

Port Acadie

Revue interdisciplinaire en études acadiennes
An Interdisciplinary Review in Acadian Studies



Patrimoine oral et transcription – exemple de deux variétés de français canadiens : acadien vs laurentien (Est et Ouest)

Pierre-Don Giancarli

Number 30, Fall 2016

Patrimoine oral et valorisation à l'ère du numérique (1^{ère} partie)

URI: <https://id.erudit.org/iderudit/1056923ar>

DOI: <https://doi.org/10.7202/1056923ar>

[See table of contents](#)

Publisher(s)

Université Sainte-Anne

ISSN

1498-7651 (print)

1916-7334 (digital)

[Explore this journal](#)

Cite this article

Giancarli, P.-D. (2016). Patrimoine oral et transcription – exemple de deux variétés de français canadiens : acadien vs laurentien (Est et Ouest). *Port Acadie*, (30), 115–135. <https://doi.org/10.7202/1056923ar>

Article abstract

Paradoxically, intangible heritage made up of (audio-/video) recorded testimony is often accessed through its written transcript. The importance of the stage of transcription is what we would like to focus on here. Drawing from thirteen Acadian and Laurentian French corpora, we aim to demonstrate that a transcript is not the faithful rendering of a segment of reality but a construct. This construct, both more or less impoverished and sometimes supplemented, results from a chain of motivated choices and resolution of problems which may have a far-reaching impact upon the interpretative work of the researcher.

Patrimoine oral et transcription – exemple de deux variétés de français canadiens : acadien vs laurentien (Est et Ouest)

Pierre-Don Giancarli
Université de Poitiers

Résumé

Le patrimoine immatériel que constituent les témoignages enregistrés ont ceci de particulier qu'ils sont souvent exploités à partir de leur transcription écrite. C'est cette étape de la transcription que, à partir de treize corpus acadiens et laurentiens de l'Est et de l'Ouest, nous souhaitons mettre ici en relief, en montrant qu'elle n'est pas le rendu à l'identique d'une tranche de « réel » mais un construit à la fois plus ou moins appauvri et éventuellement enrichi, qui est le résultat d'une succession de choix motivés et de résolution de problèmes qui ne sont pas sans conséquence sur le travail interprétatif du chercheur.

Abstract

Paradoxically, intangible heritage made up of (audio-/video) recorded testimony is often accessed through its written transcript. The importance of the stage of transcription is what we would like to focus on here. Drawing from thirteen Acadian and Laurentian French corpora, we aim to demonstrate that a transcript is not the faithful rendering of a segment of reality but a construct. This construct, both more or less impoverished and sometimes supplemented, results from a chain of motivated choices and resolution of problems which may have a far-reaching impact upon the interpretative work of the researcher.

Mots clés

corpus, transcription, patrimoine, français acadien, français laurentien

Keywords

corpus, transcription, heritage, Acadian French, Laurentian French

Les témoignages enregistrés (de façon audio ou audio + vidéo) auprès d'informateurs, qu'il s'agisse de la narration de contes, de chants, d'entretiens dirigés ou de récits de vies d'époques contemporaines ou plus anciennes, sont porteurs d'informations non seulement sur les individus et sur leurs langues mais aussi sur les communautés auxquelles ils appartiennent. Il s'agit d'informations de nature sociale, historique, culturelle, anthropologique, ethnologique, linguistique, *etc.*, qui les rendent constitutifs de notre patrimoine. C'est leur dimension linguistique qui nous intéressera ici, en gardant à l'esprit que ces témoignages enregistrés dans leur vernaculaire¹ sont souvent exploités non directement sous leur forme primaire, c'est-à-dire audio mais secondaire une fois scriptés, puis éventuellement informatisés, archivés dans des bases de données², diffusés et transmissibles aux chercheurs de demain et exploitables sous forme numérique avec des logiciels variés. Une étape de transcription doit donc avoir préalablement eu lieu. Or la transcription n'est pas une étape innocente mais le résultat d'une succession de choix qui peuvent orienter les résultats du chercheur.

Nous nous appuyerons sur des corpus de français du Canada, pays en pointe dans la francophonie dans le domaine des corpus oraux, dans ses variétés acadienne et laurentienne (voir liste des corpus en annexe 1). L'acadien, issu de ce que parlaient le petit nombre de colons originaires en grande partie du Centre-Ouest de la France, sera représenté ici par les provinces du Nouveau-Brunswick et de la Nouvelle-Écosse ainsi que par les Îles-de-la-Madeleine³. Le laurentien, issu de ce que parlaient le grand nombre de colons originaires en grande partie de l'Ouest de la France,

1 Vernaculaire au sens où les locuteurs accordent le minimum d'attention à la surveillance de leur propre discours.

2 Citons à cet égard la phonothèque de la Bibliothèque Nationale de France, qui archive aussi bien des enregistrements de langues à la survie aujourd'hui menacée, que des enregistrements de chants polyphoniques corses réalisés par Félix Quilici ou des enregistrements de terrain collectés en Corse et en Acadie par Geneviève Massignon.

3 Les Îles-de-la-Madeleine appartiennent administrativement au Québec mais sont de peuplement acadien.

sera divisé en Est et Ouest, représenté dans le premier cas par les provinces du Québec et de l'Ontario et dans le second par celles du Manitoba, de la Saskatchewan et de l'Alberta.

Au début de l'entreprise il y a la constitution du corpus oral, en fin de course il y a son exploitation, et entre les deux il y a sa transcription. Nous commencerons par les deux extrémités de la chaîne avant de nous concentrer sur l'étape clef de la transcription c'est-à-dire de la mise en forme. La forme est importante : elle l'est en soi, et aussi parce que notre interprétation du dit dépend directement de la façon dont cela est dit.

1. La constitution du corpus et son exploitation.

1.1. La constitution du corpus.

Constituer un corpus de données authentiques donne au chercheur en linguistique le sentiment qu'il est en prise directe avec le réel. Il lui permet de se rendre compte que ce qu'il croyait éventuellement non-dicible se dit puisqu'attesté, de chercher des réponses à des questions autres que celles qu'il se posait déjà, il lui évite de se fier à sa seule intuition/introspection ou d'avoir recours à de l'oral représenté c'est-à-dire fictif. Nous entendons par ce dernier l'oral rendu dans leurs œuvres par des auteurs, notamment dans des pièces de théâtre (comédies) mettant en scène le parler des classes populaires, qui reste de l'oral produit pour un certain public par une élite et qui peut se trouver dénaturé par des considérations diverses, par exemple standardisé si l'auteur est animé de la volonté commerciale d'être compris et vendu au-delà de ses frontières, ou caricaturé s'il est animé du projet idéologique de peindre telle classe sociale de façon dénigrante.

Pour autant, il serait un leurre d'opposer diamétralement un oral représenté qu'on pourrait soupçonner de non-représentativité, à un oral authentique forcément représentatif : s'ils savent/croient qu'ils seront évalués sur la forme, des locuteurs enregistrés peuvent modifier leur vernaculaire par des hypercorrections, des autocorrections en plusieurs formulations successives et une auto-

surveillance inhabituelles. Des précautions sont donc à prendre, et les conditions de constitution des enregistrements à connaître⁴.

Par exemple y a-t-il un enquêteur ou pas. S'il y a un enquêteur pour mener les entretiens, il gagne à être natif de la région, voire même membre de la communauté, voire même du cercle de connaissances du locuteur, afin de libérer la parole. Inversement avec à un universitaire francophone de France monolingue jouant le rôle d'enquêteur ou même ne disant rien mais par sa simple présence dans la pièce, tels francophones du Canada bilingues/diglosses et en situation minoritaire risquent de recourir à leur « français du dimanche », de renoncer à exploiter leurs capacités aux contacts de langues (alternance, hybridation, *etc.*) ou même de s'auto-censurer jusqu'à l'inhibition. L'enquêteur aura ensuite beau jeu de conclure à de l'insécurité linguistique en milieu minoritaire alors qu'il aura lui-même créé des conditions idéales à l'insécurité et donc vicié ses propres données de départ, ce qui nous ramène au paradoxe de l'observateur⁵.

S'il n'y a pas d'enquêteur, l'entreprise reste envisageable pour une exploitation linguistique, peut-être moins dans d'autres domaines si des questions précises sont censées amener des réponses sur des sujets précis, mais même alors des subterfuges restent imaginables. Par exemple pour la constitution du corpus chiac Perrot il a été fourni à des dyades de locuteurs un questionnaire écrit, sur des thèmes qui étaient généraux mais qui auraient pu être précis. Puis on laisse les informateurs discuter seuls, ce qui leur permet d'être en interaction sans la supervision d'un enquêteur en position d'autorité et menant les débats. Néanmoins, même sans enquêteur il reste la présence du magnétophone, qui peut jouer un

4 Une récolte d'enregistrements par entretien ne garantit de toute façon pas des conditions parfaitement naturelles (écologiques) puisque un chercheur va solliciter des locuteurs et éventuellement orchestrer les échanges, et elle favorise un certain type de production plutôt qu'un autre puisque le locuteur est dans la réflexion ou les souvenirs plutôt que dans l'action. D'un autre côté enregistrer les conversations de locuteurs en cours d'activité ou/et à leur insu pose des problèmes techniques, juridiques, et éthiques.

5 Cf. William Labov, « Some principles of linguistic methodology », dans *Language in Society*, Cambridge University Press, n° 1, 1972, p. 97-120.

rôle au niveau diaphasique (variation de style/registre en fonction de la situation de communication) dans l'obtention d'un discours excessivement surveillé.

Pour avoir un échantillonnage équilibré et se donner les moyens de vérifier la pertinence éventuelle de certains facteurs, on veille dans les corpus les plus exigeants à ce que les locuteurs, enregistrés sur un laps de temps raisonnable⁶, soient appariés. Par lieu, âge, niveau social et sexe. Appariés par lieu, variation appelée diatopique pour veiller à une certaine homogénéité topolectale, car par exemple au Canada entre provinces et à l'intérieur d'une même province il peut y avoir de grandes disparités en fonction des origines de peuplement. On vérifie donc que les locuteurs sont bien enracinés dans la communauté en question, voire même le lieu de départ et la date d'arrivée de leurs parents ou grands-parents. Appariés aussi par tranches d'âge, chaque tranche comportant le même nombre de locuteurs : on sait par exemple que les locuteurs jeunes élident plus souvent le subordonnant que leurs aînés⁷. Appariés par niveau social, variation appelée diastratique pour veiller à une homogénéité sociolectale : c'est ainsi que R. Mougeon et ses collaborateurs distinguent les locuteurs francophones non restreints (parlant français non seulement à l'école/travail mais aussi à la maison), semi-restreints (parlant français à l'école/travail mais moins à la maison) et restreints (parlant français seulement à l'école/travail). Et appariés par sexe, variation appelée diagénique, avec un nombre égal d'hommes et de femmes puisqu'on sait que les femmes ne parlent pas comme les hommes et ont tendance à favoriser les tendances les plus grammaticales et les plus standards⁸

6 Pour qu'on ait une bonne homogénéité diachronique c'est-à-dire représentant une tranche temporelle, on n'imagine pas un corpus unique s'étendant sur plusieurs décennies. Quand les informateurs du corpus Sankoff-Cedergren enregistrés à Montréal en 1971 ont été contactés treize ans plus tard pour être ré-enregistrés, cela a donné lieu à la constitution d'un autre corpus (le corpus Montréal 84).

7 Nathalie Dion, *L'effacement de que en français canadien : une étude en temps réel*, thèse de maîtrise, Université d'Ottawa, sous la direction de Shana Poplack, 2003, p. 20.

8 Parce que les femmes sont souvent chargées de l'éducation des enfants, la conservation du standard pourrait être une nécessité liée au souci de transmission

alors que les hommes préfèrent les variantes locales, non-standards et plus stigmatisées⁹. Ainsi dans cet extrait du corpus Wiesmath où, pour la Française de France présente, le mari s'auto-traduit d'anglais en français vernaculaire mais son épouse le re-traduit de français vernaculaire en français standard :

- L1 C'est juste que tant ce que ta poumme se fait, le ver
passe, pis i va dans le CORE, i va dans/ dans le mitan
L2 Ouais i va dans le/ i va dans le milieu (corpus
Wiesmath)

Cela implique donc l'établissement préalable d'informations (métadonnées) sur les locuteurs qui seront sélectionnés pour être enregistrés. Une fois les enregistrements scriptés (aujourd'hui de façon numérique), le corpus peut être exploité et servir de support à divers outils.

1.2. L'exploitation de la transcription.

La transcription une fois numérisée peut donner lieu à exploitation avec des logiciels variés – par exemple Goldvarb pour identifier et hiérarchiser les facteurs de variation statistiquement significatifs tels ceux énumérés précédemment (lieu, âge, niveau social et sexe). Ou des concordanciers à même d'extraire une chaîne assortie d'un contexte gauche et droit plus ou moins large, ou divers outils textométriques permettant de lancer des requêtes.

Les requêtes peuvent être très précises si en plus de la transcription stricto sensu un taggage/étiquetage a été effectué, consistant à indexer les mots selon leurs catégories syntaxiques (verbe, préposition, *etc.*), leur caractère composé ou dérivé, le temps et mode choisis, *etc.* Inversement la lemmatisation permettra de regrouper sous un même lemme toutes les formes qu'un mot peut prendre, par exemple la recherche à *aller* renverra à *allions, irai, va, etc.* Autant de précisions qui font gagner du temps au chercheur

inter générationnelle, mais d'autres options sont envisageables (Brouwer Dede et Van Hout Roeland, « Gender-related Variation in Amsterdam Vernacular », dans *International Journal of the Sociology of Language*, n° 94, 1992, p. 99-122).

9 Lesley Milroy et Matthew Gordon, *Sociolinguistics: Method and Interpretation*, 2^e éd., Oxford, Blackwell, 2003, 280 p.

mais dont l'établissement est très chronophage à la transcription, n'est pas garanti sans erreurs, et contribue à diminuer la lisibilité du corpus.

Le corpus permet au chercheur de vérifier la fréquence de tel phénomène, c'est-à-dire de voir s'il est généralisable et à quel point, et de s'assurer qu'il n'est pas une scorie du discours due à une erreur de performance, un hapax, ou une simple coquille. Ou au contraire que la transcription est fiable. Prenons un exemple illustrant chacun de ces trois derniers cas en nous limitant au petit domaine de la forme pronominale du verbe.

Considérons cet extrait du corpus ontarien : « Donc les grands-parents maternels du côté d'Québec pis vos parents à vous ça... s'est venu s'établir dans l'Nord de l'Ontario pis vous êtes venue au monde là ? » (Corpus Mougeon-Hébrard). Le linguiste ne connaissant pas les français parlés au Canada et rencontrant une telle formulation pourrait se dire que le verbe *venir* y connaît un emploi pronominal (*se venir*). Il vérifiera la fréquence de *se venir* dans le corpus (et éventuellement aussi dans d'autres corpus canadiens) et n'en trouvera pas. Il conclura alors à juste titre à une coquille¹⁰.

Considérons ensuite le chercheur intéressé par le choix de l'auxiliaire *être* ou *avoir* avec les verbes pronominaux, sachant qu'en acadien on trouve en général non pas *être* mais *avoir*¹¹. « *I s'ont* » graphié ainsi dans le corpus acadien Péronnet est à interpréter comme une contraction de *i s'avont*, c'est-à-dire l'auxiliaire *avoir* si la transcription est fiable. Mais vu la proximité de prononciation on pourrait penser à une assimilation du s- d'un *sont* au s- de *se* et donc à l'auxiliaire *être*. Différence d'importance pour qui s'intéresse aux auxiliaires. Pour trancher, il pourra s'appuyer sur

10 *S'en venir* par contre est fréquent. La bonne graphie ici est sans doute *Pis vos parents à vous ça... c'est venu s'établir dans...*, avec un *c'* repris du *ça* sujet personnel qui précède (emploi banal au Canada) mais réduit devant la voyelle de la copule (*est*) et conjuguant le verbe *venir* tout aussi non pronominal qu'il l'est en français de France.

11 Et conjugués avec *avoir* à l'acadienne c'est-à-dire, pour prendre l'exemple du passé composé : *je m'ai* ou *je m'as* (*je me suis*), *tu t'as* (*tu t'es*), *i s'a* / *a s'a* (*il s'est* / *elle s'est*) *on s'a* ou *je nous avons* (*nous nous sommes*), *vous vous avez* (*vous vous êtes*), *i s'avont* ou sa contraction *i s'ont* (*ils se sont*).

des exemples comme « quand qu'i s'en ont aperçu i était bâsi [il avait disparu] » (corpus Péronnet), à la même 6^e personne avec pronom *en* qui s'intercale et se substitue à l'auxiliaire en tant que support du clitique *se* dans la mesure où ils excluent toute interprétation en termes d'assimilation avec un éventuel *sont* (être), d'où la conclusion que c'est bien un *avoir* contracté et que la transcription était donc probablement fiable.

Considérons enfin l'exemple de « tout l'monde s'vont rencontrer » extrait du corpus ontarien (corpus Mougeon-Hébrard). Outre la syllepse de l'accord pluriel de *vont* avec *tout l'monde*, banale au Canada, la particularité d'un exemple comme celui-ci tient au placement du *se* en présence d'un auxiliaire, en l'occurrence avec le *aller* d'un futur périphrastique, différent du placement qu'on aurait en français de France. En effet ici le clitique monte et se place non pas devant le second verbe V2 mais devant V1 (*s'aller rencontrer* et pas *aller se rencontrer*), laissant V1 et V2 en contact. Si le linguiste rencontrant une telle formulation connaît un peu le français classique et pré-classique, il se rappellera que ceci était la syntaxe du FF jusqu'au XVII^e siècle, qui est aussi l'époque du départ des colons, et il sera tenté d'établir entre les deux une corrélation plausible en voyant dans l'usage canadien la conservation d'un usage devenu désuet en France. Cependant ce piège sera évité une fois qu'il se sera rendu compte que cet emploi, loin d'être systématique, est dans ce corpus un hapax (et dans d'autres corpus introuvable).

Ces trois conclusions ne sont possibles que si on s'appuie sur un corpus interrogeable permettant de dégager des fréquences à partir de la comparaison de variables contextuelles contrôlées.

L'exploitation du corpus est utile avant même la finalisation de sa transcription, c'est-à-dire paradoxalement en cours de constitution, pour en établir la transcription qui permettra les exploitations par les chercheurs : quand on hésite entre deux interprétations, l'emploi majoritaire peut permettre de faire raisonnablement pencher la balance d'un côté plutôt que de l'autre et donc de faire le choix de la transcription la plus raisonnablement probable. Prenons le cas où la prononciation est quasiment la même pour deux transcriptions

différentes. Ainsi comment savoir s'il faut transcrire *j'en ai pas* ou *j'en n'ai pas* ? La collecte de la totalité des formes négatives émises par le même locuteur en contexte non ambigu permettra, si elle est radicale, de choisir entre négation unaire et négation binaire.

Comment choisir entre l'imparfait *j'étais*, le passé composé de l'indicatif *j'ai été* et le passé composé du subjonctif *j'aie été* dans :

Moi je m'ai jamais travaillé avant que (j'aie été)
 (j'ai été)
 (j'étais) marié

Après un subordonnant tel que *avant que* on peut s'attendre dans de nombreuses variantes de français à trouver du subjonctif, donc *avant que j'aie été marié*. Mais rien ne dit que la variété de français en question (ici acadien de Nouvelle-Écosse à l'Isle Madame dans le corpus Hennemann) suit cet usage, et le chercheur ne doit pas céder à l'automatisme d'une transcription conforme à sa propre pratique et être attentif au champ des alternatives que lui procure sa connaissance de la variété spécifique en question. En l'occurrence le subjonctif est souvent remplacé au Canada par l'indicatif du même temps (*avant que j'ai été marié*) voire par l'indicatif imparfait « *avant que j'étais marié* », choix qui a été retenu dans ce corpus comme transcription. Seule une vérification dans le corpus des formulations non ambiguës par le locuteur des temps et modes employés après *avant que* peut permettre de sélectionner une transcription plutôt qu'une autre.

Plus le corpus est vaste (et exploitable), plus les statistiques sont fiables et plus les vérifications ont des chances de donner lieu à une transcription juste. Tout en n'oubliant pas que si à partir de la présence d'un certain nombre d'occurrences attestées le chercheur peut conclure que tel emploi existe, à partir de son absence il ne peut pas conclure qu'il n'existe pas et doit en fait s'interdire toute conclusion. Si la transcription a été bien faite.

2. La transcription : que transcrire et comment le transcrire¹².

Bien sûr sera transcrit tout ce qui est de l'ordre du lexical produit (répétitions, autocorrections, apartés, *etc.*), mais on peut aller bien au-delà. Pour rendre au mieux la dimension supra-segmentale (accentuation, intonation, *etc.*) voire paraverbale (soupirs, gestuelle, *etc.*) de la production, certaines transcriptions sont enrichies d'éléments présents lors de la production¹³ :

Par exemple les événements extérieurs qui peuvent influencer la conversation. Ainsi dans les corpus de F. Martineau¹⁴, où ils sont transcrits entre parenthèses : (rire), (le téléphone sonne), (soupir).

Ou les gestes et postures des locuteurs. Ainsi dans le CFPQ (Corpus de Français Parlé au Québec) puisque des grimaces, des clins d'œil ou l'imitation d'une autre personne peuvent changer radicalement l'interprétation de la production.

Ou les allongements de phonèmes. Ainsi dans les corpus CFPQ, Falkert et Petras l'ajout de deux points précise qu'il y a allongement de la syllabe qui précède (« Euh :, c'est sûr »).

Ou la superposition des échanges, quand l'un choisit d'intervenir avant que l'autre n'ait terminé et que l'autre choisit ou pas d'en tenir compte dans sa propre production. Ainsi dans le corpus Wiesmath les chevauchements de paroles sont notés par une accolade, et dans le corpus Petras par un soulignage.

Ou les accentuations, qui rendent un certain état émotionnel ou une prise de position par rapport au discours d'autrui. Ainsi dans le corpus Wiesmath les mots accentués sont-ils soulignés (« épouvantable »). De façon plus précise le corpus Falkert signale quelle syllabe dans le mot est accentuée par l'ajout un espace entre les lettres du mot concerné (« é p o u vantable »), et

12 Il existe des logiciels permettant, avec des pourcentages de réussite variables, d'aligner manuellement une annotation à un signal audio ou vidéo, et donc à partir de l'extrait sous forme scriptée de retrouver plus aisément son correspondant dans le fichier primaire (Anvil, Clan, Exmaralda, Praat, Ilex, Ite, Elan, Transcriber, Transana, *etc.*).

13 Si l'étiquetage est un enrichissement du corpus, la mention d'éléments supra-segmentaux et paraverbaux est plus exactement une tentative de limiter l'appauvrissement de la version écrite.

14 Cf. http://lffa.rdc.uottawa.ca/philologic/Conventions_de_transcription.pdf.

le CFPQ fait de même au moyen de majuscules (« éPOUvantable »).

De même pour le volume, la vitesse du débit et l'intonation, qui peuvent être des facteurs discriminants. Ainsi le CFPQ note-t-il si le schéma intonatif est légèrement descendant (∖), légèrement montant (/), fortement descendant (↓) ou fortement montant (↑). Si le débit est allegro (<all<vous>>), lento (<len<vous>>), accelerando (<acc<vous>>) ou rallentando (<rall<vous>>). Si le volume est fort (<f<vous>>), fortissimo (<ff<vous>>), piano (<p<vous>>), pianissimo (<pp<vous>>), diminuendo (<dim<vous>>) ou crescendo (<cresc<vous>>).

Ou encore les pauses et les silences, parfois révélateurs, et assortis de leur durée. Ainsi dans les corpus Wiesmath et Falkert chaque point équivaut-il à une pause d'environ 1 seconde, d'où autant de points que de secondes de silence.

Autant d'indications qui peuvent être parfaitement pertinentes, mais alourdissent évidemment la clarté du corpus.

Des questions supplémentaires se posent quand il s'agit de rendre les contacts de langues, une situation qui concerne les français canadiens par rapport en particulier à l'anglais. Les mots d'origine anglaise se démarquent par une transcription particulière : entre guillemets français chez F. Martineau (« Il « check »ait et re« check »ait les « hot dog »s de la « bean »erie du coin ») et en italiques dans le corpus chiac Malenfant (« i ont besoin de tchèques-un en arrière pour *back* pas faire la passe ou tcheque affaire *right* »). Mais un emprunt à l'anglais peut être intégré morphologiquement et/ou phonétiquement au système français ou pas, ce que les corpus Wiesmath et Petras précisent par des italiques dans le premier cas (« tant ce ça été *tradé* pis ça été *shaké* HAND ») et par des majuscules dans le second (« je pourrais te *caller* BACK »)¹⁵.

15 Sur *back* voir Pierre-Don Giancarli, « Back en chiac » (dans *Travaux du CLAIX*, n° 24, Presses Universitaires de Provence, 2013, p. 73-89) car son statut de terme entièrement anglais ne va pas de soi. C'est ainsi qu'en chiac il a acquis un sens (répétition) et une position qu'il n'avait pas en anglais : « I avont back annoncé sur le radio que... ». Il peut aussi se trouver en corrélation avec un second verbe à particule, ce que ne fait pas l'anglais : « je crois que c'est back supposé venir on là. »

Une influence de l'anglais par calque, c'est-à-dire reconnue comme telle malgré l'absence de tout mot anglais, est rendue chez Martineau par un placement entre chevrons (« Il a < pris une marche > ce matin »). Mais ceci doit être le résultat d'une analyse ayant pris en compte puis écarté la possibilité d'une origine française, fût-elle ancienne et disparue partout ailleurs. Une recherche géographique dans les francophonies diasporiques est donc indispensable, de même qu'une recherche diachronique dans l'histoire du français de France, en fonction des vestiges parcellaires qui en sont arrivés jusqu'à nous et malheureusement souvent limités à de l'écrit d'un niveau de langue élevé. Le(s) transcrit(e)ur(s) gagnent donc non seulement à bien connaître la communauté en question (comme l'enquêteur) mais à être également versés dans la diachronie du français (et des langues régionales et dialectes de l'Ouest du pays) ainsi que dans les usages de l'anglais en particulier dans ses variantes américaines, tout en sachant que l'identification des calques peut ne pas être parfaite : ainsi dans l'extrait graphié « i a assez hâte de travailler / i a assez hâte d'être passé le chômage » (Corpus Péronnet) on peut se demander si on n'a pas plutôt affaire à *i a assez hâte d'être pas su l'chômage* calqué sur la construction prépositionnelle anglaise *be on the dole* correspondant bien au sens du contexte gauche.

La prudence est de mise avant de prendre les apparences pour d'indiscutables évidences : y a-t-il forcément une influence de l'anglais *move* sur le vernaculaire *mouvoir* quand on sait que ce verbe existait en français de France au xvi^e siècle ? De l'anglais *haddock* sur le vernaculaire *hadec* quand l'ancien français disait *hadot / hadon* ? De l'anglais *appointment* sur le vernaculaire identique quand le français du xvi^e siècle possédait *apointement* ?

Dans le domaine syntaxique des prépositions orphelines, dira-t-on que « moi ça dépend qui je suis avec » (extrait du corpus chiac Perrot) est du français vernaculaire ou un anglicisme ? Et « quoi-ce tu vas au club pour ? » ? Et « Je sais pas j'ai assez beaucoup vu de / de movieS / oh / y en a pas vraiment comme / que je peux penser à »¹⁶ ?

16 Voir Pierre-Don Giancarli, « Les prépositions orphelines dans les propositions

Quand dans l'extrait suivant du corpus acadien Wiesmath, le mari reformule en *outer* pour la française de France présente le français vernaculaire *tirer* (tandis que son épouse le re-traduit en français standard *traire*), s'agit-il d'un emprunt à l'anglais *out* doté d'une morphologie verbale française ou plutôt du verbe *ôter* prononcé avec l'ouïsme local ? :

L1 Dans les parcs des fois i vont là-bas dans le bois hein
faulait que t'alles les chercher pour les amener à la
grange pour les tirer tu sais pour leu/ leu outer le lait

L2 Les traire. (corpus Wiesmath)

La situation est d'ailleurs plus complexe que cela puisqu'elle ne se résume pas à ce choix binaire : en effet il peut aussi y avoir convergence entre ces deux langues relativement proches si une influence de l'anglais vient ressusciter/renforcer une forme française mais morte/moribonde en France. Les précisions venant enrichir la transcription se doivent donc d'être le résultat d'une chaîne de choix motivés.

Si des emprunts se font non à une langue mais à deux, un double système de notation est mis en place. Ainsi dans le corpus Lavallée du Manitoba les emprunts et alternances de code en anglais sont insérés entre astérisques (« un *tractor* », « *in all fairness* ») et ceux en langue indienne saulteux ou crie transcrits en symboles API insérés entre crochets (« [kukun] »). On pourrait bien sûr imaginer une discrimination plus fine entre alternances codiques et emprunts, entre emprunts intégrés et non-intégrés, et ce en faisant les distinguos pour chacune des langues impliquées. Il n'en demeure pas moins qu'au-delà de ces différences rendues dans la transcription on peut avoir affaire à un flux vécu comme homogène par le locuteur et représentatif de son identité.

Des questions supplémentaires se posent quand il s'agit de rendre une prononciation du français qui n'est pas celle du français standard de France, une situation qui concerne souvent les français

en *qu-* en acadien et en laurentien (Canada) », dans *Studii de Știință și Cultură*, Université Vasile Goldis d'Arad, vol. 13, n°1, 2017, p. 55-66. Sur Internet : <http://www.revista-studii-uvvg.ro/images/stories/48/1.5.pdf>.

canadiens. La réalisation phonétique de certaines consonnes finales est souvent rendue compréhensible pour le lecteur du corpus. Soit par une apostrophe en fin de mot après la consonne comme dans les corpus Petras, Falkert et Wiesmath (« le bout' », « les gens' »), soit par une majuscule comme dans le corpus Lavallée du Manitoba (« je l'ai touT faiT », « les genS »), soit encore par l'adjonction d'un -e final comme dans le corpus Hallion de cette même province (« le boute », « je l'ai toute faite ») ce qui peut laisser croire à tort à un accord avec un nom féminin alors que celui-ci a conservé le genre masculin du français de France.

En fait la question est de savoir jusqu'où on décide d'aller pour rendre les particularités de la variété étudiée et donc modifier l'orthographe du français, car plus on modifie plus on est proche de la prononciation réelle (selon les variétés par épenthèse, méatathèse, assibilation, palatalisation, ouïsme, diphtongaison, *etc.*) mais moins la transcription est lisible, et moins la recherche informatisée sera aisée, citons *donner* graphié en « dounner » (corpus Péronnet, Wiesmath, Petras, Hennemann), *aider* en « aïder » (corpus Péronnet, Wiesmath, Petras, Falkert), *revenir* en « arvenir » (corpus Falkert), *pouvoir* en « pouoir » (corpus Péronnet, Falkert, Petras), *il est intéressé dans les sports* [il s'intéresse aux sports] en « y est intéressé din sports » (Corpus Mougeon-Hébrard). À la limite les mots modifiés seront même perdus pour une recherche informatisée, ce qui prive le chercheur d'un outil performant, ou même pire le fait aboutir à des statistiques et interprétations erronées s'il n'a pas formulé les bonnes requêtes et est passé à côté de segments présents mais non reconnus.

Il arrive aussi que les transpositeurs modifient la graphie du français, ou plus exactement sa typographie, alors que cela ne semblait pas nécessaire. Prenons l'exemple des verbes de déplacement pronominaux (*s'en aller, etc.*)¹⁷ et plus précisément de

17 Le français de France a perdu un certain nombre de ses verbes de déplacement pronominaux au cours du temps. Ainsi à côté de la douzaine au moins qu'il comptait en ancien français (*s'en aller, s'en partir, s'en fuir, s'en venir, s'en revenir, s'en tourner, s'en retourner, s'en avancer, s'en entrer, s'en rentrer, s'en exir, s'en issir*), fin XVII^e siècle le Dictionnaire de l'Académie de 1694 n'en relève plus que trois : *s'en*

l'espace entre le radical (*aller*) et le *en* adverbial. Un espace serait de mise en français de France, mais il n'est pas toujours utilisé dans certains corpus laurentiens : « ils s'envenaient » (Corpus Martineau-Mocquais), « je l'ai vu il a pris son char pis il s'est envenu en ville » (idem), « Non non non i' i' se sont bien en bien envenus » (Corpus PCR de Bonnyville).

On pourrait se dire qu'un espace ou pas est un détail sans importance, et que d'ailleurs typographiquement le français de France a lui-même renoncé à conserver un espace entre *en* et le radical pour certains verbes de déplacement tel *s'enfuir*. Il s'agit là en effet d'une évolution, partant d'une situation où en ancien français le *en* était autonome (*s'en fuir*, comme *s'en aller* à cet égard aujourd'hui vestigial) pour en arriver à une situation où il est devenu préfixe. Cependant il ne s'agit pas pour autant d'un détail sans conséquence, car la graphie unaire ou binaire, avec *en* préfixé ou *en* adverbe autonome, n'est pas gratuite mais organisée. Elle est par exemple révélatrice de la façon dont se construisent les impératifs, en France comme au Canada : *s'enfuir* avec *en* préfixé donne *enfuis-toi / enfuyez-vous* et pas **fuis-t'en / *fuyez-vous en* (comme *s'envoler* donne *envole-toi / envollez-vous* et pas **vole-t'en / *volez-vous-en*). Par contre, dotés d'un *en* graphiquement autonome, *s'en aller* ne donne pas **en va-toi / *en allez-vous* mais *va-t'en / allez-vous-en*, *s'en retourner* ne donne pas **en retourne-toi / *en retournez-vous* mais *retourne-t'en / retournez-vous-en* (voire *retourne-toi-z'en*, *retourne-toé z'en*), et *s'en venir* ne donne pas **en viens-toi / *en venez-vous* mais *viens-t'en / venez-vous-en*.

Or on trouve dans les corpus canadiens concernés non pas **enva-toi* mais *va-t-en*, non pas **enretourne-toi* mais *retourne-t-en*, non pas **enviens-toi* mais *viens-t-en*. Les graphies unaires *s'envenir* ou *envenu* qu'on y trouve impliquent donc une construction à l'impératif qui n'est pas celle que l'on trouve en réalité dans ces mêmes corpus, ni dans les français du Canada (et de France) de façon

aller, s'en retourner, s'en venir. Aujourd'hui sur ces trois le français de France n'utilise plus guère que *s'en aller* alors que l'acadien et le laurentien ont conservé entre autres ces trois verbes (avec ajout éventuel de *back* en acadien et en manitobain).

générale. Elles sont donc trompeuses dans leurs implications, et en relation de contradiction implicite avec leurs propres vernaculaires. Quels corpus canadiens sont-ils concernés ? Pas les acadiens mais les laurentiens¹⁸, plus précisément certains corpus laurentiens, et seulement les deux correspondant aux provinces les plus à l'Ouest (Saskatchewan et Alberta)¹⁹. Les transcriptions nous disent donc peut-être aussi quelque chose de l'activité épilinguistique de leurs transpositeurs, qui en sont à un stade avancé de désémantisation du *en* initialement locatif et d'attraction vers un pôle synthétique de français.

Conclusion

Un corpus oral n'est pas un observable neutre donnant lieu à une transmission de données empiriques. Même en mettant les choses au mieux, c'est-à-dire en supposant qu'il n'a pas été constitué dans l'objectif qu'on puisse à partir de lui justifier seulement telle ou telle thèse, il reste un construit calibré, subjectif à plusieurs titres : il est déjà un recueil de faits retenus à partir de la sélection de certains informateurs, en fonction d'un certain protocole ayant pu influencer ces derniers de telle ou telle façon, formant un tout déclaré à un certain moment clos, et dont l'homogénéité aura été jugée suffisante par son concepteur en fonction des variables qu'il aura lui-même choisi de considérer comme pertinentes.

Il est un construit encore plus subjectif quand on passe au stade de sa transcription, qui le plus souvent est le seul élément sur lequel les chercheurs s'appuient pour échafauder leurs interprétations. Or en consultant le corpus ces derniers ne verront que ce qu'on veut bien leur montrer, en termes de quantité et en termes de rendu

18 On aura noté qu'il y a aux temps composés une différence de placement entre laurentien et acadien : l'acadien, comme le français de France, dit *i s'est en allé* ou plus exactement *i s'en a allé* avec l'auxiliaire *avoir*, tandis que le laurentien choisit *i s'est en allé*. Aux temps composés l'acadien, qui interpose l'auxiliaire entre *en* et le radical, ne pouvait donc pas être concerné.

19 Le 3^e corpus de l'Ouest (corpus Manitoba Lavallée) va dans le même sens mais adopte une position intermédiaire au travers d'un train d'union : *en-venu* et *en-allé*.

qualitatif plus ou moins précis et plus ou moins fiable. Ils fondent donc leurs interprétations sur ce qui est déjà partiellement une interprétation.

Lors de l'élaboration de la transcription, le transcripateur se retrouve face à des choix, des alternatives à réduire, des ambiguïtés à résoudre (à moins bien sûr qu'elles n'aient été reconnues comme volontaires). Une interprétation peut se trouver privilégiée par rapport à une autre en fonction des éléments paraverbaux qu'on aura choisi de rendre accessibles (grimaces, chevauchement de tours de parole, intonation, silences, *etc.*). Ces choix peuvent nécessiter de délicates recherches et vérifications en amont, par exemple pour identifier ce qui est archaïsme ou contact de langues, et au sein de cette seconde catégorie pour identifier le type de contact. Doit-on rendre les spécificités du vernaculaire ? Si oui comment, et jusqu'où aller / ne pas aller dans le degré de finesse dans l'adéquation et donc dans la modification du français dans le domaine graphique, voire typographique, chaque choix induisant des conséquences à évaluer de façon globale dans le vernaculaire en question ?

Si ces témoignages oraux qui sont partie intégrante de notre patrimoine immatériel sont correctement conservés et archivés sur des supports pérennes, ils seront exploitables non seulement de façon transversale par plusieurs disciplines mais ils donneront aux chercheurs des générations suivantes, munis d'autres théories et d'autres outils, l'occasion d'apporter grâce à eux de nouvelles réponses, ou de poser de nouvelles questions.

Annexe 1 : les corpus

1. Acadien
 - 1.1. Nouveau-Brunswick
 - 1.1.1. Corpus acadien traditionnel Péronnet, Université de Moncton, recueilli en 1971 dans le Sud-Est du Nouveau-Brunswick auprès d'informateurs âgés et ruraux, qui sont nés et habitent dans des villages à forte majorité francophone. 75 000 mots.
 - 1.1.2. Corpus acadien traditionnel Wiesmath, recueilli en 1996-1997 dans des villages du sud-est du Nouveau-Brunswick. Environ 118 000 mots.
 - 1.1.3. Corpus acadien chiac Perrot, recueilli en 1991 à Moncton auprès de jeunes francophones du sud-est du Nouveau-Brunswick. 79 000 mots.
 - 1.1.4. Corpus acadien chiac Malenfant, recueilli à Dieppe en 1994 auprès de jeunes francophones du sud-est du Nouveau-Brunswick. 20 300 mots.
 - 1.2. Nouvelle-Écosse
 - 1.2.1. Corpus acadien de Nouvelle-Écosse Hennemann, recueilli en 2005 auprès de locuteurs plutôt âgés de l'Isle Madame. Environ 146 500 mots.
 - 1.2.2. Corpus acadien de Nouvelle-Écosse Petras, recueilli entre octobre 2004 et janvier 2006 à partir de radios communautaires francophones diffusant sur la partie sud-ouest de l'île et couvrant les communautés de la Baie Sainte-Marie et de Pubnico. Environ 110 300 mots.
 - 1.3. Îles-de-la-Madeleine
 - 1.3.1. Corpus acadien Falkert, recueilli en 2003 auprès de locuteurs d'âges divers aux Îles-de-la-Madeleine. 170 000 mots.
2. Laurentien
 - 2.1. Québec
 - 2.1.1. CFPQ (Corpus de français parlé au Québec), corpus québécois des années 2000 de Gaétane Dostie, initié en 2006 et recueilli auprès de locuteurs de 15 à 85 ans, Université de Sherbrooke. Environ 1 million de mots.

2.2. Ontario

2.2.1. Corpus Mougeon-Hébrard, recueilli à Welland en 1975 auprès de 52 adultes et 10 adolescents. Les francophones représentent 16% de la population de la ville au sein d'une majorité anglophone. 584 604 mots.

2.3. Manitoba

2.3.1. Corpus Père Lavallée, recueilli en 1987 à Saint-Laurent par le père Lavallée auprès de 50 locuteurs entre 17 et 94 ans quasiment tous d'ascendance métis/mitchif, c'est-à-dire partiellement amérindienne. 484 200 mots²⁰.

2.3.2. Corpus Hallion, recueilli entre 1995 et 1997 dans le quartier traditionnellement francophone de Saint-Boniface à Winnipeg auprès de 33 locuteurs entre 19 et 97 ans. Environ 520 600 mots.

2.4. Saskatchewan

2.4.1. Corpus Martineau-Mocquais, recueilli en 1998 et comptant à l'époque de rédaction de ce travail 17 personnes âgées. 254 500 mots.

2.5. Alberta

2.5.1. Corpus PCR, recueilli en 1976 à Bonnyville (23 locuteurs), 182.200 mots ; Falher (11 locuteurs), 44.600 mots ; et Edmonton (26 locuteurs de moins de 25 ans) 110 250 mots. Soit un total de 337 000 mots.

Annexe 2 : Références bibliographiques supplémentaires

Aebischer, Verena et Claire Forel. *Parlers masculins, parlers féminins ?*, Neuchâtel-Paris, Delachaux et Niestlé, 1983, 197 p.

Armstrong, Nigel, Cécile Bauvois, Kate Beeching et Marielle Bruynincks. *La langue française au féminin. Le sexe et le genre affectent-ils la variation linguistique ?*, Paris, L'Harmattan, 2003, 238 p.

20 Pour autant il s'agit d'un corpus de franco-manitobain, pas d'un corpus de mitchif au sens de la langue mixte franco-indienne parlée par certains Métis de l'Ouest. Précisons enfin que tous les franco-manitobains ne connaissent pas d'influence mitchif.

- Baude, Olivier. *Les corpus oraux entre science et patrimoine*, Actes du colloque international du GRESEC *La publication de la science*, Grenoble, 2004, p. 7-11.
- Baude, Olivier, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, Pascal Cordereix, Laurence Goury, Michel Jacobson, Isabelle De Lamberterie, Christiane Marchello-Nizia, Lorenza Mondada. *Corpus oraux, guide des bonnes pratiques*, Paris, CNRS Éditions, 2006, 209 p. Sur Internet : https://hal.archives-ouvertes.fr/file/index/docid/357706/filename/Corpus_Oraux_guide_des_bonnes_pratiques_2006.pdf
- Bilger, Mireille (dir.). *Corpus, méthodologies et applications linguistiques*, Paris, Champion, 2000.
- Blanche-Benveniste, Claire. « Transcription de l'oral et morphologie », dans *Romania una et diversa*. vol. 1, *Philologische Studien für Theodor Berchem zum 65. Geburtstag* sous la direction de M. Guille et R. Kiesler, Tübingen, Gunter Narr Verlag, 2000, p. 61-74.
- Chaudenson, Robert. *Français avancé, « français zéro », créoles*, Aix-en-Provence, Publications de l'Université de Provence, 1985.
- Chaudenson, Robert, Raymond Mougéon et Édouard Béniak. *Vers une approche panlectale de la variation du français*, Paris, Didier-Érudition – ACCT, coll. « Langues et développement », 1993, 139 p.
- Mellet, Sylvie (dir.). *Corpus et recherches linguistiques*, dossier thématique dans *Corpus*, n°1, 2002. Sur Internet : <http://corpus.revues.org/49>.
- Gadet, Françoise. *La variation sociale en français*, Paris, Ophrys, coll. « L'essentiel français », 2003, 135 p.
- Giancarli, Pierre-Don. *Les Auxiliaires ÊTRE et AVOIR : étude comparée corse, français, acadien et anglais*, Rennes, Presses universitaires de Rennes, coll. « Rivages Linguistiques », 2011, 400 p.
- Giancarli, Pierre-Don. « La forme pronominale en québécois et en acadien », dans *La francophonie en Acadie* sous la direction de L. Arrighi et M. LeBlanc, Sudbury, Éditions Prise de Parole, 2014, p. 81-105.
- Habert, Benoît, Adeline Nazarenko et André Salem. *Les linguistiques de corpus*, Paris, Colin, 1998, 240 p.

- Labov, William. « The intersection of sex and social class in the course of linguistic change », dans *Language variation and change*, Cambridge, Cambridge University Press, n°2, 1990, p. 205-254.
- Papen, Robert. « La diversité des parlers français de l'Ouest canadien : mythe ou réalité ? », dans *Cahiers Franco-Canadiens de l'Ouest*, vol. 16, n° 1-2, 2004, p. 13-52.