



HAL
open science

De l'émergence à l'impact social des discours : hétérogénéités d'un corpus

Georgeta Cislaru, Frédérique Sitri

► **To cite this version:**

Georgeta Cislaru, Frédérique Sitri. De l'émergence à l'impact social des discours : hétérogénéités d'un corpus. *Langages*, 2012, 187, pp.59-72. hal-00730267

HAL Id: hal-00730267

<https://hal.science/hal-00730267>

Submitted on 20 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

De l'émergence à l'impact social des discours : hétérogénéités d'un corpus

Georgeta Cislaru, Syled-EA 2290, Université Sorbonne nouvelle
Frédérique Sitri, Syled-EA 2290 et Université Paris Ouest Nanterre

Résumé

Nous discuterons dans cet article de la constitution d'un corpus de rapports éducatifs relevant du champ de la protection de l'enfance dont nous montrerons le caractère nécessairement complexe et hétérogène. Se situant dans le cadre de l'analyse du discours, notre réflexion vise à apporter quelques éléments de réponse à la question des limites et de l'homogénéité d'un corpus, notamment du point de vue des objectifs interprétatifs de la discipline. Cette question sera abordée par le biais de la contextualisation du corpus en tant que démarche de construction d'un savoir sur le domaine, de la prise en compte des déterminations interdiscursives dont la matérialité langagière constitue la trace, et enfin de l'analyse génétique de brouillons de rapports éducatifs, qui rend compte de l'inscription de contraintes diverses au fil du processus d'écriture.

Mots-clefs : analyse du discours, corpus, genre, rapport éducatif, génétique textuelle, interdiscours, contextualisation

Abstract

This paper brings under scrutiny the conditions of constituting a corpus of social reports that emanate from child protective services. It aims at showing the inexorable complexity and heterogeneity of such a corpus. From the standpoint of French Discourse Analysis, we attempt answering the question of the limits of homogeneity in corpus studies, due to the interpretive aims of the discipline. We bring evidence from three domains: i) the corpus contextualization, as a mean for building domain knowledge; ii) the analysis of the interdiscourse inasmuch as it is traceable and linguistically marked; iii) the genetic study of the drafts of social reports, pointing out the impact of various constraints all the way through the writing process.

Key words: discourse analysis, corpus, genre, child welfare report, text genetics, interdiscourse, contextualization

Nous aborderons dans cet article la question de la complexité et de l'hétérogénéité des corpus en analyse du discours (désormais AD), question abordée entre autres par Moirand (2004), Rastier & Pincemin (1999), Mayaffre (2002), etc. Nous défendrons le point de vue selon lequel l'hétérogénéité d'un corpus, loin d'altérer son caractère unitaire, contribue, au contraire, à améliorer les conditions d'analyse dans la mesure où elle est induite par les principes théoriques et méthodologiques de l'AD. Corollairement, nous serons amenées à envisager la question de l'hétérogénéité du genre.

Le corpus en AD se définit dans une double relation de dépendance par rapport à une problématique contextualisée socio-historiquement et à un système de relations textuelles : le corpus n'est qu'un point de vue qui, par des mises en relations textuelles, trouve sa dimension sociale ou discursive. Généralement, les réflexions méthodologiques ou théoriques sur la notion de corpus s'attachent à développer l'un (Moirand) ou l'autre rapport (Rastier). Or, pour saisir la complexité et la nécessaire hétérogénéité d'un corpus – qui reste néanmoins « unitaire » - il nous a semblé opportun d'articuler les deux points de vue. Pour ce faire, nous recourons à un corpus très marqué par son ancrage socio-professionnel : des rapports éducatifs visant à évaluer une situation familiale et à préconiser le cas échéant une mesure de

protection pour l'enfant en risque de danger. Ce corpus a déjà donné lieu à un certain nombre d'investigations en analyse du discours et sémantique discursive (voir Cislaru *et al.* éds, 2008) ayant permis d'appréhender toute une série de particularités sémantico-discursives le caractérisant ; une partie de ces données sera convoquée ici pour éclairer l'approche épistémologique concernant la question de l'homogénéité/hétérogénéité.

Nous adoptons une démarche croisant verticalité interdiscursive et linéarité chronologique ; notre analyse prend en effet en compte le contexte socio-professionnel encadrant la production de ces écrits et la dimension génétique (par l'étude des brouillons). Le corpus a été constitué en plusieurs temps et comprend plusieurs volets, l'objectif étant de rendre compte de la conception que l'on peut avoir d'une production discursive sous-tendue par des discours tels que les textes de loi, les discours médiatiques, les guides de rédaction, les discours de la psychologie ou de la psychanalyse, etc. et s'inscrivant dans une pratique sociale. Dans une démarche caractéristique de l'AD, la constitution du corpus suppose donc ici de se doter de savoirs socio-historico-institutionnels sur le champ dans lequel sont produits et circulent les discours analysés.

1. Hétérogénéités et homogénéité du corpus en analyse du discours

Avant d'aborder notre corpus sous l'angle des hétérogénéités gouvernant sa constitution, il convient de revenir sur le statut du corpus en AD. En effet, si la question des hétérogénéités du corpus est récurrente et abondamment traitée dans certains domaines comme la linguistique de corpus ou le traitement automatique des langues (voir Biber 1990, Illouz *et alii* 1999, Pincemin ici-même), où le couple homogénéité/hétérogénéité se définit *grosso modo* en termes de représentation statistique des données langagières¹, elle ne se pose pas de la même manière en analyse du discours, où l'unité du corpus est souvent posée comme une des conditions premières de son recueil².

1.1. Le corpus en analyse du discours

En AD, le corpus n'est pas seulement construit, comme dans la plupart des domaines de la linguistique, en fonction d'un objectif de recherche ; il est par ailleurs contextualisé et mis en relation avec des « conditions de production », avec des pratiques sociales, plus largement avec des extérieurs qui le déterminent.

Ce double aspect du corpus est donc définitoire de la discipline et justifie pleinement un questionnement sur les limites du corpus, autant vis-à-vis de l'objectif de recherche que vis-à-vis de sa contextualisation. Ainsi, un corpus doit-il répondre aux critères de pertinence pour justifier son usage en tant que lieu de vérification d'une hypothèse de travail ; il doit répondre aux critères de représentativité, d'homogénéité et de différence afin de rendre possible la mise en série des données et la généralisation/théorisation.

Quant à la contextualisation, elle va au-delà de l'herméneutique intertextuelle prônée par Rastier, qui ne semble envisager les relations intertextuelles qu'au sein d'un genre ou entre genres relevant d'un même discours – c'est-à-dire d'un même domaine d'activité (Rastier 2001 : 83).

Dans le cas de l'AD, il s'agit d'une contextualisation qui assume une porosité entre le discours et les pratiques sociales, dans la mesure où ils se co-déterminent, ou, sous un autre point de vue, entre le discours et des principes de détermination dont rendent compte les

¹ On peut citer ici Fischer-Starcke (2010: 185) : "From a linguistic perspective, this means that data is homogeneous when its language, e.g., its lexis, grammar, and phraseology, its statistically significantly similar in these categories. Conversely, data is heterogeneous when its data differs statistically significantly in these categories." L'objectif peut être alors de caractériser des corpus en tant que tels via leur homogénéité interne, comme le fait par exemple Kilgarriff (2001).

² Unité veut dire cohérence et non homogénéité : comme le souligne Maingueneau (2011) le chercheur en AD peut travailler sur des unités non topiques, c'est-à-dire des unités « construites par le chercheur, qui ne se conforment pas aux découpages préétablis par l'activité verbale » (2011 : 92).

notions de formation discursive, d'interdiscours ou de préconstruits³. Ces notions, on le verra, peuvent recouper celle de contextualisation mais ne peuvent être confondues avec elle.

Du point de vue de l'AD, le corpus ne peut donc être assimilé à une base de données, à une liste de textes, ou à un recueil d'écrits. Ce positionnement rend la question de l'homogénéité complexe car c'est à deux niveaux qu'elle devrait se réaliser : celui de la nature textuelle et celui du contexte de production. La « clôture » du corpus prônée aux débuts de l'AD a très vite, on le sait, été remise en question : le discours est conçu comme constitutivement hétérogène, et la clôture du corpus devient impossible à tenir.

Cette redéfinition du corpus par la reconnaissance de l'hétérogénéité discursive ouvre de nouvelles perspectives d'analyse avec la prise en compte de l'hétérogénéité énonciative en particulier à partir des travaux d'Authier-Revuz (1995) et la remise en cause de la linéarité langagière, notamment par les approches textométriques (Salem 1987, Viprey 2006).

1.2. Le genre, garant d'homogénéité ?

Aujourd'hui, suite, entre autres, à la diffusion des travaux de Bakhtine⁴ et, sans doute, aux questions soulevées par la constitution de « grands corpus », le genre tend à jouer un rôle important dans la constitution du corpus ; on préconise donc des corpus constitués sur la base d'une unité générique (Rastier 2002).

Si l'on peut considérer que le genre, en tant qu'il est une interface entre le discursif et le social, et qu'il cristallise d'une certaine façon les déterminations et les contraintes qui pèsent sur le discours, constitue effectivement une unité pertinente en analyse du discours, cette notion n'a pas pour autant à même d'assurer une homogénéité absolue des données.

En effet, si l'on conçoit le genre comme une construction souple et dynamique, ouverte sur l'extérieur, on peut rendre compte d'une certaine perméabilité entre genres, mais aussi entre genres et sphères d'activité – bref, on fait ressortir l'hétérogénéité des genres (Mellet, Rinck & Sitri à paraître). Par conséquent, il nous semble que la notion de genre n'est pas un garant d'homogénéité en tant que telle, comme nous allons le montrer à propos du corpus d'écrits sociaux sur lequel nous travaillons.

2. Les écrits sociaux, unité générique et hétérogénéités interdiscursives

Le corpus que nous étudions dans cet article est constitué d'écrits produits dans le cadre de la protection de l'enfance. Ce sont des rapports éducatifs ayant plusieurs traits en commun :

- Ils ont pour objectif d'évaluer une situation familiale et de préconiser le cas échéant des mesures d'intervention ;
- Ils s'inscrivent dans une pratique sociale et constituent des écrits professionnels ;
- Ce sont des écrits sous-tendus par des discours tels les textes de loi, les discours médiatiques, les guides de rédaction, etc.
- Ils sont intrinsèquement hétérogènes du point de vue énonciatif (Moirand 2007 : 12) dans la mesure où ils sont constitués de descriptions et d'évaluations produites par le travailleur social, de discours représentés de la famille, d'analyses ou diagnostics proposés par le psychiatre du service, etc.

Nous avons cherché, dans le cadre de travaux antérieurs, à décrire ces écrits afin d'en identifier les spécificités génériques d'une part (cf. Pugnère-Saavedra 2008 pour une synthèse), à observer la manière dont les contraintes linguistiques, discursives et sociales

³ La notion de formation discursive, complexe, a été en particulier reprise par Maingueneau (2011) ; la notion de préconstruit est retravaillée par Paveau (2006).

⁴ Nous rappelons ici la définition que donne Bakhtine du genre : « Tout énoncé pris isolément est, bien entendu, individuel, mais chaque sphère d'utilisation de la langue élabore ses *types relativement stables d'énoncés*, et c'est ce que nous appelons *genres de discours*. Le vouloir dire d'un locuteur se réalise avant tout dans le choix d'un genre du discours. Ce choix se détermine en fonction de la spécificité d'une sphère donnée de l'échange verbal, des besoins d'une thématique (de l'objet du sens), de l'ensemble constitué des partenaires, etc. » (Bakhtine 1984 [1952-1953] : 284).

s'inscrivent dans le discours en en repérant les traces dans la matérialité langagière d'autre part. Ces objectifs de recherche ont déterminé les principes de constitution du corpus, qui est multidimensionnel.

En effet, la constitution du corpus suppose ici de se doter de savoirs socio-historico-institutionnels sur le champ dans lequel sont produits et circulent les discours analysés. Ainsi, les écrits sociaux (volet 1) ont-ils été situés par rapport à des textes administratifs, juridiques, etc. (volet 2), de sorte que le corpus sur lequel nous avons travaillé est finalement constitué d'éléments hétérogènes par leur nature aussi bien que par leur statut dans la recherche.

2.1. Corpus et contextualisation : un corpus en deux volets

Un premier volet de corpus comprend 18 rapports correspondant à 9870 formes. Il s'agit d'écrits professionnels, produits par des services sociaux dans le cadre d'un signalement d'enfant en danger, à partir de l'observation de la situation familiale et surtout à partir d'entretiens avec l'enfant faisant l'objet du signalement, sa famille et des intervenants extérieurs (école, services sociaux, médecins, etc.).

Les 18 rapports sont issus de mesures différentes et portent des « noms de genre » différents ; on enregistre ainsi des :

- notes d'information ;
- signalements ;
- rapports d'échéance ;
- rapports éducatifs.

Se pose dès lors la question de l'hétérogénéité et de l'appartenance générique de ces écrits. Or, plusieurs traits génériques semblent justifier le regroupement de ces écrits dans un macro-genre, les « rapports éducatifs » (Cislaru et Sitri 2009a) : le contexte professionnel les situe dans le même dispositif communicationnel (textes s'adressant au juge mais aussi à la famille voire à d'autres professionnels) et leur accorde la même visée pragmatique (prise de décision en vue d'une intervention sociale). Sur le plan de la composition et des formes, on observe bien des régularités, au-delà de quelques variations de fréquence concernant des choix lexicaux ou énonciatifs.

La contextualisation des rapports éducatifs nous conduit à la mise en place d'un deuxième volet du corpus. En effet, la production de ces rapports est régie par des textes de loi qui évoluent régulièrement. L'évolution législative est elle-même corrélée à une évolution du discours « sociétal » sur la maltraitance et sur le signalement⁵, et préparée par des rapports produits par des parlementaires ou des organismes paragouvernementaux. Ces écrits sont par ailleurs encadrés par des « guides du signalement » émanant de l'éducation nationale ou des conseils généraux.

Notre volet 2 participe ainsi à la construction d'un « savoir sur le domaine » du signalement, c'est-à-dire à la compréhension des contraintes juridiques, discursives, professionnelles et sociales dans lesquelles évolue le volet 1 ; il consacre le signalement comme acte d'écriture performative, ce qui justifie la prééminence du volet 1, qui a constitué notre objet d'analyse premier. Il s'agit là d'un cadrage institutionnel *a priori*, via

- des textes de loi relatifs à la protection de l'enfance (code civil, code de la famille, arrêtés et décrets) ;
- des textes prescriptifs (guide de rédaction des écrits de signalement) ;
- des textes administratifs (rapports institutionnels et gouvernementaux : Naves-Cathala, ONED, Péresse) ;

⁵ Voir Serre (2009).

- des textes médiatiques (articles et dossiers de presse traitant de la maltraitance infantile et du signalement, tirés des titres suivants : *L'Express*, *Le Figaro*, *L'Humanité*, *Libération*, *Le Monde*, *Le Point*, *Le Sud-Ouest*).

On constate que le lexique « spécialisé » autour de lexèmes tels que « signalement », « maltraitance », « enfant » est en partie partagé entre les deux volets, ce qui constitue classiquement en linguistique textuelle un indice d'homogénéité d'un point de vue statistique. Ces champs lexicaux et les cooccurrences des lexèmes susmentionnés représentent, dans une certaine mesure, des éléments d'homogénéité transversale à l'ensemble des textes ainsi réunis. Dans un autre ordre d'idées, les notions de maltraitance et de signalement apparaissent comme des objets sociaux discursivement configurés, que la constitution du deuxième volet de corpus a permis de mieux cerner.

La connaissance des textes qui encadrent la production des rapports éducatifs et des conditions dans lesquels ils sont produits permet de rendre compte d'un certain nombre de faits textuels, comme

- la violation de la règle de non-contradiction, qui peut être rapportée aux contraintes paradoxales qui pèsent sur les scripteurs, mettant en évidence l'articulation entre déterminations contextuelles et relations textuelles (Cislaru et Sitri 2009a) ;
- la circulation de segments de discours dans les deux volets du corpus, qui peut être complexe dans le détail. Ainsi, le terme de *danger*, un des mots clefs de l'article du code civil qui organise la protection de l'enfance, apparaît régulièrement dans les paragraphes conclusifs des différents types d'écrits sous la forme *danger/être en danger/constitue un danger* et se trouve par ailleurs en lien avec l'expression de la peur, fréquente dans les rapports éducatifs (Cislaru 2008). Inversement, *signalement*, absent des textes législatifs mais présent dans les discours médiatiques et administratifs est d'un usage fréquent dans les rapports éducatifs, pour désigner un acte (le fait de signaler) ou un écrit, même après que l'évolution législative en ait restreint l'emploi (voir Mellet et Sitri 2010).

2.2. Contextualisation et relations interdiscursives ?

On pourrait nous semble-t-il parler d'intertextualité pour qualifier ces relations d'emprunt de texte à texte, même si le fonctionnement n'en est pas identique dans le cas des termes issus directement des textes de loi et dans le cas des termes plutôt issus d'un ensemble médiatique ou professionnel⁶.

Cependant, du point de vue de l'AD, le volet 2 ne peut être assimilé totalement à l'interdiscours du volet 1 : il ne permet pas de saisir clairement la façon dont le discours des écrits est traversé et constitué de discours autres dont les traces sont repérables dans des formes de langue – ce dont rend compte précisément la notion d'interdiscours :

[...] tout discours dominé est tissé de discours dominants qui lui sont intégrés, (que) les frontières discursives ne sont pas atteignables, (que) le savoir antérieur s'inscrit dans la construction d'une connaissance et se repère à travers des formes de langue. Autrement dit, hétérogénéité et antériorité de l'interdiscours s'inscrivent à l'intérieur même de l'intradiscours, elles n'en constituent pas le contexte. (Mazière 2005 : 58)

En particulier la notion d'interdiscours permet d'atteindre des énoncés non nécessairement attestés dans un corpus donné, que le discours en train de se tenir met en circulation autant qu'il les reprend. Dans cette perspective, on considère que la norme sociale

⁶ De fait la distinction faite ici entre « intertextualité » et « interdiscours » nous semble devoir être encore creusée, en relation avec le statut des « marques » ou des indices de présence d'un discours autre. La répétition à l'identique peut-elle avoir statut de marquage ?

qui sous-tend l'évaluation produite par les travailleurs sociaux peut être mise au jour dans les écrits de signalement à travers des formes telles que, par exemple, la concession ou la négation (Cislaru et Sitri 2009a).

Les guillemets constituent également une marque d'hétérogénéité : en tant que guillemets de modalisation autonymique, ils peuvent être interprétés comme indiquant que le mot ou l'expression guillemetée est « emprunté » à un autre discours. On les trouve en particulier autour des mots empruntés à un interdiscours psychologique et/ou psychanalytique, interdiscours a priori difficilement localisable via des corpus (discours-source de psychanalystes ? discours de vulgarisation dans les magazines ? discours de formation ? réunions de service ?).

Exemple 1

Sa relation « **fusionnelle** » avec sa mère laisse peu de place au père (signalement, CG)

Ainsi donc ces formes que l'on pourrait appeler « dialogiques » (au sens de Bakhtine) inscrivent bien dans le texte du rapport la trace d'un interdiscours que notre volet 2 (« corpus de contextualisation) ne permet pas d'atteindre. Cet état de fait n'est pas sans rappeler l'hypothèse de Guilhaumou (1996 : 107) concernant la contextualisation « interne » à l'énoncé : « une fois inséré, par la lecture des archives, dans une configuration signifiante, l'énoncé est à lui-même son propre contexte. »

2.3. Conclusion : un corpus contextualisé et ouvert

Cette distinction entre contextualisation par un corpus construit par l'analyste et interdiscours saisi et configuré par des formes de langue dans le discours en train de se tenir pose la question de la clôture du corpus.

On discutera ici les propositions de Mayaffre concernant la constitution d'un corpus « réflexif », que nous rejoignons dans la volonté de limiter le recours à des contextes aléatoires et subjectivement reconstitués par l'analyste dans le processus d'interprétation :

La linguistique de corpus ou la sémantique du discours ont définitivement démontré que pas plus qu'on ne pouvait comprendre un mot sans la phrase et la phrase sans le discours, on ne pouvait comprendre le discours sans l'interdiscours, le texte sans le co-texte (sans même parler ici du hors-texte), c'est-à-dire aussi et de manière plus générale, le corpus sans le hors corpus.

Tant est si bien que les efforts scientifiques qui président à la constitution des corpus textuels comme à leurs traitements rigoureux semblent anéantis au moment du bond interprétatif qui nous projette dans la lave d'un intertexte indéterminé, appréhendé intuitivement. (Mayaffre 2002)

En revanche, compte tenu de ce que nous avons montré jusqu'à présent, la clôture du corpus impliquée nous semble-t-il par sa réflexivité nous paraît contestable. Moirand (2004 : 90) souligne cet écueil en invoquant « l'impossible clôture des corpus médiatiques », où la mémoire des mots instille constamment de la verticalité dans le fil horizontal du discours.

C'est sans doute la question de la distinction entre intertexte et interdiscours qui se profile ici, celle aussi de la possibilité d'une détermination univoque du sens⁷. En effet, dans notre conception, l'interdiscours est saisi par des formes qui signalent son émergence dans le discours en train de se tenir : il ne peut être considéré comme aléatoire ou relevant

⁷ Cette position conduit à questionner la proposition de Mayaffre (2002), pour qui : « l'analyse du **sens profond** de cet extrait réclame ici, de manière évidente, le recours à des ressources co-textuelles que l'on ne saurait envisager de traiter différemment que les ressources intratextuelles du passage » [c'est nous qui soulignons].

uniquement de l'intuition de l'analyste ; cependant il ne constitue pas un extérieur toujours balisable en termes de corpus, et est donc non prévisible, non déterminable *a priori*.

Se pose dès lors également la question du rapport que le genre entretient avec l'interdiscours : le genre inclut-il toujours dans sa définition les « extérieurs » qui le traversent, comme cela semble être le cas pour un genre institutionnel comme les rapports éducatifs ? Autrement dit, nous considérons que, d'un point de vue interprétatif, le genre n'est pas le seul contexte pertinent pour l'interprétation d'une forme sauf si on inclut dans la définition du genre l'interdiscours propre à ce genre, ce qui nous conduit nécessairement à recueillir des corpus potentiellement hétérogènes.

3. Hétérogénéités en diachronie

Une deuxième phase de la recherche nous a conduits à procéder à un nouvel élargissement du corpus, permettant de mettre au jour comment, dans le processus de l'écriture, se jouent le rapport aux contraintes de divers ordres identifiés dans la première partie de la recherche⁸. La dimension diachronique est représentée par un corpus d'états successifs d'une série de rapports éducatifs. Ces rapports sont produits dans le cadre de l'accueil d'enfants dans une structure foyer/famille d'accueil (bilan, échéance⁹) et possèdent globalement la même visée que les écrits constituant le précédent corpus : description/évaluation orientée vers l'accueil possible ou pas/le maintien ou pas de la mesure, ou proposition d'orientation le cas échéant.

Les états successifs se présentent soit sous forme de brouillons « papier », les modifications étant notées sur le manuscrit, soit sous forme de versions électroniques, correspondant à différents états d'un texte enregistrés par les scripteurs volontaires pour l'expérience¹⁰. La méthode d'analyse consiste en l'observation des opérations de réécriture identifiées par la génétique textuelle (Fuchs *et alii* 2003 [1982]) : ajout, remplacement, déplacement, suppression.

3.1. Dynamiques du texte et hétérogénéités

Le processus d'écriture fait intervenir plusieurs types d'hétérogénéités.

D'une part, les corrections manuscrites en marge, ainsi que les ajouts et suppressions sur traitement de texte, constituent des données hétérogènes – l'avant-texte est hétérogène, comme le souligne Pétilon (2011). On relève

- des hétérogénéités sémiotiques (écriture en marge, biffures, etc. sur les manuscrits, majuscules et couleurs vouées à disparaître sur traitement de texte) ;
- des hétérogénéités chronologiques (opérations de réécriture échelonnées dans le temps) ;
- des hétérogénéités « autoriales » (intervention du chef d'équipe, écriture à quatre mains, etc.).

Les hétérogénéités sémiotiques et chronologiques ne seront pas abordées dans ce travail.

D'autre part, l'interprétation des opérations de réécriture s'appuie sur la prise en compte des contraintes d'ordre divers qui pèsent sur ces textes, dont

- les contraintes qui font partie de l'interdiscours des textes analysés (on parlera alors d'hétérogénéités interdiscursives) ;
- les contraintes pragmatiques déterminées par la situation de production du texte et par sa visée.

⁸ Cette étude émane du projet ANR Écritures (2011-2014), www.univ-paris3.fr/anr-ecritures/

⁹ Les travailleurs sociaux rédigent des rapports de parcours, qui constituent des bilans de la situation de l'enfant suivi, ainsi que des rapports d'échéance présentés au juge pour enfants ; ces derniers s'appuient sur les bilans rédigés régulièrement.

¹⁰ Un corpus d'écriture en temps réel, enregistré à l'aide du logiciel Inputlog (Leuten & Van Waes 2006), est en cours de constitution.

Par exemple, la gestion de la prise en charge des dires avec attribution explicite des dires au jeune renvoie aux consignes explicites des guides de rédaction demandant de bien baliser ce qui a été dit par les enfants :

Exemple 2

ajout manuscrit : pour ensuite la préparer à arrêter de fumer. Muriel n'est pas **prête** à supporter l'arrêt brutal du tabac

⇒ En effet, Muriel **dit vouloir** diminuer sa consommation du tabac, mais **dit aussi** que ce n'est pas encore...

Mais ces opérations de réécriture viennent également conforter des hypothèses que nous avons pu formuler ailleurs (cf. Cislaru et Sitri 2009b) concernant le discours de la famille et de l'enfant comme réponse à une attente sociale. Ici, l'hétérogénéité énonciative (le discours de la famille) est subordonnée à de l'hétérogénéité interdiscursive (les attentes sociales telles qu'elles circulent au sein de l'institution de la protection infantile) afin d'assurer au texte une homogénéité socio-pragmatique. C'est notamment l'analyse des segments de discours répétés *X a pu/ peut dire* et leur récurrence dans le corpus électronique qui rend compte de cette prégnance des attentes sociales (cf. Sitri 2011).

La gestion de l'interdiscours-même est parfois donnée à voir par les opérations de réécriture, à l'instar des modifications affectant les termes renvoyant à un interdiscours « psy »¹¹: ajout de guillemets dans un cas, suppression d'un terme relevant explicitement du discours « psy » dans l'autre :

Exemple 3

Aujourd'hui, nous estimons, bien que l'encoprésie soit encore présente, être moins **envahis** et pouvoir travailler avec le jeune ⇒ « **envahis** »

Exemple 4

Cependant malgré la personnalité en ~~faux-self~~ mature et responsable que Muriel veut afficher.

Ces modifications témoignent d'un travail conscient sur les formes de la présence de ce discours, discours « autre », certes, mais qui contribue à façonner d'une certaine manière le discours professionnel des travailleurs sociaux¹². Elles témoignent aussi d'une « négociation » avec la représentation que se fait le scripteur du destinataire de ces écrits.

L'analyse des segments répétés possessifs met au jour d'autres manières d'intégrer au discours les normes sociales. Beaucoup d'informations semblent venir du discours de l'enfant médié par le scripteur (*il a pu **exprimer son** désir ; Anthony nous évoque aucun élément de **son** vécu antérieur*) mais peu de constructions possessives semblent être le fait de l'enfant. Sont notamment à interpréter sous l'influence des contraintes discursives et des normes sociales les segments répétés possessifs qui formulent des attitudes et faits concernant l'enfant et pouvant servir de marqueurs de diagnostic dans l'évaluation : *son désaccord, son attachement, ses difficultés d'apprentissage*, etc. Ces segments répétés sont directement concernés par le processus de réécriture, faisant l'objet de ratures, reformulations, ajouts.

C'est dans le corpus diachronique que l'on voit plus clairement comment le discours est traversé par des hétérogénéités qui lui sont, de fait, constitutives, car le corpus de contextualisation y « laisse des traces », on repère au fil des opérations de réécriture des

¹¹ Le qualificatif « psy » renvoie à la fois à « psychologique » et à « psychanalytique ».

¹² Nous avons travaillé avec un service qui revendique une approche « psy ».

marques d'interdiscours modelant les versions successives des textes et produisant au final un objet unitaire constitué d'hétérogénéités.

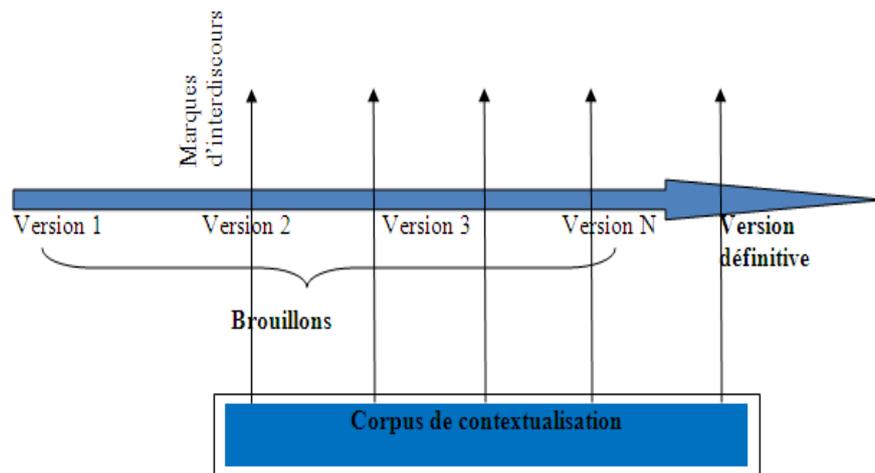


Schéma 1. Niveaux d'inscription des différentes hétérogénéités dans les brouillons

On y observe de fait la consolidation

- d'un genre et de ses normes plus ou moins locales ou générales ;
- d'un discours professionnel, avec ses routines discursives qui assurent la « pérennisation » de l'institution et de ses normes.

Conclusion

L'homogénéité et l'hétérogénéité d'un corpus peuvent être appréhendées soit dans l'optique des conditions de production, comme ce fut le cas de l'AD à ses débuts, soit dans l'optique des choix langagiers (lexicaux, syntaxiques, etc.), comme le font la linguistique de corpus ou la textométrie, soit en articulant les deux dimensions, la matérialité langagière représentant des traces des conditions de production. Cette troisième voie est la nôtre. Nous avons essayé de montrer, en croisant un regard synchronique et un regard diachronique sur les rapports éducatifs, que les hétérogénéités peuvent être constitutives de l'unité d'un corpus. En effet, un rapport éducatif ou une enquête sociale se révèle un objet textuel complexe :

- inscrit dans une pratique professionnelle qui est encadrée par des textes juridiques et administratifs et qui est prise dans une problématique sociale et politique dont les médias sont le vecteur (le signalement et la maltraitance comme objets sociaux) ;
- actualisant des normes d'évaluation accessibles par des formes telles que la négation ou la concession (entre autres) ;
- relevant d'un discours professionnel lui-même hétérogène (catégorisations relevant du discours « psy », éducatif) mais semble-t-il caractérisé par des routines d'écriture.

Pour rendre compte de cette complexité il n'est pas étonnant que l'on doive configurer un corpus lui-même « hétérogène ». Mais le caractère unitaire de ce corpus est garanti par :

- les différences de statut entre les éléments du corpus qui donnent la prééminence au volet de rapports éducatifs ;
- et par là-même l'unité générique : les différents éléments de complexité relèvent en effet de caractéristiques énonciatives et pragmatiques du macro-genre « rapport éducatif ».

Dans la tradition de la linguistique de corpus, on considère le corpus, outil homogène, comme exempt de tout indice interprétatif : « A corpus can offer evidence but cannot give information. [...] The corpus simply offers the researcher plenty of examples; only intuition

can interpret them. » (Hunston 2002 : 23). Or, la contextualisation du corpus et la prise en compte de son interdiscours nous ont permis d'identifier des éléments de cohérence/cohésion discursive qui contribuent à en proposer une lecture... homogène. En AD, d'une certaine façon, c'est l'hétérogénéité du corpus qui semble en fin de compte inéluctable si l'on admet que l'opération de contextualisation et la prise en compte des déterminations interdiscursives sont constitutives du caractère interprétatif de la discipline.

Bibliographie

- ANGENOT M. (1989), *Mil huit cent quatre-vingt-neuf : un état du discours social*, Montréal : Éditions du Préambule.
- AUTHIER-REVUZ J. (1995), *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris, Larousse.
- AUTHIER-REVUZ J. (2001), « Le discours rapporté », in R. Tommassone (dir.), *Une langue : le français*, Paris : Hachette, 192-201.
- BAKHTINE M. (1984), *Esthétique de la création verbale*, Paris, Gallimard.
- BIBER D. (1990), "Methodological issues regarding corpus-bases analyses of linguistic variation", *Literary and Linguistic Computing* 5/4, 257-270.
- CISLARU G. (2008), « L'intersubjectivation comme source de sens : expression et description de la peur dans les écrits de signalement », *Les Carnets du Cediscor* 10, 117-136.
- CISLARU G. & SITRI F. (2009a), « Texte et discours. Corpus, co-texte et analyse automatique du point de vue de l'analyse de discours », *Corpus* 8, 85-104.
- CISLARU G. & SITRI F. (2009b), « La représentation du discours autre dans des signalements d'enfant en danger : une parole interprétée ? », in J. M. Lopez Muñoz, S. Marnette, L. Rosier & D. Vincent (éds), *La circulation des discours*, Québec : Nota Bene, 57-73.
- CISLARU G., SITRI F. & PUGNIERE-SAAVEDRA F. (éds) (2008), « Analyse de discours et demande sociale : le cas des écrits de signalement » *Les Carnets du Cediscor* 10, Paris : PSN.
- FISCHER-STARCKE B. (2010), *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*, London: Continuum.
- FUCHS C., GRESILLON A., LEBRAVE J.-L., PEYARD J. & REY-DEBOVE J. (2003) [1982], *La genèse du texte. Les modèles linguistiques*, Paris : CNRS Editions.
- GUILHAUMOU J. (1996), « Vers une histoire des événements linguistiques. Un nouveau protocole d'accord entre l'historien et le linguiste », *Histoire, Épistémologie, Langage* 18/2, 103-126.
- HUNSTON S. (2002), *Corpora in applied linguistics*, Cambridge: Cambridge university press.
- ILLOUZ G., HABERT B., FLEURY S., FOLCH H., HEIDEN S. & LAFON P. (1999), « Maîtriser les déluges de données hétérogènes », in *Actes de l'Atelier thématique TALN 1999 : Corpus et Traitement Automatique des Langues : pour une réflexion méthodologique*, TALN'99, 12-17 juillet 1999, Cargèse.
- KILGARRIFF A. (2001), "Comparing Corpora", *International Journal of Corpus Linguistics* 6/1, 1-37.
- LEUTEN M. & VAN WAES L. (2006), « Inputlog: New Perspectives on the Logging of On-Line Writing », in K.P.H. Sullivan & E. Lindgren (éds), *Studies in Writing: Vol. 18. Computer Key-Stroke Logging and Writing: Methods and Applications*, Oxford: Elsevier, 73-94.
- MAINGUENEAU D. (2011), « Pertinence de la notion de formation discursive en analyse de discours », *Langage et société* 135/1, 87-99.
- MAZIERE F. (2005), *L'analyse du discours*, Paris : PUF, Que sais-je ?
- MAYAFFRE D. (2002), « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus* 1. Disponible sur : <<http://corpus.revues.org/document11.html>>

- MAYAFFRE D. (2004), « Formation(s) discursive(s) et discours politique : l'exemplarité des discours communistes *versus* bourgeois durant l'entre-deux-guerres », *Texto !* juin 2004 [en ligne]. Disponible sur : <http://www.revue-texto.net/Inedits/Mayaffre/Mayaffre_Formations.html> (Consulté le 21/10/2010).
- MELLET C. & SITRI F. (2010), « Nom de genre et institutionnalisation d'une pratique discursive : le cas de l'interpellation parlementaire et du signalement d'enfant en danger », in F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada & S. Prévost (éds) *Actes du CMLF 2010*, 1631-1646. Disponible sur : <http://www.linguistiquefrancaise.org/index.php?option=com_toc&url=/articles/cmlf/abs/2010/01/contents/contents.html>
- MELLET C., RINCK F. & SITRI F. (à paraître), « Hétérogénéité des textes, hétérogénéité des genres », journée Conscila, Modèles et théories du genre en confrontation, 21 octobre 2011, *Pratiques*.
- MOIRAND S. (2004), « L'impossible clôture des corpus médiatiques. La mise au jour des observables entre catégorisation et contextualisation », *TRANEL* 40, 71-92.
- MOIRAND S. (2007), *Les discours de la presse quotidienne*, Paris : PUF.
- PAVEAU M.-A. (2006), *Les prédiscours*, Paris : PSN.
- PECHEUX M. (inédit), « Lecture et mémoire : projet de recherche », in D. Maldidier (1990), *L'inquiétude du discours. Textes de Michel Pécheux*, Paris : Éditions des cendres.
- PECHEUX M. & FUCHS C. (1975), « Mises au point et perspectives à propos de l'analyse automatique du discours », *Langages* 37, 7-80.
- PETILLON S. (2011), « Linguistique génétique, Corpus et Style », *Cahiers de l'AFLS* 16/2, Disponible sur : <<http://www.afls.net/cahiers/16.2/6.%20Petillon.pdf>>
- PUGNIERE-SAAVEDRA F. (2008), « Quelques régularités des écrits du signalement », *Les Carnets du Cediscor* 10, 21-36.
- RASTIER F. (2001), *Arts et sciences du texte*. Paris : PUF.
- RASTIER F. (2002), « L'accès aux banques textuelles – des genres à la doxa », *Texto !* juin 2002 [en ligne]. Disponible sur : <http://www.revue-texto.net/Inedits/Rastier/Rastier_Acces.html>.
- RASTIER F. & PINCEMIN B. (1999), « Des genres à l'intertexte », *Cahiers de praxématique* 33, 83-111.
- ROUSSEAU P. (2007), *Pratique des écrits et écriture des pratiques*, Paris : L'Harmattan.
- SALEM A. (1987), *Pratique des segments répétés*, Paris : Klincksieck.
- SERRE D. (2009), *Les Couloirs de l'État social. Enquête sur les signalements d'enfant en danger*, Paris : Raisons d'agir.
- SITRI F. (2011), « Une lecture événementielle de *pouvoir* », Colloque « Langage, discours, événements », Florence, mars 2011, à paraître dans les Actes du Colloque.
- VIPREY J.-M. (2006), « Quelle place pour les sciences des textes dans l'Analyse de Discours ? », *Semen* 21. Disponible sur : <<http://semen.revues.org/1995>>