

**L'analyse de discours de la parole parlementaire dans une étude sociologique du travail de représentation des députés argentins 1998/2009. L'utilisation de logiciels d'analyse textuelle : Alceste et N-vivo.**

Américo MARIANI  
Doctorant en sociologie  
LISST-CERS (UMR 5193)  
ATER en sociologie  
Université ToulouseII-Le Mirail  
mariani@univ-tlse2.fr

L'objectif de cette communication est de présenter la mise en place d'un dispositif de recherche permettant de « faire parler » les débats parlementaires à partir de l'utilisation de logiciels de traitement de données.

Les éléments présentés ici sont issus d'une thèse en sociologie sur les pratiques de représentations des parlementaires nationaux argentins mobilisant trois ensembles de données empiriques : observations, entretiens et analyse de discours de sessions plénières à partir de retranscriptions sténographiées. C'est de ce dernier ensemble dont il sera strictement question ici.

Nous aborderons la constitution du corpus en deux étapes et le rôle des logiciels dans cette construction pour présenter ensuite l'intervention du logiciel N-Vivo dans l'analyse proprement dite. Toutefois, avant d'aller plus avant je présenterai rapidement la recherche et les présupposés théoriques.

**Présentation succincte de la recherche :**

L'intérêt pour le Parlement Argentin est né d'une interrogation générale sur la dite « crise de représentation » et sur les conséquences sur l'institution parlementaire de la crise du début du millénaire qui a atteint son climax en décembre 2001. Nous cherchons dans notre travail de thèse à saisir les évolutions dans la pratique des parlementaires notamment à travers la mise à jour des référentiels attachés à la pratique de représentation.

Après une série d'entretiens auprès des parlementaires et des observations de leurs activités à l'intérieur comme à l'extérieur du Congrès, il apparaissait nécessaire de mettre en place un dispositif permettant une saisie du corps parlementaire dans son ensemble. La question de la représentation nous semble devoir faire l'objet d'une approche collective, ne pouvant être réduite à l'activité individuelle des élus. L'élu devient législateur en apprenant un certain nombre de règles, d'attitudes, à la fois dans son groupe parlementaire mais aussi dans les travaux collectifs dont l'assemblée plénière est le moment clef, un phénomène bien illustré par M. Abélès [Abélès 2000]. Partant de ce constat il nous a semblé intéressant de mettre à jour les référentiels communs sur la démocratie qui, quelle que soit l'appartenance partisane, apparaissent dans le discours parlementaire.

Si la volonté de prendre en compte les productions textuelles existait depuis le départ, l'analyse de discours arrive dans la recherche comme une nécessité à un moment donné. Pour ne pas en rester à la petite phrase, aux effets d'annonces, une analyse de contenus risquer de rester en surface. Il fallait équiper la lecture et répondre à deux exigences principales:

- constituer un corpus adéquat. Un ensemble important de sessions ont été prélevé de manière aléatoire dans une logique d'échantillon afin d'une part de donner une place à l'ordinaire de la parole parlementaire et d'autre part de pouvoir rapporter les énoncés relevés à l'ensemble du texte.

- effectuer une lecture non linéaire du texte. Les parlementaires sont des orateurs, leur discours cherche à marquer, à interpeller, à convaincre. Il s'agissait d'échapper à une logique du sujet parlant et à la séduction de la phrase pour privilégier une approche en termes de position de sujet et d'organisation du discours.

### **Quelques précisions théoriques liminaires :**

Il convient de préciser quelques jalons de l'approche choisie pour ce travail d'analyse de discours.

Il ne s'agit évidemment pas de considérer qu'il y a un ailleurs, plus ou moins occulte où se logerait la vérité du discours, non pas que tout serait immédiatement perceptible dans une totale transparence, mais il ne s'agit pas de traquer la présence d'un autre discours « sous l'innocence de la parole (...) le discours de l'inconscient »<sup>1</sup>.

L'objectif est bien plus simple, comprendre les régularités du discours qui nous renseignent sur les visions du monde que partagent les parlementaires, ce qu'ils jugent bon et juste, ce qu'ils rejettent... Il s'agit d'adopter une démarche intégrative qui vise à articuler les diverses composantes du discours inscrit dans un espace de circulation des énoncés, un espace à la fois intertextuel et social, interne au texte donc, mais aussi entre le texte et l'institution dont il est partie prenante [Maingueneau, 1997]. « Il s'agit plutôt d'envisager comment, dans un contexte discursif donné, l'ensemble des énoncés produits tendent à s'inscrire à l'intérieur de certaines balises, à s'organiser en fonction de certaines configurations, à valider une certaine "grammaire". » [Armony, 2000] C'est à cette « grammaire » que nous nous intéressons. Écartant d'emblée la possibilité de construire une grammaire générale du discours parlementaire pris comme ensemble, nous cherchons plutôt à réaliser un découpage « topographique » centré autour de certains espaces du discours. La recherche s'intéresse aux référentiels démocratiques et à la mobilisation dans le discours du *représenté*<sup>2</sup>, il importait d'isoler des segments de textes pertinents de ce point de vue.

---

<sup>1</sup> Althusser Louis, Lire le Capital, tome I, Maspero, Paris. p. 12. [Maingueneau 1997: 21]

<sup>2</sup> Lorsque nous évoquons le *représenté*, il s'agit de l'électeur, du mandant, sous différentes dénominations qui recouvre des réalités différentes ; peuple, citoyen mais aussi la société, les

## **La construction du corpus :**

Le corpus est constitué en deux étapes :

- Une première étape consiste en un échantillonnage représentatif de la parole parlementaire en assemblée plénière.
- Une seconde étape correspond à un découpage raisonné de ce premier corpus pour constituer un texte second qui sera la cible de l'analyse. L'exigence qui préside à l'élaboration de ce second texte n'est plus la représentativité, nous ne sommes plus dans une logique d'échantillon mais de classification et découpage.

### **1ère étape :**

Les versions sténographiées des réunions plénières parlementaires sont disponibles en libre accès sur le site de la chambre des députés<sup>3</sup> depuis la période parlementaire 1998-1999. Les documents sont mis en ligne et vérifiés par « le service sténographique » (Dirección de taquígrafo) de la chambre des députés. C'est donc la production, vérifiée et certifiée de l'institution (à travers une procédure où les députés ont eux-mêmes la possibilité d'introduire des corrections) qui servira de base à la constitution du corpus. La disponibilité de captures vidéo des sessions (depuis 2004 uniquement) a permis une vérification de la fidélité des comptes-rendus sur un échantillon. Ceux-ci sont réalisés mot à mot et comportent également des indications sur les mouvements de l'assemblée et du public, tels que les applaudissements et félicitations, les prises de paroles sauvages (non consignées mais notées comme « plusieurs députés parlent en même temps »), les manifestations dans les balcons du public. Nous les nommerons « didascalies ».

Le corpus est construit sur une sélection de 50% des sessions par période parlementaire, panachées de façon à refléter la composition en types de séance de la période (minoritaire, d'information, ordinaire, etc.). Chaque séance est ensuite *réduite* aux prises de parole.

En sont exclues les lectures de texte par le secrétariat, les sommaires et ordres du jour, les « didascalies » sont maintenues dans le corps du texte mais réduites sous forme d'abréviation pour être exclues du traitement. Bien évidemment les indications d'identité des locuteurs sont maintenues afin de pouvoir le cas échéant rapporter l'énoncé à une personne physique.

Pour finir, les sessions sont réunies par période parlementaire dans un document unique. Dans un cas, lorsque la majorité change au cours de l'année parlementaire, une partition est effectuée au moment de l'intégration des nouveaux députés. De cette façon il est possible de regrouper les périodes parlementaires comprises entre deux élections législatives et de constituer des ensemble homogènes en termes de composition de la chambre.

Ce corpus en l'état ne peut constituer un ensemble analysable. Mis à part son volume considérable (quelque 9600 pages et plus de 4 000 000 de mots) il s'agit d'un ensemble hétérogène tant du point de vue des énonciateurs que du point de vue des énoncés. Il est alors nécessaire de classer et ordonner ce texte afin de construire un ensemble exploitable pour l'analyse.

---

victimes. Il ne nous sera pas possible de nous étendre ici sur la question du choix des dénominations.

<sup>3</sup> <http://www.hcdn.gov.ar>

## 2nde étape :

Cette étape est évidemment plus problématique, prise dans une difficulté en forme de paradoxe : d'une part de réaliser des découpages dans le texte permettant une analyse et que, d'autre part, ce découpage ne soit pas un placage de catégories extérieures.

En cela nous ne pouvons qu'adhérer à la remarque formulée par Julienne Brabet : «Les questions essentielles concernent le lien entre les hypothèses, le recueil et l'organisation des données, toujours hétérogènes et lacunaires, en corpus stabilisés. Le choix des outils informatisés et leur adéquation aux questions posées doit également être réfléchi ainsi que la granularité des filtres utilisés, c'est-à-dire le degré de finesse des phénomènes sémiotiques et/ou langagiers atteignables par l'outil choisi. Enfin, la capacité interprétative critique du chercheur devant les résultats produits demeure évidemment fondamentale : elle passe par la capacité à établir des corrélations plausibles et à les intégrer dans des systèmes interprétatifs globaux. »<sup>4</sup>

Notre objectif était d'objectiver méthodiquement des fragments de textes sur l'ensemble de l'échantillon préalablement constitué, et cela dans le souci de rendre possible une analyse sur deux axes :

- un axe synchronique dans une même période de deux années parlementaires par la caractérisation du registre discursif qui participe de la matrice cognitive du référentiel
- un axe diachronique à travers le repérage de l'évolution de différents éléments du registres discursif entre les périodes parlementaires.

Il reste une part de tâtonnement inévitable, ne serait ce que parce que le « choix » des outils techniques se fait dans un ensemble limité de possibilités déterminées par la situation du chercheur (géographique, institutionnelle, etc.). Cette seconde étape s'est elle même faite en deux temps : d'abord avec le logiciel Alceste étape qui s'est avérée peu concluante et ensuite avec le logiciel N-vivo. Nous passerons rapidement sur le premier temps pour nous arrêter plus longuement sur le second.

Au cours des lectures et des observations<sup>5</sup>, nous avons relevé que s'exprimaient, de manière régulière et fréquente, des opinions sur la démocratie mais de manière hétérogène. Par ailleurs le *représenté*, apparaissait dans la prise de parole sous des dénominations différentes.

Alceste est un logiciel d'Analyse de données textuelles qui quantifie le texte pour en extraire des structures signifiantes fortes à partir du postulat que ces structures sont étroitement liées à la distribution des mots dans le texte. Il apparaissait probable qu'émergent des *bassins de sens* qui contiennent ces segments du texte, d'autant plus que le logiciel reconnaît le vocabulaire et ne s'arrête pas à la reconnaissance des caractères mais prend en compte le rôle grammatical

---

<sup>4</sup> La possibilité de ce corpus est entièrement liée à des « possibilités » technologique : d'une part grâce à l'accès rapide et informatisé aux sources et, d'autre part, du fait de l'existence d'outils de traitement capables de prendre en compte de vastes corpus. Une telle situation *d'opportunité* ne peut qu'éveiller des soupçons sur le contenu scientifique d'une telle démarche. Partant du caractère construit de toute partition statistique et de toute opération cognitive de classification, il convient de rester méfiant. Il serait inapproprié d'ouvrir ici ce débat déjà largement traité par ailleurs. Citons tout de même l'ouvrage classique d'Alain Desrosières, *La politique des grands nombres. Histoire de la raisons statistique*, Paris, la découverte, 1993. Voir également [Angermüller J., Jeanpierre L ;, Ollivier-Yaniv C.,2008 :p.43 pour la citation]

<sup>5</sup> Nous avons effectué des lectures in extenso et assisté également à de nombreuses sessions lors de notre travail de terrain.

du mot. Cette dernière caractéristique s'est avérée négative au final, du fait d'un problème de dictionnaire. Le logiciel ne fonctionne tout simplement pas en espagnol dans la version qui était à ma disposition. De ce fait les résultats restaient à la superficialité du texte. J'avais pris soin de faire mes premières tentatives sur le segment du corpus concernant les séances que je connaissais le plus. Je me suis très vite rendu compte que le logiciel ne me permettait pas d'aller au-delà des lexiques propres à certains sujets. Il ne faisait que redistribuer l'ensemble du corpus sous une forme qui se révélaient presque identiques à l'organisation en session. Le texte gardait son organisation linéaire. Ce fut un échec technique, qui démontre la nécessité d'une plus grande vigilance et qui invite à avancer de manière plus méthodique.

Il fallait reprendre le travail ce qui fut fait à partir du logiciel N-vivo.

Le logiciel N-vivo, n'est pas un logiciel d'analyse de texte stricto sensu. Il est présenté par ces concepteurs comme un logiciel de recherche qualitative visant à « aider les chercheurs et toutes les personnes travaillant à partir de matériaux non structurés à compiler, à comparer et à comprendre les informations rapidement et facilement »<sup>6</sup>. Pour autant il est possible d'utiliser les fonctions de N-vivo dans le cadre de l'analyse de discours comme nous allons le démontrer en suivant. Il convient de distinguer deux moments de l'utilisation de N-Vivo celle plus liée à la constitution du corpus et une seconde qui relève plus d'une analyse proprement dite.

Nous ne sommes plus comme dans le cas précédent en présence d'un logiciel de statistique textuel. Il s'agit d'exploiter les capacités d'indexation, en utilisant la fonction de recherche lexicale, à partir de catégories que nous devons élaborer. Nous avons établi deux grands ensembles de termes :

- Le lexique qui se rapporte à des segments du texte qui convoque le *représenté* repérable par l'utilisation de différentes dénominations
- Le lexique caractéristique d'un contenu.

Dans les deux cas la sélection des termes a été faite à la fois à partir de deux prémices distinctes :

- d'une réflexion préalable à travers la recherche d'occurrences de termes, tel que « citoyenneté » ou « peuple » en ce qui concerne le premier groupe et à « démocratie » ou « institutions » pour le second. Prémices en lien avec les notions classiques de la science politique.
- mais aussi au fur et à mesure de la recherche ou à partir de constat et d'intuition née du contact avec le texte (ou de la période historique) comme les termes de « victimes » et de « travailleurs » pour le premier groupe et de « droits de l'homme » pour le second pour ne prendre que ces exemples.

Il nous a semblé important de limiter le nombre de mots dont le choix s'avère important puisqu'il conditionne l'analyse.

La recherche lexicale concernant les dénominations du mandant consiste dans les vocables suivants :

- ciudadano/s (citoyen/s), ciudadana/s (citoyenne/s), ciudadanía chaque terme constituant un ensemble à part. Il faut noter qu'en espagnol la ciudadanía désigne à la fois la citoyenneté comme attribut et l'ensemble des citoyens.
- el pueblo (le peuple)
- las organizaciones populares (les organisations populaires)

---

<sup>6</sup> D'après la présentation du logiciel : [http://www.qsrinternational.com/other-languages\\_french.aspx](http://www.qsrinternational.com/other-languages_french.aspx). L'essentiel du travail a été réalisé sous la version 8 de N-vivo

- la gente (les gens) exclusivement au singulier
- la sociedad (la société) exclusivement au singulier
- las victimas ( les victimes) exclusivement au pluriel
- trabajadores (les travailleurs) exclusivement au pluriel
- electore, electores, electorado (électeur, électeurs, électorat...)
- Argentin\* ce qui inclut donc le pays (Argentina), les habitants (Argentinos) et l'adjectif « argentine ».

La recherche lexicale concernant le contenu s'est effectuée à travers les termes suivants:

- democracia (démocratie), democrática (démocratique F), democratico (démocratique M), democratización (démocratisation)
- dictature
- institution\*
- derechos humanos (droits humains exclusivement au pluriel)
- Le registre économique : empresa\* (entreprise\*), mercados (marchés exclusivement au pluriel), economi\* (l'économie, économique), financier\* (financier et financiers masculin/féminin )
- media (mediático, mediatizado, medios...)
- elecciones (les élections)
- Política (políticos, política, político)
- luttas (conflits, grèves, revendication...)

La question du contenu a été plus problématique et nécessite quelques remarques, d'une part l'occurrence de certains termes ne suffit pas à épuiser un « thème » comme on entend généralement un sujet de discussion. Si nous prenons le thème du gouvernement autoritaire par exemple toutes les occurrences de « dictature » ne vont pas apparaître dans des zones de textes où le sujet est abordé. Qui plus est les sujets abordés à dans des segments de textes où l'occurrence dictature apparaît n'ont pas un sens homogène.

Il s'agit pour nous d'avoir des indications sur les thèmes abordés. Ces informations seront beaucoup plus précises sur le contenu *économique* ou *démocratique*. C'est lors de l'analyse que les différents ensembles constitués seront traités de manière différente.

Il convient de préciser que le logiciel nous permet de conserver en permanence le lien entre le texte source et le segment de texte. Enfin dernière précision méthodologique le logiciel nous permet de choisir la taille du segment de texte découpé avec le terme recherché. Nous avons opté dans ce travail pour retenir le paragraphe concomitant.

Une fois ce travail réalisé nous sommes en possession d'un corpus constitué de segments de texte regroupés autour d'occurrences sémantiques et situés dans un échantillon de textes représentatifs de la parole parlementaire des sessions parlementaires de mars 1998 à février 2010. C'est sur ce corpus obtenu que nous travaillerons désormais.

L'utilisation de N-vivo ne s'arrête pourtant pas là. Bien que n'étant pas dédié à l'analyse du discours, il est encore possible de mettre à profit le logiciel pour obtenir des informations sur le texte. Très immédiatement la structure même du corpus obtenu nous apporte un certain nombre de renseignements. Par la suite il est possible de continuer de « faire travailler » le logiciel sur le corpus obtenu.

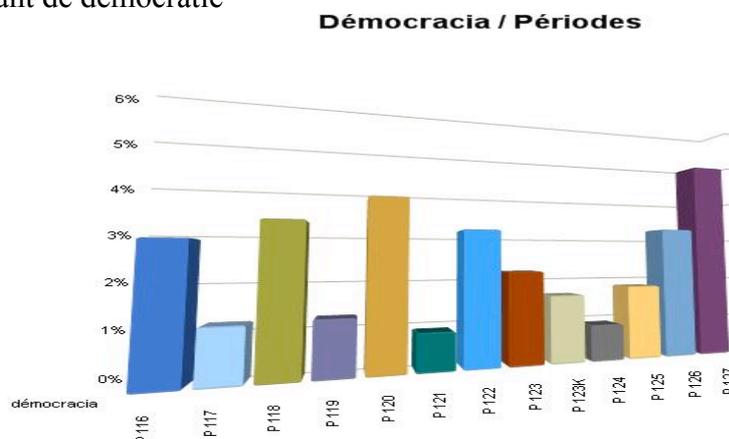
## La structure du corpus obtenus et la distribution des segments de textes.

Il est intéressant d'observer l'importance relative des différents segments de textes pour connaître la prégnance des thèmes abordés selon les sessions. C'est là qu'il convient de rester prudent quant au statut des différents fragments du corpus. Il s'agit d'information exploratoire nous renseignant sur une tendance mais qui ne constitue en aucun cas des valeurs absolues.

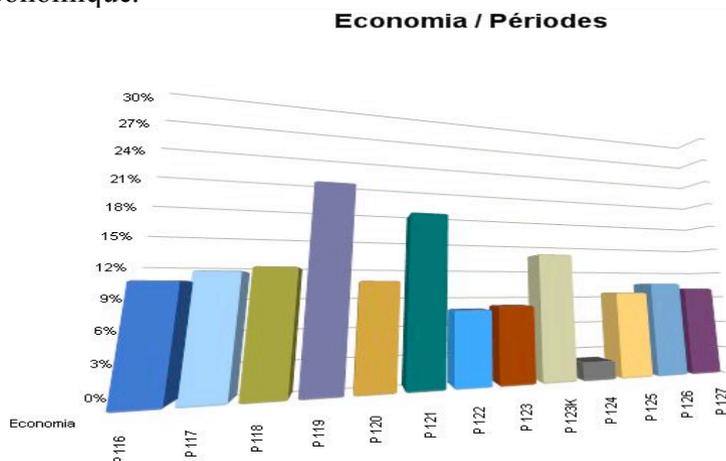
Une seconde précaution est nécessaire pour effectuer des comparaisons entre différentes périodes. En effet, la production langagière pour une session n'est pas équivalente en termes de volume selon les différentes périodes.<sup>7</sup> Le logiciel propose des mesures de recouvrement qui sont élaborées par une relation entre le nombre d'occurrence et l'ensemble du texte exprimé en pourcentage, qui nous autorise des comparaisons.

Des représentations graphiques permettent d'observer les distributions d'un seul coup d'œil. Prenons deux exemples de contenus rapportés aux années parlementaires.

Le registre traitant de démocratie



Le registre économique.



C'est deux « registres » sont particulièrement parlants sur plusieurs points.

- L'importance du registre économique qui ira jusqu'à couvrir 20% des débats alors que le registre démocratique n'atteint qu'exceptionnellement les 5% du corpus.

<sup>7</sup> Du fait notamment du nombre de sessions qui ont effectivement eue lieu dans l'année, donc in fine du nombre de fois où le quorum nécessaire a été atteint.

- Le contraste inversé entre les périodes 119 et 120 qui correspondent respectivement à l'année 2000/2001 et 2002/2003 où l'on voit très nettement que le registre économique diminue alors que le registre démocratique progresse.
- Les variations en dents de scie que l'on peut constater dans le registre démocratie avant la période Kirchneriste correspondent aux années électorales. Sur ce point nous sommes conscient que corrélation ne vaut pas explication.
- Enfin il convient de noter la structure particulière de la période 124. Correspondant à l'année 2006 marqué par une intense activité parlementaire et considéré comme la première années d'hégémonie du Kirchnerisme.

En continuant à exploiter les fonctionnalités de recherche lexicale, il est possible de rechercher des caractéristiques liées plus directement à l'énonciation, en termes d'ancrage et de modalisation.

Nous effectuerons alors des recherches portants sur les sujets et les verbes (les deux recherches sont liés par le biais des formes conjuguées) :

- première personne du singulier (yo, me, mi, mis)
- première personne du pluriel (nosotros, nos...)
- seconde et troisième personnes du pluriel (ustedes, ellos, ellas, los...)

Verbes de modalisation

- Deber [devoir] (présent, passé, futur)
- Poder [pouvoir] (présent, passé, futur)
- Saber [savoir] (présent, passé, futur)

Verbe d'attitude propositionnelle.

- Creer [croire] (présent, passé, futur)
- Pensar [penser] (présent, passé, futur)

verbes désidératifs

- Querer [vouloir] (présent, passé, futur)
- Desear [désirer/espérer] (présent, passé, futur)

C'est une étape en cours de réalisation ce qui ne nous autorise pas encore à livrer de résultats.

## **Conclusion**

Nous n'avons pas pu dans le cadre de cet article traiter de manière étendue des résultats et avons préféré nous étendre sur la démarche mise en œuvre. Construire et faire parler des données ne sont pas deux étapes séparées de la démarche, d'une part parce que la constitution d'un corpus relève d'un questionnement que l'on adresse au texte et d'autre part parce qu'à chaque étape de manipulation des données, nous sommes susceptible de recueillir des informations. C'est d'autant plus vrai avec l'outil informatique, N-vivo (mais on peut voir que l'ensemble des logiciels de traitement suivent la même dynamique) proposant des mises en formes des données qui permettent des comparaisons rapides et fructueuses comme nous l'avons vu.

Avec l'utilisation de logiciel de traitement des données, le travail n'est pas seulement simplifié ou rendu plus rapide (après une période de prise en mains fortement chronophage.), il se transforme. Et, si elle rend possible le traitement de volume important nous permettant des montés en généralité, elle introduit dans la démarche des boîtes noires que le chercheur doit apprivoiser et connaître pour en mesurer les effets et pour en maîtriser les conséquences.

- Abélès Marc, Un ethnologue à l'assemblée, ed. Odile Jacob, Paris, 2000.
- Angermüller J., Jeanpierre L., Ollivier-Yaniv C., « Analyser les pratiques discursives en sciences sociales : Journée d'études du CEDITEC à l'Université Paris XII, le 27 Avril 2007 », Bulletin de methodologie sociologique, Janvier 2008, n.97, pp. 39-47
- Armony Victor, Représenter la Nation. Le discours présidentiel de la transition démocratique en Argentine (1983-1993), l'Univers des discours, Montréal, 2000.
- Garric N. & Léglise I., « La place du logiciel, du corpus, de l'analyste : l'exemple d'une analyse de discours patronal à deux voix », in G. Williams, coord, Linguistique de corpus, Presses Universitaires de Rennes, 2005, 101-113.
- Maingueneau D., L'analyse du discours, Hachette, Paris, 1997. [éditions revue de Analyse du discours, introduction aux lectures de l'archive, 1991.]