

Les apports de l'analyse textuelle pour l'analyse électorale : les questions ouvertes du panel électoral de 2002

Mathieu Brugidou¹, Nadine Mandran², Michel Moine³, Annie-Claude Salomon²

¹CIDSP, GRETS-EDF – 92141 Clamart Cedex – France

²CIDSP, IEP – BP 48 – 38040 Grenoble Cedex 09 – France

³LabSAD-UPMF, CIDSP – BP 47- 38040 Grenoble Cedex 09 – France
mandran@cidsp.upmf-grenoble.fr, michele.moine@iutz.upmf-grenoble.fr

Abstract

During the survey « Panel électoral français 2002 », the answers to several open ended questions were collected. The analysis of the texts of the responses to questions concerning the reasons of the defeat of the left, and the reasons of the victory of the right are analyzed in this document. Two methods will be used to analyze the structure of this text. A lexical analysis offer the opportunity to extract the different speeches. This step will be improve by a thematic analysis. This methodology allows a sociological and political analysis too. In fact, each interviewed person will be characterized by the different thematics of his response.

Résumé

Lors de la dernière vague de l'enquête électorale du « Panel électoral français 2002 », les réponses à plusieurs questions ouvertes ont été recueillies. Les données analysées sont les réponses aux deux items « Selon vous, pourquoi la gauche a-t-elle perdu ? », et « selon vous, pourquoi la droite a-t-elle gagné ? ». Deux méthodes complémentaires seront utilisées pour mettre à jour la structure ce corpus. Une analyse lexicale extrait les principaux discours présents dans l'opinion. Cette étape est enrichie par une analyse thématique qui affine les classes d'énoncés. Elle permet une analyse individuelle des thèmes et une caractérisation socio-politique de ces discours.

Mots-clés : question ouverte, analyse du discours, analyse lexicale , analyse thématique, analyse électorale.

1. Présentation du Panel électoral français 2002

Les données du « **Panel électoral français 2002** » (PEF, 2002) ont été produites par le CEVIPOF, le CIDSP, le CECOP avec le soutien du Ministère de l'Intérieur et de la FNSP. Ce dispositif d'enquête a permis de mesurer les opinions et les comportements politiques des français. Cette étude a été menée en trois temps : une première vague pré-présidentielle, une seconde post-présidentielle et une troisième post-législative. Trois questionnaires comportaient principalement des questions fermées mais aussi différentes questions ouvertes. Dans la vague 1, elles portaient sur les raisons de la participation ou non au vote et sur les raisons du choix du candidat. Dans la vague 2, les questions ouvertes abordaient les problèmes d'environnement d'une part et d'autre part les raisons du vote J. Chirac, du vote J.-M. Le Pen et de l'abstention. En dernière vague, ce sont les raisons de l'abstention, de la défaite de la gauche ou de la victoire de la droite qui ont été recueillies.

2. Objectif de l'étude

Lors de la troisième vague, l'une des deux questions ouvertes suivantes était posée aléatoirement à chaque enquêté¹ :

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

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

À partir de ces deux questions, un double objectif a été défini : le premier est d'isoler les différents discours produits par les enquêtés interrogés à propos de la victoire de la droite ou la défaite de la gauche à l'issue de la séquence électorale de 2002, 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 (victoire ou défaite). 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 d'une part leurs caractéristiques sociodémographiques et d'autre part leurs opinions et leurs comportements politiques ont conduit à la conclusion que certaines relations étaient indépendantes de la question posée.

3. Analyse textuelle, définition des champs lexicaux

3.1. Analyse de discours des réponses à la question « Pourquoi la gauche a-t-elle perdu ? »

L'analyse statistique produite par le logiciel Alceste conduit à un classement des énoncés relatifs à la défaite de la gauche, une réponse étant découpée en séquence d'énoncés.

La classification descendante fait émerger deux pôles majeurs pour expliquer la défaite de la gauche : d'une part, un constat négatif sur les actions du gouvernement et des promesses non tenues, d'autre part les motifs énoncés font référence à la perception de la campagne du premier tour de la présidentielle et aux comportements électoraux lors de cette élection. Cette analyse distingue donc quatre classes de discours.

3.1.1. Une absence de programme et de leader (classe 1)

L'opinion selon laquelle la raison de la défaite de la gauche est une absence de programme et de leader s'exprime dans cette classe. Cette absence est d'autant plus forte que les candidatures à gauche lors de la présidentielle étaient nombreuses, ce qui a brouillé les signaux lors de la campagne électorale. « *car elle était mauvaise pas de programmes clairs et pas de dynamique* ». « *manque d'union manque d'information et par manque de leader digne de ce nom pas de figure charismatique* ».

3.1.2. Un vote contestataire et l'abstention (classe 2)

Les raisons de la défaite de la gauche dans cette classe sont doubles. Nous trouvons d'une part un discours qui met en avant le vote contestataire car les individus ont été déçus des actions

¹ 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.

menées par la gauche : « *il y a beaucoup de déçus qui ont voté sanction au premier tour et ça fait un vote d'extrême droite* ». D'autre part, c'est l'abstention qui est évoquée pour expliquer les raisons de la défaite de la gauche, « *parce que il y a eu beaucoup trop d' abstentions. Je vois surtout cela, les gens ont dit qu' ils n' iraient pas voter, même sans faire de politique, on peut avoir son opinion* ». Aussi, une des caractéristiques de cette classe est la présence de l'expression « *extrême droite* ». Les propos tenus ici sont une traduction des événements du 21 avril 2002 : un vote contestataire et une abstention très forte ont conduit au rejet de L. Jospin et à la victoire de l'extrême droite.

3.1.3. *Des promesses non tenues (classe 3)*

La raison de la défaite de la gauche exposée dans cette classe de discours est le fait que le gouvernement n'a pas tenu ses promesses. Ce discours fait référence à un passé lointain qui prend ancrage lors de la campagne des législatives de 1997. Cette classe est caractérisée par des marqueurs de personnalisation comme « *leurs* » ou « *ses* », qui font référence aux personnalités en place. « *ils ont menti pendant trop longtemps, ils nous ont berné ils ont pas tenu leurs promesses* ».

3.1.4. *Le bilan négatif des actions de la gauche. (classe 4)*

La caractéristique de ce discours est de lister et de dresser un bilan négatif des actions menées par le gouvernement. Les trente cinq heures arrivent en tête pour expliquer la défaite de la gauche. « *Les Français étaient pas satisfaits et notamment les trente cinq heures et la baisse du pouvoir d' achat, ils n'ont pas les moyens de profiter de la RTT* ». Ces déclarations mettent aussi en exergue les différents problèmes qui n'ont pas été abordés et/ou résolus par le gouvernement (retraite, sécurité, chômage,...). Ces actions sont parfois qualifiées de « *bêtise* », ou d'« *erreur* ».

3.2. *Analyse de discours des réponses à la question « Pourquoi la droite a-t-elle gagné ? »*

L'analyse statistique produite par le programme Alceste fait ressortir quatre classes de discours dans le corpus des réponses relatives à la victoire de la droite. Ces quatre classes s'organisent selon trois pôles : la peur de l'Extrême Droite, le refus de la cohabitation (classes 1 et 2), le ras le bol qui pousse au changement (classe 3) et la déception vis-à-vis de la gauche : bilan factuel des actions de la gauche et jugement critique de son action (classe 4).

3.2.1. *Le refus de la cohabitation, une meilleure campagne à droite (classe 1)*

Le contenu de ces énoncés renvoie à la nécessité de « *donner une majorité parlementaire au Président de la République afin d'éviter une nouvelle cohabitation* ». La victoire de la droite est présentée comme étant « *la suite logique de l'élection présidentielle* ». La campagne de la droite en 2002 est jugée meilleure – « *Ils avaient les meilleures têtes, le meilleur programme* », « *la droite était unie et non divisée* » – que celle de la gauche – « *décevante* », « *divisée* », « *mauvaise campagne* »,...

3.2.2. *La peur de l'extrême droite (classe 2)*

Le thème de la peur de l'extrême droite est très présent dans cette classe. La surprise devant les résultats du premier tour des présidentielles est évoquée. Le rôle de l'abstention au premier tour de la présidentielle, des reports des votes « *de gauche* » au deuxième tour de la présidentielle, est abordé. Les noms des candidats Chirac, Le Pen et Jospin sont particulièrement fréquents. Le thème de l'absence de choix figure dans ces énoncés, « *on n'avait plus le*

choix », « *pas assez de candidats au second tour* ». Ce discours est une représentation de l'événement du 21 avril 2002.

3.2.3. *Le ras le bol qui pousse vers une volonté de changement (classe 3)*

La volonté de changement est clairement affirmée dans les énoncés de cette classe. Les mots ou expressions « *changer* », « *changement* », « *alternance* », « *ras-le-bol de la gauche* », « *manque de confiance* » envers la gauche, « *manque d'information* » (délivrée par la gauche au pouvoir), « *promesses non tenues* » (du gouvernement de gauche) sont spécifiques de ces textes. On peut remarquer la fréquence élevée du mot Français : « *les Français aiment changer* », « *volonté de changement des Français* »,... Les extraits de discours de cette classe dénotent un point de vue très global.

3.2.4. *La politique de gauche a déçu (classe 4)*

Le thème principal de cette classe est la déception devant les actions du gouvernement de la gauche. Les problèmes relatifs aux « 35 heures », à l'immigration, aux inégalités sociales, à la sécurité, à la délinquance, à l'école, aux chômage, aux cotisations sociales, aux impôts, aux retraites... sont développés. Les jugements des enquêtés à l'égard de la gauche portent sur le manque d'écoute et de compréhension de la population, l'éloignement du terrain, l'incapacité à résoudre les problèmes, le fait qu'elle ne s'est pas montrée à la hauteur des attentes, ses erreurs, ses bêtises ...

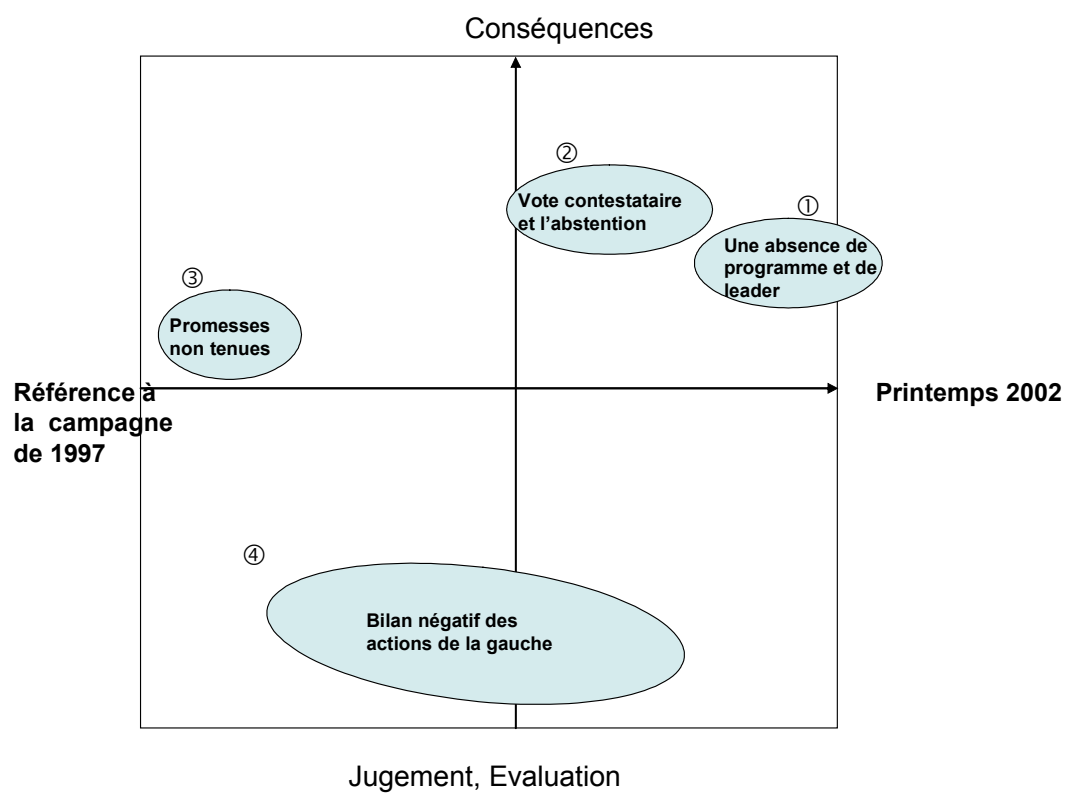
3.3. *Bilan de l'analyse lexicale*

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. Certains types sont récurrents quelle que soit la formulation de la question. Il s'agit des énoncés concernant le bilan des actions de la gauche lors des cinq dernières années, la référence aux campagnes électorales jugées meilleures pour la droite que pour la gauche. En revanche, 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.

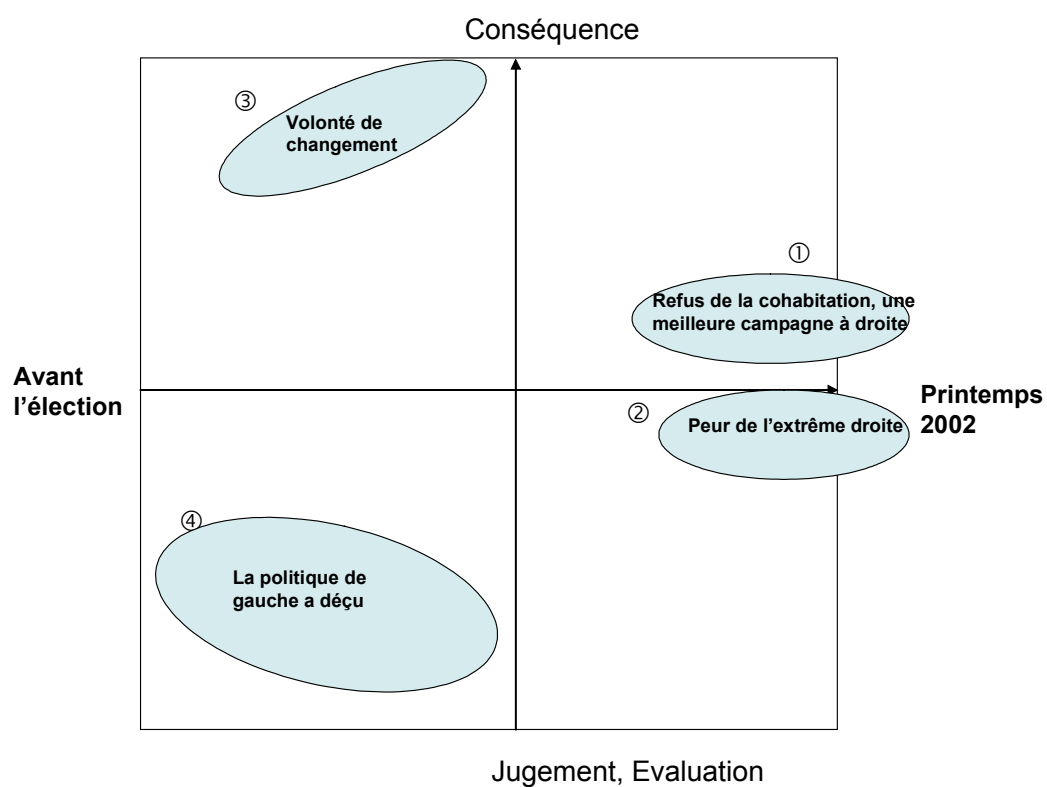
Les graphes présentés ci-dessus fournissent une image synthétique des premiers plans de l'analyse factorielle du tableau croisant le vocabulaire et les classes d'énoncés dégagées par le logiciel Alceste.

Le premier axe de chacune de ces deux analyses des correspondances 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 1. Représentation factorielle des classes Alceste
 « Raisons de la défaite » (inertie : axe 1 : 37.8%, axe 2 : 34.7%)



« Raisons de la victoire » (inertie : axe 1 : 41.7%, axe 2 : 32%)



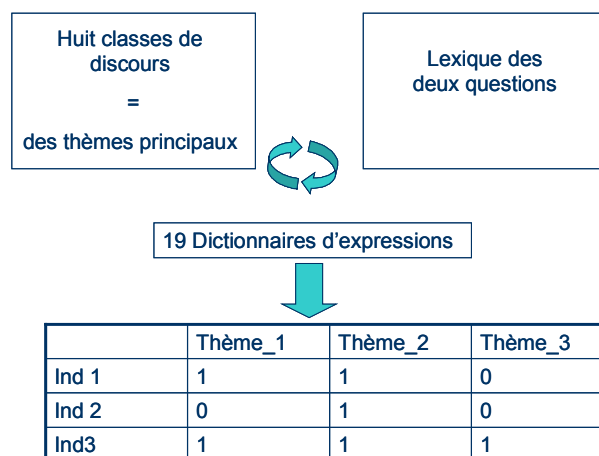
À partir de ces classes de discours, il convient maintenant, de mesurer l'influence de la question posée sur les thèmes abordés et de connaître le profil sociodémographique, les opinions et comportements politiques des individus tenant ces propos. Or, l'ensemble des verbatims n'est pas classé dans l'analyse précédente du fait de la grande variabilité de l'information recueillie et, dès lors, une étude statistique concernant l'ensemble des locuteurs s'avère délicate. De plus, les contextes dégagés précédemment sont parfois très hétérogènes (c'est le cas de la classe : « vote contestataire et l'abstention »). Ces deux raisons nous ont conduit à recourir à une analyse thématique pour affiner les classes de discours et revenir au plus près du répondant.

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

4.1. La démarche

Chacune des classes de discours produite par Alceste est caractérisée par un ensemble de mots spécifiques. À partir de ces mots, des champs lexicaux sont isolés. Ces différents champs lexicaux constituent un premier niveau de thématique. À 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.

Figure 2. Organisation de la démarche



4.2. Les thèmes et leur fréquence

Dix-neuf thèmes ont été retenus. Certains, trop peu fréquents (expression du manque d'intérêt suscité par la question, « je ne m'intéresse pas à la politique ») n'ont pas 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.

Les thèmes abordés dans les réponses ont été regroupés en six catégories :

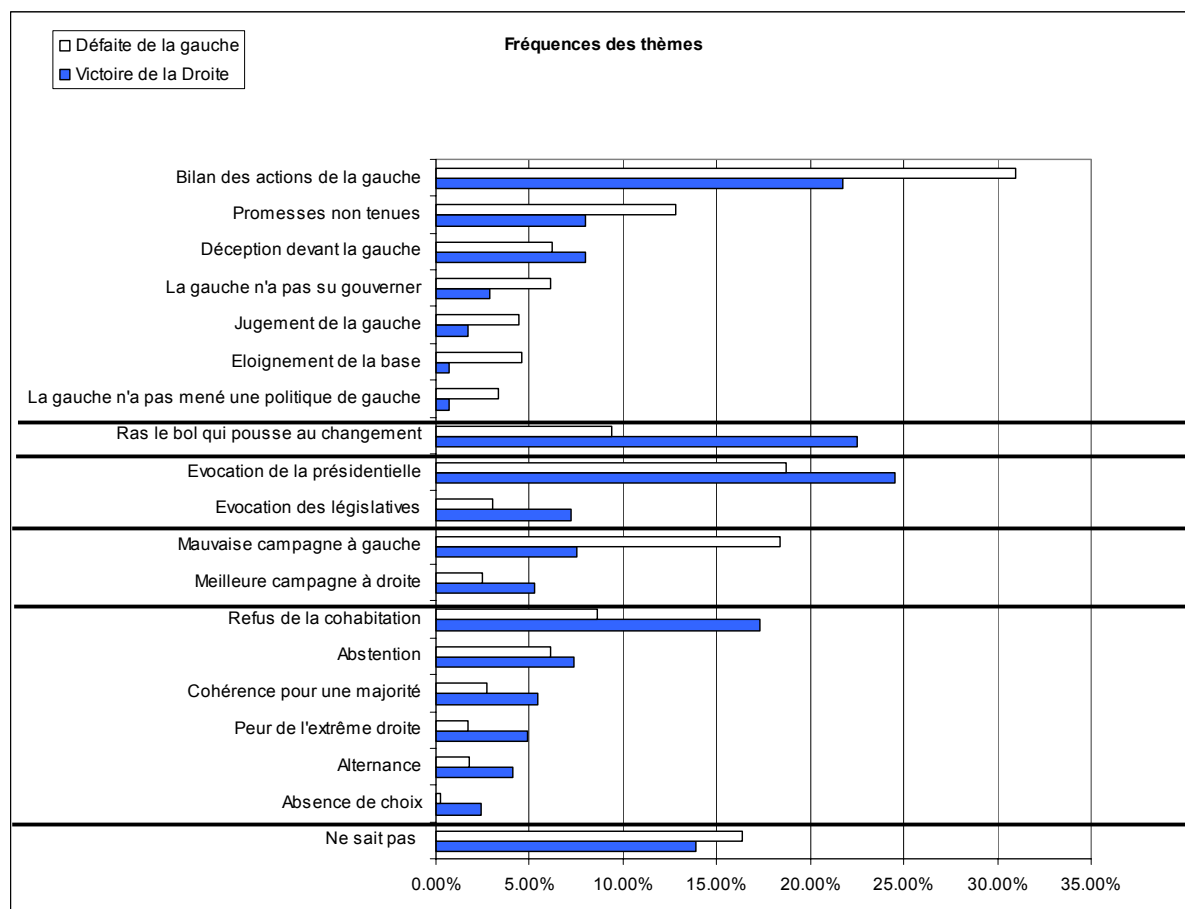
- *Le jugement sur la politique et les personnalités de gauche* constitue une première catégorie : le bilan du gouvernement de gauche, les promesses non tenues, le fait que la gauche n'a pas su gouverner, que la gauche n'a pas mené une politique de gauche, son attitude vis-à-vis de la base (le peuple, 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%).
- 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.
- 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.
- *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%).
- 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.
- 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.

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 ».

Les critiques relatives à la politique menée par la gauche, au comportement des responsables politiques de gauche et à la campagne de la gauche sont plus particulièrement présentes dans le corpus des réponses analysant les raisons de la défaite de la gauche : ce résultat était attendu.

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 électoral (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.

Figure 3. Les thèmes et leur fréquence en fonction de la formulation de la question



4.3. Le retour aux répondants

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 et comportementale a été menée. 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 trois dimensions ont été croisées avec l'âge, le sexe, le diplôme, l'intérêt pour la politique, le positionnement politique et la reconstitution du vote du premier tour de la présidentielle. 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).

4.3.1. Le profil sociodémographique des thématiques

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.

Dans le cas du niveau d'études, on a pu noter que le fait d'avoir un diplôme, le plus modeste soit-il, semble favoriser l'expression du discours. Les « sans diplômes » ont moins de certitudes pour expliquer l'issue des élections. Les plus diplômés sont plus nombreux à citer et commenter les actions de la gauche.

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. À 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.

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.

4.3.2. *Les opinions, attitudes et comportements politiques*

Le sens de la liaison entre les thèmes et l'intérêt pour la politique reste le même quelle que soit la formulation de la question. Les moins intéressés par la politique sont les plus nombreux à déclarer ne pas savoir pourquoi la droite a gagné ou la gauche a perdu. Ceux qui font le bilan des actions de la gauche lors des cinq dernières années ou ceux qui évoquent la présidentielle sont les interviewés qui se déclarent les plus intéressés par la politique.

En revanche, la liaison entre l'expression du vote au premier tour et la difficulté à trouver des raisons pour expliquer l'aboutissement de ces élections est plus forte lorsque l'on parle de la victoire de la droite. En effet, lorsque la question concerne la victoire de la droite, ce sont les extrémistes de droite ou de gauche qui déclarent ne pas savoir. Ce phénomène est beaucoup moins fréquent dans le discours des électeurs de droite. Ces électeurs sont moins hésitants, ils peuvent a minima exprimer les raisons qui ont été les leurs pour voter à droite.

| Question posée | Présence du thème... | Extrême gauche | Gauche | Droite | FN | Total |
|--|-----------------------|----------------|--------|--------|-------|-------|
| Pourquoi la gauche a-t-elle perdu ? | « Ne sait pas » | 20,4% | 12,0% | 13,1% | 14,9% | 13,8% |
| | Effectif Total | 103 | 325 | 290 | 74 | 792 |
| Pourquoi la droite a-t-elle gagné ? | « Ne sait pas » | 17,3% | 12,7% | 8,1% | 20,0% | 12,4% |
| | Effectif Total | 110 | 315 | 310 | 90 | 825 |

Fréquence du thème « Ne sait pas » selon le vote au premier tour de la présidentielle

5. Conclusion

Si l'analyse du discours a conduit à l'élaboration de huit types de discours, l'analyse thématique a permis d'affiner cette typologie en dégagant dix-neuf thèmes.

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.

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.

Après une première lecture des textes recueillis, les énoncés semblaient lapidaires, peu argumentés et comportant des informations très diffuses. Cet état des lieux laissait présager des difficultés à faire émerger une structure dans ce corpus. En fait, ce matériau s'est avéré très riche et structuré. En effet, il a fait apparaître que les raisons de la défaite font référence au programme que la gauche a proposé aux législatives de 1997, aux cinq années de politique de gauche, à la campagne du premier tour de la présidentielle et aux résultats du 21 avril 2002. Par ailleurs, les raisons de la victoire de la droite renvoient aux actions de la gauche jugées de manière négative, à la situation politique issue de l'état de choc qui a suivi le 21 avril 2002. Cette structuration du corpus permet de confirmer l'expression d'un « vote de tous les refus » (Perrineau et Ysmal, 2002) : 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. Ne s'agirait-il même pas d'un vote de tous les rejets ? .

Références

- Le Panel Electoral Français 2002. site web : http://cidsp.upmf-grenoble.fr/index_postelec.htm
- Benzecri J.-P. *et al.* (1981). Pratique de l'analyse des données, vol. (3). Dunod.
- Blot I., Hammer B. et Le roux D. (1994). Traitement des questions d'opinion « ouvertes » : utilisation d'Alceste, outil d'assistance à l'analyse. *Revue ICO Québec*, vol. (6/1-2).
- Brugidou M. (1998). Épitaphes, l'image de François Mitterrand à travers l'analyse d'une question ouverte posée à sa mort. *Revue Française de Science Politique*, vol. (48/1) : 97-120.
- Brugidou M. (2001). La combinaison des inférences statistiques, linguistiques et sociologiques dans l'analyse d'une question ouverte. *Journal de la Société Française de Statistique*, vol (142/4).
- Lebart L. et Salem A. (1994). Statistique Textuelle. Dunod.
- Muller Ch. (1993). Principes et méthodes de statistique lexicale. Slatkine-Champion.
- Perrineau P. et Ysmal C. (2002). Le vote de tous les refus. Presses de Sciences Po.
- Reinert M. (1987). Classification descendante hiérarchique et analyse lexicale par contexte : application au corpus des poésies d'Arthur Rimbaud. *Bulletin de Méthodologie Sociologique*, vol. (13).