CORPUS

Corpus

15 | 2016
Corpus de français parlé et français parlé des corpus

De l'archive de parole au corpus de référence : la base de données orales du français de Suisse romande (OFROM)

From speech archive to reference corpus: the spoken Swiss French database (OFROM)

Mathieu Avanzi, Marie-José Béguelin et Federica Diémoz

OpenEdition Journals

Édition électronique
URL : https://journals.openedition.org/corpus/3060
DOI : 10.4000/corpus.3060
ISSN : 1765-3126

Éditeur
Bases ; corpus et langage - UMR 6039

Édition imprimée
Date de publication : 15 octobre 2016
ISSN : 1638-9808

Ce document vous est offert par Université de Neuchâtel

unine UNIVERSITÉ DE NEUCHÂTEL

Référence électronique
Mathieu Avanzi, Marie-José Béguelin et Federica Diémoz, « De l'archive de parole au corpus de référence : la base de données orales du français de Suisse romande (OFROM) », *Corpus* [En ligne], 15 | 2016, mis en ligne le 15 janvier 2017, consulté le 17 août 2023. URL : http://journals.openedition.org/corpus/3060 ; DOI : https://doi.org/10.4000/corpus.3060

Ce document a été généré automatiquement le 16 février 2023.

Tous droits réservés

De l'archive de parole au corpus de référence : la base de données orales du français de Suisse romande (OFROM)

From speech archive to reference corpus: the spoken Swiss French database (OFROM)

Mathieu Avanzi, Marie-José Béguelin et Federica Diémoz

1. Introduction

Le français est, avec l'allemand, l'italien et le romanche, l'une des quatre langues officielles de la Confédération suisse. Il y est parlé par un peu plus de 22 % de la population (Lüdi & Werlen, 2005)[1], sur un territoire que l'on nomme Suisse romande, parfois Romandie. Au plan politique, la Suisse romande est composée de sept cantons, dont quatre sont officiellement unilingues (Genève, Vaud, Neuchâtel, Jura) et trois bilingues (Fribourg et Valais, à majorité francophone ; Berne, à majorité germanophone, *cf.* Schläpfer, 1985). Du fait de sa situation périphérique par rapport à la France, la Suisse romande a été, tout au long du XX^e^ siècle, sujette à l'insécurité linguistique (Jolivet, 1984). Comme l'ont montré des études sociolinguistiques menées, pour l'essentiel, dans le canton de Vaud, les Romands, bien qu'ils se déclarent parfois fiers de leur accent (Singy, 1996, 2004), entretiennent une attitude de « subordination linguistique » (Prikhodkine, 2011 : 24) par rapport aux Français, notamment parisiens, à qui ils tendent à attribuer une meilleure qualité de langue (Bayard & Jolivet, 1984 ; Singy, 1996 ; L'Eplattenier, 1998).

Le français que l'on parle en Suisse romande (désormais, pour plus de commodité, FS) n'est cependant pas uniforme (Knecht & Rubattel, 1984 ; Mahmoudian & Jolivet, 1984)[2]. Des variétés de FS ont ainsi été distinguées en fonction du canton où elles sont parlées (Voillat, 1971 ; Singy, 1996, 2004 ; Matthey, 2003 ; Andreassen, Maître & Racine, 2010) :

on a opposé la variété vaudoise à la variété fribourgeoise ou à la variété neuchâteloise, même si des distinctions plus fines peuvent être faites parmi les locuteurs d'un même canton (Voillat, 1971 ; Singy, 2004 ; Racine & Andreassen, 2012), voire à l'intérieur d'un même district ou d'une même bourgade (Métral, 1977 ; Matthey, 2003). Par ailleurs, les variétés de français parlé en Suisse ne figurent pas sur un pied d'égalité dans les représentations des Romands. Bien que le français parlé à Genève suscite des sentiments ambivalents, il semble jouir, de l'extérieur, d'un prestige plus grand que les autres variétés romandes (Schoch, 1980 ; L'Eplattenier, 1998), du fait qu'il est considéré comme plus proche du français dit « de référence » (désormais FR[3]).

Pourtant, c'est le français parlé dans la ville de Neuchâtel qui a longtemps passé pour être le « meilleur » français de Suisse romande (Jolivet & Bayard, 1984 ; L'Eplattenier, 1998 ; Racine, Schwab & Detey, 2013), notamment parce que les patois s'y sont maintenus moins longtemps qu'ailleurs (Gauchat, 1902 ; Gauchat, Jeanjaquet & Tappolet, 1925 ; Terrier, 1998 ; Kristol 1999 ; Kristol, 2013). Le français du canton de Vaud demeure associé à un accent rural, voire rustique (Knecht & Rubattel, 1984 ; Singy, 1996). Quant au français parlé dans le canton du Valais, s'il est parfois qualifié « d'incompréhensible » et de « chantant », il n'en a pas moins bonne presse chez les Romands, étant associé au sud et aux vacances…

Ces représentations reposent cependant sur des stéréotypes plus ou moins caricaturaux et rendent imparfaitement justice à la diversité des accents et des pratiques linguistiques que l'on peut observer sur le terrain. La base OFROM, que nous allons présenter dans cet article, a précisément pour vocation de permettre une observation directe des pratiques langagières en Suisse romande et de développer à ce sujet des études descriptives empiriquement fondées.

Au siècle dernier, les spécificités du français de Suisse romande ont été étudiées essentiellement sur la base de documents écrits ou d'exemples oraux recueillis à la volée (Pierrehumbert, 1926 ; Hadacek, 1983 ; Thibault, 1997). Les usages oraux, qu'il s'agisse de lexique, de syntaxe ou de phonologie, ont été moins bien documentés. De fait, on ne sait toujours pas quel crédit accorder aux stéréotypes populaires associés aux variétés de FS. Les locuteurs du canton de Neuchâtel, censés parler un français plus « pur », plus conforme à la norme, réalisent-ils *de facto* plus de *ne* de négation, d'accords en genre du participe passé et moins de dislocations du sujet que les locuteurs des autres cantons ? D'autre part, le FS a été décrit comme archaïsant. Mais qu'en est-il, de nos jours, des régionalismes attestés dans le parler des générations précédentes ? Les locuteurs actuels les utilisent-ils encore, et si oui, dans quelles conditions ? Quelles sont en outre, à date récente, les incidences linguistiques liées à la mobilité des personnes et à l'essor des nouvelles technologies de l'information et de la communication ? Et qu'en est-il, dans le parler des Romands, des néologismes observés dans les autres régions de la francophonie ? Se retrouvent-ils en Suisse, et si oui, chez quelles catégories de locuteurs ?

Le manque de données documentant le français parlé en Suisse romande[4] a longtemps freiné la recherche relative à ces différents points ; c'est la raison pour laquelle nous avons entrepris, à partir de 2011, de constituer une base de données de FS. Celle-ci a été mise en ligne en décembre 2012 sous le nom d'OFROM (*i. e.* Oral de Français de Suisse ROMande). Dans le présent article, nous présenterons d'abord les principes qui ont guidé la création de cette base. En deuxième partie, nous illustrerons les potentialités d'OFROM en examinant une série de faits relatifs au lexique et à la syntaxe[5].

2. Principes de constitution de la base

2.1 Enregistrements

Les enregistrements que la base contient actuellement sont pour les uns extraits d'entretiens guidés à dominante monologique, dans lesquels l'interviewé (un locuteur né en Suisse, et vivant en Suisse romande) était sollicité pour répondre à des questions nécessitant des réponses plus ou moins longues posées par l'intervieweur (le responsable de l'enquête) ; pour les autres, il s'agit d'interactions, impliquant généralement deux personnes (nées en Suisse, et vivant en Suisse romande) qui parlent à bâtons rompus. Les thèmes abordés concernent aussi bien le rapport à la langue, aux métiers, aux voyages, aux passe-temps des locuteurs, leurs relations de voisinage, leurs projets ou les situations incongrues auxquelles ils ont été confrontés dans leur vie. Ils peuvent également être en rapport avec le système politique ou la situation linguistique de la Suisse, voire porter sur les us et coutumes de la région où les locuteurs ont passé la plus grande partie de leur vie, etc. En moyenne, les entretiens enregistrés durent entre 30 et 40 minutes, mais seules une vingtaine de minutes sont transcrites pour chacun des locuteurs de la base (en moyenne, 10 minutes par conversation guidée, 10 minutes par discussion libre).

2.2 Locuteurs

Tous les locuteurs enregistrés sont originaires de Suisse. Ils sont en général enregistrés à l'endroit où ils ont passé la plus grande partie de leur vie. En plus de l'information géographique, cruciale dans OFROM (lieu de naissance, localité dans laquelle le locuteur a passé la plus grande partie de sa vie, localité d'habitation actuelle, nombre d'années passées dans la localité actuelle), nous récoltons pour chaque locuteur enregistré des informations sociodémographiques classiques, telles que l'âge du locuteur au moment de l'enquête, son sexe, sa langue maternelle, son métier et son niveau socio-éducatif[6].

2.3 Transcriptions

2.3.1 Support et conventions

Les transcriptions associées aux fichiers sons ont été faites soit par les étudiants en charge de l'enquête, soit par les collaborateurs scientifiques de la chaire de linguistique française et du Centre de dialectologie et d'étude du français régional de l'Université de Neuchâtel. Elles ont toutes été vérifiées, anonymisées et uniformisées par des étudiants de master et des collaborateurs scientifiques avant leur mise en ligne. Les enregistrements sont transcrits directement dans le logiciel Praat (Boersma & Weeninck, 2015), en orthographe standard, sans « trucages » ni ponctuation : nos conventions suivent en cela les recommandations du GARS (Blanche-Benveniste & Jeanjean, 1986 ; Blanche-Benveniste, 1997), reprises dans la plupart des corpus de français parlé transcrits existants (DELIC, 2004 ; Dister, Francard, Hambye & Simon, 2009 ; Baude & Dugua, ce volume ; Branca, Fleury, Lefeuvre & Pires, 2012).

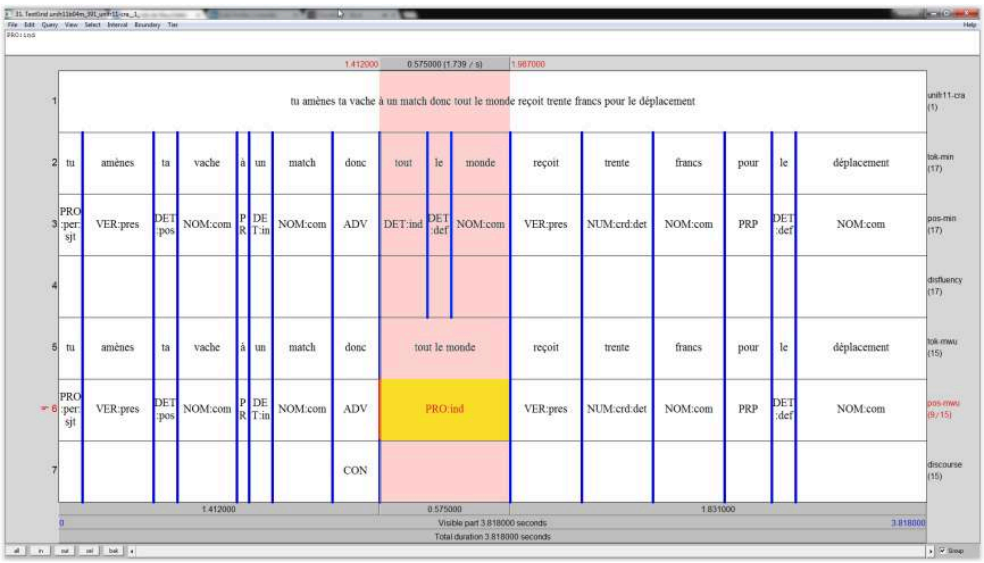
2.3.2 Anonymisation

La parole est une propriété (Baude, 2006). Les locuteurs enregistrés dans notre corpus ont signé des autorisations stipulant qu'ils donnaient leur accord pour l'enregistrement, la diffusion et l'analyse, à des fins linguistiques, de leur parole, à condition que les données soient anonymisées. Dans OFROM, nous n'avons pas procédé à une anonymisation du signal à proprement parler. Pour éviter de rendre publiques certaines informations prononcées pouvant servir à l'identification des locuteurs, nous avons simplement fait correspondre aux séquences sonores pouvant aider à l'identification du locuteur des intervalles dédiés à l'intérieur de la couche de transcription. Ces intervalles contiennent un symbole spécial (« # »), qui empêche, lors de la recherche à l'aide du concordancier, que le contenu sonore associé à l'intervalle incriminé puisse être entendu ou téléchargé. L'anonymat des locuteurs de notre corpus est ainsi préservé.

2.3.3 Annotations

En janvier 2015, les transcriptions ont été enrichies d'un codage des catégories morphosyntaxiques (*Part-of-Speech*) avec l'outil DisMo (Christodoulides, Avanzi & Goldman, 2014), qui ajoute à la transcription orthographique six couches supplémentaires d'annotation, comme on peut le voir sur la Figure 1 :

Figure 1. Annotation morphosyntaxique et discursive multi-niveau issue de l'étiquetage automatique fourni par DisMo, pour la séquence « tu amènes ta vache à un match tout le monde reçoit trente francs pour le déplacement » [unifr11-cra]



La première couche (tok-min) contient (non alignés avec le son) tous les mots séparés par un espace graphique. La seconde couche (pos-min) indique l'étiquette morphosyntaxique associée à chaque token. La troisième tire (disfluency), parallèle aux deux premières, indique les tokens disfluents. Les deux tires suivantes (tok-mwu et pos-mwu) séparent par des intervalles (non-alignés sur le son) les unités poly-lexicales, et leur catégorie morphosyntaxique. Ainsi, dans l'exemple de la Figure 1, la suite de tokens « tout le monde » (soulignée) a été traitée comme trois unités lexicales dans la tire tok-min, alors qu'elle n'en constitue qu'une seule dans la tire tok-mwu. Quant à la

dernière tire (discourse), elle peut regrouper indépendamment des unités lexicales et les catégoriser comme des marqueurs de discours. Les tires disfluency et discourse étant en cours de développement, les étiquettes que l'on peut y trouver ne seront pas présentées dans la section suivante. Les étiquettes utilisées proposées par DisMo se différencient de la plupart des étiquettes utilisées dans les taggeurs classiques (Debaisieux, Benzitoun & Deulofeu, ce volume) de par leur richesse et leur complexité.

La base de données OFROM n'est pas téléchargeable dans sa totalité : pour l'exploiter, il faut obligatoirement passer par le concordancier disponible sur le site du corpus. Celui-ci permet de chercher des chaînes de caractères et des mots, comme c'est le cas de la plupart des concordanciers associés aux autres corpus de français parlé présentés dans ce recueil. Il permet de surcroît, et c'est là son originalité, de procéder à des requêtes plus complexes, pour chercher des chaînes de constructions en croisant des critères lemmatiques et grammaticaux. Un tutoriel, téléchargeable sur le site web de la base, est disponible pour les utilisateurs.

2.4 Statistiques

Pour cet article nous faisons référence à l'état de la base de mars 2015[7] qui comprenait 407 763 tokens transcrits (soit 65 heures de parole), produits par 189 locuteurs, qui se répartissent de la façon suivante (*cf.* Figures 2 et 3) selon les cantons dans lesquels ils vivent :

Figure 2. Nombre de mots par canton dans le corpus OFROM (mars 2015), où JU = Jura, BE = Berne, GE = Genève, FR = Fribourg, VD = Vaud, VS = Valais, NE = Neuchâtel et NR = non renseigné

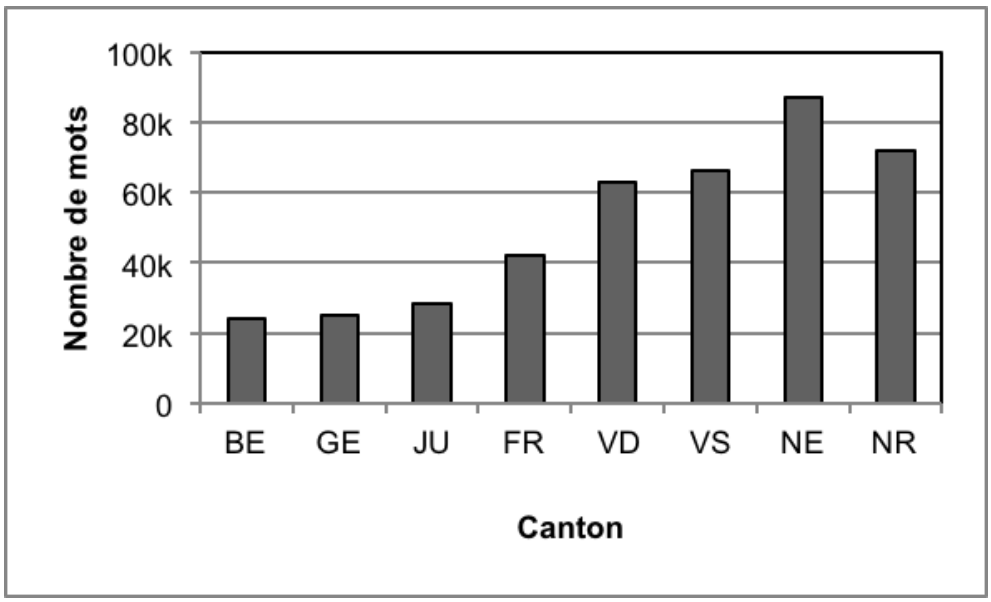
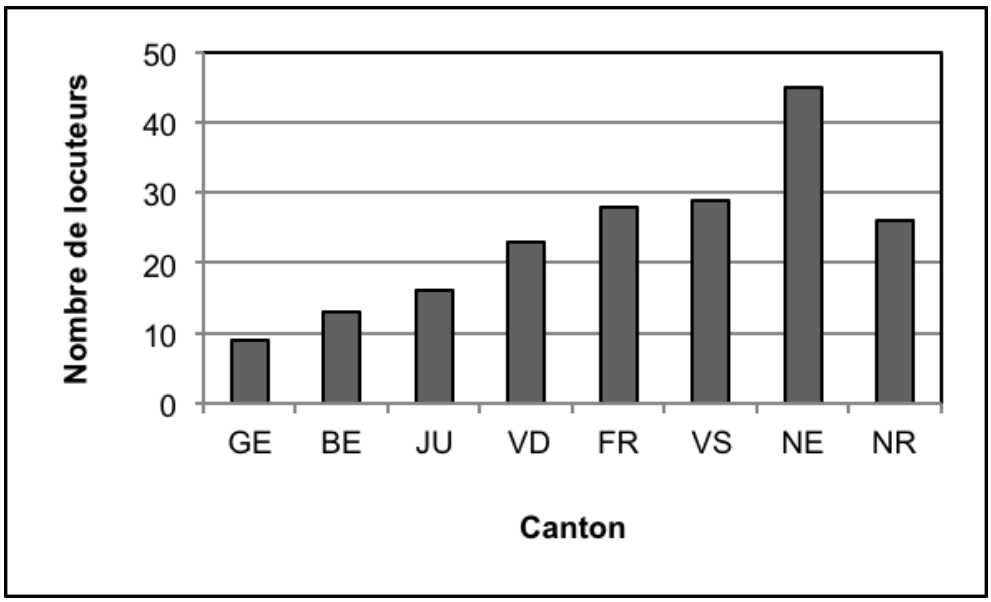


Figure 3. Nombre de locuteurs par canton dans le corpus OFROM (mars 2015), où JU = Jura, BE = Berne, GE = Genève, FR = Fribourg, VD = Vaud, VS = valais, NE = Neuchâtel et NR = non renseigné



3. Documenter la variation lexicale et syntaxique en Suisse romande

Cette partie de l'article est consacrée à la présentation de quelques recherches en cours, relatives aux particularismes lexicaux (§ 3.1) et à la variation syntaxique (§ 3.2).

3.1 Variation lexicale

Quelles sont, au début du troisième millénaire, les particularités lexicales du FS ? Parmi les régionalismes identifiés depuis longtemps par les spécialistes, lesquels sont toujours en usage, lesquels donnent des signes de déclin, du moins en l'état actuel de la base OFROM ? Qu'en est-il, par ailleurs, des néologismes repérés dans d'autres régions de la francophonie ? Appartiennent-ils ou non au français parlé en Suisse, et si oui, quels types de locuteurs sont concernés ? Est-il possible de mettre en évidence des néologismes qui seraient spécifiquement helvétiques ? Autant de questions que la base OFROM permet de documenter, ne serait-ce qu'à titre provisoire, car elle est appelée à se développer dans les années à venir.

3.1.1 Variation régionale

En guise d'illustration, nous évoquerons succinctement ci-après quelques régionalismes lexicaux dont OFROM permet de montrer le caractère bien ancré.

La lexie avoir meilleur temps de*, avec le sens d'« avoir intérêt à ».* – Cette expression passe parfois pour un helvétisme ; elle est cependant utilisée, suivant la Base de données lexicographique panfrancophone (désormais BDLP, Knecht & Kristol, 2000-2012), en Franche-Comté et dans les deux Savoie, dans le Doubs et le Jura français, voire au-delà. OFROM en livre trois occurrences à la 2e personne du singulier, dans la bouche de locuteurs différents, tous relativement jeunes :

(1) tu *as meilleur temps* de pas aller trop vite [unine08-ada, femme née en 1985]
(2) tu as tu *as meilleur temps* de pouvoir poser ton truc et pis que ça soit à ça du sol [unine12-asa, femme née en 1980]
(3) pis tu *as meilleur temps* d'apprendre sur une euh une manuelle je pense [unine15-003, homme né en 1996]

Le verbe venir *au sens de « devenir »*. – La BDLP suisse nous apprend que cette acception est ancienne et attestée dans une large partie de la francophonie, y compris le Québec ; l'ouest et le nord de la France font semble-t-il exception. Voici les exemples d'OFROM, relevés cette fois dans le parler de locuteurs d'un certain âge :

(4) je me fais du souci je me dis qu'il faut vite nettoyer pendant que c'est sec parce qu'il suffit qu'il *vienne* plus froid [unine11-jsa ; emploi impersonnel, au sens de : « il suffit qu'il se mette à faire plus froid » ; homme né en 1932]
(5) un potier qui *venait* de plus en plus vieux et puis qui pensait à sa retraite [unine08-eba, femme, âge non renseigné]

La bribe contenue dans (6) est suggestive à ce titre :

(6) euh on s'est marié on a eu quatre enfants | _ | et comme on habitait on restait en Suisse | et qu'elle est *de/ est venue* pasteure aussi en Suisse | _ | à | # | euh on a décidé que nos enfants | _ | euh parleraient le norvégien comme langue de base [unine11-rpa, homme né en 1954] [8]

Après avoir esquissé un *de/* qui peut passer pour une ébauche de la forme standard *devenir*, le locuteur de (6) se rabat finalement sur *venir* (il vaut la peine, dans le cas présent tout particulièrement, de se reporter à l'enregistrement).

La lexie familière foutre loin, *au sens de « jeter, mettre au rebut, congédier »*. – Absente du TLFi, cette expression est identifiée comme un régionalisme par la DBLP (s.v. *loin*). Elle est attestée à plusieurs reprises dans nos enregistrements :

(7) et je sais pas où a passé ce bouquin | _ | j'aimerais bien savoir où il a passé | _ | moi je l'ai jamais eu ma grand-mère l'a eu puis on l'a jamais retrouvé | _ | je sais pas ce que | et puis de qui les | _ | les filles ont fait | et compagnie elles ont tout *foutu loin* ce genre de truc | _ | je suis presque sûr elles en ont *foutu loin* plein | _ | ont *foutu loin* une partie des papiers | elles ont rien compris du tout de ce qu'elles avaient ont tout *foutu loin* bah alors non on n'a pas trouvé [unifr11-dba, homme, âge non renseigné]
(8) dis donc Germaine tu es en train de *foutre loin* des bouquins qui sont euh [unifr11-dba, id.]
(9) mais je pense que je vais arrêter de taffer là-bas quoi je sais pas ils vont me *foutre loin* mec si je bosse que six heures [unifr11-maa, étudiant, âge non renseigné]
(10) mais le problème c'est que tout à coup/ des fois tu as les paquets ils viennent pas chercher les paquets puis ils les *foutent loin* [unine11-ffa, homme né en 1990]

La BDLP mentionne également *foutre bas* au sens de « démolir » ou « abattre », dont OFROM livre une forme remotivée : *foutre en bas*.

(11) mais entre-temps ils ont *foutu en bas* toutes les archives [unine11-jsa, homme né en 1932]

L'emploi du verbe vouloir *comme auxiliaire de futur.* – En Suisse romande, comme dans les deux Savoie et le grand Est de la France (BDLP), le verbe *vouloir* est parfois employé comme auxiliaire pour marquer le futur périphrastique. Souvent considéré comme un germanisme (Pierrehumbert, 1926 : 649), cet emploi est en réalité un archaïsme. La base de données OFROM en fournit les attestions suivantes :

(12) tandis que maintenant tout est trié le fer euh | _ | alors euh ça donne beaucoup plus de commerce | _ | maintenant ben l'été je ne fais pas ça maintenant je *veux recommencer* une fois que j'ai fini au jardin | % | _ | commencer d'un petit peu trier par-là | _ | ouais ouais ouais ouais [unifr11-dba, homme, âge non renseigné]

(13) enfin on faut qu'on aille au marché de Saint Martin donc on *veut aller* au marché de Saint-Martin [unine12-asa, femme née en 1980]

Cas divers. – Signalons encore, pêle-mêle, les attestations de lexèmes tels que *pive* (« cône du sapin »), *cf.* (14)-(15) ; *huitante* (« quatre-vingts ») ; *cf.* (16)-(18) ; *cheni* (« fouillis, pagaille »), *cf.* (19)-(20) ; *souper* (« repas du soir »), *cf.* (21)-(22) ; *roillé* « fou », *cf.* (23) :

(14) on faisait des batailles de *pives* [unine15-029, femme née en 1936]
(15) oh joli avec la *pive* [...] alors ouais ouais ça c'est de la *pive* [unine15-027, femme née en 1987]
(16) on doit être présent à plus de *huitante* pourcents des cours pour que le cours nous soit validé [unine08-oca, femme née en 1988]
(17) d'ailleurs moi j'ai *huitante* ans je fais tout | mes affaires je sais tout faire [unifr11-dla, femme née en 1931]
(18) et c'est une tour on est mangé euh on a été jusqu'au *huitante-troisième* étage [unine11-nfa, femme née en 1989]
(19) parce que c'est elles qui sont venues foutre le *cheni* [unine11-sda, femme née en 1979]
(20) et y a le *cheni* habituel et tout [unine11-jma, femme née en 1954]
(21) ouais ou bien tu tu tu prends les restes du *souper* [unine12-asa, femme née en 1980]
(22) parce qu'on devait aller au bois on devait aller allumer le feu on devait faire euh | commencer le *souper* tout ça [unifr11-bga, femme née en 1931]
(23) y a un d/ un truc c'est un | _ | un chauffeur euh de à b/ de chez | # | # | | où | _ | qui devient complètement *roillé* après trente-cinq ans | qui conduit et tout ça il dit c'est affreux [unine15-024, homme né en 1928]

Le fait que, parmi ces occurrences, certaines ont été prononcées lors de l'enregistrement de jeunes locuteurs démontre que ces régionalismes sont encore bien vivants en Suisse romande. À l'inverse, la base OFROM contient des occurrences de tournures lexicales marquées comme « vieillies » dans la BDLP. C'est notamment le cas du *potager* (« cuisinière »), encore très répandu à la fin des années 60 (Voillat, 1971) :

(24) avec le l'évolution des matières et pis de de la technologie | _ | on a eu des *potagers* en | _ | à bois certainement mais | _ | déjà des *potagers* au dix-huitième siècle | _ | et pis on pouvait mettre d'autres casseroles dessus des tôles émaillées ou bien de [unine08-ebc, femme, âge non renseigné]
(25) le bain c'était dans un y avait pas de salle de bains y avait | une bassine à la cuisine | c'étaient des grandes bassines en acier | _ | % | _ | ouais | _ | à la cuisine | _ | chauffer l'eau sur le *potager* un *potager* | _ | vous savez ce que c'est | maintenant on dit une cuisinière [unine11-gaa, femme née en 1935]
(26) l'hiver tu vois | _ | et pis y avait pas | _ | y avait des fourneaux à molasse y avait *potager* à bois y avait pas tu vois de | _ | lave-linge euh tout ça | _ | non y avait une fontaine dehors [unine12-jda, femme née en 1930]

Dans ces trois exemples, *potager* est utilisé dans le cadre d'un rappel de réalités révolues ; dans (25), la locutrice prend même le soin d'en gloser le sens à l'intention de son interlocuteur. C'est aussi le cas du mot *torée* (« repas que l'on prend en plein air autour d'un feu », BDLP), dont la base ne livre qu'un seul exemple, dans la production d'un locuteur neuchâtelois :

(27) bah le c'est s/ quand on est entre amis c'est vrai que on va faire une *torée* au bord du | _ | voilà [unine15-903, homme né en 1958]

Lorsque la base sera plus développée, elle permettra, on le voit, de rendre de précieux services à la lexicographie différentielle, voire à la lexicologie en général.

3.1.2 Néologismes

Passons au cas des néologismes, qui se prêtent dans OFROM à des observations tout aussi instructives. Nous présenterons quatre exemples ; les trois derniers, comme on verra, concernent l'expression de l'intensité.

Le mot genre *utilisé en fonction d'opérateur d'approximation, d'illustration ou d'exemplification.* – Ces fonctions récemment acquises du lexème ont fait l'objet d'études de la part de Rosier (2002) et de Dufaye (2012, à par.), qui les abordent en termes de grammaticalisation. OFROM nous apprend que ces usages néologiques de *genre* sont très présents en Suisse chez les locuteurs des jeunes générations, *cf.* :

(28) parce que je me rappelle que *genre* quand j'avais sept ou huit ans | _ | j'ai commencé le tennis et je jouais au tennis avec | _ | et à l'époque il devait faire *genre* euh un mètre | _ | quarante [unine08-mba ; étudiant, âge non renseigné]
(29) c'est du brainstorming comme ils appellent ça | _ | *genre* c'est vraiment un truc mec le but c'est de les mettre pendant quatre jours ensemble [unifr11-maa, étudiant, âge non renseigné]
(30) je me suis dit je vais faire un petit truc au début tu sais je vais *genre* même commencer *genre* limite par la fin tu sais [unifr11-maa, id.]

Le Corpus suisse de SMS en français[9], auquel sont empruntés les exemples suivants, permet de montrer que l'écrit familier est également concerné :

(31) vs avez discuté *genre* de quoi? [Corpus suisse de SMS, 20413, femme, 17 ans, *sic*]
(32) Ben *genre* j'avais rendez-vous le plus tot possible avec Luigi qui repart à Genève pour lui installer Illustrator puis vers 15h pour finaliser un rapport [Corpus suisse de SMS, 15112, homme, 21 ans, *sic*]

Dans le corpus de SMS, sur 18 occurrences de *genre*, 17 sont de ce... genre. Le Tableau 1 ci-dessous offre une vue synthétique du nombre d'occurrences de *genre* par catégorie d'emploi (emplois nominaux classiques, emplois comme opérateurs d'approximation au sens large, avec, entre deux, les cas indécis). Le premier chiffre est relatif au corpus OFROM, le second au Corpus suisse de SMS en français. Les emplois néologiques de *genre* fournis par la base OFROM sont nombreux. À l'évidence, ils méritent mieux que le tri relativement grossier auquel nous les avons soumis. Nous nous proposons d'en étudier la distribution syntaxique et la répartition sociolinguistique dans une étude ultérieure (Béguelin, en prép.).

Tableau 1. Nombre d'occurrences du mot genre par catégorie d'emploi, dans OFROM d'une part, dans le Corpus suisse de SMS en français d'autre part

Type d'emploi	**Nb. d'occ. dans**		**Exemples**
	OFROM	**Corpus SMS**	
genre = N	70	1	ce *genre* de trucs un truc du *genre* des discussions de ce *genre*-là
Cas ambigus (*genre* = N apposé ou opérateur d'illustration)	14	4	un papier *genre* papier d'emballage; une petite ville forte *genre* Mont-Saint-Michel

genre = marqueur d'approximation, d'illustration ou d'exemplification	110	13	à *genre* 17 ans *genre* à 20 ans ils ont instauré des nouvelles lois | *genre* euh ils ont pas le droit d'avoir de trop grosses cylindrées
Totaux	**194**[10]	**18**	

Emplois intensifs de grave. – Zribi-Hertz (2015) a consacré une étude détaillée à l'évolution sémantico-syntaxique récente, en français informel, de cet adjectif dont le sens traditionnel est « sérieux, sévère », mais qui est utilisé désormais, dans les jeunes générations, comme marqueur adverbial de haut degré, au sens de « très, beaucoup », ou comme adjectif à valeur évaluative, avec la valeur de « fou, incroyable ». L'exemple suivant illustre successivement le second, puis le premier de ces emplois néologiques :

(33) Waw elle est *grave* cette zik je l'adore *grave*. [web < Zribi-Hertz 2015 : 65 ; « Cette musique est incroyable, je l'adore à fond »]

OFROM ainsi que le Corpus suisse de SMS démontrent que le français de suisse romande est également « grave atteint » par cette intrigante dérive fonctionnelle :

(34) En direct de Zurich, Carlo aux platines et CL au micro qui déchirent leur race *grave* ! :) [SMS 13152, homme, 24 ans ; *grave* sert ici d'intensif de prédicat verbal]
(35) [contexte de la conversation : engager quelqu'un qui filme pendant trois jours coûte cher]
ouais non *grave* même quoi [OFROM, unifr11-maa ; étudiant, âge non renseigné ; *grave* = « énormément »]
(36) je me disais que c'était peut-etre lui qui m'écrivait d'ailleurs tu vois! Je suis *grave* ! :) Bisous [SMS 21027, femme, 24 ans ; emploi en tant que « prédicat évaluatif général à orientation variable — dépréciative ou appréciative » de Zribi-Hertz 2015 : 93]

Emploi intensif de pire. – Cet usage est présent dans OFROM comme dans le Corpus suisse de SMS :

(37) et en fait je suis pas tombé amoureux tout de suite mais enfin j'ai déjà vu qu'elle était *pire* cool [unine08-eta, femme, âge non renseigné]
(38) Trop *pire* forte la fille : 21 patients et pas de retard... La classe... ;-) [Corpus suisse de SMS, 13883]
(39) Hé vieux! On a *pire* cartoné!!! \o/ Hallucinant:-D [Corpus suisse de SMS, 21841]

Emploi intensif de monstre. – Bien attesté dans la base OFROM, cet emploi est absent, de même que celui relevé de *pire* intensif, du corpus CFPP2000 (Lefeuvre & Brance-Rosoff, ce volume). Dans OFROM, *monstre* apparaît en tant qu'épithète antéposée (40), mais aussi comme intensifieur d'adjectif (42) ou de prédicat (43). Ces emplois concernent, dans la base suisse, 11 occurrences sur 13 du lemme en question, contre 2 exemples seulement de l'emploi nominal classique, style *le monstre du Loch Ness*. *Cf.* :

(40) ça fait des ça fait des *monstres* dégâts euh [unine08-mba ; 6 ex. de ce type dans la base, homme, âge non renseigné]
(41) ils sont *monstre* haut dans le euh dans l'organigramme quoi [unifr11-maa, homme, âge non renseigné ; 4 ex. de ce type dans la base]
(42) il a *monstre* poussé le gazon [unine11-ffa, homme né en 1990 ; seul exemple ad-verbal]

L'espace à disposition nous contraint à clore provisoirement cet inventaire. Les exemples présentés ci-dessus n'avaient d'autre ambition que de montrer le potentiel de la base OFROM (ainsi bien sûr que du Corpus suisse de SMS), en vue d'une meilleure

connaissance non seulement des particularités lexicales du français en Suisse, mais aussi des évolutions qui concernent la langue française en général.

3.2 La variation morphosyntaxique

Cette section est organisée en deux parties. La première est consacrée aux régionalismes syntaxiques (§ 3.2.1), dans la seconde nous analysons dans le corpus la fréquence de (non-)réalisation du *ne* de négation (§ 3.2.2).

3.2.1 Variation régionale

Sur le plan syntaxique, la mise à disposition de corpus oraux de plus en plus riches, OFROM compris, aidera à déterminer la représentation géographique, très souvent transfrontalière, de variantes telles que (43)-(50), qui passent à tort dans certains travaux pour des spécificités du français parlé en Suisse. Ces variantes concernent l'ordre des mots (comme dans (43)-(45)), le mode de réalisation de certaines valences verbales (comme dans (46)-(48)), et l'usage de certaines tournures grammaticales particulières (comme dans (49)-(50))[11] :

(43) je ça regarde [Bürgi 1999 : 149]
(44) j'ai personne vu [Redard 1971 : 3]
(45) je lui le donne [Tuaillon 1983 : 234]
(46) aider à quelqu'un [Lüdi 1981 : 90]
(47) demander après quelqu'un [Knecht & Rubattel 1984 :141]
(48) ça, j'y veux [Tuaillon 1983 : 230]
(49) le chien m'est venu contre [Voillat 1971 : 224]
(50) il a eu fumé [Walter 1981 : 28]

De ces tours, on ne sait que peu de choses. D'un point de vue diatopique, certains de ces phénomènes s'étendent à l'ensemble du domaine francoprovençal (notamment l'antéposition de *personne*), voire au-delà (l'usage du passé surcomposé en principale), d'autres ne sont pas connus en dehors de certaines régions bien spécifiques (l'antéposition de *ça* est généralement décrite comme un phénomène typiquement vaudois, Bürgi, 1999 ; l'usage du pronom neutre *y*, bien connu dans les deux Savoie, en Isère et dans le Rhône et en Bourgogne ne semble pas être employé en Suisse romande ailleurs qu'à Genève, Tuaillon, 1983). De leur vivacité dans les conversations contemporaines, on ne sait pour le moment que peu de choses. Jusqu'à présent, on l'a dit, les spécificités lexicales et syntaxiques supposées du FS ont surtout été étudiées sur la base de documents écrits, à partir d'exemples oraux recueillis au cours d'enquêtes ponctuelles ou des jugements de quelques informateurs. À ce jour, aucune étude systématique visant à vérifier la validité empirique de ces remarques sur un échantillon d'informateurs plus large (qui permettrait de tenir compte non seulement de l'origine géographique des informateurs, mais aussi de leur âge, de leur sexe et de leur statut socio-économique), n'a jamais été conduite. Nous avons cherché à pallier cette lacune en interrogeant la base de données OFROM. Cependant, la recherche des contextes syntaxiques exemplifiés sous (43)-(50) n'a pas donné de résultats vraiment concluants. Nous n'avons en effet trouvé qu'un seul emploi de *personne* dans un emploi de type appositionnel relativement particulier, mais différent de (44), *cf.* (51). Il a été prononcé par une locutrice âgée, qui parle encore le patois de la région de Fribourg :

(51) ils ont *personne* le même patois mais c'est assez près [unifr11-dla, femme née en 1931]

Nous avons trouvé une seule attestation du tour pro_{datif} *V contre* (*cf.* (49) *supra*), avec le sens de « arriver sur soi », prononcé par une locutrice âgée d'une vingtaine d'années, originaire du canton de Fribourg :

(52) un coup de chaud qui *nous* arrivait *contre* [unine12-avb, femme née en 1991]

Le corpus contient un pronom *y*, faisant office d'objet direct renvoyant à un référent propositionnel, prononcé par un locuteur vivant dans le canton de Genève :

(53) on peut le chauffer avec la lampe à souder alors euh ils *y* font [unine15-017, homme né en 1941]

En ce qui concerne le passé surcomposé, nous avons trouvé les sept attestations suivantes (54)-(60), dont quatre s'insèrent dans des subordonnées à valeur temporelle, ce qui confirme qu'il s'agit d'un contexte d'apparition privilégié pour ces formes (Jolivet, 1984). Ici aussi, les locuteurs sont tous originaires de cantons différents :

(54) quand *il a eu fini* l'école un qui travaillait à l'UBS il lui a dit toi tu vas venir à la à l'UBS et tu peux travailler [unifr11-dla, femme née en 1931]
(55) et dès qu'on *a eu tourné* le dos ils ont remis le les petites midinettes euh qui se trémoussaient dans tous les sens donc c'est c'est encore ce côté que je trouve très très hypocrite [unine09-lba, femme née en 1976]
(56) et puis quand *j'ai eu fini* ma euh ma formation de nurse [unine14-smc, femme née en 1938]
(57) déjà avant de faire son diplôme ce qui fait que quand il *a eu fini* on est retourné à # qu'on connaissait bien [unine11-lva, femme née en 1933]
(58) *j'ai eu été* jouer au volley avec des copains au badminton [unine11-fdb, homme née en 1987]
(59) j'*ai eu été* euh avec euh en en sortie avec euh des amis à Europa Park [id.]
(60) on s'est jamais chicané | mais on *a eu été* vingt minutes | trente minutes sur un mot [unine15-033, femme née en 1936]

Sur le plan de la variation régionale, les exemples que nous avons recensés demeurent trop peu nombreux pour que l'on puisse parler de véritable variation dans l'espace, ou pour que l'on puisse faire des hypothèses sur les facteurs sociodémographiques qui les motivent. Quant aux formes non attestées, on ne se risquera pas non plus à tirer argument de leur absence dans la base OFROM pour dire qu'elles ont disparu des usages. On sait en effet que des tournures syntaxiques pourtant courantes dans les conversations de tous les jours ne le sont pas forcément dans les corpus oraux (Bilger & Cappeau, 2004 ; Cappeau & Gadet, 2007). C'est pourquoi d'autres méthodes doivent être envisagées afin de documenter la vivacité et la répartition effectives de ces tours dans les variétés de français de Suisse romande.

3.2.2 Note sur la (non-)réalisation du ne de négation

Cette section est consacrée à l'alternance *ne/0* dans le marquage de la négation. Leur nombre étant suffisant dans le corpus, nous testons également les effets des variables sociodémographiques des locuteurs sur cette alternance (âge, niveau socio-éducatif et origine cantonale).

Dans un premier temps, nous avons réalisé une estimation du pourcentage de double négation dans le corpus en divisant le nombre d'occurrences des *ne* et *n'* taggées comme adverbes de négation par le nombre total d'occurrences des formes *pas/personne/rien/jamais/aucun/plus*, taggées comme adverbes de négation[12]. Sur les 5 857 négations extraites du corpus, seulement 427, soit 7,3 %, contenaient le morphème *ne*. Pour étudier la répartition sociale et géographique de ces formes dans

l'espace, nous n'avons retenu que les formes pour lesquelles on disposait des informations relatives à l'origine, l'âge et au statut social du locuteur. Nous avons exclu les locuteurs pour qui le français n'est pas la langue maternelle, et avons regroupé ceux du Jura et de Berne dans un seul et même groupe. Sur les 4 345 formes restantes, 234 formes (soit 5,3 %) contiennent un *ne* de négation. Pour tester l'impact de variables sociodémographiques sur la présence ou l'absence de *ne*, nous avons effectué trois modèles linéaires généralisés à mesures répétés (Ghisletta & Spini, 2004), avec la présence de *ne* (VRAI/FAUX) comme variable dépendante et le locuteur comme variable aléatoire[13]. Dans un premier modèle, le canton dans lequel le locuteur a passé la plus grande partie de sa vie a été entré comme variable indépendante. Les résultats ont permis de montrer que la variable diatopique avait un effet sur le choix du type de négation (Wald χ^2 (5) = 16.592, $p < 0.01$). Les tests *post-hoc* ont cependant montré que parmi les différences que l'on observe sur la figure 4, seuls les locuteurs de Genève produisent moins de *ne* de négation que les locuteurs de Fribourg, du Jura et du Valais ($p < 0.05$). Dans un second modèle, l'année de naissance du locuteur a été entrée comme variable indépendante. Les résultats ont permis de montrer que cette variable avait un effet sur le choix du type de négation (Wald χ^2 (1) = 10.340, $p < 0.001$). Comme on peut le voir sur la figure 5 plus bas, plus le locuteur est jeune, plus il a tendance à ne pas double-marquer ses négations.

Figure 4. Pourcentage de négations comportant le morphème ne, en fonction du canton dans lequel le locuteur a passé la plus grande partie de sa vie

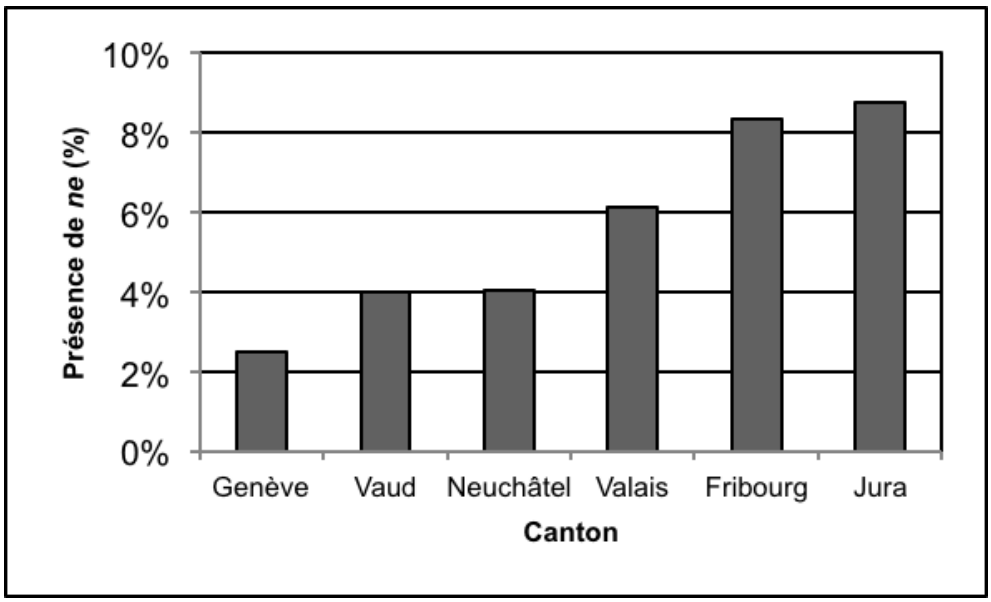
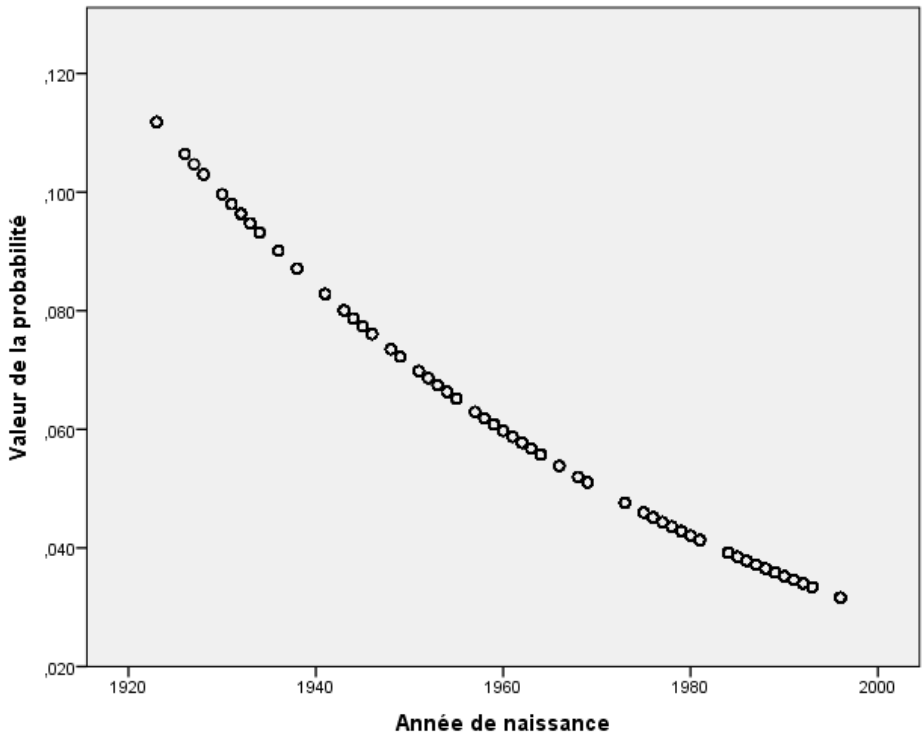


Figure 5. Probabilité que la négation contienne le morphème ne, en fonction de l'année de naissance du locuteur



Sur le plan distributionnel, les résultats que nous avons obtenus confirment ceux de Fonseca-Greber (2007) et de Meisner (2013), qui observent des taux de réalisation très bas du *ne* de négation dans les corpus de FS qu'elles étudient. Ces résultats sont assez proches des taux calculés après l'analyse de productions de locuteurs vivant dans le Nord de la France (moins de 10 % chez les locuteurs analysés par Hansen & Malderez, 2004 et Torreira, Adda-Decker & Ernestus, 2010), alors qu'avec plus de 50 % de réalisation la présence de cette marque est encore bien vivace chez les locuteurs du Midi (Diller, 1983) et de Belgique (Moreau, 1986). En ce qui concerne la variation inter-cantonale, nous avons observé une propension plus grande à faire chuter le *ne* de négation à Genève qu'ailleurs, mais nous n'avons pas trouvé un taux de maintien plus haut dans les productions des Neuchâtelois, ce qui permet de répondre au moins provisoirement, à la question posée en introduction sur le français des Neuchâtelois. Dans notre analyse, les résultats relatifs à l'âge des locuteurs confirment ce qui avait déjà été montré par d'autres que nous (*cf.* notamment Ashby, 1967, 1981 ; Coveney, 1998 ; Armstrong, 2002 et Hansen & Malderez, 2004). Nous avons en effet observé que plus l'âge des locuteurs augmentait, plus le taux de maintien de *ne* augmentait[14]. Quant à l'effet de niveau socio-éducatif, sans surprise celui-ci ne s'est pas révélé significatif, ce qui confirme les observations de Meisner, Robert-Tissot & Stark (à par.). On conclura en rappelant que ces résultats doivent être appréhendés avec précaution : nous n'avons pas pris en compte les effets possibles des indices syntaxiques (position du *ne* par rapport au verbe), sémantiques (portée de la négation) et prosodique (vitesse de parole, nombre de syllabes dans le groupe accentuel hôte, etc.), or, on le sait, ceux-ci jouent un rôle important dans la distribution de *ne* (Meisner, 2013).

4. Conclusion

Dans cet article, nous avons présenté la base de données orales de français de Suisse romande, OFROM. Nous avons dans un premier temps rappelé les hypothèses et les principes fondamentaux qui ont guidé sa constitution, ainsi que les aspects relatifs à la transcription et à l'annotation des données. La base, d'une taille approximative de 400 000 mots au moment où a été réalisée cette étude, en compte plus de 800 000 au moment où elle paraît. C'est l'une des premières à avoir été entièrement annotée en parties du discours, et à disposer d'un concordancier qui permet à n'importe quel utilisateur de faire en ligne des recherches complexes. Même si elle ne permet pas encore de tester solidement des hypothèses sur la répartition des particularismes lexicaux et syntaxiques, elle permet déjà d'illustrer certaines tendances, de renouveler les données d'ordinaire utilisées pour approcher la variation, et surtout de poser les bases d'une description systématique du français parlé en Suisse romande à l'orée du XXIe siècle. D'ici quelque temps, la base devrait être enrichie de nouvelles annotations, notamment pour étudier la variation phonologique et son interface avec la syntaxe. L'ajout de données de genres discursifs variés est également envisagé. La base devrait ainsi progressivement devenir un corpus de référence, au sens classique du terme (Habert, 2000).

Remerciements

La confection de la base OFROM n'aurait jamais été possible sans le soutien financier du programme Campus virtuel suisse, de la Faculté des Lettres et Sciences humaines et du Rectorat de l'Université de Neuchâtel, ainsi que du Fonds national suisse de la recherche scientifique (subsides n° P300P1_147781 et n° P3P3P1_161040). Nous remercions Pierre Ménétrey (http:// www.webox-it.com/), webmestre, pour le travail de confection du site. Merci également à Sandra Schwab (universités de Genève et de Zurich) pour ses conseils et pour la confection des scripts Praat qui ont permis la mise en ligne des premières données sonores et des transcriptions associées. George Christodoulides nous a en outre fourni le logiciel pour tagger la base de données et créer des fichiers xml pour la charger. François Delafontaine (Université de Neuchâtel) a réalisé un travail colossal de révision et de correction des transcriptions. Il a été rejoint récemment par Maude Ehinger et Julie Rothenbühler (Université de Neuchâtel). Christophe Benzitoun a nettoyé, dans le cadre du projet ANR ORFEO, certains des fichiers présents dans la base. Qu'ils soient toutes et tous remerciés très sincèrement. Enfin nous remercions de leur généreux engagement l'ensemble des collaborateurs scientifiques, des étudiants et des locuteurs qui ont participé aux diverses campagnes d'enquête.

BIBLIOGRAPHIE

Andreassen H., Maître R. & Racine I. (2010). « La Suisse », *in* S. Detey, J. Durand, B. Laks & C. Lyche (éd.) *Les variétés du français parlé dans l'espace francophone : ressources pour l'enseignement*. Paris/ Gap : Ophrys, 201-212.

Armstrong N. (2002). « Variable deletion of french *ne* : a cross-stylistic perspective », *Language Sciences* 24 : 153-173.

Ashby W. (1976). « The loss of the negative morpheme *ne* in Parisian French », *Lingua* 39 : 119-137.

Ashby W. (1981). « The loss of the negative particle *ne* in French : a syntactic change in progress », *Language* 57 (3) : 674-687.

Avanzi M. (2012). *L'interface prosodie/syntaxe en français. Dislocations, incises et asyndètes*. Bruxelles : Peter Lang.

Avanzi M., Béguelin M.-J. & Diémoz F. (2012). « Présentation du corpus OFROM – corpus oral de français de Suisse romande ». Manuscrit, Université de Neuchâtel.

Avanzi M., Schwab S., Dubosson P. & Goldman J.-P. (2012). « La prosodie de quelques variétés de français parlées en Suisse romande », *in* A. C. Simon (éd.) *La variation prosodique régionale en français*. Louvain-la-Neuve : De Boeck/ Duculot, 89-120.

Bardiaux A. & Boula de Mareüil P. (2012). Allongements vocaliques en français de Belgique : approche perceptive et expérimentale. *Actes des 29es JEP*, non paginé.

Bayard C. & Jolivet R. (1984). « Des Vaudois devant la norme », *Le français moderne* 52 : 151-158.

Béguelin M.-J. (en prép.). « Les emplois de *genre* dans la base OFROM ». Manuscrit, Université de Neuchâtel.

Blanche-Benveniste C. (1997). « La notion de variation syntaxique dans la langue parlée », *Langue Française* 115 : 19-29.

Blanche-Benveniste C. & Jeanjean C. (1986). *Le français parlé. Transcription et édition*. Paris : Didier Érudition.

Blasco-Dulbecco M. (1996). « Pour une approche syntaxique des dislocations », *Journal of French Language Studies* 7 : 1-21.

Boersma P. & Weenink D. (2015). Praat, doing phonetics by computer, v. 5.4, http:// www.fon.hum.uva.nl/praat/.

Branca-Rosoff S., Fleury S., Lefeuvre F. & Pires M. (2009). « Discours sur la ville. Corpus de français parlé parisien des années 2000 (CFPP2000) », http://cfpp2000.univ-paris3.fr/.

Bürgi A. (1999). « Le pronom *ça* en français vaudois », *Vox Romanica* 58 : 149-171.

Cappeau P. & Gadet F. (2007). « Où en sont les corpus sur les français parlés ? », *Revue française de linguistique appliquée* 12 : 129-133.

Carton F., Rossi M., Autesserre D. & Léon P. (1983). *Les accents des Français*. Paris : Hachette.

Christodoulides G., Avanzi M. & Goldman J.-P. (2014). « DisMo : a morphosyntactic, disfluency and multi-word unit annotator. An evaluation on a corpus of French spontaneous and read speech ». *Proceedings of the 9th International Conference on Language Resources and Evaluation*, 3902-3907.

Coveney A. (1998). « Awareness of linguistic constraints on variable *ne* omission », *Journal of French Language Studies* 8 : 159-187.

DELIC (2004). « Présentation du Corpus de référence du français parlé », *Recherches sur le français parlé* 18 : 11-42.

Deshaies D. (1991). « Contribution à l'analyse du français québécois : études des pronoms personnels », *Revue québécoise de linguistique théorique et appliquée* 10/3 : 11-40.

Detey S., Durand J., Laks B. & Lyche C. (2010). *Les variétés du français parlé dans l'espace francophone. Ressources pour l'enseignement*. Paris/Gap : Ophrys.

Detey S. & Le Gac D. (2008). « Didactique de l'oral et normes de prononciation : quid du français "standard" dans une approche perceptive », *Actes du 1er CMLF*, 475-487.

Diller A.-M. (1983). « Subject NP structure and variable constraints : the case of french *ne* deletion », *in* R. Fasold (éd.) *Variation in the Form and the Use of Language*. Washington : Georgetown University Press, 167-174.

Dister A., Francard M., Hambye P. & Simon A. C. (2009). « Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de la banque de données textuelles orales VALIBEL (1989-2009) », *Cahiers de l'Institut de linguistique de Louvain* 33 : 113-129.

Dufaye L. (2012). « *Genre* : Trace d'un transfert de prise en charge », *in* M. Birkelund & H. Nølke (éd.) *La Linguistique énonciative*, Aarhus, 49-63.

Dufaye L. (à par.). « *Genre* ou le scénario d'une grammaticalisation », *Linx*.

Durand J., Laks B. & Lyche C. (2002). « La phonologie du français contemporain : usages, variétés et structure », *in* C. Pusch & W. Raible (éd.) *Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, 93-106.

Durand J., Laks B. & Lyche C. (éd.) (2009). *Phonologie, variation et accents du français*. Paris : Hermès.

Fonseca-Greber B. (2007). « The Emergence of emphatic *ne* in conversational Swiss French », *Journal of French language Studies* 17 : 249–275.

Francard M. (1997). « Le français en Wallonie », *in* D. Blampain, J.-M. Goosse, J.-M. Klinkenberg & M. Wilmet (éd.) *Le français en Belgique. Une langue, une communauté*. Louvain-la-Neuve : Duculot, 229-237.

Gadet F. (1992). *Le français populaire*. Paris : PUF.

Gauchat L. (1902). « Nos patois romands », *Bulletin du Glossaire* 1 : 3-24.

Gauchat L., Jeanjaquet J. Tappolet E. (1925). *Tableaux phonétiques des patois suisses romands. Relevés comparatifs d'environ 500 mots dans 62 patois-types*. Neuchâtel : Attinger.

Ghisletta P. & Spini D. (2004). « An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals », *Journal of Educational and Behavioral Statistic* 29/4 : 421-437.

Habert B. (2000). « Des corpus représentatifs : de quoi, pour quoi, comment ? », *Cahiers de l'Université de Perpignan* 31 : 11-58.

Hadacek C. (1983). *Le suisse romand tel qu'on le parle. Lexique romand-français*. Lausanne : P.-M. Favre.

Hambye P. & Simon A. C. (2009). « La prononciation du français en Belgique », *in* J. Durand, B. Laks & C. Lyche (éd.) *Phonologie, variation et accents du français*. Paris : Hermès, 95-130.

Hansen A. B. & Malderez, I. (2004). « Le *ne* de négation en région parisienne. Une étude en temps réel », *Langage et société* 107 : 5-30.

Jolivet R. (1984). « L'acceptabilité des formes verbales surcomposées », *Le Français moderne* 52 : 159-176.

Knecht P. (1979). « Le français en Suisse romande : aspects linguistiques et sociolinguistiques », *in* A. Valdman (éd.) *Le français hors de France*. Honoré Champion, 249-258.

Knecht P. (1985). « La Suisse romande », *in* R. Schläpfer (éd.) *La Suisse aux quatre langues*. Genève : Éditions Zoé, 125-169.

Knecht P. (2000). « Le français en Suisse romande », *in* G. Antoine & B. Cerquiglini (éd.) *Histoire de la langue française 1945-2000*. Paris : Éditions du CNRS, 719-727.

Knecht P. & Kristol A. (2000-2012). « Base de données lexicographique panfrancophone (la Suisse romande) », www. bdlp.org.

Knecht P. & Rubattel C. (1984). « À propos de la dimension sociolinguistique du français en Suisse romande », *Le français moderne* 52 : 138-150.

Kristol A. (1999). « Histoire linguistique de la Suisse romande : quelques jalons », *Babylonia* 3/99 : 8-13.

Kristol A. (2013). « Regards sur le paysage linguistique neuchâtelois (1734-1849) : le témoignage sociolinguistique des signalements policiers », *in* A. Gendre *et al.* (éd.), *Des mots rayonnants, des mots de lumière : mélanges de littérature, d'histoire et de linguistique offerts au professeur Philippe Terrier*. Neuchâtel : Université de Neuchâtel, Faculté des Lettres et Sciences humaines ; Genève : Droz, 277-295.

Laberge S. (1977). *Étude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal*. PhD Thesis, Université de Montréal.

L'Eplattenier C. (1998). *A Perceptual Dialect Study of French in Switzerland.* PhD Master, Lausanne.

Lüdi G. (1981). « Sémantique, syntaxe et forme casuelle. Remarques sur la construction *aider à qn* en français romand », *Vox Romanica* 40 : 85-97.

Lüdi G. & Werlen I. (2005). « Le paysage linguistique en Suisse. Recensement fédéral de la population 2000 », manuscrit, http://www.bfs.admin.ch/bfs/portal/fr/index/themen/01/22/publ.html?publicationID=1738.

Lyche C. (2010). « Le français de référence : éléments de synthèse », *in* S. Detey, J. Durand, B. Laks & C. Lyche (éd.) *Les variétés du français parlé dans l'espace francophone : ressources pour l'enseignement*. Paris/Gap : Ophrys, 143-165.

Mahmoudian M. & Jolivet R. (1984). « L'accent vaudois », *in Encyclopédie illustrée du Pays de Vaud*, Éditions 24 Heures, 294-307.

Matthey M. (2003). « Le français langue de contact en Suisse romande », *Glottopol* 2 : 92-100.

Meisner C. (2013). *La variation pluridimensionnelle : une analyse de la négation en français*. Thèse de doctorat, Universität Zürich.

Meisner C., Robert-Tissot A. & Stark E. (à par.). « L'absence et la présence du *ne* de négation », *in* M.-J. Béguelin, A. Berrendonner, C. Blanche-Benveniste, J. Deulofeu et D. Willems (éd.) *Encyclopédie grammaticale du français*.

Métral J.-P. (1977). « Le vocalisme du français en Suisse romande. Considérations phonologiques », *Cahiers Ferdinand de Saussure* 31 : 145-176.

Moreau M.-L. (1986). « Les séquences préformées entre combinaisons et idiomatismes. Le cas de la négation avec ou sans *ne* », *Le français moderne* 54 : 137-160.

Morin Y. (2000). « Le français de référence et les normes de prononciation », *Cahiers de l'Institut linguistique de Louvain* 26 : 91-135.

Noailly M. (1990). *L'adjectif épithète*. Paris : PUF.

Pierrehumbert W. (1926). *Dictionnaire historique du parler neuchâtelois et suisse romand*. Neuchâtel : Attinger.

Prikhodkine A. (2011). *Dynamique normative du français en usage en Suisse romande*. Paris : L'Harmattan.

Pustka E. & Vordermayer M. (2006). « Le français parlé en Haute-Savoie et les corrélats prosodiques d'un accent perçu comme "traînant" », *Bulletin PFC* 7 : 273-281.

Racine I. & Andreassen H. (2012). « A phonological study of a Swiss French variety : data from the canton of Neuchâtel », *in* R. Gess, C. Lyche & T. Meisenburg (éd.) *Phonological Variation in French : Illustrations from Three Continents*. Amsterdam : John Benjamins, 173-207.

Racine I., Schwab S. & Detey S. (2013). « Accent(s) suisse(s) ou standard(s) suisse(s) ? Approche perceptive dans quatre régions de Suisse romande », *in* A. Falkert (éd.) *La perception des accents du français hors de France*. Mons : CIPA, 41-59.

Redard, G. (1971). « Sur le français de Suisse romande », *Revue Neuchâteloise* 54, 2-6.

Rittaud-Hutinet C. (1978). « Le français régional de Besançon », *Revue de linguistique romane* 42 : 123-148.

Rosier L. (2002). « *Genre* : le nuancier de sa grammaticalisation », *Travaux de linguistique* 44 : 79-88.

Schläpfer R. (éd.) (1985). *La Suisse aux quatre langues*. Éditions Zoé.

Schoch M. (1980). « Résultats d'une enquête phonologique en Suisse romande », *Bulletin de la Section de linguistique de la Faculté des lettres de Lausanne* 2 : 1-38.

Schüle E. (1971). « Documents de français régional actuel », *Revue neuchâteloise* 54 : 11-23.

Singy P. (1996). *L'image du français en Suisse romande. Une enquête sociolinguistique en Pays de Vaud*. Paris : L'Harmattan.

Singy P. (2004). *Identités de genre, identités de classe et insécurité linguistique*. Berne : Peter Lang.

Terrier Ph. (1997). « 100 ans d'enseignement du français langue étrangère à l'Université de Neuchâtel (1892-1992) », *in* G. Kahn (éd.) *L'Apport des centres de français langue étrangère à la didactique des langues*. Paris : SIHFLES, 127-140.

Thibault A. (1997). *Dictionnaire suisse romand*. Carouge : Éditions Zoé.

Thibault A. (à par.). « Le français en Suisse romande », *in* U. Reutner (éd.) *Manuel des francophonies*. Mouton de Gruyter.

Torreira F., Adda-Decker M. & Ernestus M. (2010). « The Nijmegen Corpus of Casual French », *Speech Communication* 52 : 201-212.

Tuaillon G. (1983). « Régionalismes grammaticaux », *Recherches sur le français parlé* 5 : 227-240.

Voillat F. (1971). « Aspects du français régional actuel ». *Actes du colloque de dialectologie francoprovençale organisé par le Glossaire des patois de la Suisse romande*, Genève : Droz, 216-246.

Walter H. (1981). « Le surcomposé dans les usages actuels du français », *Actants, voix et aspects verbaux*. Université d'Angers, 24-44.

Walter H. (1986). « Un sondage lexical en marge de l'enquête phonologique sur les français régionaux ». *Actes du XVII^e Congrès de linguistique et de philologie romanes*, Université d'Aix-en-Provence, 261-268.

Warnant L. (1997). « Phonétique et phonologie [du français en Belgique] », *in* A. D. Blampain, A. Goosse, J.-M. Klinkenberg & M. Wilmet (éd.) *Le français en Belgique. Une langue, une communauté*. Louvain-la-Neuve : Duculot, 163-174.

Zribi-Hertz A. (2015). « De l'évolution des propriétés du mot *grave* en français européen moderne », *in* G. Dostie & P. Hadermann (éd.) *La dia-variation du français*. Berne : Peter Lang, 63-98.

NOTES

1. Selon les dernières estimations de l'Office fédéral de la statistique, la Suisse a passé le cap des 8 millions d'habitants à la fin de l'année 2013.

2. Dans cet article, nous nous servons de l'abréviation « FS » pour renvoyer à l'ensemble des productions de français écrites et parlées dans l'aire géopolitique que constitue la Suisse romande, et non pour désigner une (illusoire) variété dont les frontières coïncideraient avec les frontières politiques de la Suisse romande. Comme l'ont souligné de très nombreux auteurs avant nous (Knecht 1979, 1985, 2000 ; Terrier, 1997 ; Andreassen, Maître & Racine, 2010), il n'y a en effet que très peu de traits linguistiques qui soient propres au FS (cela ne concerne en fait que quelques statalismes, comme *bancomat* « distributeur d'argent » ou *natel* « téléphone portable », voir sur ce point Thibault, 1997 et à paraître). Sur le plan lexical et syntaxique, certaines particularités que l'on considère traditionnellement comme des romandismes se retrouvent en France voisine (*bobet* « niais, idiot », Thibault, 1997 ; *j'ai personne vu*, Tuaillon, 1983), dans le domaine francoprovençal (*panosse*, « serpillière », Walter 1986), dans la partie sud de la France (l'usage du passé surcomposé dans les principales, Walter, 1981 ; les adjectifs déverbaux dits « tronqués » (*trempe* pour « trempé », Tuaillon, 1983), voire en Belgique (*septante* pour « soixante-dix », *nonante* pour « quatre-vingt-dix », *boiler* pour « chauffe-eau », Francard, 1997). Sur le plan phonologique, l'accentuation des syllabes pénultièmes de groupe et la lenteur du débit, que l'on décrit comme typiques de l'accent romand (Avanzi, Schwab, Dubosson & Goldman, 2012), sont des traits qui caractérisent également le français parlé dans le Jura français (Rittaud-Hutinet, 1978 ; Carton, Rossi, Autesserre & Léon, 1983 : 42), en Haute-Savoie (Pustka & Vordermayer, 2006) ou en Belgique (Warnant, 1997 ; Hambye & Simon, 2009 ; Bardiaux & Boula de Mareüil, 2012).

3. L'expression « français de référence » se substitue de plus en plus dans les travaux sur la variation régionale à la notion de « français standard ». Pour une problématisation de ces notions, nous renvoyons le lecteur à Morin, 2000 ; Detey & Le Gac, 2008, 2010 ; Lyche, 2010.

4. Jusqu'il y a peu, seule la base de données Phonologie du français contemporain (PFC, *cf.* Durand, Laks & Lyche, 2002, 2009, ainsi que Racine, Andreassen & Durand, ce volume) hébergeait également des enregistrements de locuteurs romands (originaires de Genève, de Neuchâtel et de Nyon).

5. Nous laissons volontairement de côté dans cet article la variation phonologique, dont l'étude nécessite des annotations et des outils d'investigation supplémentaires qui sont en cours de développement.

6. Pour ce dernier point, nous nous sommes basés sur la catégorisation socio-éducative opérée par Racine & Andreassen (2012) pour les locuteurs et locutrices du point d'enquête neuchâtelois du corpus PFC. Nous différencions ainsi les personnes ayant atteint le niveau de l'école

obligatoire avec, à l'issue, un apprentissage plutôt technique (maçons, ouvriers, coiffeurs, etc., Niveau I) de celles ayant atteint le niveau de l'école obligatoire avec apprentissage plutôt administratif (employés de banque, employés administratifs, etc., Niveau II), celles avec une maturité (équivalent français du baccalauréat, Niveau III), enfin celles ayant suivi des études universitaires (Niveau IV). Nous sommes bien conscients que ces catégories sociolinguistiques ne sont pas suffisantes (Cappeau & Gadet, 2007 ; Gadet & Guérin, ce volume), mais elles permettent d'opérer au moins un premier tri. Soulignons enfin que ces renseignements ne sont pas tous disponibles pour les enregistrements réalisés avant 2009.

7. Lors de la dernière mise à jour (avril 2016) la base comprenait 809 828 tokens transcrits (soit 83 heures de parole), pour un total de 269 locuteurs.

8. Dans les transcriptions, les barres verticales « | » indiquent une frontière d'intervalle dans l'alignement Praat, le symbole « _ » une pause silencieuse, le symbole « # » un segment anonymisé et le symbole « % » un segment inaudible (*cf.* Avanzi, Béguelin & Diémoz, 2012).

9. Corpus suisse de SMS, version 2013.04.04. Distribué par l'Université de Zurich, au nom de sms4science. URL : http://www.sms4science. uzh.ch.

10. La recherche brute du lemme *genre* fournit, via le concordancier, 198 entrées. Pour calculer le nombre d'occurrences de *genre*, nous avons été conduits à en retrancher 7 : 4 contenant une erreur probable de transcription, plus 3 formes du N pluriel *genres*, non pertinentes en l'occurrence. Nous avons dû en revanche rajouter 3 exemples qui, apparaissant en cooccurrence avec un autre, n'avaient pas été décomptés par le concordancier.

11. Dans les productions des locuteurs du FS, les variantes régionales et standards coexistent : il ne faudrait donc surtout pas croire que les Suisses romands parlent et écrivent un français régional homogène et standardisé. Knecht a fait naguère à ce sujet, dans la préface du *Dictionnaire suisse romand* (Thibaut, 1997), une mise en garde très utile.

12. Nous avons pris soin, lors de la recherche de *ne*, d'exclure les cas où le morphème était précédé du pronom *on*, car lorsque ce morphème est suivie d'un mot à initiale vocalique, on ne peut pas savoir si le *ne* a été prononcé ou pas (*on (n')en sait rien*). Nous avons exclu également, lors de la recherche de *n'*, les contextes où le morphème était suivi de *importe Q*, *empêche Q* et *est-ce pas*. Pour la recherche des forclusifs *pas/personne/rien/jamais/aucun/plus*, nous avons exclu du contexte antérieur (5 mots avant) les occurrences des morphèmes *ne* et *n'*.

13. Le test a été conduit sur un total de 129 locuteurs, avec pour chacun un nombre minimum de 2 observations et un nombre maximum de 110 observations (soit 33,6 observations en moyenne par participant). Le fait de mettre le locuteur comme un facteur aléatoire permet de tenir compte du fait que le nombre d'observations n'est équilibré ni d'un groupe à l'autre, ni d'un locuteur à l'autre. Compte tenu du faible nombre d'occurrences contenant *ne*, il n'était pas possible de faire un seul modèle et de tester les interactions entre les variables.

14. Il reste encore à voir si la différence d'âge entre l'enquêteur et l'enquêté ne joue pas aussi un rôle. En d'autres termes, que font les personnes plus âgées, quand elles sont entre elles et dans un contexte de communication familier, entre pairs ?

RÉSUMÉS

La base de données orales de français de Suisse romande (OFROM) contient des transcriptions d'enregistrements auxquels ont participé des locuteurs nés et vivant en Suisse. Elle a été créée par des linguistes soucieux de documenter les usages oraux du français en Suisse romande, et de

combler ainsi le retard dans la description des usages oraux de la langue parlée dans cette partie de la francophonie, notamment sur le plan du lexique, de la syntaxe et de la phonologie. La base de données OFROM a été mise en ligne en décembre 2012 ; elle est encore jeune et d'ampleur modeste (65 heures de parole pour près de 410 000 mots transcrits, au moment où nous écrivons), mais il est prévu de l'enrichir et de la diversifier régulièrement au cours des années à venir. La première partie de cet article présente les principes fondamentaux qui ont guidé la constitution de la base (choix des locuteurs, des supports et conventions de transcription, annotations, etc.). La seconde partie contient les résultats obtenus à l'issue d'études en cours ou de sondages prospectifs, portant sur la variation lexicale et la variation syntaxique. Ils permettront d'illustrer quelques-unes des potentialités qu'offrent la base et son moteur de recherche.

The Swiss French oral database of Switzerland (OFROM) hosts recordings and transcriptions of productions of speakers born and living in Switzerland. It was created by linguists who wanted to document the oral uses of French in Switzerland, and thereby bridge the gap in the description of the description of the spoken aspects of the lexicon, syntax and phonology. The database OFROM was opened in December 2012; it is still young and modest (65 hours of speech for nearly 410.000 transcribed words, at the time of writing), but it is planned to enrich and diversify regularly over the coming years. The first part of this article presents the fundamental principles that drove the basic constitution (choice of speakers, conventions of transcriptions, annotations, etc.). The second part presents the results obtained in ongoing or prospective surveys on the lexical specificities and syntactic variation. They illustrate some of the potential offered by the database and its search engine.

INDEX

Mots-clés : corpus, Suisse romande, français parlé, OFROM, lexique, variation syntaxique, régionalismes
Keywords : corpus, French speaking Switzerland, OFROM, lexicon, syntactic variation, regionalisms

AUTEURS

MATHIEU AVANZI

Universités de Genève et de Zurich

MARIE-JOSÉ BÉGUELIN

Université de Neuchâtel

FEDERICA DIÉMOZ

Université de Neuchâtel