

Méthodologie pour l'analyse comparée avec *Lexico3* et résultats : l'étude des pronoms chez les verts français et italiens

Béatrice Fracchiolla

Université de Toulon, UFR de lettres et sciences humaines, laboratoire Babel

Laboratoire SYLED, Université de Paris 3 Sorbonne Nouvelle

bearfrac@yahoo.com

Résumé

Nous présentons ici dans un premier temps les difficultés rencontrées lors de comparaisons inter-idiomatiques entre deux corpus oraux en français et en italien (entretiens de militants écologistes) dans le cadre de la statistique textuelle, et la méthodologie suivie pour résoudre certaines de ces difficultés, ainsi que certains des résultats obtenus en suivant cette méthodologie, dans un deuxième temps. L'observation part de la différence des emplois pronominaux dans les deux langues tout en montrant comment il est malgré tout possible d'aboutir à des données comparables, en particulier concernant la répartition entre discours des 'élus' et des 'non élus'.

Abstract

This is first a presentation of the difficulties met when using statistical text analysis in comparing two oral corpora of interviews with ecologists activists in two different languages (French and Italian). We then present the methodology followed to overcome these difficulties. We start with observing the difference in the use of pronouns in both languages, to then show how we nevertheless reach comparable data, especially regarding the division between 'elected' and 'non elected' speakers.

Mots clés : Logiciels d'analyse textuels (Lexico3) - Analyse de réponses à des questions ouvertes - Traitement automatique du langage naturel (corpus bilingue) - Analyse de discours

1. Introduction

L'originalité de l'étude présentée réside dans l'utilisation de *Lexico3* pour l'analyse de corpus **oraux, dans deux langues distinctes**. Ces corpus ont été recueillis et utilisés dans le cadre de notre travail de thèse : la question posée à l'origine de la thèse était d'étudier les figures de l'altérité dans le discours des Verts français et italiens. Une analyse de leurs discours met en évidence que les écologistes passent insensiblement d'un discours sur les valeurs à une revendication de droits. Le discours écologique se trouve pris en étau entre une universalité des intentions et une singularité des contextes et pose à sa manière la question de savoir si les droits de l'homme associés à une écologie politique peuvent ou non fonder une éthique de l'altérité. C'est à ce propos que l'on a observé une répartition particulière de l'emploi des pronoms, et en particulier du « on », du « nous », et du « je » selon deux types de discours, et observé que cette répartition recoupe en partie celle entre 'élus' et 'non élus'. L'observation d'éléments comparables dans le corpus italien, permet de voir qu'il existe en effet, plutôt qu'une répartition marquée entre 'élus' et 'non-élus', une répartition naturelle entre deux ensembles de locuteurs qui ont, d'un côté un discours 'impliqué' et de l'autre, un discours 'distancé' ou encore, 'critique'. Partant d'entretiens oraux, nous avons utilisé deux logiciels

différents de dictée vocale pour la retranscription écrite des entretiens afin de pouvoir les soumettre à l'exploration de *Lexico3*¹.

2. Les Corpus

Le corpus total est composé de 44 entretiens **oraux** réunis auprès de militants écologistes Verts français et italiens dont 24 entretiens sont en français et 20 en italien. La période de recueil des données pour l'ensemble des entretiens s'est échelonnée sur une année entière, de juillet 2000 à août 2001². La partition naturelle du corpus en 24 et 20 locuteurs – qui constituent autant de parties – a été complétée par l'étude d'autres types de partitions, pour distinguer par exemple ici les discours des 'élus' de celui des 'non élus'.

Corpus Total	Français	Italien
Nombre d'occurrences	162388	116300
Nombre de formes	8868	9945
Nombre d'hapax	4080	5100
Fréquence maximale	4787 (<i>est</i>)	3939 (<i>che</i>)

Tableau n°1 : Les corpus en chiffres

Le tableau ci-dessus donne diverses informations concernant :

le nombre total d'unités comprises entre deux délimiteurs (ou espaces blancs) comptabilisées dites *occurrences*,

le nombre total de *formes* distinctes comptabilisées parmi ces occurrences,

le nombre total d'*hapax*, soit de formes n'apparaissant qu'une seule fois, quelles que soient ces formes et enfin

la fréquence *maximale* où apparaît la forme comptabilisée le plus grand nombre de fois pour chaque corpus.

Cette différence est due essentiellement au nombre supérieur de flexions différentes en italien ; en particulier : marque du masculin/féminin par variations régulières tant pour les substantifs que pour les adjectifs, alors que le français compte un grand nombre d'épicènes, avec l'unique variation du déterminant comme marqueur de genre, tant au singulier qu'au pluriel (ex : le politique/la politique/les politiques *vs il politico, la politica, i politici, le politiche* ; ou encore : riches/riche *vs ricchi, ricche, ricco, ricca*). On remarque également que la variation syntaxico-morphologique de l'emploi des pronoms personnels compléments augmente le nombre de formes par rapport au français. Ainsi par exemple dira-t-on en français : « dites-le moi » *vs* « ditemelo » ou « ditelo a me » ou « me lo dite (o no) ». Le nombre d'hapax comme le nombre supérieur des formes présentes en italien relativement au nombre d'occurrences n'est donc pas surprenant.

¹ C'est ainsi la version française du logiciel *Dragon Naturally Speaking* pour PC qui a été utilisée pour la retranscription des entretiens français, et la version italienne du logiciel *Via Voice* d'IBM pour PC pour la retranscription des entretiens italiens. Voir le tableau récapitulatif de la méthodologie suivie pour le travail préparatoire aux explorations présenté en annexe.

2.1. Les difficultés posées par une analyse de discours en deux langues distinctes

Les comparaisons inter idiomatiques dans le cadre de la statistique textuelle ne sont pas sans poser de nombreux problèmes. On veut présenter rapidement ici certains d'entre eux, et comment, ensuite, on parvient à contourner les difficultés. Par exemple :

Partie	Nb occurrences	Nb formes	Nb hapax	Fréq. Max	Forme
<i>Andrea</i> (italien)	4662	1278	763	156	che
<i>Alexandre</i> (français)	4907	1031	586	154	je
<i>Alberto</i> (italien)	9197	1925	1064	303	che
<i>Barbara</i> (français)	9309	1510	819	293	est
<i>Aurélia</i> (français)	9323	1440	803	315	je

Tableau n°2 : Comparaison statistique des corpus par locuteurs

Les entretiens de *Barbara* et *Aurélia*, qui sont d'une longueur comparable, présentent plus d'occurrences, moins d'hapax et moins de formes que l'entretien d'*Alberto*. De même, l'entretien d'*Alexandre* compte plus d'occurrences et moins de formes différentes et d'hapax que celui avec *Andrea*. Le fait que les entretiens soient de longueurs différentes et que le corpus français considéré compte quatre entretiens supplémentaires semble donc n'avoir aucune incidence du point de vue de la comparaison statistique, dans la mesure où, comme présenté ci-dessus, les deux corpus sont globalement analysés selon des partitions (par locuteur, âge, sexe en particulier) pour lesquelles la longueur du corpus n'est pas un critère discriminant.

2.2. Propositions méthodologiques pour dépasser les difficultés

Comme le montre le tableau 1, on trouve en fréquence de rang 1 dans le corpus français la forme « est » (4787 occurrences), c'est-à-dire une forme verbale servant à la fois à décrire et définir et posant donc les éléments dans leur état. La forme de fréquence maximale est « *che* » pour l'italien (3940 occurrences). Or, en italien « *che* » représente de manière indifférenciée trois éléments grammaticaux qui sont distincts en français : « qui » et « que » relatifs et « que » introducteur de complétive. On constate que si l'on additionne dans le corpus français la somme des « qui » (2365 occurrences, en fréquence de rang 13) et des « que » — relatifs et complétifs confondus sous la même forme homonyme en français — (3424 occurrences en fréquence de rang 4) on trouve le nombre 5789. Il est alors possible de dire que les deux corpus mettent en évidence en position de fréquence maximale des éléments de nature et fonction grammaticale *équivalente* — et non identique. C'est-à-dire que si l'on veut comparer des corpus constitués dans deux idiomes distincts, l'exemple de « *che* » fait apparaître qu'il est nécessaire de se situer sur un autre plan que le plan grammatical, ou même que sur le plan d'équivalences lexicales. **Il s'agit plutôt d'envisager une comparaison dans une perspective idiomatique, qui prenne en considération la perspective de la traduction des représentations dans l'une et l'autre langue**³. C'est pourquoi nous avons d'abord défini les

³ Dans cette perspective, on constate par exemple que les rangs de fréquence se rapprochent dans les deux corpus puisque le corpus italien comptabilise respectivement dans les quatre premiers rangs de fréquence, décroissante : *che*, *di*, *non*, *è*, là où le corpus français donne *est*, *de*, *c*, *que*. Mais si l'on compte : [*che* → [*qui* + *que*] où

formes recouvrant dans chacune des langues des réalités équivalentes, en les comptant pour les comparer statistiquement par rapport à ce qu'elles représentent dans l'une et l'autre langue et l'une par rapport à l'autre à l'aide de la fonction 'groupe de formes' de Lexico3⁴.

Nous proposons à la suite un exemple des résultats obtenus dans le cadre de l'analyse de discours politique, et d'une comparaison entre les partitions 'élus' / 'non élus' suivant cette méthodologie.

3. L'analyse du corpus : une carte des pronoms dans les deux corpus

Une analyse des spécificités des entretiens selon la partition *élus/non élus* révèle un positionnement spontané — qu'il soit ou non délibéré — différent selon les personnes en fonction de leur rôle et de leur place au sein du parti, en tant que simple militant ou en tant qu'élu. On remarque que les formes déictiques l'emportent (de loin) sur toutes les autres formes de pronoms. Or, dans les entretiens d'enquête avec les militants, chaque militant dispose pour se définir de trois possibilités qui correspondent à trois types d'attitude face aux autres personnes avec lesquelles il/elle milite, exprimées de trois manières différentes dans l'énonciation :

- Soit en opposition ;
- Soit en relation de proximité ;
- Soit par l'indifférence.

Chacun de ces types de relation se définit selon un certain schéma construit en fonction de chaque possibilité offerte au *je* pour se situer relativement aux autres pronoms (en exclusion ou en inclusion) sur le plan de l'énonciation. Nous avons donc constitué des 'groupes de formes'⁵ afin d'étudier ensemble toutes les déclinaisons morphologiques possibles d'un même pronom. Ainsi, le groupe « nous+ » correspond au regroupement de l'ensemble des formes de la première personne du pluriel (sans que l'on puisse différencier, d'ailleurs, au moyen des fonctionnalités de Lexico3, la forme « nous » pronom sujet de la forme « nous » pronom objet. Nous avons ensuite lancé une recherche de concordances sur ces groupes de formes dans les deux corpus. Les résultats obtenus sont les suivants :

Dans le corpus italien :

- « *voi+* » (*voi, vostro, vostra*) : 15 contextes
- « *loro+* » (*lui, lei, loro, suo, sua, suoi*) : 383 contextes
- « *tu+* » (*ti, te, tu, tuoi*) : 416 contextes

que=(relatif®+complétif©)], alors les rangs de fréquence du corpus français deviendraient *qui+que* (*relatif+complétif*), *est, de, c*. Dans la perspective d'une comparaison des spécificités des deux corpus, on peut ainsi noter :

« Equivalences »	Rang 1 (français. max.)	Rang 2	Rang 3	Rang 4
Corpus Italien	<i>che</i> : 3940	<i>di</i> : 3471	<i>non</i> : 2479	<i>è</i> : 2460
Corpus français	<i>qui+que®</i> et © : (5789)	<i>est</i> : 4787	<i>de</i> : 3663	<i>c</i> : 3797

⁴ Cette fonction permet en effet d'établir des listes de mots par lesquelles se trouvent regroupés divers lexèmes qui se réfèrent à un même champ sémantique, et/ou lexical, ce qu'il semblait nécessaire de faire : « Le sens en contexte est réfractaire à toute systématité. On ne peut, au départ d'une étude lexicométrique, que cumuler, confronter, réunir ou opposer des occurrences de formes textuelles. [...] Un texte étant un tout solidaire en système d'oppositions avec d'autres, l'emploi d'un terme ne fonctionne pas isolément, mais, à l'intérieur d'un vocabulaire, dans un univers de rapports aux autres emplois et aux autres termes. » (Bonnafous, S., Tournier, M. 1995. Analyse du discours, lexicométrie, communication et politique. in *Langages*, n° 117 : 69-70.)

⁵ À l'aide de l'outil du même nom, présent dans Lexico3.

« *noi+* » (*noi, nostro, nostri, nostra*) : 429 contextes
 « *io+* » (*io, mio, me mi, m', miei, mia*) : 2140 contextes

Dans le corpus français :

« *je+* » (*je, moi, me, ma, mon, mes, m'*) : 5214 contextes
 « *on+* » (*on*) : 3253 contextes
 « *tu+* » (*tu, t', toi, te, ta, ton, tes, tien, tiens, tienne, tiennes*) : 1333 contextes
 « *vous+* » (*vous, votre, vos*) : 120 contextes
 « *nous +* » (*nous, notre, nos*) : 505 contextes
 « *il+* » (*il, elle, ils, elles, eux, leur, leurs, son, sa, ses*) : 4037 contextes

Ces résultats offrent une vision d'ensemble des emplois des pronoms dans les deux corpus, et font apparaître la différence de langue entre le français et l'italien quant à l'emploi des pronoms. On notera, au vu de ces résultats, que la langue italienne emploie peu les pronoms personnels comparée à la langue française⁶, ainsi que la valeur représentative du *on* par rapport au *nous* et au *je* dans le corpus français, distinction qui n'existe pas dans le corpus italien dans la mesure où la langue italienne ne dispose que d'une forme unique : « *noi* »⁷.

3.1. Les pronoms personnels et la définition de deux discours distincts : le corpus français

L'analyse des entretiens avec les militants des Verts et des *Verdi* conduit à identifier dans leur discours deux domaines de références distincts que sont d'une part *je*⁸ et d'autre part « les autres » sur lesquels *je* tient un discours. De même que *je* peut aussi parfois tenir un discours sur lui-même ou bien sur la somme de *je+eux* correspondant à une nouvelle unité-entité *on* ou *nous*. En tant que 'militant qui adhère', il s'agglutine alors à autrui pour former un ensemble, le parti, où chacun fait donc partie de, c'est-à-dire construit, constitue, définit le parti les Verts, individuellement et collectivement. Il se trouve dès lors aussi en mesure de tenir un discours sur *on* et *nous* soit en s'y associant par une attitude inclusive : il tient alors un discours sur *je*= [*je+eux = on/nous*] ou bien en s'en dissociant : il se désolidarise alors par l'intermédiaire d'un discours exclusif du *je*, où le *je* devient observateur extérieur, distancié : *je* ≠ [*je+eux = on/nous*], mais où *je* continue de faire partie du *on/nous*.

L'identité est de fait divisée et la notion d'altérité intervient alors, soit que le *je* se considère autre par rapport au parti, soit qu'il considère le parti autre par rapport au *je*. Ainsi, on retrouve la dialectique de l'*ipse* et du différent propre à l'altérité, soit par le **discours d'adhésion**, soit par le **discours qui se désolidarise**. On dira en termes de positionnement

⁶ Une rapide recherche des formes des pronoms enclitiques verbaux en italien à l'aide de la fonction « groupe de formes » et l'élimination manuelle de toutes les formes non pertinentes, montre en effet qu'ils sont utilisés en proportion assez négligeable, même s'ils sont présents avec 175 occurrences pour les formes combinées Vb-li, Vb-mi, Vb-la, Vb-lo. Après avoir effectué une recherche de formes nous n'avons pas inclus ici la forme Vb-ti car sa forme est en italien épïcène du participe passé (*dichiarati* : déclare-toi/ *si sono dichiarati* : ils se sont déclarés) et elle apparaît très peu comparée au participe, en particulier dans un contexte d'entretien semi dirigé.

⁷ Les valeurs du *on* français sont exprimées en italien de diverses manières. Elles sont en particulier présentes dans la formulation « si + 3e personne du singulier » : « si può dire che... » : on peut dire que ou « si +3e personne du pluriel » : si possono trovare molte idee interessanti in questo libro » : on peut trouver beaucoup d'idées intéressantes dans ce livre ; le verbe à la 2e personne du singulier ou à la 3e personne du pluriel peut parfois aussi se traduire par un *on*.

⁸ Le *je* étant toujours en énonciation décrit comme étant à la fois le centre et le porteur du discours, qui organise le réel en fonction de sa parole et impose à son ou ses interlocuteurs sa propre vision du monde. Voir entre autres Maingueneau D. (1999). *L'Énonciation en linguistique française*. Paris, Hachette et Kerbrat-Orecchioni C. (1999) *L'Énonciation*. Paris, Armand Colin.

énonciatif, que c'est là le propre du **discours critique**. Le discours critique est celui où *je* se désolidarise du groupe. Il est formulé par l'énonciateur par l'intermédiaire de phrases telles que « comme parti, on a ce type de discours tout le temps, mais moi je pense que... ». Le discours inclusif, que l'on peut appeler d'adhésion étant donné qu'il s'agit d'un discours politique, se caractérise au contraire par une présence filée du *on* ou du *nous* tout au long de l'entretien. On remarquera que ce type de discours est globalement plutôt caractéristique des personnes élues à des postes de responsabilité au sein du parti, et ayant l'habitude d'assumer une fonction de représentation. Une interlocutrice, Catherine, qui est députée européen en est l'exemple le plus représentatif dans le corpus français, car elle est la seule dont l'entretien présente pour forme de fréquence maximale le pronom personnel *on*.

Partie	Nb occurrences	Nb formes	Nb hapax	Fréq. max.	Forme
Catherine	10187	1611	872	291	on

3.2. L'expression du pronom personnel en italien

En italien en revanche, l'expression du pronom personnel se trouve normalement morphologiquement contenue pour tous les verbes dans la désinence verbale de chaque conjugaison. Par exemple, pour les verbes être (*essere*) et chanter (*cantare*) au présent de l'indicatif :

1 ^{ère}	son-o cant-o	(io) /je
2 ^e	se-i cant- i	(tu) /tu
3 ^e	è cant-a	(lui, lei) / lui, elle
1 ^{ère} pluriel	s-iamo cant-iamo	(noi) / nous
2 ^e pluriel	si-ete cant-ate	(voi) / vous
3 ^e pluriel	s-ono cant-ano	(loro) / ils, elles

Du fait de cette différence quant à l'expression des pronoms dans les deux langues, il est difficile d'effectuer exactement le même type d'analyse du corpus. Néanmoins, étant donné que *Lexico3* permet de mesurer les emplois des formes proportionnellement à l'ensemble du corpus, vérifier l'emploi des pronoms exprimés dans le corpus italien reste assez pertinent. On ne peut pas cependant en tirer de conclusions définitives du même ordre que celles déduites des résultats obtenus pour le corpus français du fait que sont ignorées, par le seul comptage des pronoms explicitement exprimés, toutes les formes verbales morphologiquement porteuses de la personne. Il est en effet tout aussi grammaticalement correct de dire en italien : « *io sono andato* » que « *sono andato* » alors que le français ne peut en aucun cas dire « *suis allé » en faisant l'économie du « je ». Plus que l'énonciation *stricto sensu*, c'est donc plus exactement l'expression spécifique des pronoms qui est ici étudiée.

3.3. La spécificité de « io » critique dans le corpus italien

On trouve ainsi 863 occurrences de « *io* » dans le corpus italien, auxquelles il faut rajouter les marques complément du pronom personnel de la première personne : « *mi, me, m'* ». Du fait même qu'elle ne soit pas normalement exprimée en italien, cette représentation explicite de la première personne du singulier est remarquable et relève sans aucun doute de la situation

particulière de l'entretien qui conduit l'énonciateur à accentuer la modalisation de son discours et à affirmer, plus qu'ailleurs, les objets du discours dans une temporalité qui réfère à l'« ego », centre de l'énonciation en même temps qu'il traduit son attitude, son opinion et sa position à l'égard de ce qu'il énonce et de son interlocuteur.

Il en est de même pour l'expression des autres pronoms, et donc en particulier ceux de la deuxième personne du singulier, dont on trouve 138 occurrences dans le corpus italien. Cet emploi récurrent de « *io* » équivaut à un discours d'individus, où la narration d'expériences personnelles et l'affirmation du locuteur en général est caractéristique et caractérisée par rapport aux « autres ». On comptabilise en revanche seulement 138 « *tu* » et 12 « *voi* », avec un pic de « *voi* » chez l'un des locuteurs, Luigi, ce qui peut s'interpréter comme une manière de se démarquer des *Verdi*, dont il a fait partie, mais dont il cherche à se désolidariser à présent, y compris dans son discours. Ce qui nous amène à parler ici aussi d'une forme de discours *critique*, mais 'à l'italienne'.

Le fait que les pronoms apparaissent dans le discours induit un effet d'insistance qui n'équivaut pas cependant à l'effet redondant et renforcé du « moi je » ou « toi tu » français, mais apparaît comme un phénomène plus naturel. On remarque par ailleurs que cette tendance au renforcement par l'expression du pronom est plus marquée en italien à l'oral qu'à l'écrit.

L'analyse des spécificités du corpus italien en fonction des élus/non élus, et du regroupement 'naturel' par locuteurs, permet de faire les mêmes remarques que celles faites sur le corpus français : plus qu'un discours spécifique d'élus et de non élus, on est en présence de deux discours distincts⁹. L'un, plutôt caractérisé par l'expression d'une réflexion personnelle fondée sur la subjectivité assumée du locuteur : voilà où *je* suis, voilà ce que *je* pense. Dans ce discours, l'adhésion au discours des *Verdi*, la cohésion avec le parti est signifiée par l'emploi de « *noi* » et de nombreux verbes conjugués à la première personne du pluriel. L'autre discours, d'après le lexique se résume plutôt par une description de comment les choses étaient (emploi du verbe « *essere* » à l'imparfait) et sont : « *c'è, tutto sommato, nel senso* » par rapport aux *Verdi* (ou aux *Giovani Verdi*), et de ce que l'on peut dire des *Verdi*. Lorsque le *je* s'affirme, il le fait alors plus pour se désolidariser que dans la perspective d'un discours d'adhésion. Le *je* énonciateur qui parle, bien que représentant une identité collective, tient cependant à conserver son individualité. Il refuse l'amalgame du « *nous* » et signifie plutôt « toi et moi ». C'est pourquoi nous nous sommes plus particulièrement intéressés, dans le corpus français, au *je* adhérent au *nous* ou *on* collectif, et au lien qui existe du *je* au collectif.

4. Conclusions

D'après la recherche effectuée par Pascal Marchand sur le discours des militants socialistes, plus le discours est engagé (le discours des élus en particulier), moins il y a de *je* et de verbe d'opinion¹⁰. Ces remarques se vérifient également dans nos résultats, puisque l'on trouve effectivement, statistiquement, dans le groupe des interlocuteurs français défini comme porteur du discours collectif, les cinq élus nationaux interrogés, soit quatre députés européens et un à l'Assemblée nationale. L'identité du discours Vert se construit sur les bases même d'une contradiction toujours possible et même recherchée. Il n'y a pas de charte de l'être Vert, à laquelle chaque personne désireuse d'adhérer devrait préalablement adhérer, mais une acceptation *a priori* d'autrui comme Vert dès lors qu'il se dit lui-même être Vert. Aussi,

⁹ Ce que permettent de visualiser l'analyse factorielle des correspondances élus/non élus et par locuteurs Italiens et français (à joindre en annexes).

¹⁰ Marchand, P. (1998). *L'Analyse du Discours Assisté par Ordinateur*. Paris, Armand Colin : 98-100.

l'identité Verte est en constante construction, redéfinition, parce que toujours nécessairement somme d'un *je* plus les *autres*, sans que le *je* ait jamais aucun pouvoir sur l'identité individuelle de « ces autres », qu'il ne peut appréhender que collectivement comme étant les « autres Verts », en cela semblable à lui. Chaque individu Vert choisit ainsi en disant *nous* et *on* à chaque énonciation, de réaffirmer son adhésion et son engagement dans le discours et se présente comme porteur de ce discours. Le *nous* et le *on* apparaissent essentiellement symboliques de l'engagement militant, car représentatifs de l'adhésion individuelle à l'ensemble du parti comme collectif. Ce constat est également patent dans le fait que chacune des personnes interrogées a un discours 'vert' auto-référent, qui ne nécessite jamais la citation ou le renvoi spécifique aux dires de quelqu'un d'autre ou d'une quelconque autorité «Verte » pour valider ses propres dires.

Références

- Bolasco, S. (1993). Choix de lemmatisation en vue de reconstructions syntagmatiques du texte par l'analyse des correspondances. In *Actes des 2ndes Journées Internationales d'Analyse statistique de Données Textuelles à Montpellier*, ENST, Paris.
- Bolasco, S., et al. (1995). *IIIe Giornate internazionali di analisi statistica dei dati testuali*, Rome, CISU, II.
- Bolasco, S. (1996). Il lessico del discorso programmatico di governo. in Villone, M., Zuliani, A. (dir.), *L'attività dei governi della Repubblica italiana : (1948-1994)*, Bologne, Il Mulino, 1996.
- Bonnafous, S., Tournier, M. (1995). Analyse du discours, lexicométrie, communication et politique. In *Langages*, n°117 : 69-70.
- Boy, D., Roche, A., Villalba, B. (1999). *L'enquête sur les adhérents Verts*. Paris, Les Verts, Confédération Ecologiste – Parti Ecologiste, Secrétariat national, service documentation.
- Boy, D. (1991). Enquête auprès des écologistes : les Verts en politique, 235-250, in Duhamel, O., Jaffré, J. (dir.), *L'Etat de l'Opinion*, Paris, Le Seuil (Sofres).
- Fracchiolla, B., Kuncova, A., Maisondieu, A. (2003). *Manuel d'utilisation de Lexico3. Outils de statistique textuelle* (Lamalle, C., Martinez, W., Fleury, S., Salem, A.). Version 3.41, SYLED – CLA2T, Université de la Sorbonne Nouvelle – Paris 3.
- Kerbrat-Orecchioni, C. (1999). *L'Énonciation*, Paris, Armand Colin.
- Lafon, P. (1980). *Dépouillements et statistiques en lexicométrie*, Genève-Paris, Slatkine-Champion.
- Le Bart, C. (1998). *Le discours politique*. Paris, PUF, Que Sais-Je ?.
- Maingueneau, D. (1999). *L'Énonciation en linguistique française*, Paris, Hachette.
- Marchand, P. (1998). *L'Analyse du Discours Assisté par Ordinateur*, Paris, Armand Colin.
- Habert, B., Nazarenko, A., Salem, A. (1997). *Les linguistiques de corpus*. Paris, Armand Colin.
- Salem, A., Lebart, L. (1994). *Statistique textuelle*. Paris, Dunod.
- Salem, A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris, Klincksieck

ANNEXES

Annexe n°1. Principales caractéristiques (PCLP) : Locuteurs français

Partie	Nb occurrences	Nb formes	Nb hapax	Fréq. Max	Forme
Adeline	1487	383	212	52	est
Alexandre	4907	1031	586	154	je
Aude	11161	1612	873	325	de
Aurélia	9323	1440	803	315	je
Barbara	9309	1510	819	293	est
Catherine	10187	1611	872	291	on
Charles	3268	795	475	135	de
Christophe	7929	1352	730	345	de
Danièle	2855	706	402	84	de
Félix	4461	960	594	148	est
Geneviève	4872	828	445	219	est
Hélène	6934	1250	719	223	de
Jean	10275	1474	783	350	est
Josiane	5223	1018	569	153	est
Karine	3310	652	365	116	est
Marie	5224	1006	553	121	je
Marina	7006	995	520	240	que
Martin	3543	834	482	137	de
Patrick	10813	1627	852	339	est
Paula	10555	1600	884	278	de
Philippe	6773	1165	641	250	est
Pierre	7862	1257	666	243	est
Thierry	7679	1316	708	256	de
Vincent	7437	1459	838	232	est

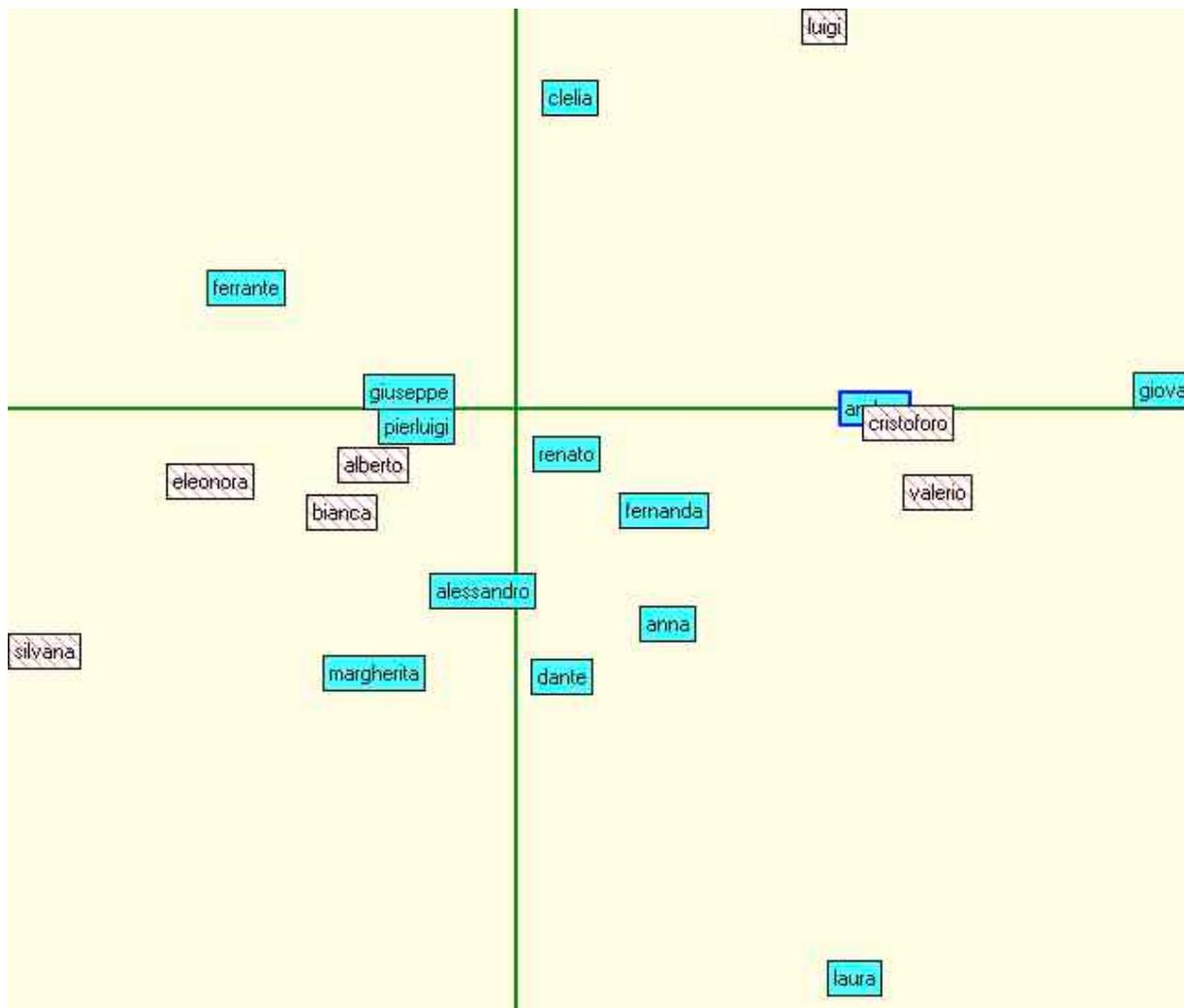
Annexe n°2. Principales caractéristiques (PCLP) : Locuteurs italiens

Partie	Nb occurrences	Nb formes	Nb hapax	Fréq. Max	Forme
Alberto	9197	1925	1064	303	che
Alessandro	5695	1559	972	217	che
Andrea	4662	1278	763	156	che
Anna	6046	1483	903	208	che
Bianca	3971	983	575	139	e
Clelia	9997	1857	1031	390	che
Cristoforo	4071	1110	650	125	è
Dante	3390	1082	683	148	di
Eleonora	6871	1363	774	311	che
Fernanda	4457	1202	756	151	che
Ferrante	16110	2613	1497	557	che

Giovanni	5435	1397	851	238	di
Giuseppe	2580	780	470	75	che
Laura	5908	1405	807	202	di
Luigi	8262	1634	880	222	di
Margherita	1412	517	315	50	che
Pierluigi	6115	1533	922	258	che
Renato	2485	799	468	101	di
Silvana	3960	1004	569	119	non
Valerio	5687	1382	806	199	di

Corpus : ITALIEN	Locuteurs 'élus' sélectionnés en surbrillance en vue du calcul des spécificités	Partition = locuteur
---------------------	--	-------------------------

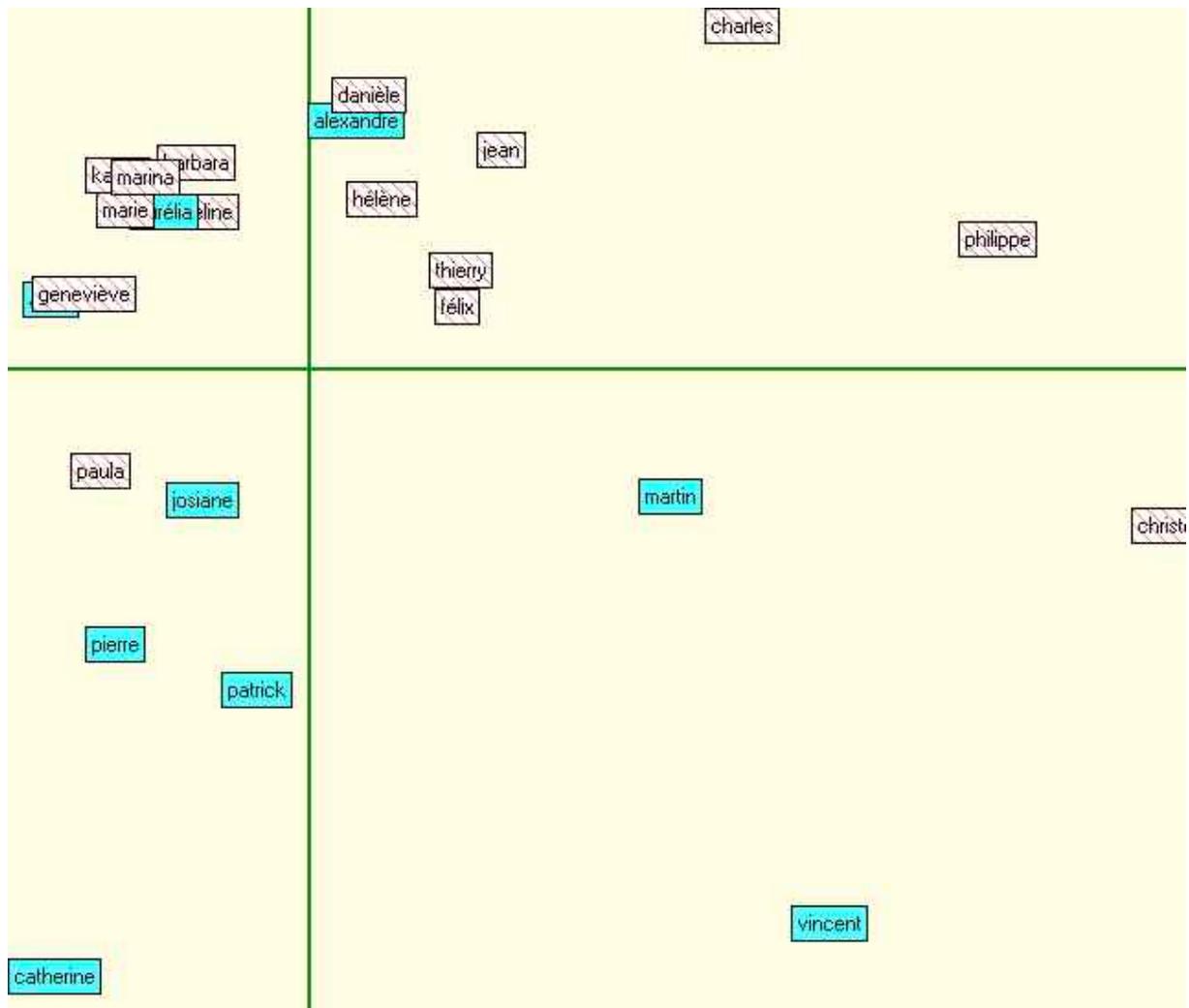
Nombre de parties : 20



Analyse factorielle des correspondances - locuteur

Corpus : FRANCAIS	Elus : en bleu / Non élus : barrés	Partition = locuteur
-------------------	------------------------------------	----------------------

Nombre de parties : 24



Analyse factorielle des correspondances – locuteur

