

CONSTRUIRE ET MAÎTRISER L'HÉTÉROGÉNÉITÉ PAR LA VARIATION DES DONNÉES, DES CORPUS ET DES MÉTHODES

Nathalie Garric

Armand Colin | « [Langages](#) »

2012/3 n° 187 | pages 73 à 92

ISSN 0458-726X

ISBN 9782200927769

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-langages-2012-3-page-73.htm>

Pour citer cet article :

Nathalie Garric, « Construire et maîtriser l'hétérogénéité par la variation des données, des corpus et des méthodes », *Langages* 2012/3 (n° 187), p. 73-92.
DOI 10.3917/lang.187.0073

Distribution électronique Cairn.info pour Armand Colin.

© Armand Colin. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Construire et maîtriser l'hétérogénéité par la variation des données, des corpus et des méthodes

Depuis quelques années, la recherche interdisciplinaire en Sciences Humaines et Sociales connaît des manifestations plus nombreuses qui impliquent l'analyse de discours et obligent à de réguliers retours réflexifs. La réflexion, centrée sur la problématique de l'hétérogénéité, est orientée sur la nature des données, la construction des corpus et les méthodes d'exploration et de lecture du sens pour la compréhension de différents phénomènes sociaux. Cette contribution aborde, à partir d'une étude de cas, ces questionnements. Elle s'appuie sur une recherche interdisciplinaire menée par des psychologues et des linguistes (Garric & Capdevielle-Mougnibas 2006, 2009) qui vise à repérer les processus à l'origine des ruptures de contrat d'apprentissage dans le parcours de formation d'apprentis et de lycéens.

La collaboration établie a consisté pour le linguiste à prendre connaissance des objectifs de la recherche et, dans une interaction constante, à déterminer selon quelles modalités il pouvait exploiter les données recueillies, sans contradiction avec une conception discursive de ces productions expérimentales. Nous rappelons d'abord les types de données disponibles pour les interroger par rapport aux pratiques de l'analyste de discours. Constatant leur hétérogénéité, nous proposons ensuite de la saisir sous la forme d'un principe méthodologique de variation par la mobilisation de différentes sources d'hétérogénéisation. Nous présentons enfin quelques résultats permis par la méthodologie préconisée pour tenter de montrer comment le principe de variation fournit une méthodologie opératoire pour atteindre l'objectif de M. Pêcheux (1975) et pour répondre efficacement aux enjeux de la recherche : mettre en œuvre dans l'analyse de discours les concepts de *formations discursives* et d'*interdiscours*, en les associant à certains apports de la sociologie du langage (Achard 1993).

1. DES DONNÉES AU PRINCIPE DE VARIATION

Depuis 1990, le taux d'échec des jeunes en situation d'apprentissage est croissant si bien que la rupture contractuelle avoisine 25 %. Elle se manifeste par l'abandon de la formation et/ou par l'entrée dans un nouveau contrat avec ou sans réorientation vers un autre métier. L'objectif de l'étude menée est d'identifier les causes de ces différentes issues du processus de formation. Il s'agit de comprendre un phénomène psychosocial – la rupture – par l'étude de ses manifestations langagières. L'analyse de discours ne constitue qu'un des volets de l'ensemble de la recherche. Elle est convoquée de façon complémentaire avec pour hypothèse que, sa saisie discursive du matériau, sa méthodologie et ses observables peuvent fournir des clés d'interprétation pour confirmer, rectifier ou invalider les résultats obtenus selon les méthodologies et les présupposés de la psychologie sociale.

Deux types de recueils des données ont été réalisés en différents temps du processus de formation :

- des entretiens semi-directifs menés auprès de 24 dyades¹ jeunes en formation/maîtres d'apprentissage qui produisent :
 - 24 entretiens semi-directifs avec les jeunes en formation ;
 - 24 entretiens semi-directifs menés avec les maîtres d'apprentissage ;
- une question ouverte qui a produit 346 textes rédigés par les jeunes en formation : « Depuis que tu es né, tu as appris beaucoup de choses, chez toi, dans ton quartier, à l'école et ailleurs... Qu'as-tu appris ? Avec qui ? Qu'est-ce qui est important pour toi dans tout ça ? Et maintenant, qu'est-ce que tu attends ? ».

Située dans l'interdisciplinarité et surtout articulée à des données fournies par des protocoles familiers aux psychologues, la recherche menée a suscité des questionnements portant explicitement sur le matériau dont nous disposons.

1.1. Problématiser les corpus

P. Charaudeau (2009), dans un article évocateur pour notre objet « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », montre la diversité des objectifs occupant le champ des analyses de discours, en les articulant aux diverses modalités de sélection des données et de construction des corpus. Il distingue, en fonction des caractéristiques et des visées associées aux corpus, trois problématiques :

1. En situation de formation professionnelle, chaque jeune, lycéen ou apprenti, réalise un stage en entreprise sous la responsabilité d'un professionnel qui détient alors le titre de *maître d'apprentissage*. En parlant de dyade, il s'agit d'indiquer que la passation des entretiens opère systématiquement sur le couple formé/formateur, sur des questions néanmoins distinctes et dans des conditions également distinctes, au lycée ou au centre de formation pour le jeune formé, au sein de l'entreprise pour le maître d'apprentissage.

- une problématique dite *cognitive* et *catégorisante* pour laquelle l'objectif est de décrire des mécanismes discursifs à partir de corpus aléatoires dans le sens où ils ne sont pas construits en fonction de conditions de production. Ces analyses ne visent pas l'identification du sens social, elles s'intéressent néanmoins à la contextualisation linguistique des fonctionnements étudiés. Les données langagières sont plus ou moins larges selon le phénomène observé, de l'énoncé au texte en passant par l'extrait. L'homogénéité du corpus est établie par la structure ou le fonctionnement régulier analysé ;
- une problématique dite *communicative* et *descriptive* qui décrit des situations discursives dans une visée typologique à partir de corpus génériques. Les corpus sont constitués de productions attachées à un dispositif situationnel spécifique, en général doté d'une forte institutionnalisation : textes publicitaires ou textes journalistiques, par exemple. L'homogénéité du corpus est établie par le dispositif de production. Ces analyses visent la création d'une typologie des discours qui énonce la régularité de chacun des genres identifiés ;
- une problématique dite *représentationnelle* et *interprétative* pour laquelle il s'agit de décrire et surtout de formuler des hypothèses interprétatives de représentations socio-discursives à partir de corpus de textes. L'objectif affirmé de ces approches est l'identification du sens social, associé à un objectif de théorisation plus ou moins marqué. Le texte fonctionne comme un médiateur (Mellet 2002) pour la compréhension d'un fait social, duquel il obtient sa condition d'homogénéité. Ses fonctionnements énonciatifs – cohérents ou conflictuels –, en tant que pratiques, sont un lieu de construction de la réalité étudiée : ils disposent d'une représentativité socio-historique que cherche à atteindre l'analyse.

La recherche décrite entre dans la troisième de ces problématiques où elle rencontre deux difficultés, l'une relevant de l'objectif même de cette problématique, l'autre de la nature expérimentale de notre matériau dans le cadre de l'analyse de discours.

1.2. Propriétés des données pour l'interprétation

P. Charaudeau, précisant la définition de la problématique interprétative, mentionne que les corpus nécessaires sont d'ordre discursif. Leur objectif est de mettre en œuvre les conditions d'une construction interdiscursive du sens pour formuler des hypothèses interprétatives. Partant, cette problématique s'expose à une difficulté essentielle qui :

réside justement dans le fait qu'il faut faire ces hypothèses et que, pour ce faire, on doit s'appuyer sur des manifestations discursives extrêmement diverses, non toujours explicites, voire diffuses et même floues, ce qui est un problème pour la constitution de corpus. (Charaudeau, 2009 : 51).

Ces réflexions expriment une difficulté méthodologique introduite par l'hétérogénéité des données pourtant requise pour l'activité interprétative. Elles suggèrent cependant simultanément la possibilité d'exploiter cette hétérogénéité en une opportunité méthodologique. Nous posons que la valeur heuristique de cette dernière ne peut être obtenue que par un principe régulateur que nous définissons en termes de *principe de variation*. Il s'agit d'une opération que l'on pourrait qualifier de pluralisation généralisée, qui actualise la variation par différentes pratiques d'analyse et à différents niveaux d'analyse. Ce principe est mis au service du concept, introduit par B. Pincemin (1999) d'*interprétabilité*. Il repose sur trois ensembles de conditions de bonne formation du corpus, des conditions de signifiante, des conditions d'acceptabilité et des conditions d'exploitabilité. Toutes visent à fournir un corpus, susceptible de sous-catégorisations, qui intègre les outils ou les interprétants nécessaires à la construction du sens par contextualisation et contextualisation. Et, comme le souligne l'auteur, le corpus « est susceptible d'une portée interprétative étendue », « il donne la matière concrète pour représenter un axe de lecture, et de fait à la multiplicité des lectures répond la non moindre multiplicité des corpus possibles » à envisager comme autant de mises en relation interdiscursives du texte. La méthodologie adoptée a donc consisté à exploiter de façon systématique le principe de variation, mis en œuvre par les données, les corpus, les outils et les pratiques.

1.3. Nature des données discursives

P. Charaudeau distingue deux types de corpus utilisés pour la problématique interprétative. Le premier s'apparente au corpus de textes-archives qui se compose de textes fondateurs ou, selon la terminologie de F. Cossutta et D. Maingueneau (1995), de textes constituants. Le second est constitué d'extraits textuels que l'auteur propose de nommer « signes-symptômes » : « ils représentent de façon emblématique des systèmes de valeurs ». Notre matériau est apparemment foncièrement différent des données mentionnées et, plus généralement, il ne se révèle pas caractéristique des données habituellement manipulées par l'analyse de discours de tradition française dont les travaux se sont longtemps limités à l'analyse des discours politiques ou tout au moins de discours institutionnalisés et spontanés. Le groupe *Langage et Travail* recourt cependant abondamment à des données comparables aux nôtres :

L'analyse de discours s'est constituée, en France, dans un rapport étroit au corpus, dont elle s'est attachée à définir les conditions méthodologiques de validité. Dans la mesure où le corpus renvoyait à des repérages extra-linguistiques préalables, [...] il était généralement constitué à partir de discours préexistants et ne posait guère qu'à la marge des problèmes de choix et de découpage. [...] Situation bien différente de celle qu'on rencontre aujourd'hui dans le domaine du travail, où discours, textes et autres matériaux de l'analyse sont pour une large part à produire par le mouvement de la recherche elle-même. [...], le chercheur ne peut échapper à la question de la construction des données. [...] La question du corpus se voit replacée dans une problématique plus générale de l'enquête. (Boutet, Gardin & Lacoste, 1995 : 13-14)

Dans un autre cadre théorique, M. Brugidou (2005) explicite sa conception et son utilisation du corpus. Rompant avec la tradition imposant de travailler à partir de données authentiques si l'objectif visé est la réalité sociale, il adopte un dispositif expérimental. L'auteur argumente ce choix par un certain nombre d'évolutions méthodologiques et théoriques. Elles sont autorisées par le rapprochement interdisciplinaire entre l'Analyse de données textuelles, la Sémantique textuelle, les Statistiques et l'Informatique :

Ces avancées sur le front des méthodes permettent ainsi d'envisager des dispositifs d'enquête hybrides donnant une plus large place aux questions ouvertes, voire à des séquences mêlant questions fermées et ouvertes [Brugidou 2003 ; Brugidou *et al.* 2004]. Ils rendent ainsi possible la constitution de véritables corpus « d'énoncés d'opinion publique ». (Brugidou 2005)

Pour justifier la construction de corpus à partir de données suscitées, l'auteur introduit la variation dans le protocole de recueil des données. Toutefois, subordonnant également la faisabilité de ces corpus aux avancées technologiques, plus particulièrement aux logiciels de statistique textuelle, il adopte la variation jusque dans les procédures d'analyse. L'hybridation dont il est question renvoie au principe de variation, conçu comme différentes opérations de contextualisation des textes, méthodologiquement primordiales dans le cadre de toute étude utilisant des données expérimentales sans les dissocier de l'unité texte.

L'auteur souligne cependant que l'adoption ou le refus de corpus expérimentaux ne relève pas exclusivement d'un questionnement méthodologique, mais aussi théorique. Il consiste de façon générale à accepter que ce type de données artificielles puisse constituer une réelle voie d'accès à la réalité sociale. Ainsi s'attelle-t-il, pour le matériau qu'il recueille, à défendre qu'il occupe, en tant qu'acte de langage doté d'une valeur performative, une place à part entière dans l'ensemble des pratiques discursives. Il note que les questionnaires sont des pratiques sociales régulières. En effet, comme toute production discursive, l'entretien est soumis à des déterminations dont les caractéristiques génériques sont très contraignantes à l'image de tout discours fortement institutionnalisé. Nous le concevons donc comme une interaction sociale délimitant des places sociales en relation avec une formation discursive. Selon D. Demazière,

Le discours de l'interviewé est une prise de position sur la question posée, est l'expression d'un sujet parlant qui signifie et qui agit en disant des choses, est un acte de construction en situation d'une vision du monde. (Demazière, 2007 : 95-96)

Selon le protocole de passation, le discours sollicité peut prendre place au sein de conditions de production plus ou moins effacées, augmentant les précautions méthodologiques, mais il ne peut pas se déployer en dehors d'une formation discursive. La mise en place d'un protocole de recherche dans tout espace institutionnel prend du temps, donne lieu à des négociations, à l'écriture de documents et à la diffusion d'informations qui sont autant d'éléments qui actualisent la formation discursive. Par ailleurs, au moment de cette recherche, le gouvernement avait entrepris une modification du système de formation en

alternance l'introduisant sur la scène médiatique. Les élèves, mais également les professionnels impliqués par le dispositif, n'y trouvaient qu'un statut externe. Interpellés, ils ne participaient pas au débat ; les entretiens menés leur donnaient l'opportunité d'y entrer par une voie singulière.

Les données expérimentales, si elles sont traitées en tant que telles, si elles sont exploitées à l'aide des évolutions méthodologiques décrites *supra*, dont la principale propriété est de permettre différentes opérations de contextualisation des textes, constituent un matériau pertinent pour l'analyse de discours. Elles constituent des textes « suggérés mais authentiques », qui ne deviennent discours « que par le travail de structuration des données » (Garric & Capdevielle-Mougnibas 2009).

2. INTRODUCTION DES SOURCES DE VARIATION

La mise en œuvre du principe de variation se révèle donc décisive. Elle s'effectue à plusieurs niveaux par la mobilisation de différentes sources d'hétérogénéisation.

2.1. Hétérogénéité des données

Adossée à deux types de recueil des données, l'hétérogénéité est inhérente au matériau analysé sans détruire la condition d'homogénéité nécessaire à l'élaboration de tout corpus. Celle-ci relève, d'une part, de la thématique générale des entretiens et du bilan de savoir qui délimite un corpus de référence² ; d'autre part, du dispositif de recueil qui délimite trois corpus d'étude, tels qu'ils sont mentionnés sous § 1.

2.2. Hétérogénéité de la population

Deux populations homogènes sont impliquées dans la recherche, les formés et les formateurs professionnels. Chacune présente plusieurs variables susceptibles de constituer autant de sources de différenciation des places sociales pour la formulation d'hypothèses interprétatives. Ces variables – *sexe, *CSP, *métier, *âge, *expérience, *formation, *issue et *statut – ont systématiquement renseigné les données. Nous avons essentiellement exploré les variables *statut et *issue que nous déclinons sous la forme de schémas synthétiques qui présentent les

2. Selon Rastier (2005 : 33) :

« 1/ L'*archive* contient l'ensemble des documents accessibles. Elle n'est pas un corpus, parce qu'elle n'est pas constituée pour une recherche déterminée.

2/ Le *corpus de référence* est constitué par l'ensemble de textes sur lequel on va contraster les corpus d'étude.

3/ Le *corpus d'étude* est délimité par les besoins de l'application.

4/ Enfin le *sous-corpus de travail* varie selon les phases de l'étude et peut ne contenir que des passages pertinents du texte ou des textes étudiés ».

La notion de *sous-corpus de travail* sera illustrée par les analyses consacrées à certains marqueurs construits dans l'interdiscursivité des corpus d'étude.

divisions susceptibles de définir des variables pertinentes pour la délimitation de positionnements représentatifs de groupes sociaux :

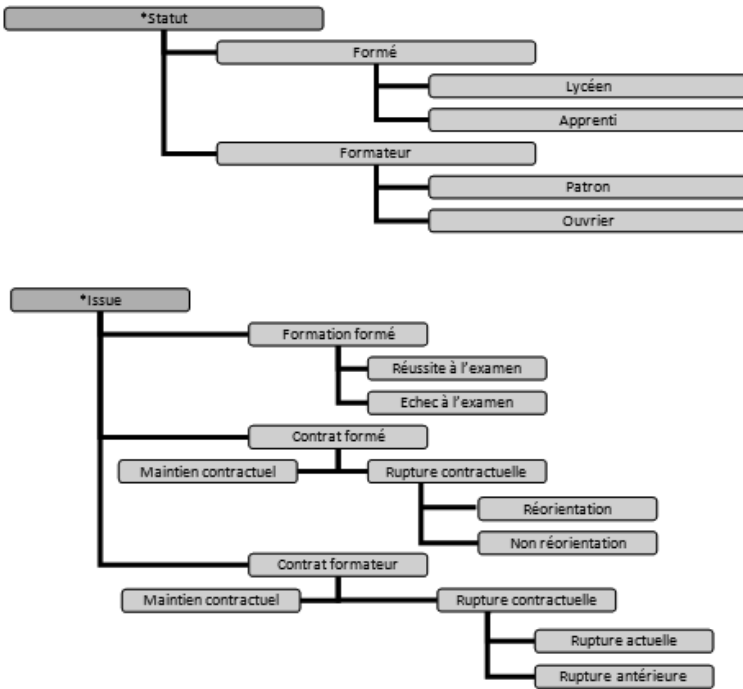


Figure 1 : Déclinaison des variables *statut et *issue

2.3. Hétérogénéité des pratiques

Enfin, l'hétérogénéité est introduite au cœur de la méthodologie par l'utilisation des principes de la lexicométrie ou textométrie (cf. Pincemin, ce volume). La textométrie renvoie à un ensemble d'opérations de comptage appliquées sur un corpus sur la base d'un principe différentiel et proportionnel établi entre les parties constitutives d'un corpus (Lebart & Salem 1994). Les calculs sont effectués à l'intérieur des parties en relation avec le vocabulaire de l'ensemble du corpus de référence. Les logiciels de lexicométrie peuvent participer, selon leur appropriation, à la fois à la procédure de construction dynamique et raisonnée des corpus et aux procédures d'interprétation contextualisées.

Le logiciel *Lexico3*, que nous adoptons dans le cadre de cette étude, permet de tester l'élaboration de différents corpus en sélectionnant tour à tour les variables introduites par des clés de partition dans le corpus de référence. La pertinence des variables sur lesquelles l'analyse formule des hypothèses de causalité est ainsi évaluée en amont. Si le balisage de tout corpus relève nécessairement

d'une opération *a priori*, et donc d'une opération d'interprétation préalable, les expérimentations autorisées par le logiciel permettent de rectifier *a posteriori* ces catégories prédéfinies en leur accordant le statut de participant illustratif. Ces opérations lexicométriques sont conçues comme des procédures de contextualisation distinctes susceptibles, par la présence ou l'absence d'indices langagiers – des régularités ou des divergences formelles – de susciter la formulation d'hypothèses interprétatives.

3. INTERPRÉTER DANS LA VARIATION

La mise en œuvre du principe de variation relève des corpus définis et des méthodes d'analyse utilisées, chacun instaurant différents réseaux intertextuels pour autant de parcours interprétatifs. Le principe de variation consiste à multiplier ces réseaux afin d'instaurer des dynamiques intertextuelles plurielles qui informent et guident l'interprétation de l'analyste (Tognini-Bonelli 2001).

Ces modalités d'application du principe de variation décrites, nous nous intéressons à l'étude de certains marqueurs énonciatifs pour montrer comment la variation en favorise l'interprétation. Ces marqueurs sont identifiés en corpus par le traitement lexicométrique et par leur participation à des réseaux de cohérence interprétative construits dans la textualité. En tant que constituants discursifs de la réalité sociale, ils sont traités comme manifestation d'un certain positionnement énonciatif, lequel témoigne d'un rapport singulier au processus de formation.

3.1. Configurations plurielles de corpus

L'interface de *Lexico3* autorise plusieurs possibilités de découpage ou de regroupement des données par la sélection, en cours d'analyse, des variables que l'on souhaite actualiser en neutralisant certaines d'entre elles. Il est ainsi possible d'initier des corpus d'étude établis en de nouvelles sources de données pour la construction de nouveaux corpus requis par la progression de l'analyse. En première étape, nous utilisons l'analyse factorielle des correspondances (désormais AFC) pour formuler des hypothèses de positionnements et les mettre à l'épreuve. La projection qu'elle propose visualise, par des nuages de points, l'existence ou l'absence de réseaux de causalité éventuels. Les discours partageant des affinités fréquentielles sont contigus sur la représentation et l'intersection des axes établit le profil moyen du vocabulaire.

Nous avons, lors de la recherche sur le corpus [Bilans de savoir], formulé l'hypothèse que les variables *sexe, *apprenti/lycéen, *rupture/non rupture, *échec/réussite à l'examen délimitaient de potentielles-clés interprétatives.

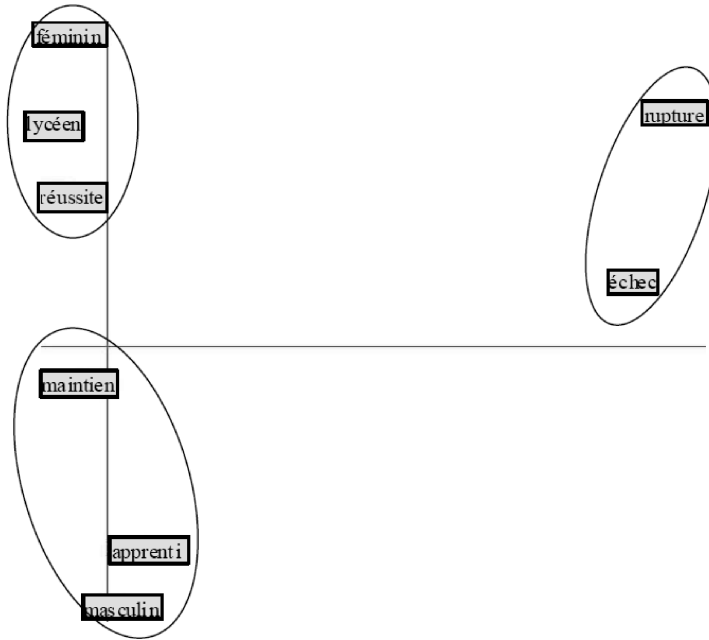


Figure 2 : Corpus [Bilans de savoir], AFC (1a)

Cette représentation présente une répartition des différentes classes de locuteurs avec des délimitations spatiales très nettes. Trois nuages de points, corrélables avec des ensembles discursifs, sont identifiables : 1) <Féminin + Lycéen + Réussite> ; 2) <Rupture + Échec> ; 3) <Maintien + Apprenti + Masculin>. Ils contiennent autant d'indices pour la construction interprétative. Les contours des ensembles ainsi figurés sont toutefois larges et marqués par le positionnement, plus ou moins périphérique, de certains sous-corpus. La périphérie notamment des classes <féminin>, <masculin>, <salarié>, <rupture> et <échec> pourrait invalider les délimitations précédentes. De nouvelles procédures sont donc requises pour tester la pertinence d'autres regroupements constitués de certains des sous-corpus retenus à l'exclusion d'autres. L'AFC (1b) suspend la participation de la variable *sexe et (1c) suspend celle des variables *sexe et *statut (différenciées par la couleur).

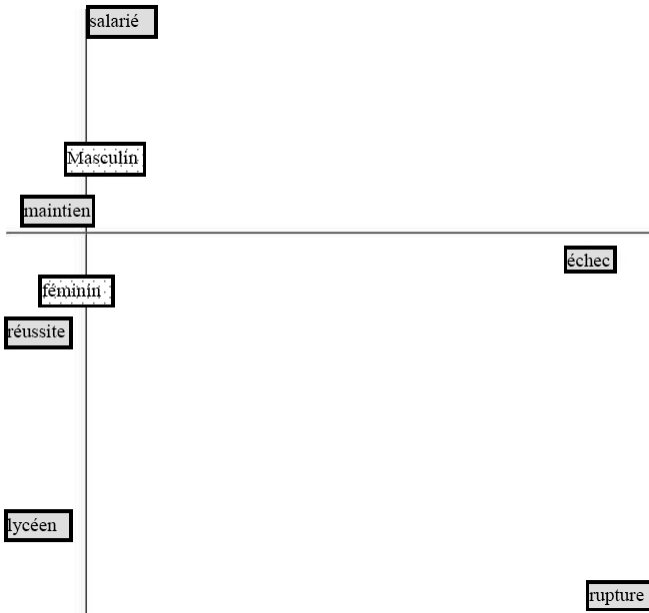


Figure 3 : Corpus [Bilans de savoir], AFC (1b)

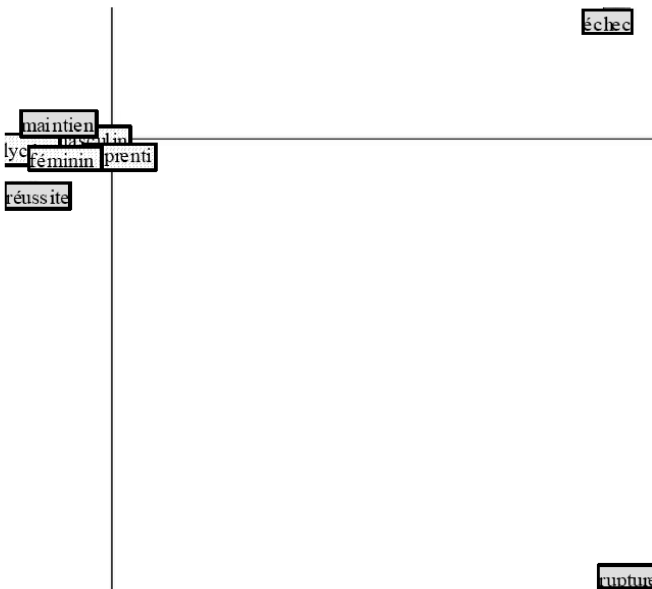


Figure 4 : Corpus [Bilans de savoir], AFC (1c)

La proximité entre <maintien> et <réussite> persiste et s'accroît lors des deux reconfigurations, alors que les sous-corpus <échec> et <rupture> se désolidarisent progressivement l'un de l'autre : deux positionnements distincts sont identifiables, tandis que la neutralisation des variables renforce la cohérence discursive des deux premiers. On peut également souligner le maintien, au cours des différentes opérations de contextualisation, des sous-corpus <maintien> et <rupture>. L'hypothèse de deux modalités discursives peut être croisée avec le corpus des [Entretiens avec les apprentis] dont l'AFC (2) définit deux, voire trois registres : les <apprentis sans rupture>, les <apprentis en rupture sans changement de métier> et les <apprentis en rupture avec changement de métier>.

Enfin, on remarque sur l'AFC (3) qui résulte du traitement du corpus des [Entretiens avec les maîtres d'apprentissage] que l'hypothèse d'un registre propre à la situation de maintien se confirme : les locuteurs connaissant une situation de réussite contractuelle actuelle avec leur apprenti, <AnPn> et <AoPn>, se rassemblent à droite de l'axe vertical, indépendamment de leurs expériences antérieures pourtant discriminantes dans le cas des locuteurs <AoPo> et <AnPo> confrontés, lors du recueil des données, à la rupture.

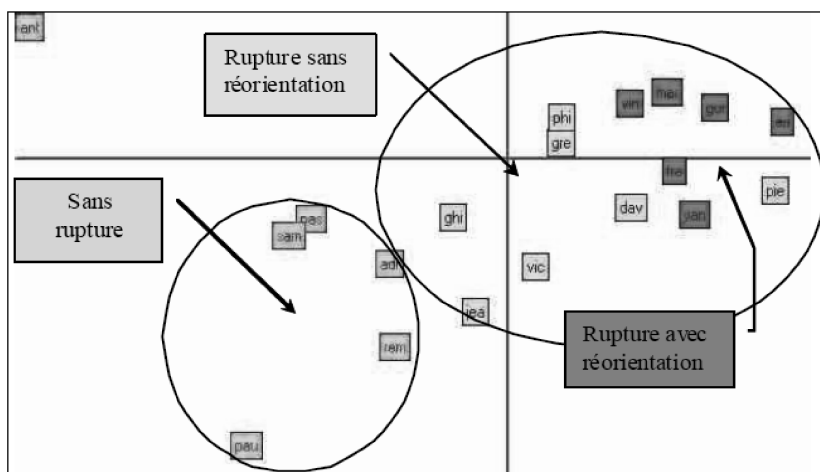


Figure 5 : Corpus [Entretiens avec les apprentis], AFC (2)

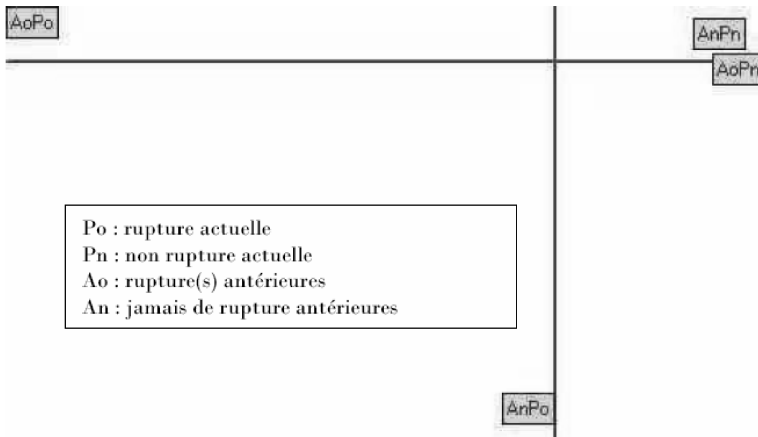


Figure 6 : Corpus [Entretiens avec les maîtres d'apprentissage], AFC (3)

La variation permise par *Lexico3* va au-delà de la procédure de sélection des variables pour la constitution progressive de nouveaux corpus de travail. Elle résulte également de la variété des outils disponibles pour déconstruire et reconstruire les corpus et la textualité. Par exemple, la complémentarité observée entre les AFC des corpus [Entretiens avec les apprentis] et [Bilans de savoir] interroge l'opposition qui se dessine entre les locuteurs <en rupture> et ceux <sans rupture>. L'analyse des spécificités participe à cet objectif. Par un calcul statistique de type probabiliste, elle formule, en référence à l'ensemble du corpus, un jugement sur la fréquence de chaque unité constitutive des sous-corpus. Une unité est ou *banale* ou *spécifique positive* ou *spécifique négative*. Dans ces deux derniers cas, elle est définie comme caractéristique du sous-corpus dans lequel elle est sur-représentée ou sous-représentée. Les pistes d'interprétation ainsi favorisées sont d'autant plus argumentées que cette procédure établit les calculs, non seulement à partir de la partition initiale sur laquelle s'appuie l'AFC, mais également, à partir de nouvelles partitions sélectionnées sur l'AFC elle-même qui fournissent des corpus d'étude inédits. On peut, par exemple, en s'appuyant sur l'AFC des [Bilans de savoir] regrouper les deux parties <rupture> et <échec> en une partie plus large <rupture + échec>. Ce sont ces différentes disponibilités que nous nous approprions *infra*.

3.2. Analyse de quelques marqueurs énonciatifs

Par l'étude de certains marqueurs énonciatifs, nous montrerons comment la variation des corpus introduit une interprétation construite dans la complémentarité des différentes mises en série textuelles permises par les diverses configurations de corpus.

3.2.1. Corpus [Entretien avec les apprentis]

L'analyse du corpus [Entretiens avec les apprentis] témoigne de deux cohérences discursives à partir desquelles nous formulons l'hypothèse d'une corrélation entre la place sociale discursivement construite des acteurs impliqués dans le processus de formation et l'engagement contractuel des apprentis (Garric & Capdevielle-Mougnibas 2006). Alors que les apprentis en échec et s'orientant vers un autre métier présentent une énonciation débrayée, notamment marquée par l'emploi de l'imparfait, les apprentis en réussite produisent une énonciation embrayée, inscrite dans l'actualité de la formation. Différents indices témoignent de cette relation de causalité. On observe sur le graphe suivant une répartition marquée des pronoms personnels.

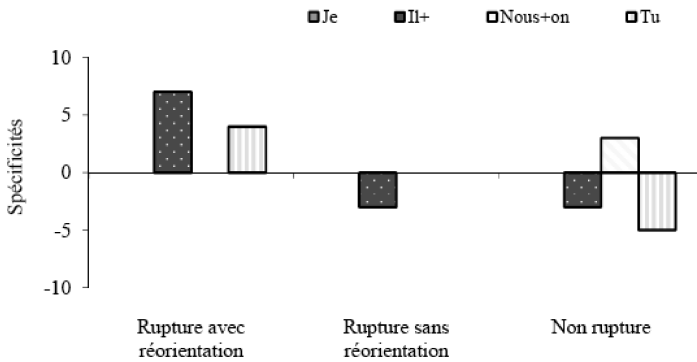


Figure 7 : Graphe de ventilation des spécificités personnelles

Les <locuteurs sans rupture avec réorientation> se caractérisent par des formes délocutives, l'utilisation du *il* impersonnel et celui de la non-personne, ceux <en rupture> sont majoritairement marqués par l'occurrence d'un *nous* et *on* collectifs à valeur communautaire. Le pronom *je*, non spécifique, est identifié comme forme générique de ces discours. La spécificité du pronom *tu* pour la première classe peut sembler contradictoire. L'examen toutefois de ses occurrences confirme la délocutivité : la forme *tu* s'impose ou avec une valeur de datif éthique qui généralise le procès [(1)-(2)] ou en tant que forme de discours rapporté [(3)-(4)] :

- (1) ouais, ouais, ils délirent plus aussi. **tu** rigoles avec eux, tu... voilà quoi. Que ce soit pas tout le temps en train d'écrire au tableau
- (2) en entreprise si, ici on voit plus de choses, en entreprise, tu vois tout ce que **tu** fais en entreprise mais à l'école tu vois un peu tout c'est ça qui est mieux
- (3) quoi il veut toujours sa main d'œuvre, quoi je lui ai dit oui, mais bon il m'a dit **tu** sais il faut que tu prouves
- (4) lui il était pour, il me disait c'est bien, si **tu** veux apprendre un métier c'est là ou **tu** dois aller quoi, c'est pas dans un lycée quoi il était pour la chambre des métiers

Les ensembles mentionnés sont également construits par la morphologie verbale. Les locuteurs <avec rupture> se distinguent de ceux <sans rupture> par l'emploi récurrent de l'imparfait, alors que les seconds lui substituent le présent ou encore, le futur. L'occurrence banale de *je* permet au sujet, associée à la valeur durative de l'imparfait, de se raconter dans le passé en tant qu'acteur autre ; associée au présent, elle actualise sa valeur déictique et met en scène un comportement élocutif. Le traitement des autres corpus, par l'analyse notamment de la répartition des verbes de la modalité déontique et celle de certaines formes d'hétérogénéité montrée (Authier-Revuz 1995) ou d'hétérogénéité suggérée (Moirand 2007), complète l'interprétation.

3.2.2. Corpus complémentaire des [Bilans de savoir]

L'analyse des parties <échec> et <rupture>, ou encore <échec + rupture>, issue de la reconfiguration du corpus [Bilans de savoir], révèle une répartition des formes personnelles différentes. Elle rappelle ainsi le poids des contraintes génériques liées au protocole et l'impact des différentes mises en séries, mais elle présente également des complémentarités.

La Figure 8 montre que la position excentrée des sous-corpus <échec> et <rupture> se justifie essentiellement par la spécificité positive de *je* (<échec> +30 ; <rupture> +13), associée à celle de *il* (<échec> +7 ; <rupture> +3). Les formes personnelles disposent d'une répartition comparable pour les autres catégories de locuteurs. On relève également que le pronom *on* réalise une spécificité positive pour les locuteurs en <échec>. Cette fréquence, dans le contexte des [Bilans de savoir], qui interpelle directement le sujet sur son expérience, signale une attitude de retranchement derrière une figure indéterminée. On peut formuler l'hypothèse que son occurrence est comparable à celle du datif éthique : il permet de formuler des règles comportementales génériques, endossées par une responsabilité autre que l'apprenti lui-même, mais construite par lui.

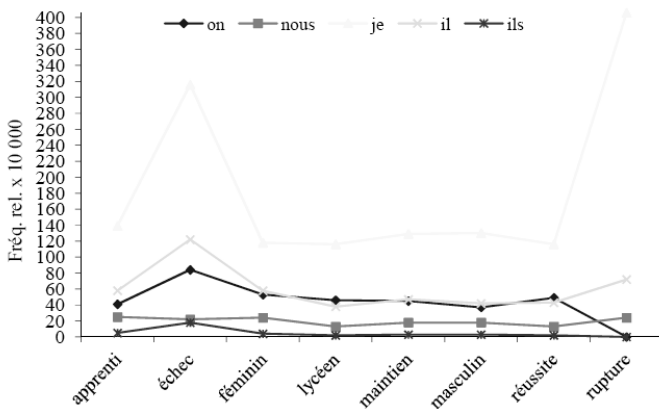


Figure 8 : Ventilation des fréquences personnelles

Sur le graphe 9, on constate que les verbes de modalité contribuent aussi au positionnement sur l'AFC des locuteurs en difficulté. Le modal *devoir*, qui appelle une réalisation actancielle, est peu représentatif du corpus. En revanche, *falloir* dispose d'un emploi fréquent qui introduit une obligation impersonnelle. Cet emploi, disposant de réalisations morphologiques contrastées en fonction des locuteurs, est discriminant. Ceux <en situation de réussite> et/ou de <maintien contractuel> en présentent une occurrence banale quel que soit le temps : le déontique n'est donc pas caractéristique de leur discours diminuant ainsi la perception de toute contrainte dans leur projet de formation. Ceux <en échec> mobilisent, quant à eux, les segments répétés *il faut* et *il fallait*, tandis que ceux <en rupture> sous-emploient le présent et sur-utilisent l'imparfait.

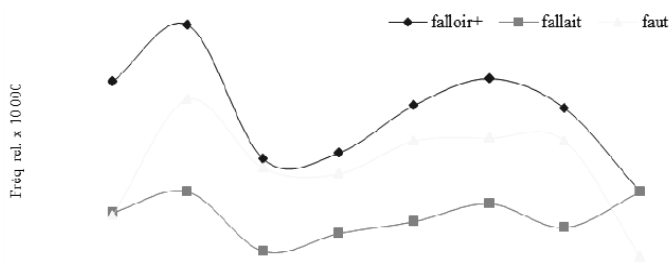


Figure 9 : Ventilation des fréquences de l'obligation

La répartition de la modalité en fonction du morphème temporel suscite un rapprochement avec le corpus [Entretiens avec les apprentis]. Les locuteurs <en rupture> – pour lesquels nous avons souligné la présence originale de l'imparfait – intègrent l'obligation à l'espace de référence passé ainsi construit. La contrainte n'appartient pas à l'actualité du locuteur, comme le signale la fréquence faible du lemme *falloir+*. L'échec, quant à lui, ne s'apparente pas à une occultation de la contrainte, puisque celle-ci s'exprime tant au présent qu'au passé, justifiant probablement de la non-rupture contractuelle pour ces sujets. La rupture pourrait donc être envisagée comme une forme de démission qui consiste à se libérer des choix extérieurs qui ont été imposés à l'apprenti avant sa formation.

3.2.3. Corpus complémentaire des [Entretiens avec les maîtres d'apprentissage]

Nous avons mis en relation ces marqueurs avec le corpus [Entretiens avec les maîtres d'apprentissage] en nous consacrant à la mise en scène de l'apprenti. Les locuteurs familiers avec la rupture (<Rup. Ant. Rup. Act.>) utilisent des formes de décrochage énonciatif qui construisent la parole de l'apprenti comme dire autre dont ils n'assument pas la prise en charge : on relève le sur-emploi de

soi-disant (+2), du terme *guillemets* (+3), du verbe de citation *dit* (+7) et de la citation guillemetée (+37).

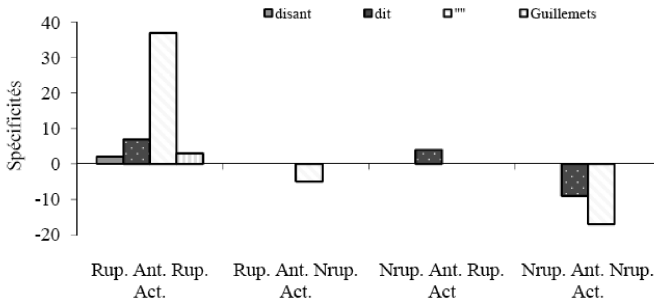


Figure 10 : Graphe de ventilation des indices d'hétérogénéité

Le dire montré peut être également construit en opinion stigmatisée et dévalorisée de l'apprenti à l'aide notamment du segment *ils pensent que* :

- (5) C'est-à-dire que... ils **pensent**, ils **pensent** que finalement c'est... la facilité.
- (6) ben le mec, il est en vacances. Ils **pensent** qu'ils vont gagner de l'argent, d'ailleurs, té
- (7) ils sont plus motivés, qu'ils en ont rien à faire. Ils **pensent** que... quand vous avez des copains qui vous disent

En contraste, les maîtres d'apprentissage en situation de réussite avec les apprentis (<Nrup. Ant. Nrup. Act.>) négligent ces indices et mettent en scène leur propre parole sous la forme du discours rapporté ou de la modalité autonymique à l'aide de *je dis* (+7) ou encore de *disons* (+10). Ce marqueur de modalisation, qui indique la non-coïncidence interlocutive, signale au co-énonciateur un dysfonctionnement de la nomination. Il exprime une prise de distance vis-à-vis du dire, mais celle-ci marque simultanément une volonté de s'approprier le discours et l'univers de l'apprenti.

Dans l'hétérogénéité construite par la variation et révélée par les traces énonciatives se dessine cet « ailleurs » (Authier-Revuz 1995), cette zone d'interdiscursivité (Pêcheux 1975), qui détermine les pratiques sociales. Par exemple, en confrontation avec les autres indices d'hétérogénéité, la voix générique instituée par le datif éthique dans le corpus [Entretiens avec les apprentis] dans le cas des jeunes en rupture résonne avec celle mise en scène par les maîtres d'apprentissage en difficulté avec leurs apprentis, qui, en outre, ne désignent les apprenants qu'avec un *ils* collectif, illustré par les exemples *supra* et des termes pantonymes tels *mec(s)* (+8/+5), *gosse* (+5), *môme* (+3). Il apparaît ainsi que l'échec contractuel est corrélé à une modalité énonciative spécifique qui fait entendre un sujet universel assénant des préconstruits et véhiculant un certain type d'idées reçues, négatives et dévalorisantes, sur l'orientation professionnelle que ne s'approprient pas les apprentis, mais que les maîtres d'apprentissages

récupèrent pour juger les apprentis. La reconfiguration du corpus [Bilans de savoir], construit avec l'association <échec + rupture>, détermine une place sociale qui peut être corrélée avec celle que leur attribuent les maîtres d'apprentissage. Isolés dans leurs discours et leurs agissements comme des « mecs » ou « gosses », et ainsi dépourvus de statut dans l'entreprise, ils développent un discours où ils se présentent seuls face à un autre indéterminé associable à la voix générique identifiée. On relève en particulier l'occurrence spécifique de la forme pronominale *s(e)* (<échec + rupture> +5) et *m(e)* (<échec + rupture> +10), qui apparaît essentiellement dans les syntagmes verbaux suivants : *s'en sortir, se débrouiller, se contrôler, se bouger, se faire respecter*.

[Or.] la construction pronominale est un opérateur de diathèse qui intervient dans la distribution des participants au procès. L'énonciateur est construit comme agent et patient, [...]. Cette caractéristique, associée à « je », délimite un énonciateur saisi dans son individualité, voire dans sa solitude, comme en témoigne la forme « seul » (+3). (Garric & Capdevielle-Mougnibas, 2009 : 121)

Les élèves <en échec à l'examen> et ceux <en rupture> utilisent un dire égocentré qui exprime, par la clôture du procès, une coupure vis-à-vis des acteurs impliqués dans le processus de formation et auxquels est attribuée la dévalorisation de l'apprentissage.

L'échec se manifeste ainsi comme la difficulté à se positionner, tant pour les apprentis que pour les maîtres d'apprentissage, vis-à-vis des stéréotypes sur la formation en alternance. Pour chacun, la situation de formation n'est pas construite dans l'actualité de l'expérience vécue, mais en fonction des représentations véhiculées dans l'espace public sur ce type de formation. Le formateur prive ainsi discursivement l'apprenti de tout statut social, l'apprenti rejette son parcours de formation à une contrainte subie de laquelle il tente de s'extraire.

4. ARTICULATION THÉORIQUE DU PRINCIPE DE VARIATION

Selon P. Achard, « un texte donné ne constitue jamais seul un discours, il prend sens par l'hypothèse qu'il fait partie d'une série, distincte de telle autre » (1986 : 15). La série ainsi mentionnée correspond au registre – à une hypothèse de place sociale – qui pour être expérimentée nécessite la constitution d'un corpus. Partant de cette observation, M. Sassier (2008) propose de redéfinir le registre par la formulation compacte suivante :

Un registre discursif est une accumulation de textes dans un même voisinage ; celui-ci, sous la dépendance d'un point de vue formulable en termes de type d'actes, relève du système régulé des rapports sociaux. (Sassier, 2008 : 43)

Il s'agit de retenir de cette définition que le registre est nécessairement un point de vue de l'analyste, il peut être suffisant pour l'objectif, notamment descriptif, de certaines analyses de discours, mais il reste incomplet pour une visée interprétative. Par ailleurs, il n'est pas un « descripteur directement applicable à des

données empiriques » (*op. cit.* : 41) : il est la condition de pertinence du corpus qui lui permet d'acquérir une visibilité.

Un registre est un ensemble empirique (ouvert) de discours assignables à une même catégorie de sources d'énonciation, et donc suffisamment voisins dans la pratique pour partager une même indexicalité. Construire un corpus, c'est faire l'hypothèse que l'ensemble (fermé) d'événements discursifs attestés et documentés (extraits d'une "archive") que l'analyste a sélectionnés peuvent témoigner du fonctionnement catégoriel du *registre par renvoi aux genres associés*. (Achard 1997, cité par Sassier, 2008 : 46).

Cependant, la notion de *registre* est indissociable de celle de *formation discursive*. P. Achard est peu prolix à son sujet. Il l'emprunte à M. Foucault (1969) et à M. Pêcheux (1969) et la définit comme « la structuration de l'espace social par différenciation des discours » (Achard, 1995 : 84). Elle est une « structure implicationnelle des affectations d'effets de sens aux registres » (*op. cit.* : 91), une structure d'attracteurs qui exerce, comme le souligne M. Sassier, ses contraintes sur le genre, le registre et le sens. Mais P. Achard indique également qu'un « énoncé attesté interpelle l'ensemble des interdiscursivités possibles » (*op. cit.* : 90). Il n'appartient donc pas à l'espace unique et clos d'une formation discursive, qui elle-même est relationnelle. Il est susceptible de prendre place dans plusieurs séries, supposant chacune une indexicalité régulée, dans plusieurs paradigmes interprétatifs, que l'on peut saisir par la construction de corpus pluriels. La variation des données et des corpus constitue autant de points de vue d'interprétation des textes, elle est une forme d'insertion des textes dans l'interdiscours. Le corpus donne corps au registre, mais plus encore sa pluralisation, par les divers réseaux qu'elle engendre, permet d'établir des dynamiques dialogiques nécessaires à la discursivisation textuelle :

Se centrer sur un lieu particulier permet de décrire (cartographier) un registre comme structure d'attracteurs en s'appuyant sur les genres comme observatoires. Décrire une formation suppose de cumuler des résultats, c'est-à-dire de raccorder les cartes. Pour raccorder, il faut effectuer diverses opérations : homogénéiser les échelles (largeur de cadre et précision des descriptions) ; raccorder les directions (rendre les "axes" parallèles) ; rendre compatibles les intersections (recollements) en faisant apparaître les déformations (afin de respecter le dialogisme). (Achard, 1995 : 92)

À partir des notions développées, l'auteur situe l'intérêt principal de l'analyse de discours dans l'étude des distorsions entre genre et registre. Cette priorité analytique résulte des propriétés de la formation discursive et de l'interdiscours. Elles ont pour conséquence que :

[...] chaque attesté, comme acte n'a pas un sens propre (n'est pas pré-interprété dans une série particulière) mais est à la fois acte d'appropriation des registres auxquels il prétend, et acte dont l'interprétation est objet d'affrontement entre appropriations possibles. (Achard, 1995 : 90)

La variation, en tant que principe méthodologique, est une instrumentation de maîtrise des données hétérogènes et de manipulation de ces données pour les reconstruire en discours, eux-mêmes par définition hétérogènes. Ainsi structurées en corpus établissant des réseaux intertextuels, les données permettent de

reconfigurer l'hétérogénéité en indices énonciatifs des enjeux sociaux de l'apprentissage. Elles sont aussi la trace d'affrontements entre différents registres et montrent que l'issue du processus de formation est déterminée par des enjeux de places sociales dans les espaces éducatifs et professionnels.

5. CONCLUSION

L'interprétation de tout discours bénéficie donc de la création expérimentale de la stabilité/instabilité des textes que nous mettons sous la dépendance du principe de variation, dont la mise en œuvre systématique est dépendante de la procédure d'hétérogénéisation décrite. Ce principe est la condition de leur animation et de leur dynamisation dans l'espace social. Un texte est à la fois un matériau empirique et une réserve d'interprétants pour saisir le sens de certains éléments de la réalité sociale à condition qu'il soit mis en relation avec d'autres textes, socio-idéologiquement apparentés ou dissemblables, qui constituent son contexte (Van Dijk 2008).

La notion de *corpus* est centrale : elle détient les conditions de l'interprétation qui repose sur une « approche multidimensionnelle de la co(n)textualité » (Cislaru & Sitri, 2009 : 101). Selon ces auteurs, ce volume, « le corpus se définit dans une double relation de dépendance par rapport à une problématique contextualisée socio-historiquement et un système de relations textuelles », mais ces points de vue – linéarité chronologique et verticalité interdiscursive – sont en général dissociés. Les articuler passe par une construction des corpus adaptée, fondée sur une hétérogénéité de nature des données et de statut des corpus, et par l'utilisation d'un traitement lexicométrique. Le principe de variation met en œuvre et systématise l'hétérogénéité requise ; le principe différentiel lexicométrique plonge le texte dans différents réseaux interdiscursifs qui l'informent, le déstructurent pour le restructurer autrement.

Références

- ACHARD P. (1986), « Discours et sociologie du langage », *Langage & société* 37, 5-60.
- ACHARD P. (1993), *La sociologie du langage*, Paris : Presses Universitaires de France.
- ACHARD P. (1995), « Formation discursive, dialogisme et sociologie », *Langages* 117, 82-95.
- ACHARD P. (1997), « La sémantique discursive », Communication par affiche au *XVI^e Congrès International des Linguistes*, Paris, juillet 1997.
- AUTHIER-REVUZ J. (1995), *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris : Larousse.
- BOUTET J., GARDIN B. & LACOSTE M. (1995), « Discours en situation de travail », *Langages* 117, 12-31.
- BRUGIDOU M. (2005), « Vers une représentation discursive de l'opinion publique », *Corpus* 4. [<http://corpus.revues.org/index311.html>]

- CHARAUDEAU P. (2009), « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », *Corpus* 8, 37-66.
- CISLARU G. & SITRI F. (2009), « TEXTE et DISCOURS. Corpus, co-texte et analyse automatique du point de vue de l'analyste de discours », *Corpus* 8, 85-104.
- CISLARU G. & SITRI F. (2012), « De l'émergence à l'impact social des discours : hétérogénéités d'un corpus », *Langages* 187, Paris : Larousse/Armand Colin (ce volume).
- COSSUTTA F. & MAINGUENEAU D. (1995), « L'analyse des discours constituants », *Langages* 117, 112-125.
- DEMAZIÈRE D. (2007), « À qui peut-on se fier ? Les sociologues et la parole des interviewés », *Langages & Société* 121/122, 85-100.
- FOUCAULT M. (1969), *L'Archéologie du savoir*, Paris : Gallimard.
- GARRIC N. & CAPDEVIELLE-MOUGNIBAS V. (2006), « Intérêts et limites de l'analyse de discours pour la recherche interdisciplinaire et la coproduction de connaissances scientifiques. Le cas d'une analyse lexicométrique d'entretiens avec *Lexico3* », *JADT* 1, 439-450.
- GARRIC N. & CAPDEVIELLE-MOUGNIBAS V. (2009), « La variation comme principe d'exploration de corpus. Intérêts et limites de l'analyse lexicométrique interdisciplinaire pour l'étude de discours », *Corpus* 8, 105-128.
- LEBART L. & SALEM A. (1994), *Statistique Textuelle*, Paris : Dunod.
- MELLET S. (2002) « Corpus et recherches linguistiques : introduction », *Corpus* 1, 5-13.
- MOIRAND S. (2007), *Les discours de la presse quotidienne. Observer, analyser, comprendre*, Paris : Presses Universitaires de France.
- PÊCHEUX M. (1969), *Analyse automatique du discours*, Paris : Dunod.
- PÊCHEUX M. (1975), *Les vérités de la Palice*, Paris : Maspéro.
- PINCEMIN B. (1999), « Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative », in A. Condamines, M.-P. Péry-Woodley & C. Fabre (éds), *Atelier Corpus et TAL : pour une réflexion méthodologique* (TALN 99), Cargèse (Corse, France), 26-36.
- PINCEMIN B. (2012), « Hétérogénéité des corpus et textométrie », *Langages* 187, Paris : Larousse/Armand Colin (ce volume).
- RASTIER F. (2005), « Enjeux épistémologiques de la linguistique de corpus », in G. Williams (éd.), *La linguistique de corpus*, Rennes : Presses Universitaires de Rennes, 31-45.
- SASSIER M. (2008), « Genre, Registre, Formation discursive et corpus », *Langage & Société* 124, 39-57.
- TOGNINI-BONELLI E. (2001), *Corpus Linguistics at Work*, Amsterdam: John Benjamin's Publishing.
- VAN DIJK T.-A. (2008), *Society in Discourse. How Context Controls Text and Talk*, Cambridge: Cambridge University Press.