

Construire un corpus pour des façons de parler non standard : « Multicultural Paris French »

Constructing a corpus for non standard ways of speaking: MPF

Françoise Gadet et Emmanuelle Guerin



Édition électronique

URL : <http://journals.openedition.org/corpus/3049>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 15 octobre 2016

ISSN : 1638-9808

Référence électronique

Françoise Gadet et Emmanuelle Guerin, « Construire un corpus pour des façons de parler non standard : « Multicultural Paris French » », *Corpus* [En ligne], 15 | 2016, mis en ligne le 15 janvier 2017, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/corpus/3049>

Ce document a été généré automatiquement le 8 septembre 2020.

© Tous droits réservés

Construire un corpus pour des façons de parler non standard : « Multicultural Paris French »

Constructing a corpus for non standard ways of speaking: MPF

Françoise Gadet et Emmanuelle Guerin

- 1 L'observation de parlars non standard suppose un recueil de données illustrant des pratiques réglées selon des normes locales, en marge des règles reconnues comme légitimes. Afin de constituer un tel corpus, il faut commencer par questionner les conditions auxquelles ces pratiques prennent place. L'opposition traditionnelle entre situation formelle vs informelle (en supposant qu'on sache la circonscrire), suffit-elle à garantir l'émergence de faits de langue marginaux ? Peut-on déterminer en amont les types de locuteurs concernés ? Pour s'assurer que l'on a affaire à des données non standard, la sélection de locuteurs et de situations d'interaction (entretiens avec un chercheur, échanges entre pairs, prises de parole publique...) doit ainsi être problématisée. Autrement dit, il n'est pas certain que l'identité des locuteurs dans certains contextes suffise pour actualiser un type de parler donné. Partant, un corpus de données non standard semble une entreprise à peu près irréalisable.
- 2 Nous présentons ici un corpus tenant compte de ces difficultés, MPF (*Multicultural Paris French*). Si, comme les autres corpus, il n'est pas exempt d'*a priori* catégoriels (choix de territoires socialement et géographiquement déterminés, choix de locuteurs), ceux-ci sont réévalués au fur et à mesure de l'observation des données recueillies. Le projet a été mis en place dans le cadre d'une ANR (09 FRBR-037-01, avec une contrepartie britannique, *Multicultural London English*). Après la fin de l'ANR, le projet continue en s'étant recentré sur le français et non plus sur la comparaison¹. Le corpus ainsi constitué en région parisienne récuse la sélection des enquêtés à travers les seuls critères de catégories externes tels qu'ils sont assurés par la tradition de corpus antérieurs ou par des points de vue privilégiant des quadrillages sociodémographiques, où les questionnements sociolinguistiques ne sont pas constitutifs. On a ainsi fait en

sorte que la réalisation d'un grand corpus soit un moment de réflexion théorique et méthodologique, accordant une place centrale à la sociolinguistique.

1. Le corpus MPF, les principes à ses fondements, ses modalités de recueil

- 3 Suite à l'immigration qu'a connue la région parisienne depuis la fin du XIX^e siècle (Noiriel, 2002 ; Blanc-Chaléard, 2001 ; Rabaud, 2014), on peut aujourd'hui parler d'une « Île-de-France multiculturelle », qui se matérialise linguistiquement à travers des effets de contacts de langues sur le français tel qu'il y est parlé. On peut ainsi considérer ces effets comme relevant de formes non standard du français. Dans le droit fil de réflexions de Labov (1972) ou de Kerswill (2010) sur les locuteurs *innovateurs*, on considère que les phénomènes langagiers issus des effets du contact sont observables dans les pratiques de locuteurs répondant aux critères : « jeunes », « de milieu modeste ou populaire », « connaissant des contacts multiculturels réguliers ».

1.1 Des corpus antérieurs

- 4 Toutes les métropoles occidentales ont aujourd'hui affaire à une telle immigration et à la présence (que l'on peut désormais considérer comme pérenne) de populations influencées plus ou moins directement par d'autres cultures. De fait, des études (socio)linguistiques sur ce thème ont été pratiquées dans plusieurs des métropoles européennes, surtout d'Europe du Nord, où le phénomène est un peu plus ancien qu'au sud (voir les travaux de Cheshire *et al.*, 2011, sur Londres ; Wiese, 2013, sur Berlin ; Kotsinas, 1998, sur Stockholm ; Quist, 2008, sur Copenhague ; Jaspers, 2008, sur Anvers...²). La France, quant à elle, a pris un gros retard dans l'étude des effets linguistiques pouvant ainsi être induits sur le français, alors même qu'elle est l'un des plus anciens pays d'immigration d'Europe (depuis la seconde moitié du XIX^e siècle, où la plupart des autres pays européens étaient des terres d'émigration). Le projet MPF a été constitué avec l'objectif de contribuer à combler ce retard.
- 5 Ce n'est pas qu'il n'existe aucun corpus sur Paris, mais d'une part il y en a peu, d'autre part ceux qui existent n'intègrent pas la question des contacts comme un facteur de dynamique langagière. Parmi les quelques corpus sur la région parisienne (petits ou de taille moyenne), les recueils de données ont en général été construits pour cibler un seul aspect de la langue : soit le niveau phonique – voir Jamin (2004), Fagyal (2010) –, soit le plan lexical – voir les nombreux dictionnaires, dont aucun ne s'appuie réellement sur un corpus (en tous cas visible et publié), même si, dans le meilleur des cas, ils mettent en jeu des relevés et des observations systématiques. Parmi les corpus plus grands sur Paris, on peut évoquer le récent CFPP2000 (Branca-Rosoff *et al.*, 2011). Bien que constitué « en premier lieu pour étudier les modifications qui interviennent dans ce qu'on peut considérer comme un parisien véhiculaire en tension entre le pôle du standard et le pôle du vernaculaire », il exclut volontairement la question du contact. Le corpus MPF adopte une autre position, puisque les effets du contact y sont considérés comme constitutifs d'une forme de vernaculaire. Des corpus à objectifs sociolinguistiques ont aussi été recueillis dans d'autres grandes villes françaises, comme Grenoble (Billiez, 1992), Orléans³, Marseille ou Lille (Pooley, 2012)⁴ ; voir aussi Jamin & Trimaille (2008), Jamin *et al.* (2006) pour des comparaisons entre villes.

- 6 MPF s'inscrit ainsi dans le paysage des corpus de français avec un positionnement original puisqu'il est multi-objectif et situe les contacts de langue comme intégrant la réflexion sur la dynamique des pratiques langagières.

1.2 La méthodologie adoptée dans MPF

- 7 Pour constituer le corpus MPF, on a renoncé à la problématique du vernaculaire théoriquement abordée dans l'opposition entre situations formelles et informelles (Labov, 1972) pour privilégier la qualité des interactions entre les protagonistes, qui correspondait mieux à nos hypothèses. Les options sous-jacentes à ce choix concernent les facteurs susceptibles de produire de la variation et de la diversification langagières dans une même langue : les facteurs décisifs relèvent-ils des situations ? Des genres ? Des caractéristiques sociodémographiques des locuteurs ? Des interactions ? Ces questions ne sont pas indépendantes les unes des autres, et la méthodologie adoptée tient compte de leur imbrication.
- 8 Nous étions conscients du caractère partiel des réponses apportées par un quadrillage sociodémographique (Eckert, 2012) et nous voulions acter dans le corpus même le fait que la réalité sociolinguistique est bien plus complexe (Gadet & Guerin, 2012 ; Gadet, 2013). Les informateurs ont ainsi été sélectionnés sur réseaux, soit préalables aux enregistrements, soit construits pour l'occasion à travers une immersion prolongée, avec l'idée de se rapprocher de l'immédiat communicatif authentique, c'est-à-dire d'interactions entre locuteurs entretenant une réelle connivence. La réflexion méthodologique a ainsi débouché sur une démarche innovante, privilégiant la qualité des échanges conversationnels et le contenu de ce qui est dit, plutôt qu'une illusoire représentativité sociodémographique, quadrillant surtout des facteurs aisément quantifiables en dichotomie ou en continuum, comme le sexe, l'âge, la classe sociale, et éventuellement l'origine ethnique – tous facteurs dont il reste à montrer qu'ils constitueraient bien une clef de la production de diversité langagière (voir Gadet & Wachs, 2015, pour une critique de ce point de vue).
- 9 Bien que MPF joue aussi sur les identités pré-assignées par le chercheur (jeunes, de milieux populaires, multiculturels), le profil des informateurs tient compte de métadonnées éclairant notamment la relation préalable enquêteur/enquêté, et les circonstances de l'enregistrement. C'est l'identité en interaction (Greco *et al.*, 2014) qui prévaut sur les caractéristiques objectivables et quantifiables. Quand cela a été possible, on a fait faire, parallèlement à l'entretien (dit « de proximité » quand l'enquêteur est parvenu à établir une interaction convivente), un enregistrement *écologique*⁵ afin de disposer de matériel pour étudier la variation intra-locuteur : il s'agit d'enregistrements effectués en l'absence de l'enquêteur – parfois en sa présence, mais sans qu'il se trouve alors en posture d'enquêteur (conversations entre pairs, non provoquées). Toutefois, différents obstacles, plus ou moins prévisibles du point de vue méthodologique, ont conduit à ce que le nombre d'enregistrements écologiques soit plus réduit que celui des entretiens.
- 10 Le recueil du corpus a été soutenu d'une mise au point du vocabulaire, la plupart des termes s'avérant problématiques, qu'ils soient issus de la réflexion linguistique, de la reprise sociale (en particulier médiatique) ou encore d'ordre sociologique/ethnographique : *jeunes* (si l'on convient que ce qui aboutit à de la différenciation sociolinguistique dépasse largement le sociodémographique – voir Cheshire, 2005), la

catégorisation ethnique qui tend à ne pas se dire (voir une critique de ce point de vue bien français dans Lagrange, 2010), la caractérisation par le type d'habitat (*banlieue, quartier, cité, ghetto...* et même le fréquent *langue de la rue*) ; mais aussi quant à la façon de parler : *langue des jeunes* ou *parlers jeunes...* (voir Gadet & Hambye, 2014).

2. Le traitement du corpus

- 11 Il n'y a pas lieu de distinguer les exigences pour le recueil et celles des étapes du traitement. Elles sont toutes intégrées à la réflexion qui oriente les choix méthodologiques. Ces derniers étant le reflet du cadre théorique adopté.

2.1 En amont du recueil : les options méthodologiques

- 12 Tout recueil d'une certaine envergure confronte au fait que la qualité des données n'est pas en dichotomie (entretiens en face à face vs enregistrements écologiques) : il y a des continuums entre ces deux pôles, de plusieurs types, d'une complexité telle que les effets n'en sont pas toujours convergents ni prévisibles.
- 13 Constituer un corpus multi-objectifs dont l'exploitation pourra concerner et des phénomènes linguistiques et des contenus susceptibles d'une analyse de discours, impose de songer non seulement à la qualité phonique des enregistrements (objectif facile à atteindre avec les moyens actuels), mais aussi à des qualités intrinsèques du discours, linguistiques, énonciatives, langagières et communicatives. On rend ainsi possible le traitement de certains phénomènes, en croisant les points de vue (voir p. ex. Guerin & Paternostro, 2013 ; ou Moreno & Paternostro, 2014).
- 14 Ce sont les exploitations morphosyntaxiques et discursives qui s'avèrent les plus exigeantes pour la nature des données. Au-delà de l'évidente nécessité de diversifier les genres discursifs pour garantir la diversité des productions (voir Blanche-Benveniste, 2010), on prendra les deux exemples d'exigences quant à la longueur et quant à la complexité des séquences : pourquoi et comment chercher à obtenir du « discours long » et du « discours complexe » ?
- 15 La sociolinguistique et l'analyse de la conversation ont établi qu'il y avait deux occasions sociales privilégiées de production de discours long : les récits, où une suspension conversationnelle autorise le locuteur en place à conserver la parole (voir p. ex. Gülich & Mondada, 2001), et les argumentations. Une condition nécessaire mais non suffisante est la longueur des tours de parole, réduite quand la parole circule vite, ce qui est justement un indice de naturel dans la conversation ordinaire. Aussi, si l'on veut étudier des phénomènes qui ne se manifestent qu'en discours long, il faut une collecte ciblée. Maintenant, qu'est-ce que la longueur des séquences assure ? En tous cas pas la diversité de phénomènes, et on peut de ce point de vue opposer les récits et les argumentations/explications. Dans beaucoup d'échanges ordinaires, en effet, les récits manifestent dans le format long des structures monotones, peu complexes. Ce sont surtout les argumentations/explications qui peuvent receler des structures complexes et des imbrications. Mais il y en a peu dans la conversation ordinaire, car elles répondent à des situations très particulières : discours didactique, recherche d'explications spontanées ou sollicitées, provocation amenant à se justifier. On peut voir à ce propos Blanche-Benveniste (2008) sur des spécialistes expliquant leur expertise, et de nombreux exemples dans Blanche-Benveniste (2010). Lorsque l'on

cherche à recueillir de la parole d'interactions en proximité communicationnelle, obtenir des argumentations ou des explications est difficile, puisque ces situations supposent une connivence et le partage d'un ensemble dense de savoirs et d'expériences à l'origine de nombreux implicites.

- 16 Les entretiens n'excluent pas ces deux genres discursifs, surtout les récits. Thibault & Vincent (1990 : 49), qui ont pratiqué des interviews pour le corpus de Montréal (1984), soulignent que celles-ci tendent à produire des données « monologiques », contraignant l'occurrence de certaines formes. Cependant, il est difficile de corréliser cette tendance avec d'autres éléments en rapport avec la proximité/distance communicationnelle, puisque le corpus est constitué d'entretiens traditionnels, avec des locuteurs inconnus. Ainsi, si le genre discursif favorise l'apparition de tours de parole longs, il reste à déterminer si d'autres paramètres peuvent rectifier la tendance : la reconnaissance d'une connivence (comme la découverte au cours de l'entretien d'une passion commune) aurait-elle un effet sur la production de monologues/dialogues ?
- 17 Quoi qu'il en soit, les modalités d'obtention des données sont liées aux attentes du chercheur, une même propriété pouvant s'avérer qualité ou limite. Nous avons obtenu beaucoup plus de récits, en particulier sur les rapports des jeunes avec la police, que d'argumentations, ce qui n'est pas inattendu.

2.2 La transcription et l'état actuel du corpus MPF

- 18 La transcription a été effectuée sous Praat, par exigence d'alignement son-transcription⁶. En droite ligne des travaux de Blanche-Benveniste, synthétisés en particulier dans un article de 2003, on a renoncé à toute forme de truchage orthographique⁷ (Cappeau *et al.*, 2011 pour des précisions sur les choix de transcriptions dans MPF). On peut dire, comme Baude & Dugua (2011 : 107) :
- Nous avons adopté des principes de base généralement partagés à savoir une transcription orthographique qui conserve les spécificités de l'oral (amorces, disfluences, répétitions, etc.), sans usage de la ponctuation, et avec la segmentation des tours de paroles. (p. 105)
- 19 La transcription est ainsi regardée comme une notation facilitant l'accès à la source sonore. C'est pourquoi la transcription de MPF ne cherche pas à représenter de spécificités phoniques (voir aussi note 8).
- 20 Des conventions ont été établies pour les termes à la graphie peu ou pas du tout stabilisée, en particulier pour les mots empruntés à l'arabe, pour lesquels on rencontre jusqu'à trois ou quatre graphies différentes (par exemple *zaama*, aussi écrit *zarma*, *zaâma*) : ces termes ont été répertoriés dans un petit lexique, utile tout autant pour les transcrip-teurs que pour les réflexions sur le lexique (voir le point 3.1).
- 21 Le corpus comporte, en avril 2016, dans les 820 000 mots transcrits, révisés, anonymisés, traduits quand il y a lieu, ce qui correspond à un peu plus de 60 heures enregistrées. Sont considérées comme « finalisées » les données qui ont parcouru les trois étapes de la transcription, la révision par un chercheur autre que le transcrip-teur – parfois plusieurs, pour les enregistrements particulièrement complexes comportant beaucoup de recouvrements de parole, et de l'anonymisation, pour le son et pour la transcription. À terme, les données ainsi traitées seront mises en ligne à disposition de la communauté scientifique. S'y ajoute une « traduction » de tous les termes considérés comme « non-standard », qu'il s'agisse de verlan, d'argot ou d'emprunts, ces derniers

provenant surtout de l'arabe (mots, parfois alternances codiques, quand l'enquêteur parle lui aussi arabe) ainsi que de l'anglais – mais aussi des néologismes.

- 22 D'autres enregistrements ont été recueillis, et en sont actuellement à différents stades de traitement. Il n'y a pour le moment pas d'annotations. Une vitrine en ligne montre un extrait de 34 enregistrements⁸ : elle offre la possibilité d'écouter et de visualiser une minute (son + transcription) d'une partie des enquêtes.

2.3 Les métadonnées et leur mise en fiches

- 23 Comme il est désormais de coutume, les enregistrements sont accompagnés des consentements des enquêtés et de fiches de métadonnées, concernant à la fois les locuteurs et les modalités d'enregistrement. Elles donnent accès à des renseignements de différentes sortes : sociodémographiques, linguistiques et sur les circonstances de l'enregistrement. Mais elles insistent surtout sur des facteurs autres que sociodémographiques, qui renseignent sur la connivence des interactants lors de l'échange. Elles donnent ainsi une idée, faute de pouvoir l'évaluer plus finement, de la proximité communicationnelle entre interactants (informateur/enquêteur pour les entretiens, informateurs entre eux pour les enregistrements écologiques).
- 24 On cherche ainsi à préciser les caractéristiques objectives des locuteurs et des situations, de même que des spécificités de l'interaction. Dès lors, tous les items n'appellent pas le même traitement dans les fiches de métadonnées : certaines rubriques sont des questions ouvertes, laissant à l'enquêteur la liberté d'évaluer la pertinence de certaines informations. Par exemple, indiquer que l'enquêteur partage avec l'informateur un intérêt pour un style de musique, sans qu'ils entretiennent d'intimité, peut s'avérer déterminant pour la proximité communicationnelle si le thème de la musique est abordé. Inversement, un sujet inconnu de l'un des interactants favorise la distance communicationnelle, même dans un échange entre proches (voir Koch & Esterreicher, 2001).

2.4 Classifier les enregistrements

- 25 Les enregistrements ont finalement été classifiés en 3 groupes, sur la base de la qualité sonore perceptible (telle qu'elle peut être reconnue par un groupe d'évaluateurs de l'équipe) : A) « entretiens traditionnels », B) « entretiens de proximité », et C) « corpus écologiques » – à quoi ont été ajoutés des entretiens avec des « grands témoins »⁹.
- 26 La qualité du lien interpersonnel entre les protagonistes l'emportant sur tout autre critère pour la sélection des enquêtés, les localisations géographiques ne sont pas regardées comme prioritaires, et les enquêtés proviennent de 4 arrondissements parisiens et de 27 villes de la couronne parisienne. On est ainsi très loin aussi bien des monographies que des études de cas.
- 27 Si l'on peut distinguer entre ces trois types d'enregistrements, les frontières entre eux ne sont pas étanches, et il a parfois été difficile de catégoriser, malgré les qualifications fines des interactions. On peut citer l'exemple d'une conversation entre un membre du projet, son conjoint et la sœur de ce dernier. *A priori*, cette situation aurait pu donner lieu à un enregistrement de type B ou C. Mais l'incongruité de la présence de l'enregistreur et du questionnement de l'enquêteur dans un contexte intime a conduit à une production contrôlée, relevant plutôt du type A. De plus, dans un même

enregistrement, il n'est pas rare que différents moments tendent vers des catégorisations différentes. C'est pourquoi, si l'on maintient ce type de classement par commodité, on n'ignore pas son caractère approximatif dans un continuum.

- 28 Le programme se prolongera avec la constitution d'une banque de données (en cours de réalisation) qui retiendra des critères d'interrogation de trois ordres : 1) paramètres, difficilement quantifiables, concernant la proximité/distance entre les interactants, 2) phénomènes linguistiques et langagiers, 3) aspects sociodémographiques (ville, âge, sexe des interactants, origine ethnique...). L'architecture même de cette banque de données reflète ainsi le fait que la priorité ne va pas aux paramètres sociodémographiques en tant qu'ils constitueraient une clé de la diversification des façons de parler : la complexité et l'intrication des différents facteurs doivent être reflétées.

3. Quelques exploitations pour un corpus

- 29 Quant aux exploitations d'ores et déjà permises pour ce corpus, on en prendra quatre exemples, en diversifiant les domaines et les auteurs. Ils concernent des thèses, soutenues ou en cours, et/ ou des communications et articles. Les trois premiers exemples portent sur des formes linguistiques, le dernier sur les représentations sur les langues.

3.1 Plan lexical

- 30 Un premier domaine concerne le lexique, avec un inventaire et un classement de termes d'argot, de verlan, d'emprunts et de néologismes. Il s'agit, au-delà de la collection de mots, de tenter de dégager les matrices qui les sous-tendent et de les comparer avec les dictionnaires, en particulier le *Dictionnaire de la zone* (collaboratif), *Keskiladi*, *Langue française bob*, le *Dico-des-mots* (collaboratif), ou le *Petit Momo*.
- 31 L'étude d'un premier échantillon a montré que, selon le dictionnaire retenu pour la comparaison, il n'y avait que de 35 % à 70 % des mots de notre corpus qui y figuraient, certains mots du corpus (comme *tchipage*) n'apparaissant dans aucun dictionnaire. Ce qui permet d'évaluer comparativement la qualité des dictionnaires, avec l'intérêt d'exemples en usage pris dans un contexte large.
- 32 L'une des particularités du lexique ainsi recueilli concerne le traitement des mots d'origine étrangère, qui ne sont pas regardés comme une classe à part. Ce choix reflète un positionnement théorique quant aux effets des contacts de langues, renforcé par l'étude du corpus. En effet, en écoutant les informateurs, il apparaît que l'usage de termes d'origine étrangère ne relève le plus souvent pas de phénomènes de *code-switching* ou de *crossing*, qui impliqueraient la conservation de leur valeur « ethnique » initiale. Des termes comme *wesh* ou *zaama*, employés fréquemment par des informateurs sachant l'interlocuteur arabophone (ou ayant un rapport avec l'arabe), peuvent dans un même discours relever du *code-switching* ou constituer des marqueurs discursifs d'une certaine forme de français, le « parler jeune ». En ce cas, *wesh* ou *zaama* employés par des non-arabophones n'illustreraient du *crossing* que si les locuteurs n'appartiennent pas au groupe des jeunes.

- 33 Il semble plutôt s'agir d'usages relatifs à des éléments de l'environnement des informateurs. En d'autres termes, ils ne sont pas à comprendre d'abord comme des manifestations d'identité « ethnique » (réelle ou fantasmée), mais ils indexent l'appartenance à un groupe de locuteurs du français, les jeunes, dont les normes autorisent des formes non standard et innovantes, produits de la « distorsion » de formes attestées (par exemple, le verlan), de la créativité de formes inédites ou du détournement de formes provenant d'une autre langue ou propres à d'autres groupes de locuteurs (comme le vieil argot). Le travail de Sandrine Wachs, Nacer Kaci, Nawal Boussouira, Joanne Kanguara, Magali Floren et Zakia Ayadi a notamment été présenté dans Wachs (2014), tentant d'établir l'extension de termes rencontrés dans MPF, de points de vue à la fois régional et social.

3.2 Plan phonique

- 34 Un deuxième exemple concerne la prosodie de contours intonatifs montant-descendant, réputés typiques d'un « accent banlieue ». Après l'annotation manuelle d'environ 600 contours intonatifs extraits du corpus (ce qui est loin de constituer la totalité de ces contours), Roberto Paternostro a pu montrer que, plutôt qu'à un contour banlieue, on a affaire à un contour déjà repéré et bien attesté en français standard et dans les usages ordinaires, où il est associé à l'expression de l'emphase. Ce qu'il y a de particulier dans les usages des jeunes réside non dans la spécificité, mais dans la fréquence, et dans les lieux d'apparition inhabituels de ce contour (voir Paternostro, 2013 et 2014 ; Paternostro & Goldman, 2014).
- 35 Une telle analyse a été rendue possible parce qu'on a pris des distances par rapport à la précatégorisation des locuteurs, qui aurait poussé à intégrer ce contour à l'ensemble des manifestations phoniques traditionnellement relevées pour caractériser les parlers jeunes (voir Jamin *et al.*, 2006), comme une certaine prononciation du *r*, qui selon certaines interprétations aurait à voir avec la langue arabe. De même que ce *r* renseignerait sur les influences multiculturelles des locuteurs concernés, la surreprésentation des contours emphatiques pourrait être interprétée comme l'expression d'une certaine violence verbale à l'image de la violence associée communément aux pratiques des jeunes de banlieue.
- 36 Cependant, qu'il s'agisse d'influence de l'arabe (voir ce que nous avons vu pour le lexique) ou de violence verbale dans une posture de mise en scène de « langue de la rue », on peut se demander si ces interprétations ne risquent pas d'être orientées par des *a priori* liés aux catégories sociales et/ou ethniques. Le corpus MPF a ainsi l'intérêt d'étayer la possibilité d'autres interprétations, sur la base d'analyses objectives.

3.3 Morphosyntaxe et discours

- 37 Pour la morphosyntaxe et le discours, plusieurs exploitations sont en cours, comme le recensement des formes verbales non conjuguées, à la fois l'inventaire (*ken, marave, tètj, il va graille...*), le contexte où elles apparaissent et les contraintes que l'on peut dégager. D'autres phénomènes syntaxiques notables concernent l'absence de clitique (comme dans les exemples (1) ou (11)), des usages élargis de *que* comme en (2), ou son absence, comme en (3) et (4), tous phénomènes déjà répertoriés dans différentes variétés de français... D'autres phénomènes, que l'on qualifiera d'« émergents » faute de

dénomination plus adéquate, concernant en particulier *total*, *direct*, *obligé*, *grave*, *même* (*pas*), *wesh*, *zaama*, *vas-y*, *perso* (voir les exemples de (5) à (10)) :

- (1) – vous faites la différence entre l’arabe littéraire et l’arabe marocain ?
– non je fais pas (Wajih 4, 464¹⁰)
 - (2) c’est bizarre que un français qui dit ça / que normalement ça doit être un arabe (Emmanuelle 2, 246)
 - (3) moi le bled moi franchement ça fait longtemps je suis pas parti (Wajih 4, 244)
 - (4) c’était la première fois ils y allaient (Wajih 4, 367)
 - (5) parce que vas-y c’est un peu cher les billets et tout (Wajih 4, 254)
 - (6) même pas je sais parler rebeu (Wajih 4, 953)
 - (7) je sais pas c’est quoi zaama on fait trop de bruit je sais ap (Wajih 4, 1396)
 - (8) moi perso comme j’ai dit moi ça sert à rien des longues études de toute façon (Wajih 4, 1914)
 - (9) elles par contre ça les a grave aidées au niveau de l’anglais quoi (Aristide 2a, 292)
 - (10) obligé à la rentrée d’octobre je me donne au moins une semaine pour oublier (Emmanuelle 4, 746)
- 38 On peut prendre l’exemple de *genre*, dans des emplois comme (11). Le nombre élevé d’occurrences permettra d’affiner la connaissance des contraintes gouvernant ses emplois : on en trouve par exemple 55 dans un enregistrement pris au hasard, Nacer 2, qui dure 86 minutes :
- (11) si une fois genre quelqu’un il te demande quelque chose et il voit que tu as de l’argent et tu dis non j’ai pas ou je peux pas nananinana après genre tu as une réputation tout le monde sait que tu es tu es une radine (Nacer 2, 1651)
- 39 On rencontre aussi des phénomènes qui impliquent des tronçons d’énoncés plus vastes, comme le discours rapporté (mode d’introduction, continueurs – voir Moreno, 2014 ; Guerin & Moreno, 2014, et ici-même les exemples (11) ou (12)), des interrogatives indirectes sur le schéma des interrogations directes par intonation (exemples (7), (13) ou (14)). MPF permet aussi de s’interroger sur des tendances à plus long terme de la langue française, comme la gestion de certains accords morphologiques. Ainsi, un travail a porté sur des cas de non-accord entre le sujet et le verbe (Benzitoun & Cappeau, à paraître – ici, exemple (15)), supposant que le corpus MPF pourrait être regardé comme du « français avancé » :
- (12) une fois elle m’a dit euh ah non elle m’avait dit elle est où ta racli (Nacer 3, 393)
 - (13) je sais même pas moi ça veut dire quoi (Wajih 4, 1004)
 - (14) à chaque fois il me voit je sais ap il a quoi contre oim (Wajih 4, 1289)
 - (15) genre les meufs qui dit je m’en bats les couilles (Roberto 2c, 386)
- 40 D’autres exploitations en morphologie et en syntaxe sont en cours de réalisation ou en projet.

3.4 Les représentations sur la/les langue(s)

- 41 Enfin, du côté du contenu des enregistrements, et dans les enregistrements écologiques et dans certains entretiens, nous nous sommes intéressés aux représentations sur les langues et les identités qu’elles manifestent, qui ont d’évidents effets sur les pratiques langagières : les discours tenus sur le français – en particulier sur la notion de « parler normal », comme en (16), et ceux sur les langues d’origine, quel que soit le degré de maîtrise par l’informateur :
- (16) wallah je les ai pas traités je leur parlais normal wesh (Wajih 4, 4122)

- 42 Ces pratiques sont crucialement adressées à un interlocuteur spécifiquement identifié, comme le montre l'exemple d'un enregistrement où l'enquêteur, lui-même Algérien, se trouve de plain-pied avec un Beur qui se dit algérien bien qu'il ne parle pas arabe et ne soit jamais allé en Algérie (voir Gadet & Kaci, 2012 [2015]). Plusieurs des enquêteurs de MPF étant des Maghrébins, il y a des effets pour des jeunes dont les origines maghrébines appuient une identité en partie fantasmée (ils se disent souvent « rebeus » et non « beurs »). On a ainsi exploré l'intrication de discours tenus sur les langues, en particulier quant aux relations entre arabe et français. De façon générale, il apparaît que les jeunes locuteurs, quelle que soit leur propre façon de parler, sont souvent porteurs d'idéologie du standard dans leurs représentations du français, tout en paraissant s'y opposer (ce qui revient à réasserter quelque chose par le fait même de le mettre en cause – à la marge d'ailleurs ; voir Gadet & Guerin, 2015 ; Jaspers, 2011, sur des jugements à propos du flamand dans une école de Gand).
- 43 Ces différents travaux et d'autres qui sont en cours visent à exploiter désormais les possibilités de croisements entre les domaines.

4. Remarques conclusives : qu'est-ce que MPF manifeste de spécifique ?

- 44 Le corpus MPF permet des réflexions qui n'auraient pas nécessairement pu prendre place sur la base d'autres corpus.
- 45 Tout d'abord, à une étape de mise en place des conditions du recueil, MPF a été l'occasion d'une réflexion sur le recueil de données, une préoccupation que les linguistes ont trop souvent, dans l'élaboration de grands corpus, traité avec quelque légèreté, ou comme une évidence ne méritant pas d'élaboration ; au contraire du souci constant manifesté par les ethnologues dans l'élaboration de leurs données (voir p. ex. les réflexions de Lepoutre, 2001, qui revient après coup sur son travail de 1997).
- 46 Au niveau de la collection de données et d'un point de vue avant tout formel, MPF permet de disposer d'une ample documentation sur des catégories « émergentes » ou « innovantes », en tous cas encore peu décrites – le terme *émergent* étant d'ailleurs à manier avec précaution car difficile à définir, mais n'étant certainement pas à confondre avec « nouveau » (qui ne serait guère adapté que pour le lexique, et encore...). Il permet aussi de mesurer des contraintes sur des phénomènes déjà connus. La pratique de recueil d'immédiat communicatif sur base de réseaux ouvre-t-elle l'accès à des données différentes de celles qui sont recueillies dans des entretiens traditionnels ? C'était en tout cas notre hypothèse. Maintenant, à quel(s) niveau(x) linguistique(s) des effets se manifestent-ils ? Il apparaît impossible de répondre à cette question – ce qui n'empêche pas de la soulever –, les entretiens n'étant pas réitérables *ad libitum* afin de mesurer un éventuel « effet enquêteur »¹¹.
- 47 L'empreinte de la sociolinguistique et de l'ethnolinguistique dans la réflexion sur MPF est plus nette encore pour ce qui est de la considération des effets du contact des langues. S'il apparaît en effet certain que les langues peuvent emprunter du lexique ou des intonations d'une autre langue, il est moins évident – et cela demeure à documenter – qu'il soit possible de « copier » des phénomènes phoniques segmentaux, des faits syntaxiques et, probablement, encore moins des faits morphologiques. Il faut ainsi se garder de sauter trop vite à l'hypothèse qu'on a affaire à des effets du contact,

et confronter les phénomènes qui pourraient être des candidats à être traités comme des emprunts à d'autres hypothèses, dont celle de l'évolution endogène. Toutefois, une telle démonstration exigerait une documentation, à travers toute la francophonie, bien plus solide que celle dont on dispose actuellement.

- 48 Enfin, MPF permet aussi de s'intéresser à ce qui est dit dans les enregistrements, au contenu, au-delà du corpus conçu comme un « réservoir de données » pour pratiquer des études linguistiques. Ainsi, on peut entrevoir la possibilité d'aborder des questions sociolinguistiques, didactiques et même politiques (sur le système d'enseignement et au-delà – voir Sayad, 2014¹²), avec un éclairage sur la posture de mise en scène de soi des jeunes, en rapport à l'intensité, à l'emphase et à la culture des rues (Lepoutre, 1997 ; Gadet & Hambye, 2014). Des analyses plus fines à ce propos permettront sans aucun doute de mieux comprendre la difficile quête d'identité de beaucoup de ces jeunes.

BIBLIOGRAPHIE

- Baude O. & Dugua C. (2011). « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? », *Corpus* 10 : 99-118.
- Benzitoun Ch. & Cappeau P. (2014). « Quel accord ? Retour sur les cas de 'non accord' entre sujet et verbe à l'oral », Colloque *Les relations d'accord dans la syntaxe du français*. Fribourg, 25-26 septembre 2014.
- Billiez J. (1992). « Le 'parler véhiculaire interethnique' de groupes d'adolescents en milieu urbain », in E. Gouaini and N. Thiam (éd.) *Des langues et des villes*. Paris : Didier-Érudition, 117-126.
- Blanc-Chaléard M.-Ch. (2001). *Histoire de l'immigration*. Paris : La Découverte.
- Blanche-Benveniste C. (2003). « Réflexions sur les transcriptions de corpus de français parlé », *Revue PAroles* 22-23-24 : 91-116.
- Blanche-Benveniste C. (2008). « Le français parlé au XXI^e siècle. Réflexions sur les méthodes de description : système et variations », in M. Abecassis, L. Ayosso, E. Vialleton (éd.) *Le français parlé au XXI^e siècle : normes et variations dans les discours et les interactions*. Volume 2. Paris : L'Harmattan, 17-39.
- Blanche-Benveniste C. (2010). *Le français : Usages de la langue parlée*. Leuven : Peeters.
- Branca-Rosoff S. et al. (2011). « Constitution et exploitation d'un corpus de français parlé parisien », *Corpus* 10 : 81-98.
- Cappeau P. & Gadet F. (2010). « Transcrire, ponctuer, découper l'oral. Bien plus que de simples choix techniques », *Cahiers de linguistique* 35/1 : 187-202.
- Cappeau P. & Gadet F. (2014). « Quand l'œil écoute... Que donnent à lire les transcriptions d'oral ? ». *Actes du colloque CILPR de Nancy*, en ligne : <http://www.atilf.fr/cilpr2013/>.
- Cappeau P., Gadet F., Guerin E. & Paternostro R. (2011). « Réflexions sur les incidences de quelques aspects de la transcription outillée », *LINX* 64-65 : 85-100.

- Cheshire J. (2005). « Age and generation-specific use of language », in U. Ammon, N. Dittmar, K. Mattheier & P. Trudgill (éd.) *Sociolinguistics : An Introductory Handbook of the Science of Language and Society*. Berlin : Mouton de Gruyter, 1552-1563.
- Cheshire J., Kerswill P., Fox S. & Torgersen E. (2011). « Contact, the feature pool and the speech community : the Emergence of Multicultural London English », *Journal of Sociolinguistics* 15-2 : 151-196.
- Dico-des-mots <http://dico-des-mots.com/liste-de-mots>.
- Dictionnaire de la zone <http://www.dictionnairedelazone.fr/?index=lexique&let=0>.
- Eckert P. (2012). « Three waves of variation study : the Emergence of meaning in the study of sociolinguistic variation », *Annual Review of Anthropology* 41 : 87-100.
- Fagyal Z. (2010). *L'Accent de banlieue*. Paris : L'Harmattan.
- Gadet F. (2008). « L'oreille et l'œil à l'écoute du social », in M. Bilger (dir.) *Données orales. Les enjeux de la transcription. Les cahiers* 37. Presses universitaires de Perpignan, 35-48.
- Gadet F. (2013). « Collecting a new corpus in the Paris area : intertwining methodological and sociolinguistic reflections », in D. Hornsby & M. Jones (éd.) *Language and Social Structure in Urban France*. Oxford : Legenda, 162-171.
- Gadet F. (2015). « Le style et les corpus : réflexions à partir d'un corpus recueilli en région parisienne », in K. Jeppesen Kragh & J. Lindschouw (éd.) *Les Variations diasystématiques et leurs interdépendances dans les langues romanes*. Société de linguistique romane, série TraLiRo, collection *Sociolinguistique, dialectologie, variation*.
- Gadet F. & Guerin E. (2012). « Des données pour étudier la variation : petits gestes méthodologiques, gros effets », *Cahiers de linguistique* 2012-1 : 41-65.
- Gadet F. & Guerin E. (2015). « Le français en contact en région parisienne : le poids des représentations sur les langues », in Actes du colloque de Raguse, Repères DoRIF. Revue en ligne.
- Gadet F. & Hambye Ph. (2014). « Contact and ethnicity in 'youth language' description : in search of specificity », in R. Nicolăi (dir.) *Questioning Language Contact. Limits of Contact, Contact at its limits*. Leiden/Boston : Brill, 183-216.
- Gadet F. & Kaci, N. (2012 [2015]). « Identification en première personne. Le discours d'un 'jeune de banlieue' en entretien », *Cahiers de praxématique*, n° 59, 29-44.
- Gadet F. & Paternostro R. (2013). « Un accent multiculturel en région parisienne ? », *Repères DORIF* 3. <http://www.dorif.it/ezine/>.
- Gadet F. & Wachs S. (2015). « Comparer des données de corpus : évidence, illusion, ou construction ? », *Langage & Société*, n° 154, 33-49.
- Greco L., Mondada L. & Renaud P. (dir.) (2014). *Identités en interaction*. Limoges : Lambert Lucas.
- Guerin E. & Moreno A. (2014). « Le discours rapporté dans les interactions orales et écrites. Au-delà d'une opposition de surface », *Actes du colloque CILPR de Nancy*. En ligne : <http://www.atilf.fr/cilpr2013/>
- Guerin E. & Paternostro R. (2014). « What is langue des jeunes and Who speaks it », in H. Tyne et al. (éd.) *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*. Actes du colloque AFLS de Nancy 2011. Cambridge Scholars Publishing.
- Gülich E. & Mondada L. (2001). « Analyse conversationnelle », in *Lexikon der Romanistischen Linguistik*. Tübingen : Max Niemeyer Verlag, 196-250.

- Jamin M. (2004). « 'Beurs' and accent des cités : a case study of linguistic diffusion in La Courneuve », *Contemporary French and Francophone Studies* 8-2 : 169-176.
- Jamin M. & Trimaille C. (2008). « Quartiers pluriethniques et plurilingues en France : berceaux de formes supra-locales (péri-)urbaines ? », in M. Abecassis, L. Ayosso & V. Alleton (éd.) *Le français parlé au XXI^e siècle. Normes et variations géographiques et sociales*, Volume 1. Paris : L'Harmattan, 225-246.
- Jamin M., Trimaille C. & Gasquet-Cyrus M. (2006). « De la convergence dans la divergence : le cas des quartiers pluriethniques en France », *Journal of French Language Studies* 16-3 : 335-356.
- Jaspers J. (2008). « Problematizing Ethnolects : Naming linguistic practices in an Antwerp secondary school », *International Journal of Bilingualism* 12 (1-2) : 85-103.
- Jaspers J. (2011). « Talking like a 'zerolingual' : Ambiguous linguistic caricatures at an urban secondary school », *Journal of Pragmatics* 43 : 1264-1278.
- Kerswill P. (2010). « Youth Languages in Africa and in Europe : Linguistic subversion or emerging vernaculars ? ». http://www.lancaster.ac.uk/fass/doc_library/linguistics/kerswill/Kerswill-African-Studies-19-10-10.pdf
- Keskiladi <http://www.keskiladi.com/>.
- Koch P., Esterreicher W. (2001). « Langage oral et langage écrit », in G. Holtus, M. Metzeltin & C. Schmitt (éd.) *Lexikon der Romanistischen Linguistik*, Tome 1. Tübingen : Max Niemeyer Verlag, 584-627.
- Kotsinas U.-B. (1998). « Language Contact in Rinkeby – an immigrant suburb », in J. Androutsopoulos & A. Scholz (éd.) *Jugendsprache, langue des jeunes, Youth Language*. Frankfurt am Main : Peter Lang, 125-148.
- Labov W. (1972). *Sociolinguistic Patterns*. Philadelphia : University of Pennsylvania Press.
- Lagrange H. (2010). *Le Déni des cultures*. Paris : Le Seuil.
- Langue française bob. <http://www.languefrancaise.net/bob/>.
- Lepoutre D. (1997). *Cœur de banlieue*. Paris : Odile Jacob.
- Lepoutre D. (2001). « La photo volée. Les pièges de l'ethnographie en cité de banlieue », *Ethnologie française* XXXI-1 : 89-101.
- Moreno A. (2014). « Le discours rapporté dans l'interaction : proximité et variabilité », in *Actes du Congrès mondial de linguistique française de Berlin*.
- Moreno A. & Paternostro R. (2014). « Les parlers jeunes en banlieue parisienne au croisement de la syntaxe et de la prosodie », Communication au Colloque de Nanterre (5-6-7 juin 2014), *Les Métropoles francophones en temps de globalisation*.
- Noiriel G. (2002). *Atlas de l'immigration en France*. Paris : Éditions Autrement.
- Paternostro R. (2013). « La langue des jeunes parisiens : une forme actualisée dans la proximité ? », *Cahiers de l'École doctorale de Brescia* 7 : 9-19.
- Paternostro R. (2014). *L'Intonation des jeunes en région parisienne : aspects phonétiques et sociolinguistiques, implications didactiques*. Thèse non publiée des universités de Brescia et de Paris Ouest Nanterre la Défense.

Paternostro R. & Goldman, J.-Ph. (2014). « Modeling of a rise-fall intonation pattern in the language of young Paris speakers », *Actes du colloque Speech Prosody 7*. Trinity College de Dublin, 20-23 mai 2014.

Petit Momo. http://michel.buze.perso.neuf.fr/lavache/petit_momo.htm.

Pooley T. (2012). « Code-crossing and multilingualism among adolescents in Lille », *Journal of French Language Studies* 22-3 : 371-394.

Quist P. (2008). « Sociolinguistic approaches to Multiethnolect : Language variety and stylistic practice », *International Journal of Bilingualism* 12 (1-2) : 43-61.

Rabaud A. (2014). « Une histoire de l'immigration en France », in J.-Y. Blum Le Coat et M. Eberhard (dir.) *Les Immigrés en France*. Paris : La Documentation française, 15-36.

Rickford J., McNair-Knox F. (1994). « Addressee – and topic –influenced style shift : a quantitative sociolinguistic study », in D. Biber & E. Finegan (éd.) *Sociolinguistic Perspectives on Register*. New York : Oxford University Press, 235-276.

Rougé J.-L. (2013). « Faire le portrait linguistique d'une ville. Présentation du projet de recherche langues en contact à Orléans (LCO) », *Langage & Société* 145 : 123-129.

Sayad A. (2014). *L'École et les enfants de l'immigration*. Paris : Le Seuil (La couleur des idées).

Thibault P. & Vincent D. (1990). *Un corpus de français parlé*. Québec : Recherches sociolinguistiques 1.

Wachs S. (2014). « Paris, créativités lexicales et frontières géographiques », Communication au Colloque de Nanterre (5-6-7 juin 2014), *Les Métropoles francophones en temps de globalisation*.

Wiese H. (2013). « What can new urban dialects tell us about internal language dynamics ? The power of language diversity », *Linguistische Berichte* 19 : 208-245.

NOTES

1. Il bénéficie actuellement de soutiens par la DGLFLF (budget 2015) et par le GTRC canadien *Le français à la mesure d'un continent* (France Martineau, directrice).
2. Pour chacun de ces auteurs, on ne donne ici qu'une référence afin de ne pas surcharger la bibliographie. Pour une bibliographie plus complète sur l'Europe, voir Gadet & Hambye, 2014. Nous n'envisageons pas ici les comparaisons avec l'Afrique ou l'Amérique « francophones ».
3. Pour les corpus sur Orléans, le contact est traité de façon spécifique dans un projet parallèle aux corpus des ESLOs (« Étude SocioLinguistique sur Orléans ») : le projet LCO, « Langues en contact à Orléans », dont on trouve une présentation dans Rougé, 2013.
4. Nous adoptons ici le même choix que celui exposé à la note 2, en renvoyant à la même publication pour une bibliographie plus ample. On pourrait certes évoquer d'autres corpus, mais nous nous arrêtons là, du fait de la modestie de taille ou de ce que l'option de qualité interactive empêche de s'appesantir sur des entretiens faits sans beaucoup de préoccupations sociolinguistiques.
5. Le terme *écologique* appuie dans MPF la distinction entre les auto-enregistrements et les entretiens. Cependant, il ne s'agit nullement d'une dichotomie, et les entretiens peuvent tendre plus ou moins vers l'écologique, compte tenu de l'attention portée aux spécificités contextuelles et au rôle des réseaux liant les interactants.
6. Cappeau *et al.*, 2011 tente de justifier le choix qui a été fait de Praat pour MPF, parmi les différents logiciels de transcription offrant le même type de propriétés.

7. Tous les chercheurs qui travaillent sur des parlers non standard ont eu à se demander comment éviter de stigmatiser le parler qu'ils étudient, dès l'étape de la transcription. MPF a suivi la perspective de Blanche-Benveniste (2003, 2010) pour laquelle seul un écrit standardisé prenant acte de l'incommensurabilité sémiotique des ordres de l'oral et de l'écrit peut satisfaire cette exigence. On ne cherche ainsi pas à représenter graphiquement des particularités phoniques comme la chute d'un *e* muet ou une liaison, mais les mots non prononcés ne sont pas restitués (voir aussi Gadet, 2008 ; Cappeau & Gadet, 2014).

8. <http://mpfvitrine.modyco.fr>

9. Il s'agit d'entretiens avec des « personnalités » ayant un rapport, de types variés, avec les pratiques langagières visées. Si ces enregistrements peuvent parfois constituer un matériau intéressant pour l'analyse linguistique, ils sont avant tout recueillis pour l'intérêt des propos qui y sont tenus. Ces « grands témoins », en général un peu plus âgés, sont des responsables d'association de quartier, des enseignants en ZUS, des acteurs de la culture urbaine...

10. Les exemples sont ici présentés avec le prénom de l'enquêteur et le numéro de l'enquête. Le chiffre qui suit est le minutage en secondes.

11. Des tentatives ont pourtant été faites en ce sens (p. ex. Rickford & McNair Knox, 1994). Elles permettent certes de montrer quelque chose quant au fonctionnement du style dans les langues, mais elles ont selon nous l'inconvénient de négliger le point de vue de l'enquêté, pourtant crucial pour la qualité des données obtenues. Comment un enquêté qui se trouve sollicité pour réitérer le même type d'entretien, avec des enquêteurs différents, mais avec lesquels il n'a toujours aucune histoire conversationnelle antérieure, peut-il construire une interprétation vraisemblable de la situation ?

12. Ces inédits d'un auteur décédé en 1998 reproduisent des textes datant du début des années 80, permettant ainsi de mesurer que ce sont à peu près les mêmes problèmes qui se posent aujourd'hui, en particulier à l'école, demeurés intraités ou négligés, et donc qui n'ont fait entre-temps que s'aggraver.

RÉSUMÉS

Dans cet article, nous présentons le corpus « Multicultural Paris French » (MPF), en montrant les enjeux théoriques et méthodologiques, ainsi que quelques directions d'exploitation. En visant des données non-standard illustrant ce qu'il est de coutume d'appeler « parler jeune », MPF repose sur une réflexion quant à la façon d'appréhender les situations propices à leur émergence, au-delà d'informations généralement retenues, d'ordres sociodémographiques ou ethniques, ou du degré de formalisme de l'échange. Si ces considérations ne sont évidemment pas évacuées, elles sont combinées à d'autres qui concernent les effets de la relation entre locuteurs au moment de l'enregistrement. De ce fait, MPF apparaît comme un corpus original dans le champ des corpus de français parlé contemporain, tant du point de vue de la nature des données recueillies que de celui du traitement rendu possible par cette spécificité.

This article presents the “Multicultural Paris French” (MPF) corpus, focusing on the theoretical and methodological issues and possible research avenues that it opens up. The MPF corpus comprises non-standard speech, commonly referred to as ‘youth language’. A key aspect of the corpus design is that it inquires into the situational factors conducive to the emergence of this type of speech, beyond the features generally taken into consideration, such as socio-

demographic and ethnic data or the degree of formality of the exchange. While these features are obviously not ignored, they are combined with others concerning the effect of the relationship between speakers during the recording. This specificity distinguishes MPF from other corpora of contemporary spoken French, making it original both as regards the type of data collected and the possible avenues for investigation that it offers.

INDEX

Mots-clés : corpus, non-standard, langue parlée, parlers jeunes, multiculturalisme

Keywords : corpus, non-standard, speech, youth language, multiculturalism

AUTEURS

FRANÇOISE GADET

Université Paris Ouest & MoDyCo

EMMANUELLE GUERIN

Université d'Orléans & LLL