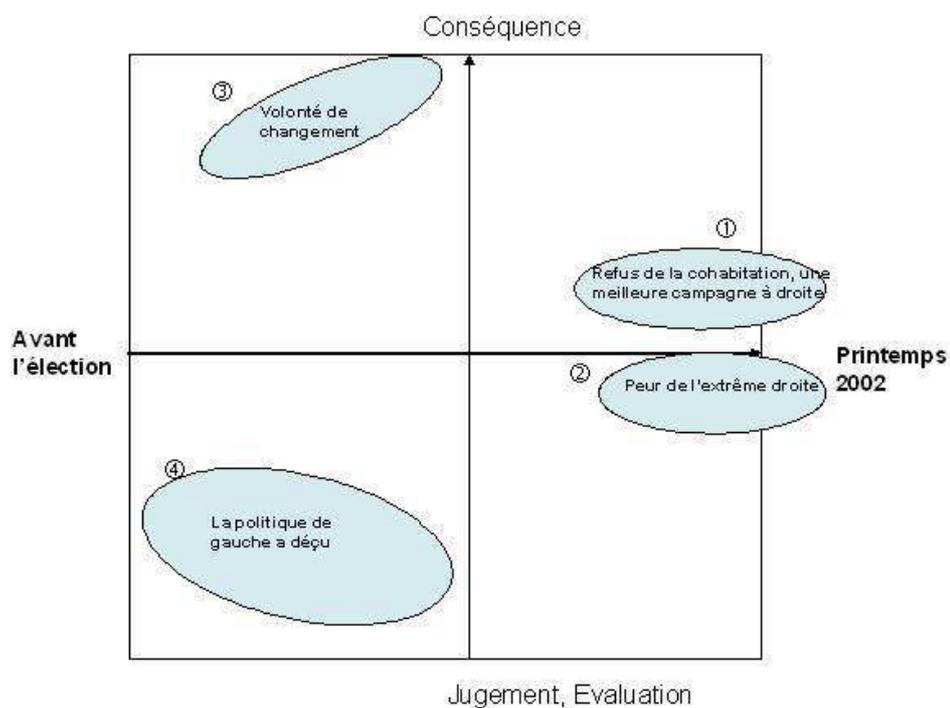


Figure 3 (suite)



- 57 A partir de ces classes de discours, il convient maintenant, de mesurer l'influence de la question posée sur les thèmes abordés et de connaître le profil sociodémographique des individus. Une analyse thématique a permis d'affiner les classes de discours et revenir au plus près du répondant.

L'analyse thématique pour revenir aux répondants et à leur profil

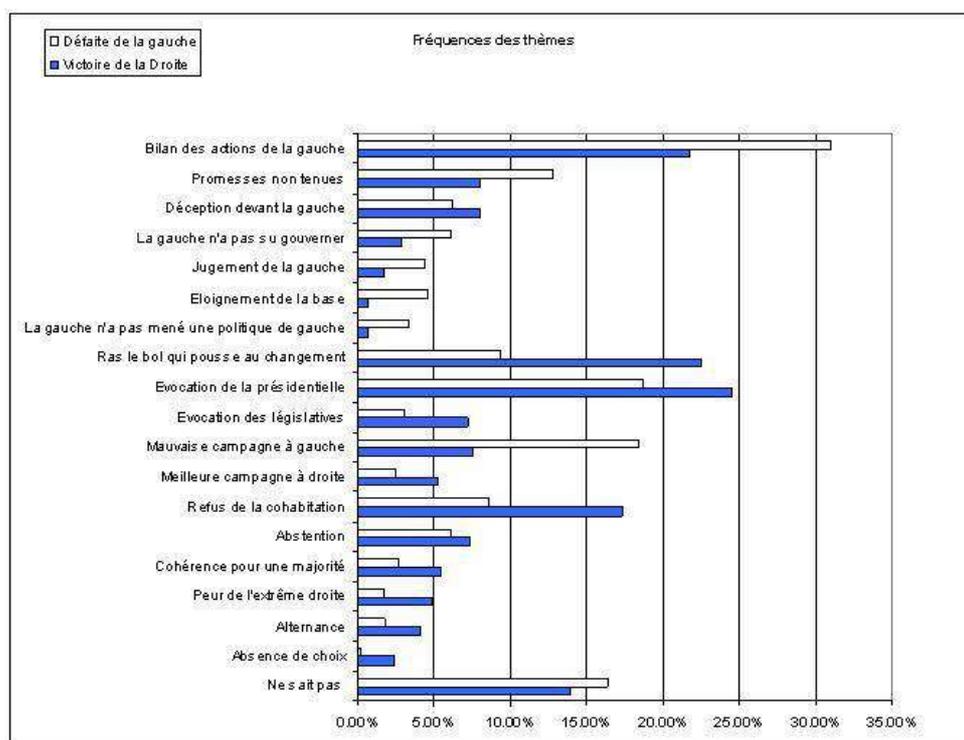
La démarche

- 58 Chacune des classes de discours produite par Alceste est caractérisée par un ensemble de mots spécifiques. A partir de ces mots, des champs lexicaux sont isolés. Ces différents champs lexicaux constituent un premier niveau de thématique. A partir des champs lexicaux relatifs à un thème, les verbatims de réponses des individus sont explorés de façon à constituer un dictionnaire de mots et d'expressions associés à ce thème. La dernière étape consiste à créer, pour tout thème, une variable binaire prenant la valeur 1 lorsque l'enquêté évoque ce thème et 0 sinon. Cette phase de l'analyse a été facilitée par l'utilisation du logiciel Sphinx. Par cette démarche qui a nécessité un va et vient entre les classes d'énoncés, le lexique (ou les segments répétés) et leur contexte d'utilisation, nous avons affiné les thématiques. Par exemple, la classe de discours "bilan négatif des actions de la gauche" a été divisée en cinq sous-thèmes. Par ailleurs, cet approfondissement a permis d'observer l'importance de la thématique "je ne sais pas" non extraite lors de l'analyse lexicale précédente.

Les thèmes et leur fréquence

- 59 Dix-neuf thèmes ont été retenus. Dans 9% des cas, la réponse d'un individu n'est affectée à aucun des thèmes retenus, car aucun des mots ou expressions ne correspond aux dictionnaires thématiques créés. Ce sont pour la plupart des réponses peu claires ou trop ambiguës.
- 60 Les thèmes abordés dans les réponses ont été regroupés en six catégories.
- 61 *Le jugement sur la politique et les personnalités de gauche* constitue une première catégorie : le bilan du gouvernement de gauche, les promesses non tenues, le fait que la gauche n'a pas su gouverner, que la gauche n'a pas mené une politique de gauche, son attitude vis-à-vis de la base (le peuple, les électeurs, les Français). Le bilan négatif des actions de gauche, qui couvre entre autres les problèmes relatifs aux 35 heures, est un thème largement évoqué (26,2% des personnes interrogées). Il faut noter aussi l'importance de la thématique des promesses non tenues (9,2%).
- 62 L'expression du *ras-le-bol de la gauche qui pousse au changement* est plus particulièrement rencontrée dans le corpus des raisons de la victoire de la droite (15,9% des personnes interrogées). Ce n'est pas la droite qui a gagné, c'est la gauche qui a perdu.
- 63 La *référence à l'élection présidentielle* est très importante (21,6% des personnes interrogées). Ce thème est plus présent que celui des élections législatives (7,3%) Cette différence permet de penser que l'élection présidentielle est la référence lorsque l'on évoque cet ensemble d'élections. Les législatives ont été mises en sourdine.
- 64 *L'évocation de la campagne de la gauche, jugée mauvaise*, est plus fréquente dans le texte des réponses associées aux raisons de la défaite de la gauche (12,8%). Le thème relatif à la *campagne de droite, jugée meilleure*, est plus fréquemment cité pour expliquer les raisons de la victoire de la droite. Il constitue cependant un thème très peu fréquent (3,8%).

Figure 2 : Les thèmes et leur fréquence en fonction de la formulation de la question



- 65 Les thèmes visant à expliquer les enchaînements logiques sous-jacents au déroulement des élections et faisant référence au contexte électoral d'après le 21 avril 2002, (*l'abstention des électeurs - 7,5%, la peur de l'extrême droite - 5,2%, l'absence de choix au second tour de l'élection présidentielle - 1,3%, le refus de la cohabitation 12,9%, l'alternance - 3,3%, donner une majorité présidentielle - 3,4%*) sont plus largement évoqués dans le corpus des enquêtés interrogés sur les raisons de la victoire de la droite.
- 66 La proportion des personnes exprimant leur embarras à expliquer les résultats des élections est importante (14,9%). Cela est vrai quelle que soit la question posée. Parfois, la réponse se limite à ces quatre mots "je ne sais pas" et dans d'autres cas, elle est ponctuée par cette expression.
- 67 Les relations entre les thèmes et la formulation de la question sont toutes significatives à l'exception des thèmes relatifs à "la déception devant la gauche", "l'abstention" et aux "je ne sait pas".
- 68 Pour ce qui concerne les causes de la victoire de la droite, le contenu du programme de la droite est peu commenté, les principaux arguments avancés par les locuteurs sont relatifs à l'articulation de l'ensemble des événements de la séquence électorale (le rôle de l'abstention, la peur de l'extrême droite, l'absence de choix). Ces éléments ont marqué les campagnes électorales et ont déclenché certains processus électoraux.

Le retour aux répondants

- 69 L'analyse a permis d'améliorer l'homogénéité des thèmes et d'associer une liste de thèmes à chaque répondant. La matrice individus-thèmes est l'outil avec lequel l'analyse sociodémographique est menée. Pour illustrer la démarche qui vise à analyser la liaison entre un thème donné et le profil sociodémographique de l'énonciateur, nous

commenterons le croisement des thèmes avec les variables de sexe et d'âge. Les thèmes les plus fréquents ont été conservés : le bilan des actions de la gauche, l'évocation de la présidentielle et les "ne sait pas". Ces liaisons ont été étudiées en contrôlant la formulation de la question. Le premier constat que nous pouvons faire est que des liaisons existent mais sont dans la majorité des cas de faible intensité (V de Cramer < 0.2).

- 70 Le discours des femmes s'oppose à celui des hommes. L'absence d'opinion politique est plus fréquente chez les femmes. En revanche, les hommes sont plus nombreux à énumérer et à critiquer les actions de la gauche. Cette énumération est précise et témoigne d'un degré de connaissance supérieur de la vie politique. De plus, les hommes sont les plus nombreux à évoquer l'élection présidentielle. Cette situation confirme les résultats d'autres études selon lesquelles le niveau de politisation des hommes est supérieur à celui des femmes.
- 71 Les verbatims des moins de 25 ans révèlent que ce sont eux qui ont le plus de difficulté à expliquer ces événements, à l'instar des plus de 65 ans. A l'inverse ce manque d'argumentation se retrouve rarement chez les personnes âgées de 50 à 65 ans. Ce résultat confirme que la politisation et l'argumentation politique atteint un optimum dans cette classe d'âge.
- 72 L'ensemble des relations qui viennent d'être décrites sont les mêmes, que l'interviewé soit sollicité sur les motifs de la victoire de la droite ou sur ceux de la défaite de la gauche.

L'intérêt de l'analyse lexicale et thématique

- 73 Cet enchaînement méthodologique offre la possibilité de mesurer l'importance des thèmes présents dans l'opinion publique pour expliquer l'issue de la séquence électorale de 2002. Ainsi l'évocation de l'élection présidentielle est prépondérante sur l'évocation des élections législatives. Le bilan des élections de 2002 semble majoritairement associé aux résultats de l'élection présidentielle. Les élections de référence ne sont pas les mêmes pour expliquer la victoire ou la défaite, tout découle du premier tour de la présidentielle et tout est en place au soir du 21 avril 2002.
- 74 Le deuxième point fort de cette démarche est le retour aux locuteurs caractérisés par l'ensemble des thèmes qu'ils ont énoncés. Ainsi nous pouvons éclairer les motifs présents dans les réponses par le profil sociodémographique, l'opinion et le comportement politique des répondants.
- 75 Le matériau initial comportant des informations diffuses s'est avéré riche et structuré. Cette structuration du corpus permet de confirmer l'expression d'un "vote de tous les refus"¹⁴ : un bilan négatif de la gauche qui conduit à son désaveu, un rejet de la vie politique par une abstention en forte hausse, un refus de l'extrême droite à l'issue du premier tour de la présidentielle et un rejet de la cohabitation, expérience jugée non convaincante.

Conclusions

- 76 *L'apport de questions ouvertes dans le dispositif d'enquêtes électorales et le recours à des approches d'analyses des données textuelles permettent incontestablement de proposer de nouveaux angles d'analyses et d'enrichir les approches de la sociologie électorale. Toutefois, elles suscitent aussi de*

nombreuses questions qui constituent un véritable programme de recherche dont on se contentera ici d'indiquer les grandes lignes :

- 77 du point de vue des stratégies de questionnement, il s'agit de construire une véritable typologie des questions ouvertes – esquissée ici – différenciant des niveaux de réponse (citation, mémorisation, discours) mais aussi des types de discours (question ouverte « projective », « argumentative » etc.). Cette étude est inséparable d'une réflexion approfondie sur les séquences de questions – reprenant les éléments concernant les effets d'ordre, de contexte – en s'interrogeant par exemple sur des séquences incluant questions fermées et ouvertes.¹⁵ Au-delà, il s'agirait de reprendre à nouveaux frais le débat amorcé dans les années quarante entre Rensis Likert et Elmo Wilson sur les questions ouvertes et les questions fermées¹⁶ afin de promouvoir des approches hybrides et de nouveaux dispositifs.¹⁷
- 78 du point de vue des « terrains », ce travail après d'autres nous invite à apporter une extrême vigilance aux interactions enquêteurs/enquêté et à la qualité du recueil des « données » si sensibles quand il s'agit de discours,¹⁸
- 79 concernant les stratégies d'analyse, il s'agit d'approfondir selon les cas les approches en termes d'analyse de contenu ou de discours, voire à parfois cumuler ces stratégies. Celles-ci posent et articulent entre elles de nombreuses questions linguistiques (et de manière plus générale sur la place et l'importance à accorder au langage dans l'analyse de ces questions) mais aussi statistiques (de simples croisements avec des variables agrégées aux procédures de classification exploratoires en passant par le traitement de plusieurs questions ouvertes etc.). Elles impliquent en outre une réflexion sur les logiciels d'analyse des données textuelles et leurs attendus théoriques.¹⁹
- 80 Le « retour aux répondants » implique une réflexion méthodologique sur les protocoles (dont certains ont été ici présentés) mais aussi une recherche épistémologique plus large sur la combinaison d'approches compréhensive et explicative,
- 81 Enfin il convient de s'interroger sur certains des « objets » construits par la théorie tant il est vrai que le renouvellement théorique, ainsi que le rappellent Jean-Claude Passeron et Michel de Fornel, est aussi le résultat « de la mise au point de nouvelles modalités de l'enquête empirique » qui transforment « l'intelligibilité des « faits » »²⁰. Ainsi l'étude des questions ouvertes issues de ce type de dispositif peut-elle par exemple contribuer à « décentrer » la réflexion théorique sur l'opinion publique ou sur la notion d'attitude, en proposant des représentations discursives de ces objets y compris au niveau agrégé.

NOTES

*. Cet article reprend le texte d'une communication au congrès de l'AFS, RTF 20- Figures du pluralisme méthodologiques, 24 février 2004.

1. Ce dispositif d'enquête a notamment permis la réalisation d'une série d'études de sociologie électorale à paraître dans Bruno Cautrès, Nonna Mayer (dir.), *le nouveau désordre électoral*, Presses de Science Po, PUG, 2004.

2. Cf. Brugidou, Mathieu. « Note récapitulative sur l'analyse des questions ouvertes du PEF2002 », 5 mars 2003.
3. En vague 1, 5 questions ouvertes sont introduites parmi les 66 questions posées. Les vagues 2 et 3 en comportent deux chacune, soit neuf au total.
4. Reinert, M. (1995) Quelques aspects de choix des unités d'analyse et de leur contrôle dans la méthode Alceste, in *JADT 1995* ; vol I p.27-34.
5. Le programme crée un tableau croisant les mots-pleins en colonne et les réponses à la question ouverte en ligne. Une classification descendante hiérarchique part d'une seule classe réunissant l'ensemble des réponses et la partitionne en deux classes regroupant les verbatims similaires à partir du critère discriminant du χ^2 . Puis ces deux classes sont à leur tour divisées en deux classes stables et ainsi de suite jusqu'à l'obtention de groupes homogènes de verbatims. Le résultat est une classification hiérarchisée qui peut être schématisée sur un graphique en arbre. (cf. Graphique 1).
6. Brugidou, M. , « La combinaison des inférences statistiques, linguistiques et sociologiques dans l'analyse d'une question ouverte », *Journal de la Société Française de Statistique*, , 2001, 4.
7. Nous avons étudié l'univers de ceux qui disent vouloir voter au premier tour (Q9=oui) (n=3667).
8. Seuls les mots pleins sont traités. Les mots-outils seront considérés ultérieurement comme des variables illustratives.
9. Les deux classes ont été agrégées ici, d'une part, pour des raisons statistiques (cf. sur le dendrogramme, la proximité des classes 1 et 2) et, d'autre part, pour des raisons « herméneutiques », la signification étant construite par le réseau des relations nouées entre quatre pôles.
10. Mais la qualité de la classification est pour cette classe inégale, elle agrège en effet des réponses parfois hétérogènes.
11. Le logiciel Alceste permet aussi de créer une variable continue qui mesure la qualité du classement de l'individu (qui est plus ou moins au centre de sa classe). Il est particulièrement intéressant, compte tenu de nos effectifs importants de travailler sur deux ou trois univers différents : les « bien classés » dans Alceste et les « mal classés ». On peut ainsi travailler sur le « noyau » de chaque classe et un tri croisé « classe Alceste » par variables explicatives contrôlées par la variable « bien/mal classé » montre que l'on améliore, souvent, les liaisons existantes. La valeur du V de Cramer passe de 0.126 (sig. : 0.000) à 0,203 (sig. : 0.000).
12. La place nous manque ici pour détailler ces points abordés dans la communication suivante : M. Brugidou, N. Mandran, M. Moine, A.-C. Salomon « Les apports de l'analyse textuelle pour l'analyse électorale », *JADT 2004*.
13. Le nombre de réponses à cette question ouverte est de 1003 pour la première question, de 1005 pour la seconde. Le taux de non-réponse à ces deux questions est très faible, puisque le nombre total d'enquêtés de la troisième vague est de 2013. Avant de procéder à cette étude, nous avons vérifié que la structure par âge, sexe, niveau d'études, vote au premier tour est indépendante du mode de questionnement. Les questions ouvertes ont été recueillies et retranscrites directement par l'enquêteur lors du terrain téléphonique.
14. P. Perrineau, C. Ysmal (dir.), *Le Vote de tous les refus, Chroniques Electorales*, Paris, Presses de Sciences Po, 2003.
15. Brugidou M. « Argumentation and Values : an analysis of ordinary political competence via an open-ended question », *International Journal of Public Opinion Research*, V15 n°4, winter 2003.
16. Converse, J.-M. « Strong arguments and weak evidence : The open/closed questioning controversy of the 1940s. *Public Opinion Quarterly*, 48, 267-282.
17. Par exemple, *La démocratie à l'épreuve. Une nouvelle approche de l'opinion des Français*. Sous la direction de Gérard Grunberg, Nonna Mayer et Paul M. Sniderman, Presses de Sciences Po, mars 2002, 352 p.

18. Voir notamment le n° spécial du *Journal de la Société Française de Statistique*, 2001, 4.
 19. Voir sur ce sujet le travail pionnier mené en France par Jacques Jenny à l'origine du réseau ARCATI (Atelier-Réseau Coopératif pour Analyses Textuelles Informatisées).
 20. de Fornel M., et Passeron J.-C., *L'argumentation, preuve et persuasion, Enquête, éditions de l'école des Hautes Etudes en Sciences Sociales*. 2002, p. 10.
-

RÉSUMÉS

Cette article se propose de décrire un dispositif d'enquête électoral qui comporte de nombreuses questions ouvertes, puis de détailler les stratégies d'analyses retenues et enfin, à partir de quelques exemples de traitement, de montrer les avancées méthodologiques – et leurs limites – des approches mêlant des dimensions quantitative et qualitative. Lors de la séquence électorale du printemps 2002, un dispositif d'enquête comprenant un panel a permis de mesurer les opinions et les comportements politiques des français en trois vagues d'enquêtes. La première partie aborde les différents problèmes posés par ce type de dispositif qui s'avèrent plus ou moins bien traités actuellement : élaboration d'une typologie de questions ouvertes, place de la question ouverte dans le questionnaire, interactions enquêteur/enquêté... Les réponses à une partie de ces questions ont été intégralement retranscrites et ont été analysées par le logiciel d'analyse textuelle Alceste. Le choix de ces questions ainsi que l'approche en terme d'analyse des données textuelles seront explicités dans un deuxième temps. Pour revenir à une approche plus explicative, il est possible de croiser les classes produites par Alceste avec les questions fermées. Cette approche pose toutefois une série de questions épineuses tant du point de vue technique qu'épistémologique que nous aborderons dans une troisième partie.

Method and Strategy of Analysis of Open Questions of the French Electoral Panel: This article presents a voting survey method which includes many open questions, the detailed strategies of analysis employed and also, based on analyzed examples, the methodological advances – and their limits – of such strategies which mix quantitative and qualitative approaches. During the spring 2002 elections, a survey apparatus was set up including a panel which would permit the measuring of French political opinions and behavior over three waves of the survey. The first part of this article addresses problems engendered by this survey apparatus and how they have been handled. This involves the creation of a typology of open questions, the place of open questions in a questionnaire, interviewer/interviewee interactions, etc. Replies to part of the questions were transcribed and analyzed using the Alceste program. The choice of these questions and the use of a text analysis approach are discussed in the second part of the article. To return to a more explanatory approach, it is possible to cross-tabulate the classes produced by Alceste and the closed questions of the questionnaire. This procedure encounters several difficult problems that are both technical and epistemological that we address in the third part of the article.

INDEX

Keywords : Open Questions, Analysis of Textual Data, Discourse Analysis, Themes, Alceste

Mots-clés : Questions ouvertes, Analyse des données textuelles, Analyse de discours, Thème

AUTEURS

MATHIEU BRUGIDOU

CIDSP, EDF R&D, 92141 Clamart Cedex, France ; mathieu.brugidou@edf.fr

NADINE MANDRAN

CIDSP – IEP, 38040 Grenoble Cedex 09, France

MICHÈLE MOINE

CIDSP, LabSAD, 38040 Grenoble Cedex 09, France

ANNIE-CLAUDE SALOMON

CIDSP – IEP, 38040 Grenoble Cedex 09, France