



6^{ème} Congrès des Associations Francophones de Science Politique

ST 20 : Comprendre les programmes électoraux : comparaison des méthodes d'encodage manuel et automatique

Les auteurs :

Grégory Piet, Boursier de doctorat ARC, Université de Liège (BE), Département de science politique, Spiral – Gouvernance et Société.

Régis Dandoy, Chercheur Prometeo, FLACSO Ecuador, Université Libre de Bruxelles.

Jeroen Joly, Chercheur Postdoctoral, University of Toronto, Universiteit Antwerpen.

Résumé :

En science politique, l'analyse de documents politiques est cruciale afin de tester toute une série d'hypothèses concernant les priorités des acteurs et institutions politiques, mais les méthodes utilisées afin d'analyser ces textes politiques sont (trop) nombreuses. Les techniques d'analyse de textes basées sur le codage humain sont les plus répandues et fournissent les meilleures données en ce qui concerne le contenu de ces documents. Toutefois, les techniques automatisées présentent également d'incontournables avantages. Cet article entend comparer les résultats obtenus via deux méthodes distinctes d'analyse d'un même corpus de textes politiques, à savoir une méthode d'encodage manuel et une méthode d'encodage automatisé. Les programmes électoraux des principaux partis politiques belges entre 1987 et 2007 seront utilisés comme textes de base. L'identification et comparaison des principales thématiques abordées dans ces documents permettront de déterminer les similarités et différences obtenues via ces deux méthodes. Au final, cet article discute ces deux méthodes et les enseignements qui peuvent en être retirés pour de futures analyses de documents politiques.

Cinq mots clés:

Programmes électoraux, Belgique, méthodologie, Prospéro, Comparative Agendas Project

Introduction

Afin de comprendre les systèmes politiques des sociétés occidentales, il est crucial de prendre en compte les priorités de leurs partis politiques. Celles-ci sont ainsi déterminantes pour comprendre la mise en place des politiques publiques et certains processus politiques. Partant, la science politique mobilise les priorités politiques partisans – souvent sur base des programmes électoraux – comme variable indépendante afin d’expliquer certaines politiques publiques. Les priorités politiques des partis permettent ainsi d’appréhender les budgets et dépenses étatiques, la législation, ou encore certaines politiques socio-économiques. Elles permettent également de comprendre certains phénomènes et processus politiques, comme les résultats électoraux, la formation des gouvernements et des coalitions ou le contenu de certains accords entre partis ou autres déclarations gouvernementales.

Dans le cadre de cet article, la priorité politique d’un parti concernant une thématique publique est définie comme étant l’importance relative que ce parti attribue à cette thématique. En effet, si un parti prête attention à de nombreux enjeux de politiques publiques, cette attention n’est pas infinie et doit donc être distribuée à travers les différentes thématiques de politiques publiques. Elle ne peut donc être identique pour tous ces enjeux. Dépendantes du contexte politique, des stratégies du leadership du parti ou tout simplement de l’idéologie partisane, certaines thématiques vont être mises en avant, au détriment d’autres. Ainsi, il paraît vraisemblable qu’un parti écologiste va mettre l’accent sur les politiques environnementales tandis qu’un parti socialiste va se focaliser sur les politiques sociales.

Dans le sillage de Robertson (1976) et de Budge et Farlie (1983), la science politique s’est penchée sur la *saliency theory*. L’argument de base de cette théorie est que le point de vue des partis politiques sur certaines thématiques publiques peut être analysé sur base de leurs priorités politiques plutôt que sur base de leurs positions. Les partis choisissent ainsi stratégiquement de mettre l’accent sur certaines thématiques plutôt que d’autres, afin de se différencier des autres partis, de séduire une nouvelle frange de l’électorat, etc. Ces préférences politiques différentes selon les partis ont mené à l’observation de phénomènes politiques tels que la crédibilité et la possession d’une thématique (les partis essaient de développer et faire reconnaître par l’électeur leur expertise sur une thématique précise), la compétition partisane (les partis entrent en concurrence afin de mettre en avant un nouveau thème ou reprendre celui d’un autre parti), la mise en avant d’enjeux politiques incontournables (démocratie, chômage, etc.) plutôt que de thématiques plus polémiques, etc.

La base empirique de la *saliency theory* est, comme son nom l’indique, la mesure de la place relative prise par une thématique dans l’idéologie et le discours officiel. Les priorités politiques d’un parti peuvent ainsi être appréhendées via la place (et le rang) que prennent certaines thématiques parmi toutes les thématiques abordées lors d’une campagne électorale.

Afin de mesurer ces priorités politiques, le programme électoral est le document le plus souvent analysé en science politique. À la suite du *Comparative Manifesto Project*, l’utilisation des programmes électoraux s’est imposée dans de nombreuses recherches comparatives sur les partis politiques (Budge, 2001). Les programmes électoraux présentent de nombreux avantages, en comparaison d’autres techniques parfois utilisées (enquêtes d’opinion, documents parlementaires, vote dans les assemblées, analyses de discours de campagne, etc.) : il s’agit de documents officiels, écrits et avalisés par les structures du parti ; ils engagent le parti, ses candidats et ses différentes factions lors de la campagne et lors de la législature ; ils traitent de toutes les thématiques de politiques publiques que le parti désire mettre en avant ; ils sont rendus publics et sont destinés aux électeurs, à la société civile et aux autres

partis ; ils sont disponibles à intervalle régulier (lors de chaque élection) et chaque parti en rédige un, etc. Au demeurant, le principal avantage des programmes électoraux est qu'ils permettent une comparaison dans le temps et dans l'espace – entre partis – sur toute une série de thématiques de politiques publiques.

Les méthodes utilisées en science politique afin d'analyser un texte politique sont nombreuses. Les techniques d'analyse de textes basées sur l'humain sont les plus répandues et fournissent vraisemblablement les meilleures données en ce qui concerne le contenu des programmes électoraux (Budge, 2001 : 50). Toutefois, les techniques automatisées présentent également d'incontournables avantages. Cet article entend comparer les résultats obtenus via deux méthodes distinctes d'analyse d'un même corpus de textes, à savoir une méthode d'encodage manuel et une méthode d'encodage automatisé. Les programmes électoraux des principaux partis politiques belges entre 1987 et 2007 seront utilisés comme textes de base. L'identification et comparaison des principales thématiques abordées dans ces documents permettront de déterminer les similarités et différences obtenues via ces deux méthodes. Enfin, cet article discutera ces deux méthodes et les enseignements que nous pourrions tirer pour de futures analyses de documents politiques.

Les programmes électoraux en Belgique

La Belgique est prise comme cas d'étude¹ pour la comparaison de ces deux méthodes d'analyse des priorités politiques partisans sur base des programmes électoraux. L'avantage du cas belge repose sur le fait que, au sein d'un même pays, il est possible de comparer deux systèmes de partis distincts. Dans la seconde moitié du 20^e siècle, l'histoire politique belge a mené à la création de deux systèmes de partis imperméables et basés sur les langues néerlandaise et française. Entre 1968 et 1978, les partis traditionnels se sont scindés en ailes linguistiques, menant à la création de partis politiques totalement indépendants. Tous les partis politiques qui furent créés après cette période ont également fait le choix de ne se présenter à l'électeur que dans une seule langue. Il n'y a aujourd'hui en Belgique plus de véritables partis nationaux mais bien deux ensembles de partis régionaux (Dandoy, De Decker, 2009). Bien que le cas belge ait été très souvent inclus dans les études comparatives portant sur l'analyse du contenu des programmes électoraux, cette spécificité a rarement été mobilisée par la littérature (voir Hearl, 1987; Walgrave, De Winter, Nuytemans, 2005 pour une exception). Qui plus est, certaines études quantitatives récentes (voir par exemple Walgrave, De Swert, 2007; Bouteau, 2009) se focalisent exclusivement sur les programmes électoraux flamands, laissant de côté la comparaison avec les programmes francophones.

Le cas belge est d'autant plus intéressant que les programmes électoraux sont exclusivement rédigés dans la langue du parti. Depuis les années 1970, les programmes électoraux ne sont même plus traduits dans les autres langues nationales. Cela concerne également les partis qui se présentent à Bruxelles, région pourtant bilingue. Le défi de l'analyse des programmes électoraux dans ce pays est donc posé : comprendre les préférences politiques partisans dans un système politique et des institutions fédérales

¹ À l'opposé des nombreuses études comparatives internationales qui peinent à fournir des résultats cohérents ou qui réduisent les priorités politiques partisans à un simple positionnement sur un axe gauche-droite, Warwick (2005 : 4) affirme que les analyses de cas basées sur un pays sont plus adéquates pour la compréhension des priorités politiques partisans. A contrario, l'analyse d'un seul cas réduit fortement la mobilisation d'outils quantitatifs pour le traitement statistique du contenu des programmes électoraux.

identiques pour tous les partis, mais au sein de systèmes partisans séparés et sur base d'une langue de travail différente.

Etant donné le considérable travail nécessaire pour l'encodage des programmes électoraux sur base de la méthode CAP (voir ci-dessous), une sélection des documents à analyser a été effectuée. Il n'est ainsi pas possible de mesurer les priorités politiques pour tous les partis politiques, pour toutes les élections et à tous les niveaux de pouvoir. Afin de diversifier au mieux les documents analysés, nous avons décidé de nous pencher sur un segment temporel assez large, allant de 1987 à 2007. L'avantage de cette période est que la structure des deux systèmes de partis (flamands et francophones) est très stable puisqu'aucun nouveau parti politique n'est entré de manière durable au parlement fédéral dans cette période. Mais cette stabilité des systèmes de partis ne va pas de pair avec une stabilité gouvernementale puisque, entre 1995 et 2007, toutes les formules de coalitions qui furent mises en place au niveau du gouvernement fédéral ont été différentes. Le segment temporel allant de 1987 à 2007 nous permet ainsi de diversifier les partis au pouvoir et dans l'opposition tout en conservant un ensemble stable sur les partis étudiés.

La Belgique est composée de cinq niveaux de pouvoir (communal, provincial, régional et communautaire, fédéral, et européen) où se déroulent des élections et où, par définition, des programmes électoraux sont rédigés. Le niveau de pouvoir le plus intéressant pour prendre en compte les différences linguistiques est le niveau fédéral. En effet, même si les partis flamands et francophones se présentent principalement dans des circonscriptions électorales distinctes, de nombreux aspects de ces élections et de la campagne restent communs : même calendrier électoral, mêmes institutions publiques, mêmes compétences et même contexte international et socio-économique. Les enjeux spécifiquement régionaux et communautaires sont discutés lors des campagnes pour les élections à ces niveaux de pouvoir. À remarquer que les élections fédérales de 1995 et 1999 ont lieu le même jour que les élections régionales et communautaires². Enfin, bien que les élections fédérales peuvent être organisées de manière anticipée, les élections fédérales entre 1987 et 2007 ont eu lieu à chaque fois à la fin du terme législatif (terme de quatre ans). Par conséquent, les partis politiques ont produit un programme électoral pour ces élections à un intervalle régulier³.

Les contenus des programmes électoraux pour les six élections fédérales organisées entre 1987 et 2007 seront analysés pour un maximum de partis. La sélection de ceux-ci se base sur un critère simple : avoir obtenu au moins un siège au parlement fédéral (Chambre et/ou Sénat) lors de deux élections parmi les six scrutins analysés. Selon ce critère, pas moins de 12 partis politiques seront identifiés et répartis de manière symétrique dans chaque système de partis : les deux partis sociaux-chrétiens (le flamand CD&V⁴ et le francophone cdH⁵), socialistes (le flamand sp.a⁶ et le francophone PS), libéraux (le flamand Open VLD⁷ et le francophone MR⁸), écologistes (le flamand Groen!⁹ et le francophone Ecolo),

² Pour être complet, signalons que les élections nationales de 1987 et 1991 ont eu lieu le même jour que les élections provinciales tandis que les élections de 1999 ont également eu lieu le même jour que les élections européennes.

³ Cet intervalle entre deux points de mesure (4 ans) est par ailleurs plus court que celui aux niveaux régionaux, communautaires et européens (5 ans) et que celui aux niveaux communaux et provinciaux (6 ans).

⁴ Anciennement CVP, devenu CD&V en 2001.

⁵ Anciennement PSC, devenu cdH en 2002.

⁶ Anciennement SP, devenu sp.a en 2001.

⁷ Anciennement PVV, devenu VLD en 1992 et Open VLD en 2007.

⁸ Anciennement PRL, devenu MR en 2002.

⁹ Anciennement Agalev, devenu Groen! en 2003.

d'extrême-droite (le flamand VB¹⁰ et le francophone FN) et régionalistes (le flamand N-VA¹¹ et le francophone FDF). Au total, ces partis ont rédigé pas moins de 64 programmes électoraux (28 francophones et 36 néerlandophones) pour ces six élections successives. Ces 64 documents, dont la longueur varie entre 3.031 mots dans le cas du VB en 1987 et 208.803 mots dans le cas d'Ecolo en 2007, constituent notre corpus de textes comptant 2.263.209 mots.

L'approche par les « agendas » comparés

À la suite de la *saliency theory* (voir plus haut), l'approche par les « agendas » comparés (*Comparative Agendas Project* – CAP) met l'accent sur le concept d'attention (Baumgartner, Jones, 1993 ; Soroka, 2002 ; Jones, Baumgartner, 2005 ; Brouard, *al.*, 2009 ; Wilkerson, *al.*, 2009 ; Green-Perdersen, Walgrave, 2014). L'attention nous permet de tracer l'évolution d'un enjeu de politique publique à travers le temps, mais également d'examiner la manière dont divers acteurs s'influencent et la façon dont une question politique passe d'un agenda politique ou partisan à un autre. En termes de recherche, l'attention est conceptualisée comme la saillance d'une thématique politique sur un agenda, et comparée aux autres thématiques, ce qui nous donne une mesure de l'importance relative de cette question. Afin de tracer l'évolution des nombreux enjeux abordés par différents acteurs politiques, Baumgartner et Jones (1993) ont conçu un *codebook* thématique qui permet de catégoriser tous ces enjeux politiques selon des thèmes spécifiques et généraux. Bien que créé initialement pour coder la législation et les budgets nord-américains, ce *codebook* est désormais utilisé pour encoder diverses activités politiques dans de nombreux pays (voir le site internet du projet : <http://www.comparativeagendas.info/>).

Ce *codebook* comprend quelques 249 thématiques politiques spécifiques, regroupés en 21 ensembles de politiques publiques, et permet de tracer l'évolution temporelle des priorités politiques pour chaque acteur partisan. La méthode CAP, utilisée ici pour mesurer les priorités politiques au sein d'un programme électoral, est très simple. Dans un premier temps, chaque phrase du document se voit attribuer jusqu'à trois codes thématiques. Ensuite, le nombre de références attribuées à chaque thématique est comptabilisé. La mesure de l'importance (relative) de chaque enjeu politique pour chaque acteur est calculée sur base de la proportion de chaque thématique par rapport à l'entièreté du programme électoral – ce que l'on appelle communément un « agenda partisan ».

Le premier défi pour composer l'agenda des partis politiques consiste à obtenir le matériel brut, c'est-à-dire les programmes électoraux tels qu'ils ont été présentés aux électeurs et discutés dans les médias lors de la campagne. Ensuite, ces documents, faisant parfois plusieurs centaines de pages ont été digitalisés. Cette digitalisation est une étape cruciale puisque les étudiants formés et rémunérés encodent directement en ligne ces programmes au moyen du logiciel AmCat, conçu spécialement pour ce type d'encodage par l'Université d'Amsterdam. L'encodage des programmes se fait phrase par phrase. Les phrases contenant plusieurs idées sont divisées en sous-unités, appelées quasi-phrases selon les techniques appliquées par le *Comparative Manifestos Project* (CMP). Les étudiants attribuent alors jusqu'à trois codes thématiques par quasi-phrase, selon la thématique mis en avant dans la phrase – d'abord le moyen, puis le but de la politique mis en exergue¹².

¹⁰ Anciennement Vlaams Blok, devenu Vlaams Belang en 2004.

¹¹ La Volksunie (VU) disparaît en 2001. La N-VA est considérée comme son héritière.

¹² Afin de pouvoir comparer les priorités à travers différents agendas politiques, on encode d'abord le moyen, puis le but de la promesse électorale ou de la politique à mettre en place. Si l'on veut, par exemple, baisser la vitesse maximale des voitures pour des fins écologiques, on encodera d'abord le thème « transport » et ensuite le thème « écologique » vu qu'une législation sera encodée selon le code « transport ».

Une fois le programme électoral encodé dans son entièreté, des statistiques descriptives nous permettent d'observer combien de fois une thématique a été mise en avant par programme et par parti. En calculant la proportion d'attention par thème politique, il nous est possible de mesurer les priorités politiques de chaque parti. Comme l'indique son nom, l'approche CAP est une approche comparative à plusieurs niveaux : entre partis pour une même élection ou pour un même parti à travers le temps. De cette manière, il est possible de comprendre la manière dont les priorités politiques partisanes ont évolué et quand/comment un thème spécifique s'est imposé sur l'agenda d'un parti ou de plusieurs partis. Par la suite, il est possible de positionner les partis sur différents clivages (par exemple sur une échelle gauche-droite) selon les enjeux politiques qui sont mis en avant dans leurs programmes. Il est dès lors possible de voir quels programmes – et quels partis – sont les plus compatibles au niveau de leurs priorités politiques.

Enfin, le concept d'attention et l'utilisation de la saillance relative permettent de comparer les priorités de divers acteurs de la société civile ou au sein d'un même type d'acteur comme les partis politiques, et ce, à travers différents stades du processus décisionnel. Le choix de se concentrer sur la saillance et non sur une sélection de propositions concrètes se justifie dans cette optique comparative. De cette manière, plusieurs auteurs ont constatés les effets des partis politiques sur les politiques des gouvernements dans plusieurs pays ou l'effet des discours publics sur les politiques suivantes (Laver, Budge, 1992 ; Klingemann, *al.*, 1994 ; McDonald, Budge, 2005).

L'approche par la trajectoire des enjeux

Si la méthode CAP dispose de son propre système de traitement d'analyse et de classification des enjeux politiques, dans le cadre de cet article, le *codebook* CAP a été adapté par Grégory Piet à un logiciel informatique de traitement de données (Prospéro). La méthode automatique diffère, toutefois, de la méthode manuelle puisqu'il s'agit ici de travailler par répertoires thématiques et occurrences pré-encodées (mots et expressions) et non au départ de données identifiées manuellement par un codeur, phrase par phrase ou quasi-phrase par quasi-phrase. À la différence de la méthode CAP, les thématiques construites au sein du logiciel sont « formées de plusieurs représentants [mots et expressions], dont la composition est sous le contrôle de l'utilisateur » (Chateauraynaud, 2003 : 215). Une fois le codage des représentants réalisé, il est alors facile d'identifier les mots et expressions présents et absents dans l'ensemble du dossier étudié. Dans le cadre de cet article, nous avons travaillé sur une base de quelques 16.862 représentants qualifiant les 21 thématiques et les 249 sous-thématiques du *codebook*.

Si les deux méthodes sont différentes, la manière d'appréhender les enjeux politiques ou les problèmes publics l'est nettement moins. S'il s'agit avant tout d'étudier l'évolution temporelle des enjeux, de leurs attentions et leurs trajectoires, les questions posées par les deux approches convergent. En effet, d'une part, le logiciel Prospéro (*PROgrammes de Sociologie Pragmatique Expérimentale et Réflexive sur Ordinateur*), développé par Chateauraynaud (2003), est un outil informatique qui vise à étudier « des processus historiques, des controverses, des affaires, des crises, des polémiques, autant de configurations que nous rangeons sous l'appellation commode de "dossiers complexes" » (Chateauraynaud, 2003 : 17-18). D'autre part, le projet CAP repose sur trois questions majeures : (1) « Comment les problèmes parviennent-ils sur l'agenda gouvernemental ? » ; (2) « Comment les agendas gouvernementaux ont-ils évolué au cours du temps et pourquoi ? » ; (3) « Dans quelle mesure les actions des gouvernements reflètent-ils les priorités des citoyens ? » (Wilkerson, *al.*, 2009 : 369-370).

Nous retrouvons ces mêmes interrogations dans le développement du logiciel Prospéro et de la sociologie argumentative à travers plusieurs débats théoriques, épistémologiques et méthodologiques depuis la fin des années 1980.

Au niveau théorique, l'objectif de Chateauraynaud est de répondre à un besoin de la sociologie contemporaine en dépassant une certaine forme de dualité entre « sociologie de la critique » et « théorie de l'acteur-réseau » en proposant une méthode d'analyse des changements et turbulences au travers des interactions entre les acteurs, les objets, les controverses, les crises, etc. (Chateauraynaud, Torny, 1999 ; Chateauraynaud, 2011). Prospéro répond à cette volonté (Chateauraynaud, 2003 : 22). Cette démarche, complétée d'une prise au sérieux des récits et des arguments des différents acteurs, s'inscrit donc aisément à la suite des travaux de la théorie de la saillance et de l'étude des évolutions des enjeux politiques telle qu'envisagée, notamment, par Carmines et Stimson (1989) et Baumgartner et Jones (1993).

Au niveau épistémologique, le débat porte sur la dualité au sein des sciences sociales entre l'approche quantitative et qualitative. Le statut de l'interprétation (Chateauraynaud, 2003 : 137-210) est au cœur de la réflexion sociologique entamée à travers le logiciel Prospéro. Il se voit donner un rôle d'appui à l'interprétation des changements, ruptures et turbulences conjoncturelles d'un dossier complexe étudié par l'utilisateur à travers les jeux d'arguments et les formes de la critique sur lesquels les acteurs prennent appui. L'analyse quantitative préalable travaille, d'une part, à la mise en contexte des dossiers complexes et longs pour lesquels nous pouvons par la suite produire des analyses qualitatives discursives et argumentatives. Le logiciel permet, d'autre part, de dépasser l'opposition désuète entre approches quantitatives et qualitatives (Chateauraynaud, 2003 : 63). Les deux analyses complémentaires permettent également de mesurer statistiquement le corpus et de qualifier ces thématiques par rapport au contexte – politique, par exemple – du dossier étudié en dégagant ses évolutions et changements (nouvelles thématiques, nouveaux paradigmes, nouveaux arguments, nouvelles formes de la critique, nouveaux acteurs, etc.), et ce, en replaçant le chercheur et l'étude des jeux d'arguments au cœur du processus d'interprétation. Si le développement du logiciel Prospéro a été pensé à l'origine dans une démarche qualitative, nous pouvons donc aisément l'utiliser à travers une démarche quantitative, comme le montrera cet article. Ce logiciel nous permet ainsi de passer d'une démarche à l'autre et de produire des analyses complémentaires sur les trajectoires des enjeux et leur argumentation (Piet, 2013). Nous retrouvons également dans ce débat épistémologique les tenants de la méthode CAP permettant de passer de l'analyse du contenu à l'analyse du discours pour étudier tant les évolutions des enjeux que la manière dont chaque acteur politique s'y positionne (enjeu consensuel, positionnel ou propriétaire ; voir par exemple Budge, *al.*, 2001).

Le débat est, enfin, méthodologique. Le logiciel Prospéro propose six niveaux d'analyse différents : au niveau de l'ensemble du corpus, des acteurs, des textes, des discours et des arguments. À nouveau, les deux méthodes ont des points de convergence. Tout d'abord, il est possible de travailler sur une grande variété de textes et sources différentes (programmes électoraux, médias, législation, etc.). L'analyse peut, ensuite, se concentrer sur les thématiques et sous-thématiques qui émergent dans un dossier, et ce, soit texte par texte, soit acteur par acteur. Les deux méthodes envisagent, également, d'étudier l'« environnement » lexical dans lesquels se développent les entités et les acteurs au sein des énoncés (défini ici comme une phrase), et ce, même si la manière de les analyser diffère d'une approche à l'autre. Les différences essentielles se concentrent donc au niveau de la considération des enchaînements argumentatifs et de l'approche argumentative (Chateauraynaud, 2011) absente dans la méthode CAP.

Comparaison des méthodes manuelle et automatique

Dans le cadre de cet article, nous nous concentrerons donc sur la comparaison des analyses produites par les deux méthodes (manuelle CAP et automatique Prospéro), et ce, au niveau de l'identification des acteurs, des thématiques qu'ils défendent et de l'évolution temporelle des enjeux identifiés dans les programmes électoraux. Plusieurs analyses sont proposées pour juger de la qualité des résultats et de leur proximité : l'analyse des similarités (Green-Pedersen, Mortensen, 2009), l'analyse des corrélations ainsi que l'écart des rangs et leur corrélation. La première nous informe sur la similarité des thématiques présentes dans chaque texte et la faible variabilité des données. La deuxième nous informe sur la liaison entre la distribution des enjeux de chaque programme électoral tandis que les troisième et quatrième analyses nous indiquent la variation de la hiérarchie ou du rang des thématiques pour chaque programme. Le tableau 1 reprend l'exemple du programme électoral du parti libéral francophone (PRL) en 1987 et ces quatre mesures.

Dans notre corpus de 64 documents, seuls 59 programmes sont comparables. En effet, cinq programmes électoraux différents ont été utilisés par l'une des deux méthodes : chacune ayant utilisé soit la version complète du programme, soit une version résumée ou synthétisée produite par les partis politiques. Il s'agit des programmes électoraux du cdH (2003 et 2007), d'Ecolo (2003 et 2007) et du PS (2007). L'analyse repose donc, finalement, sur un ensemble de 23 documents francophones (625.736 mots) et 36 néerlandophones (658.732 mots).

De manière générale, la similarité moyenne entre les résultats des deux méthodes est de 83,61% sur l'ensemble du corpus. Autrement dit, un peu plus de 16% des attentions politiques attribuées aux 21 ensembles de politiques publiques fluctuent entre les deux méthodes. La corrélation moyenne (.86) entre les 59 textes confirme le précédent résultat et la forte ressemblance entre l'encodage des mêmes programmes électoraux au travers des deux méthodes. Nous relevons également des similarités moyennes relativement proches entre les programmes francophones et néerlandophones, et ce, même si les versions francophones révèlent une similarité moyenne légèrement plus élevée (84,26%) par rapport aux programmes électoraux des partis flamands (83,19%). Les corrélations moyennes des versions francophones (.88) et néerlandophones (.85) confirment également cette tendance. Ces résultats donnent déjà une première indication positive quant à la qualité et la validité des deux méthodes.

Nous constatons également que la manière de rédiger un programme électoral ou le vocabulaire particulier utilisé par certains partis politiques joue certainement un rôle dans la proximité des résultats entre les deux méthodes manuelle et automatique. Du côté des partis francophones, par exemple, les programmes du FN testés par les deux méthodes sont les plus éloignés : 80,38% de similarité moyenne et .77 de corrélation moyenne. Du côté néerlandophone, les programmes électoraux du parti écologiste (Groen) connaissent la plus faible corrélation moyenne (.70). Ce résultat est toutefois à nuancer, d'une part, en indiquant que cette faible corrélation est essentiellement due à deux liaisons très faibles entre les programmes de 1987 (.58 de corrélation) et les programmes de 1991 (.42) et, d'autre part, par une similarité moyenne de 81,29% qui reste fort acceptable pour l'analyse d'un même texte. Nous constatons également que, tant chez les partis politiques francophones que néerlandophones, les partis de la famille démocrate-chrétienne présentent les niveaux les plus hauts en termes de similarité moyenne (86,59% pour les programmes du CD&V et 86,48% pour ceux du cdH).

En dehors de ces quelques cas particuliers, l'ensemble des similarités moyennes par parti se situe entre 80,38% (FN) et 86,48% (cdH), du côté francophone, et entre 81,01% (VB) – suivi directement de Groen! (81,29%) – et 86,59% (CD&V), du côté néerlandophone. Au regard de cette analyse, les familles politiques écologiste et d'extrême-droite révèlent le plus fort taux de redistribution des attentions politiques (19%) entre les deux méthodes pour un même texte. L'analyse des corrélations moyennes confirme en partie ce résultat : entre .70 (Groen!) et .92 (N-VA), chez les partis néerlandophones, et entre .77 (FN) et .94 (FDF), chez les partis francophones. Ces importantes différences ne sont ainsi pas spécifiques aux petits partis et aux partis « de niche », puisque les corrélations maximales sont observées dans le cas de partis correspondant aux mêmes caractéristiques. Entre ces extrêmes, nous retrouvons tous les principaux partis traditionnels belges.

Enfin, l'analyse de l'écart moyen de rang et leur corrélation confirme les précédents constats sur les deux méthodes. L'écart moyen des rangs est de 2,24 et a une corrélation moyenne de .87. Les thématiques varient en moyenne de deux rangs entre l'analyse produite par les deux méthodes, et ce, tant au sein des programmes électoraux francophones (2,34) que néerlandophones (2,17). Les corrélations moyennes des rangs sont, quant à elles, identiques (respectivement .86 et .87).

Exemple de comparaison des résultats produits : les enjeux écologiques

Si les deux méthodes testées ici semblent proches sur l'ensemble des textes d'un même corpus, produisent-elles pour autant les mêmes résultats et les mêmes trajectoires lorsque nous étudions l'évolution temporelle d'un même enjeu politique ? En d'autres termes, si nous testons les trajectoires d'attention politique portée à un enjeu de politique publique entre 1987 et 2007, suivront-elles les mêmes tendances pour tous les partis politiques, et ce, quelle que soit la méthode utilisée ?

L'hypothèse que nous soulevons à ce stade est la suivante : si les méthodes d'encodage automatique et manuelle s'équivalent et valident chacune les résultats produits par l'autre, elles devraient produire les mêmes évolutions des enjeux de politique publique, les mêmes trajectoires des attentions politiques et, donc, les mêmes résultats. Nous proposons de tester une trajectoire en particulier : celle des enjeux écologiques. Ces enjeux englobent trois thématiques : l'environnement, la politique énergétique et l'agriculture. Au niveau des partis politiques, suite aux limites de notre corpus préalablement identifiées, nous testerons, du côté francophone, ces trajectoires au MR (1987-2007), au PS (1987-2003), au PSC (1987-1999) et chez Ecolo (1987-1999). Les programmes des FDF et du FN ont été laissés de côté pour cette analyse puisqu'ils n'ont présentés un programme électoral qu'à deux reprises (1987 et 1991 pour les FDF et 2003 et 2007 pour le FN) – leurs données ne permettent donc pas d'établir une quelconque trajectoire. Du côté néerlandophone, l'ensemble des programmes électoraux des six partis politiques est utilisé, et ce, sur toute la période allant de 1987 à 2007. Au total, dix trajectoires des attentions politiques portées aux enjeux écologiques issues de dix partis politiques belges sont analysées.

Cinq des dix trajectoires d'attention politique suivent les mêmes évolutions électorales au sp.a (Graphe 1), au CD&V (Graphe 2), à la N-VA (Graphe 3), chez Ecolo (Graphe 4) et au MR (Graphe 5). Deux trajectoires sont très similaires, mais rencontrent juste une évolution différente entre deux élections spécifiques selon les résultats issus des deux méthodes : au VLD (1995-1999 – Graphe 6) et au VB (1987-1991 – Graphe 7). Enfin, trois trajectoires sont relativement peu semblables entre les deux méthodes, même si elles restent dans un ordre de magnitude relativement similaire : au PS (Graphe 8), au cdH (Graphe 9) et chez Groen! (Graphe 10).

Le relevé des trajectoires est, en soi, très intéressant car il permet de confirmer la qualité et la validité externe des deux méthodes et de leurs résultats. Certaines différences persistent entre les deux approches automatique et manuelle, mais elles ne remettent ni en cause les tendances des trajectoires, ni l'évolution des enjeux. Elles se situent davantage au niveau de l'un ou l'autre programme électoral. Tout comme le démontrait déjà l'analyse des similarités, des corrélations et des écarts de rangs, 11 programmes électoraux sur 55 utilisés pour l'analyse des trajectoires des attentions politiques portées aux enjeux écologiques affichent des tendances différentes : soit à la baisse ou soit à la hausse, en fonction des deux méthodes utilisées.

Discussion

Les diverses analyses (similarité, corrélation, écart des rangs et trajectoires d'enjeux) ont démontré que les méthodes d'encodage manuel et automatisé sont relativement proches sur l'ensemble du corpus, même si quelques différences et variations sont relevées. Ainsi, en moyenne, 16% des attentions politiques sont redistribuées différemment d'une méthode à l'autre. De même, l'analyse des trajectoires montre qu'un programme électoral sur cinq présente des tendances sensiblement différentes.

Plusieurs éléments permettent certainement d'expliquer ces écarts. Premièrement, l'élément principal justifiant ces variations dans les résultats repose sur la méthode d'encodage elle-même. En effet, même si les 21 thématiques traitées sont similaires, comptabiliser des données au regard de phrases ou de quasi-phrases est clairement distinct d'un codage sur base de mots et d'expressions repris dans des répertoires construits par le chercheur. Le nombre total de données encodées par chaque méthode varie d'ailleurs, en moyenne, de 41% sur l'ensemble des 59 programmes de notre corpus.

Deuxièmement, le vocabulaire, les métaphores ou les expressions particulières utilisées par certains partis politiques limite la capacité de la méthode automatique à identifier certaines données que la méthode manuelle, par l'intervention d'un codeur, n'a aucune difficulté à prendre en compte.

Troisièmement, la démarche manuelle ou automatique influe sur la production des résultats. Par exemple, le grand nombre de codeurs dans la méthode CAP multiplie les risques d'un codage différent d'un programme électoral à l'autre. Pas moins de 17 codeurs ont participé à l'encodage de l'ensemble des programmes électoraux allant de 1987 à 2007¹³, à la différence de la méthode automatique qui repose sur un seul codeur – le logiciel. Par ailleurs, selon nos analyses préliminaires, la fiabilité entre codeurs est équivalente au niveau de fiabilité relevée entre les deux méthodes. De plus, la systématique de l'approche par logiciel tranche, d'une part, immanquablement par rapport aux choix laissés aux chercheurs dans la méthode manuelle d'encoder ou non une donnée dans une thématique. Elle a, d'autre part, des conséquences sur les résultats produits. À titre d'exemple et de comparaison, nous avons analysé, pour valider notre méthode automatique (dans sa version française et néerlandaise), un programme produit dans les deux langues par un petit parti politique régionaliste (ProBruxsel) qui se présentait aux élections régionales bruxelloises de mai 2014. Les cantons électoraux de la Région Bruxelles-Capitale étant bilingues, ce parti avait décidé de rédiger une version française et une version néerlandaise du même programme électoral. Il ressortait alors une corrélation de .99, une similarité de 92,69% et un écart des rangs de 0,71 entre les deux versions du même programme.

¹³ Une majorité des programmes électoraux ont été encodés par deux codeurs différents et leurs encodages comparés.

Quatrièmement, la variation entre certains résultats s'explique par les éléments des programmes électoraux qui ont été effectivement encodés. Dans certains cas, les titres, les tables de matières, les introductions, les conclusions, les « mots du président », etc. ont été laissés ou retirés de l'encodage des programmes en fonction de la méthode choisie. Les textes traités ne sont donc pas à 100% identique lors de leur traitement.

Enfin, cinquièmement, la méthode automatique cherche impérativement à éviter les doublons. Autrement dit, le chercheur cherche à éviter qu'un même mot ou expression se retrouve dans deux catégories différentes. Par exemple, des mots et expressions comme « développement durable », « environnement » ou « climat » ont donc été réintroduits, au cas par cas, dans les thématiques correspondantes. Il peut donc subsister quelques erreurs minimales dans le recodage de certains mots et expressions.

L'addition de ces distinctions méthodologiques influe alors sur les résultats et explique ces quelques écarts sur certains textes. L'analyse des trajectoires montre, par exemple, clairement que les différences se situent au niveau de quelques programmes et non au niveau des tendances générales et de l'évolution des enjeux politiques étudiés.

Conclusion

L'objectif de cet article était de comparer les résultats de l'analyse d'un même corpus de textes, à savoir les programmes électoraux des principaux partis politiques belges entre 1987 et 2007, au moyen de deux méthodes distinctes : une méthode d'encodage manuel et une méthode d'encodage automatisé. Les principales thématiques abordées dans ces documents ont été identifiées et comparées et nous ont permis de déterminer quelles étaient les similarités et différences essentielles obtenues via ces deux méthodes. En guise de conclusion, nous présentons les enseignements que nous pouvons tirer de ces analyses.

Chacune de ces deux méthodes comporte des avantages et des inconvénients. Basée sur un *codebook* et une méthodologie communs relativement invariants dans le temps et l'espace, la méthode CAP permet la comparaison de textes issus de nombreux pays et de différentes périodes. Son principal avantage est qu'elle permet ainsi de concilier une analyse systématique d'un document avec une expertise et une connaissance approfondie d'un parti et/ou d'une thématique.

Mais cette méthode présente également d'importantes faiblesses. Tout d'abord, elle requiert un investissement intensif en temps et en ressources humaines et financières. Ensuite, elle peut parfois poser des problèmes en termes de fiabilité entre différents codeurs, voire entre le même codeur à deux points dans le temps. Ce désaccord peut porter non seulement sur le découpage du texte en phrases et en quasi-phrases mais également sur le ou les codes à attribuer à chacune de ces unités. Qui plus est, la définition des catégories utilisées pour cet encodage est relativement subjective, parfois imprécise, et pas toujours mutuellement exclusive.

La méthode automatique permet, quant à elle, un gain de temps considérable, surtout pour de longs documents et/ou pour un large ensemble de documents. Une fois les répertoires de mots et d'expressions établis, il ne reste plus au chercheur qu'à vérifier la qualité numérique du texte, à le transformer au bon format (.txt) et à extraire les résultats produits par le logiciel. Cette méthode permet aussi un gain d'argent puisqu'il n'est plus nécessaire d'engager de nombreux codeurs à chaque nouvelle élection. De

plus, tout comme la méthode CAP, d'autres types de sources politiques (questions/réponses parlementaires, accords de gouvernement, etc.) ou médiatiques (articles de presse, communiqués, etc.) peuvent être introduits dans le corpus et traités de la même manière. La méthode automatique réduit fortement toute erreur humaine et donne également la possibilité aux chercheurs de revenir à tout moment sur leurs codages, d'améliorer leurs répertoires en introduisant de nouveaux mots et expressions, de revoir leurs codes, etc. Enfin, si les répertoires sont disponibles dans différentes versions linguistiques (français et néerlandais, dans notre cas), il est tout à fait possible d'introduire des documents bilingues ou multilingues dans un même corpus – voire, de travailler sur un même document bilingue comme les déclarations gouvernementales belges, par exemple.

L'effet pervers de cette possibilité de recodage perpétuel est, toutefois, de devenir « codeur boulimique ». En effet, une question légitime qui a traversé notre recherche était de savoir à quel moment estimons-nous que les répertoires thématiques sont suffisamment stabilisés pour qu'ils puissent produire de « bons » résultats ? Dans notre cas, la transposition du *codebook* CAP a nécessité quelques 18 mois d'encodage et de phases de tests. Il a fallu, tout d'abord, transformer les programmes qui étaient soit scannés pour les plus anciens, soit en version .pdf – voire parfois protégés en écriture. Le nettoyage des textes (vérifier et corriger les accents, la ponctuation, l'orthographe, réécrire certains paragraphes ou textes protégés, etc.) a été une première étape essentielle dans cette démarche afin de rendre les documents « lisibles » par le logiciel. Ensuite, les répertoires thématiques étaient totalement vides, étant donné que nous partions d'une page blanche pour cette recherche (voir Lejeune, 2010, pour une présentation intéressante des familles de logiciels). D'innombrables « aller-retour » entre les textes et les répertoires ont donc été nécessaires. Enfin, il nous fallait compter par texte un nombre d'occurrences suffisamment important afin d'éviter que l'introduction de nouveaux mots ou expressions ne fasse basculer les résultats dans le sens d'une thématique ou d'une autre. Dans cette démarche, le fait qu'un seul chercheur ait pris cette tâche à son compte a été tant un inconvénient qu'un avantage. Le temps a certainement été le plus important inconvénient tandis que l'avantage reposait certainement sur le fait que le chercheur a pu conserver un regard transversal sur l'encodage de toutes les thématiques et sous-thématiques.

Au demeurant, ces deux méthodes sont différentes, mais parfaitement complémentaires. Elles se partagent l'étude des mêmes enjeux politiques et des mêmes thématiques, mais elles ont leurs propres avantages, leurs propres contraintes, et, surtout, leurs propres logiques. Toutefois – et c'est de loin l'essentiel –, elles produisent des résultats équivalents à bien des égards et confirment des tendances similaires dans l'étude des attentions politiques et des évolutions des enjeux mis en exergue par les partis politiques. Les enseignements que l'on peut tirer de cet exercice concernent de futures analyses des programmes électoraux, voire d'autres documents politiques (accords de gouvernements, débats parlementaires, agendas médiatiques, etc.). Étant donné que ces deux méthodes produisent des résultats similaires, le chercheur peut indifféremment choisir la méthode qui correspond le mieux aux intérêts de sa recherche et à ses ressources humaines, financières, etc. Qui plus est, ces deux méthodes parviennent à analyser des documents rédigés dans des langues différentes, ouvrant la porte à une application de ces méthodes dans une perspective comparée internationale.

Références

BAUMGARTNER F. R., JONES B. D., 1993, *Agendas and Instability in American Politics*, Chicago, University of Chicago Press.

BOUTECA N., 2009, « Ideological convergence on the socioeconomic dimension. A position-based analysis of the Flemish liberal and social democratic party between 1978 and 2007 », *Dutch-Flemish Politicologenetmaal*, Nijmegen.

BROUARD S., CHAQUES L., GREEN-PEDERSEN C., TIMMERMANS A., 2009, « Comparer les agendas politiques : une nouvelle voie pour la politique comparée », *Revue internationale de politique comparée*, vol. 16, n°3, p. 361-364.

BUDGE I., FARLIE D., 1983, *Explaining and Predicting Elections: Issue Effects and Party Strategies in Twenty-Three Democracies*, Londres, George Allen & Unwi.

BUDGE I., ROBERTSON D., HEARL D. éd., 1987, *Ideology, Strategy and Party Change: Spatial Analyses of Post-war Election Programmes in 19 Democracies*, Cambridge, Cambridge University Press.

BUDGE I., 2001, « Validating the Manifesto Research Group approach. Theoretical assumptions and empirical confirmations », M. LAVER éd., *Estimating the Policy Positions of Political Actors*, Oxford/New York, Routledge, p. 50-65.

BUDGE I., KLINGEMAN H.-D., VOLKENS A., BARA J., TANENBAUM E. éd., 2001, *Mapping Policy Preferences. Estimates for Partis, Electors and Governments*, Oxford, Oxford University Press.

CARMINES E. G., STIMSON J. A., 1989, *Issue Evolution. Race and the Transformation of American Politics*, Princeton, Princeton University Press.

CHATEAURAYNAUD F., TÖRNY D., 1999, *Les sombres précurseurs. Une sociologie pragmatique de l'alerte et du risque*, Paris, Éditions de l'EHESS.

CHATEAURAYNAUD F., 2003, *Prospéro. Une technologie littéraire pour les sciences humaines*, Paris, CNRS Editions.

CHATEAURAYNAUD F., 2011, *Argumenter dans un champ de forces. Essai de balistique sociologique*, Paris, Éditions Pétra.

DANDOY R., DE DECKER N., « Peut-on encore parler de partis frères en Belgique ? », J.-B. PILET, J.-M. DE WAELE et S. JAUMAIN éd., 2009, *L'absence de partis nationaux : menace ou opportunité ?*, Bruxelles, Editions de l'Université de Bruxelles, p. 19-35.

GREEN-PEDERSEN C., MORTENSEN P. B., 2009, « Issue Competition and Election Campaigns : Avoidance and Engagement », *Potsdam, ECPR Conference*, consulté le 15 août 2014, [en ligne] <http://www.agendasetting.dk/papers/uploaded/115201044905PM.pdf>.

GREEN-PEDERSEN C., WALGRAVE S., 2014, *Agenda Setting, Policies, and Political Systems. A Comparative Approach*, Chicago, University of Chicago Press.

HEARL D. J., 1987, « Belgium, 1946-1981 », I. BUDGE, D. ROBERTSON et D. HEARL éd., *Ideology, Strategy and Party Change: Spatial Analyses of Post-war Election Programmes in 19 Democracies*, Cambridge, Cambridge University Press, p. 230-253.

JONES B. D., BAUMGARTNER F. R., 2005, *The politics of attention: How government prioritizes problems*, Chicago, University of Chicago Press.

KLINGEMAN H.-D., HOFFERBERT R. I., BUDGE I. éd., 1994, *Parties, Policies and Democracy*, Boulder, Westview Press.

LAVER M., BUDGE I., 1992, *Party Policy and Government Coalitions*, Londres, Macmillan.

LEJEUNE C., 2010, « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches Qualitatives*, Hors Série, n°9, p. 15-32.

MCDONALD M. D., BUDGE I., 2005, *Elections, parties, democracy: conferring the median mandate*, Oxford, Oxford University Press.

PIET G., 2013, « Derrière les mots du politique. *Le-pire-est-à-venir*, un marqueur temporel particulier dans le discours politique en Belgique », *Mots. Les langages du politiques*, n°103, 2013, p. 121-136.

PIET G., DANDOY R., JOLY J., 2014, « PS, MR, cdH et Ecolo : analyse comparée des priorités politiques des quatre partis politiques traditionnels », *Blog Elections belges 2014*, consulté le 15 août 2014, [en ligne] <http://electionsbelges2014.blogspot.be/2014/05/ps-mr-cdh-et-ecolo-analyse-comparee-des.html>.

ROBERTSON D., 1976, *A Theory of Party Competition*, Londres/New York, Wiley.

SOROKA S. N., 2002, *Agenda-setting Dynamics in Canada*, Vancouver, UBC Press.

WALGRAVE S., DE WINTER L., NUYTEMANS M. éd., 2005, *Politieke-agendasetting in België – Mise à l'agenda politique en Belgique (1991-2000)*, Gand, Academia Press.

WALGRAVE S., DE SWERT K., 2007, « Where does Issue Ownership come from ? From the Party or from the Media ? Issue Identifications in Belgium, 1991-2005 », *The Harvard International Journal of Press/Politics*, vol. 12, n°1, p. 37-67.

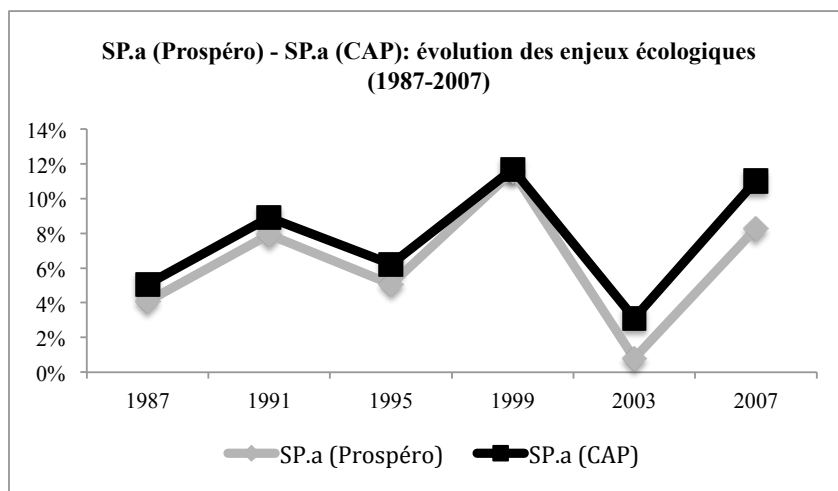
WILKERSON J., BAUMGARTNER F. R., BROUARD S., CHAQUES L., GREEN-PEDERSEN C., GROSSMAN E., JONES B. D., TIMMERMANS A., WALGRAVE S., 2003, « Le Projet Agendas Comparés : objectifs et contenus », *Revue internationale de politique comparée*, vol. 2, n°10, p. 365-379.

Annexes

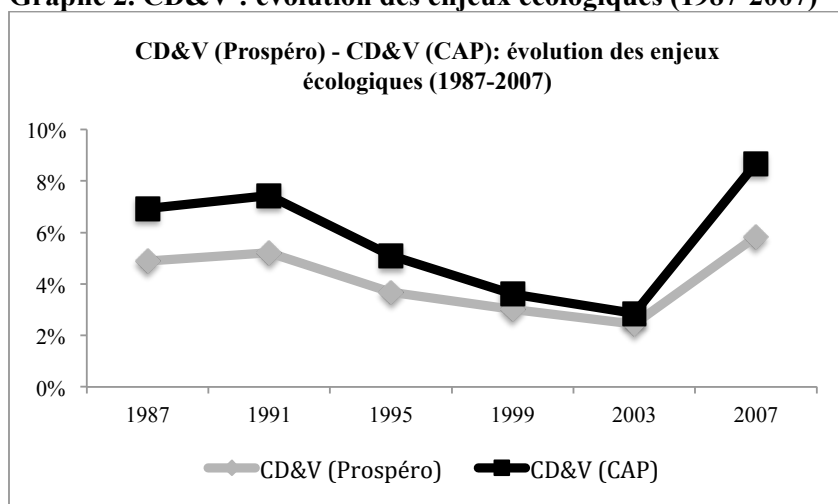
Tableau 1. Exemple : le programme électoral du PRL (1987)

	Prospéro (N= 2201)	CAP (N = 2344)	Différence	Rang Prospéro	Rang CAP
Macroéconomie	12,31%	12,40%	0,09%	2	2
Droits et libertés	5,27%	4,32%	-0,95%	6	6
Santé	2,64%	4,19%	1,55%	11	7
Agriculture	1,50%	2,38%	0,88%	14	16
Travail	6,54%	3,57%	-2,97%	5	11
Enseignement	5,04%	6,84%	1,80%	7	3
Environnement	1,14%	0,35%	-0,79%	19	21
Energie	0,45%	1,32%	0,87%	20	19
Immigration	1,27%	2,87%	1,60%	18	14
Mobilité	2,45%	3,00%	0,55%	12	12
Justice	4,95%	5,47%	0,52%	8	5
Affaires sociales	3,23%	3,00%	-0,23%	10	13
Développement local	1,41%	1,90%	0,49%	16	17
Entreprises	8,91%	6,40%	-2,51%	3	4
Défense	1,50%	4,06%	2,56%	15	9
Recherche scientifique	1,32%	1,77%	0,45%	17	18
Commerce extérieur	2,32%	2,87%	0,55%	13	15
Affaires étrangères	7,50%	4,15%	-3,35%	4	8
Administration publique	26,31%	24,89%	-1,42%	1	1
Territoire	0,18%	0,53%	0,35%	21	20
Culture	3,77%	3,71%	-0,06%	9	10
Total	100,00%	100,00%			
Similarités des attentions politiques					87,75%
Corrélation des attentions politiques					.97
Ecart des rangs					2,19
Corrélation de l'écart des rangs					.89

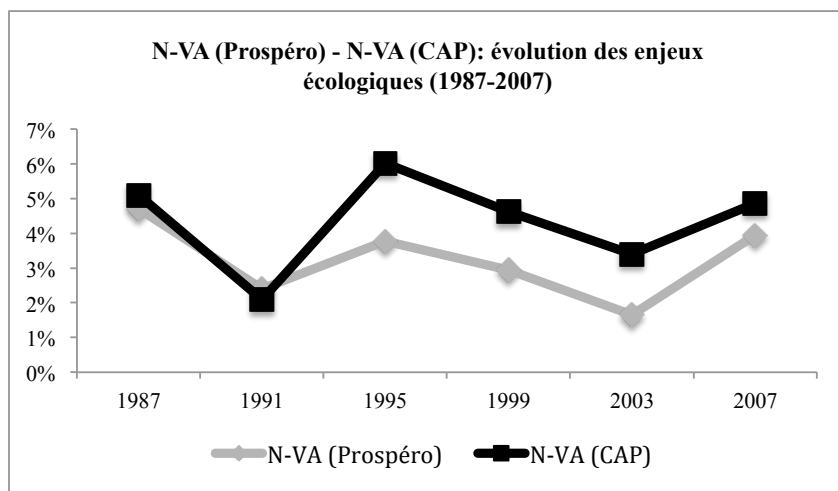
Graphe 1. SP.a : évolution des enjeux écologiques (1987-2007)



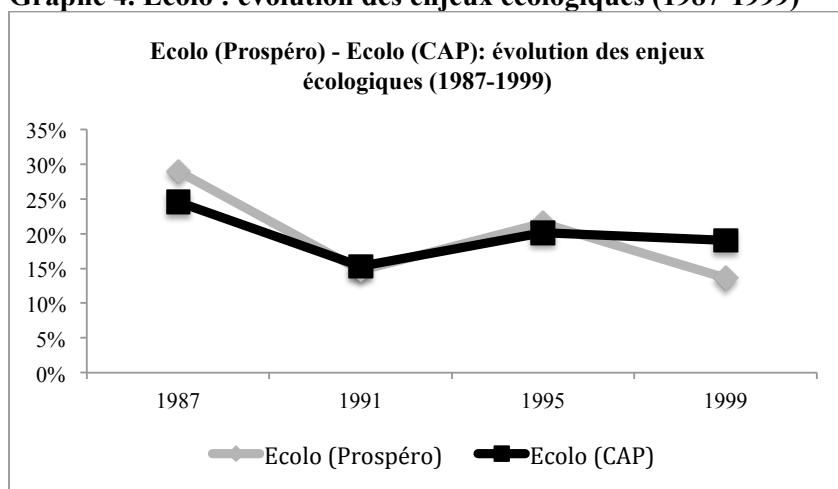
Graphe 2. CD&V : évolution des enjeux écologiques (1987-2007)



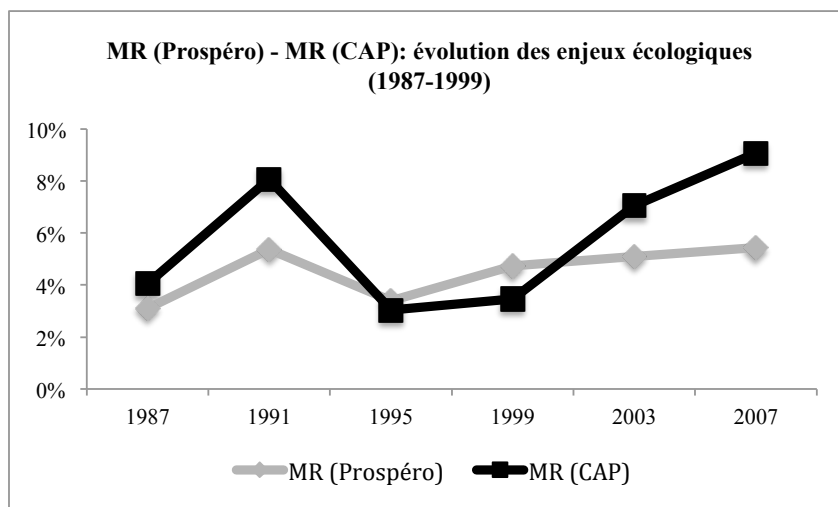
Graphe 3. N-VA : évolution des enjeux écologiques (1987-2007)



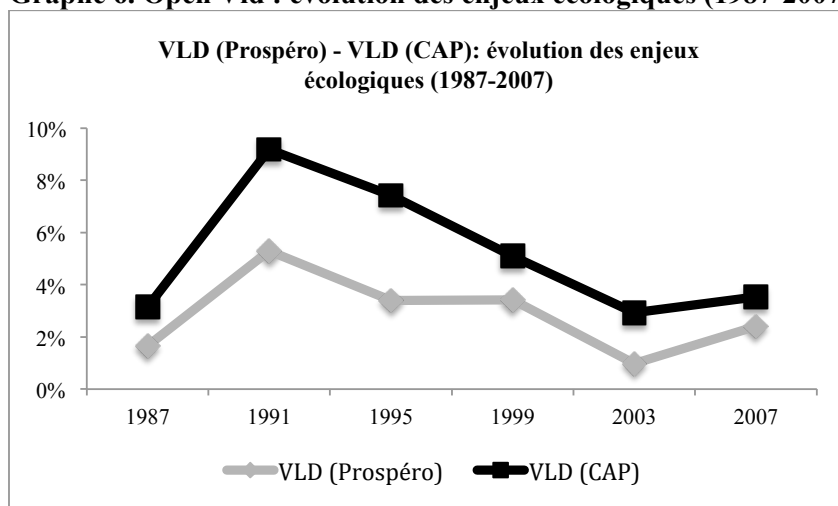
Graphe 4. Ecolo : évolution des enjeux écologiques (1987-1999)



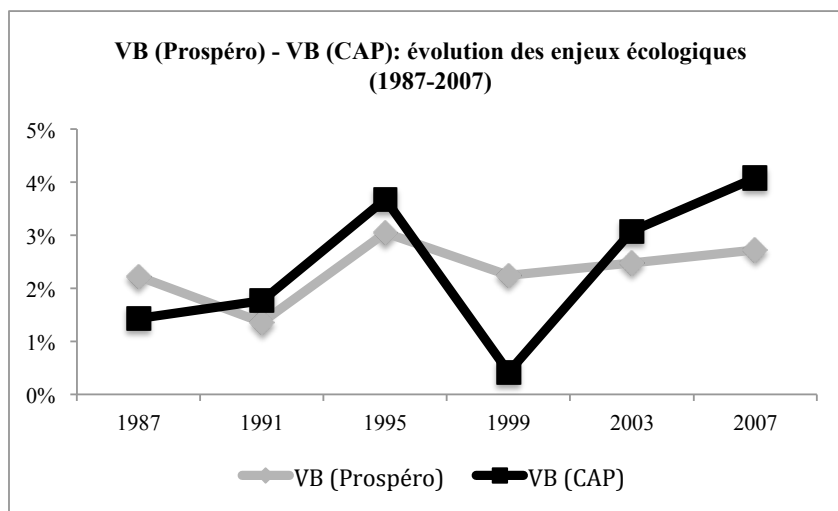
Graphe 5. MR : évolution des enjeux écologiques (1987-2007)



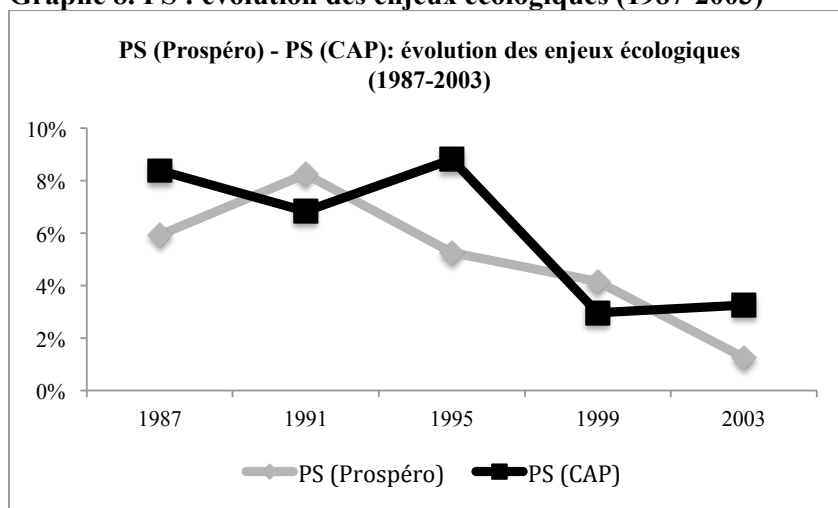
Graphe 6. Open Vld : évolution des enjeux écologiques (1987-2007)



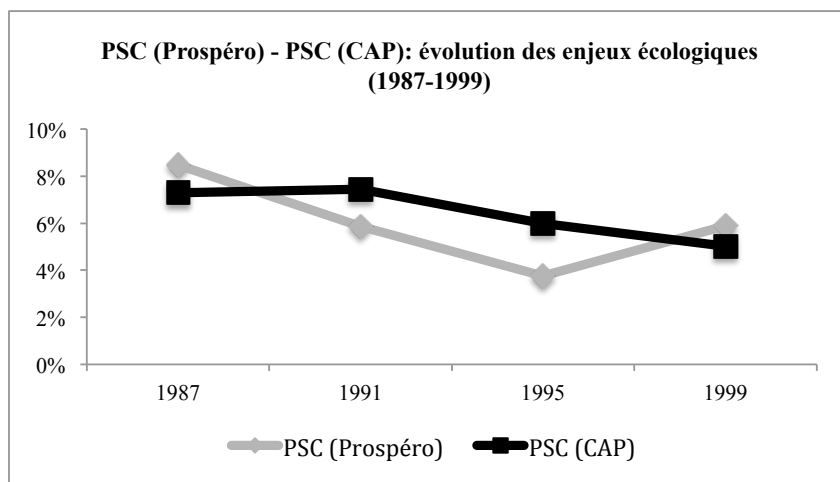
Graphe 7. VB : évolution des enjeux écologiques (1987-2007)



Graphe 8. PS : évolution des enjeux écologiques (1987-2003)



Graphe 9. PSC (cdH) : évolution des enjeux écologiques (1987-1999)



Graphe 10. Groen! : évolution des enjeux écologiques (1987-2007)

