

*Corpus de français parlé  
et français parlé des corpus*

Numéro coordonné et présenté par

**Mathieu AVANZI, Marie-José BÉGUELIN  
et Federica DIÉMOZ**



**Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales :  
Exemple de mise à profit des données à partir d'un  
examen lexico-sémantique de la séquence *je sais pas***

Gaétane DOSTIE

Département des lettres et communications  
Faculté des lettres et sciences humaines  
Université de Sherbrooke, Québec, Canada

Cet article se déroule en deux temps. Il présente d'abord le contexte général ayant conduit à l'élaboration du Corpus de français parlé au Québec (CFPQ) et les principes méthodologiques ayant présidé à sa confection (section 1). Il illustre ensuite l'intérêt que représente cette ressource documentaire pour l'étude de la langue parlée en contexte informel par le biais d'un examen lexico-sémantique d'une séquence qui y est fréquente, à savoir *je sais pas* (section 2).

## **1. Présentation du corpus**

### ***1.1 Les corpus lexicaux québécois***

En 1997, le Secrétariat à la politique linguistique du Québec lançait une vaste entreprise afin de donner accès, à partir d'un portail commun, aux principaux corpus de langue qui avaient été constitués jusque-là dans les universités québécoises. Le projet, du nom de « Corpus lexicaux québécois », prenait fin en 2007. Le site Internet créé reliait désormais entre eux 15 corpus reflétant la langue, d'époques différentes, en usage dans des genres communicatifs extrêmement variés (p. ex. : textes littéraires parus entre 1837 et 1919, entrevues dirigées et semi-dirigées réalisées entre 1960 et 1990 dans le cadre d'enquêtes sociolinguistiques...).

Huit ans plus tard, ce site est toujours accessible et les objectifs poursuivis par ses promoteurs y sont énoncés dans les termes suivants :

- Mettre en œuvre des actions concertées visant l'emploi et la qualité du français en usage au Québec.
- Instrumenter les chercheurs en vue de permettre une description scientifique du français en usage au Québec.
- Favoriser l'élaboration d'ouvrages de référence qui peuvent tenir compte de la réalité linguistique québécoise (faune et flore, géographie, institutions et organisations sociopolitiques, environnement, etc.).
- Créer, dans Internet, un réseau de banques de données textuelles représentatives du français en usage au Québec à l'intention des chercheurs québécois et autres internautes de la Francophonie.
- Accroître la participation et la visibilité du Québec au sein de la francophonie internationale.

Site : Secrétariat à la politique linguistique,  
Gouvernement du Québec<sup>1</sup>.

La consultation simultanée des corpus ici regroupés se veut simple et efficace. « On peut taper [peut-on lire], par exemple, les mots *bozo*, *bouette*, *érablière* ou *ouananiche*, et découvrir pour chacun ce qu'en révèlent les différentes sources de ces riches corpus » (site : Secrétariat à la politique linguistique, Gouvernement du Québec).

Conformément à ses objectifs, la base de données « Corpus lexicaux québécois » constitue une référence précieuse pour décrire une partie non négligeable du lexique propre à la culture québécoise – notamment, les unités à valeur référentielle comme celles données en exemple ci-dessus. Cependant, elle est moins propice à soutenir les analyses axées sur la langue contemporaine usitée dans les conversations familiales. Cette forme de langue, on le sait, foisonne de marqueurs à valeur grammaticale et pragmatique, soit absents des ouvrages de référence, soit décrits de manière plutôt sommaire. À ce

---

<sup>1</sup> <http://www.spl.gouv.qc.ca/languefrancaise/corpuslexicaux>

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales*

chapitre, un exemple récent dans les annales lexicographiques québécoises est celui du dictionnaire Usito (2012-) où, à titre illustratif, les expressions et mots suivants, typiques de la langue de tous les jours, n'ont pas droit de cité : *coudon, pis, mets-en, fait que, ben / ben ben...*

En un sens, on ne peut reprocher à Usito d'ignorer tout un pan du lexique courant, dans la mesure où il vise à « décrire le *français standard* en usage au Québec », indique-t-on sur son site internet<sup>2</sup>. Voilà donc qui explique en partie maintes exclusions<sup>3</sup>, auxquelles il apparaît tout aussi légitime d'accorder priorité.

**1.2 Le CFPQ et la langue des conversations familiales**

C'est dans le contexte exposé ci-haut que le projet relatif à la création du Corpus de français parlé au Québec (CFPQ) a vu le jour. Son but, clairement affiché, était de venir appuyer les études sémantiques portant sur le lexique caractéristique de la langue orale familière, notamment les études axées sur les marqueurs à valeur grammaticale et pragmatique. Les travaux entourant la mise en place du corpus ont démarré en 2006 dans le cadre des activités du Centre d'analyse et de traitement informatique du français québécois (CATIFQ) de l'Université de Sherbrooke<sup>4</sup>.

Le CFPQ regroupe aujourd'hui 30 sous-corpus de conversations à bâtons rompus enregistrées sur support audiovisuel ; chacun d'entre eux dure approximativement 1 heure et demie. Au total, 45 heures d'enregistrement ont ainsi été effectuées entre 2006 et 2013, dans diverses régions du Québec. Les différents sous-corpus mettent en scène 3 ou 4 locuteurs qui se connaissent très bien. En tout, 109 locuteurs, dont l'âge va de 15 à 95 ans, ont pris part au projet ; ils discutent librement, dans

---

<sup>2</sup> C'est nous qui plaçons en italiques le terme *français standard*.

<sup>3</sup> Il en va de même pour le traitement plus que minimal de certains mots courants, notamment ceux d'origine anglaise (entre autres, Poirier, 2015).

<sup>4</sup> Le projet, dirigé par la signataire du présent article, a reçu l'appui financier de l'Université de Sherbrooke, du Fonds de recherche québécois sur la société et la culture (FRQSC, 2007-2014 ; subvention d'équipe accordée au CATIFQ) et du Conseil de recherche en sciences humaines du Canada (CRSH, subvention individuelle, 2008-2012).

un lieu familier, de sujets divers (p. ex. : le travail, la maladie, les loisirs, les autres...).

Une trentaine d'étudiants assistants ont collaboré aux travaux du CFPQ (enregistrements, transcription, révision, support technique...), et un analyste-informaticien à l'emploi de l'Université de Sherbrooke a assuré la mise en place de la base de données (accessible en ligne) à laquelle nous revenons un peu plus bas.

Les transcriptions sont effectuées à l'aide du logiciel Transana, qui permet un alignement du son, de l'image et du texte. Les conventions de transcription, présentées sur le site du corpus, sont relativement standards pour la langue orale. Celles-ci visent à refléter, au mieux, le caractère multimodal des interactions verbales en face-à-face. Ainsi, en plus de noter le matériel verbal (c'est-à-dire les mots selon l'orthographe habituelle), les transcrip-teurs prennent également en compte le matériel paraverbal (p. ex. : les pauses, la vitesse et le volume de la voix) et le matériel non verbal (les gestes significatifs sur le plan communicationnel, comme hocher la tête négativement ou faire un clin d'œil en signe de complicité).

La base de données, créée en 2011, permet actuellement d'effectuer des recherches dans 21 sous-corpus, ce qui correspond à 31 heures et demie de conversations à bâtons rompus. Dans leurs versions transcrites, ces sous-corpus équivalent à 471 575 unités graphiques au sein desquelles figurent 21 016 unités graphiques différentes. Les transcriptions des 9 derniers sous-corpus sont déjà bien amorcées et devraient être disponibles prochainement<sup>5</sup>. La taille finale du corpus sera alors d'environ 675 000 unités graphiques.

Le site Internet du CFPQ présente les métadonnées relatives aux enregistrements (dates, lieux, principaux thèmes abordés...), aux participants (âges, professions/occupations, scolarité...) et aux transcriptions (dates où elles ont été effectuées, nom des transcrip-teurs...). À l'exception des informations portant sur les transcriptions, les métadonnées peuvent faire l'objet d'une recherche croisée avec des données dans les fichiers-

---

<sup>5</sup> Il est prévu que les transcriptions restantes soient accessibles en ligne au plus tard à l'automne 2015.

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familières*

textes (tous téléchargeables en format PDF). En guise d'exemple, il est possible d'effectuer des recherches combinant l'âge et la scolarité, d'un côté, et tel ou tel marqueur, de l'autre.

Par ailleurs, le site présente, par ordre décroissant, les unités graphiques les plus fréquentes sur une base individuelle, de même que les séquences les plus fréquentes où 2, 3, voire 4 unités graphiques apparaissent en cumul. Ainsi, on découvre que *c'est ça, parce que* et *fait que* se logent respectivement au premier, deuxième et troisième rangs pour ce qui concerne les séquences les plus fréquentes du corpus composées de 2 unités graphiques. De même, on repère, de manière automatique sur le site, les unités graphiques les plus souvent répétées en contiguïté (de 2 à 5 fois). Sur ce point, il n'y a pas de surprise : *non non non non non* est la séquence la plus fréquente dans laquelle une même unité figure 5 fois de suite. Elle est suivie de près par *oui oui oui oui oui*.

Ces deux derniers exemples illustrent bien le fait que le CFPQ est un corpus de langue orale typique des interactions verbales spontanées. De ce fait, il constitue un bon repère pour identifier les marqueurs et les séquences de mots à étudier en priorité, si l'objectif est d'accorder une attention spéciale à la langue des conversations familières, en vue notamment de la voir un jour mieux représentée dans les dictionnaires... Cela dit, dans la mesure où ce dernier point relève pour l'immédiat davantage de l'utopie que de la réalité, nous le laisserons en suspens pour nous attaquer maintenant à notre étude de cas à partir des données prélevées dans le corpus.

## **2. Je sais pas : de séquence verbale à marqueur discursif**

Nous présentons dans ce qui suit quelques données quantifiées qui font ressortir l'intérêt d'examiner la séquence morphologiquement complexe *je sais pas*, en français québécois spontané (section 2.1). Par la suite, nous portons notre attention sur ses emplois verbaux (section 2.2) afin d'établir comment celle-ci a pu en venir à acquérir une valeur qui l'apparente, dans certains contextes, à un marqueur discursif (MD), tel que souligné notamment dans Gauvin 1999, de Sève 2005 et Pop 2009 (voir aussi sur *I don't know* en anglais, qui est proche de *je sais pas*,

Diani 2004, Grant, 2010 et Aijmer, 2014). Nous focalisons notre attention, en terminant, sur la valeur discursive associée à l'expression considérée et insistons sur le double rôle qu'elle assume alors : ce rôle est de l'ordre à la fois de la collaboration et de la protection de soi (section 2.3).

L'étude d'une séquence telle *je sais pas* ramène vite à la problématique classique relative à la dissymétrie entre la première personne du présent de l'indicatif et les autres personnes de la conjugaison propre à certains verbes appelés diversement, entre autres, *verbes assertifs*, *verbes parenthétiques* et *verbes d'attitude propositionnelle*, comme *je crois que P*, *je trouve que P* et *je sais que P* (notamment, Urmson 1952 ; Récanati 1984 ; Schneider 2007a, 2007b et 2013). Nous y revenons plus loin.

### **2.1 Quelques données quantifiées**

Notre intérêt pour *je sais pas* (prononcé [ʃepa]) vient d'abord d'un constat : sa fréquence d'emploi est particulièrement élevée dans le CFPQ, eu égard à sa morphologie complexe. Ainsi, il s'agit de la séquence la plus fréquente du corpus composée de 3 unités graphiques. On en trouve 705 occurrences, ce qui la place devant les 4 séquences complexes suivantes, elles aussi formées de 3 unités graphiques se succédant dans le texte : *en tout cas* (576 occurrences), *oui c'est ça* (352 occurrences), *fait que là* (332 occurrences) et *tout le temps* (313 occurrences). De plus, le CFPQ contient 97 occurrences d'une forme équivalente à *je sais pas*, transcrite *je le sais pas* (prononcée [ʒəlʃepa]), si bien que le nombre total d'occurrences relatif à la séquence en question est en réalité encore plus important (on en dénombre 802 en tout).

L'importance quantitative de *je sais pas* dans le CFPQ et donc, vraisemblablement, dans les conversations à bâtons rompus, se vérifie encore de deux façons :

- le plus souvent, *(le) sais pas* est précédé, dans le corpus, de *je* et non pas de *tu*. Il y a en effet 1 050 occurrences de *(le) sais pas*, ce qui signifie que dans 76 % des cas c'est *je* qui introduit *(le) sais pas* et non pas *tu*. Les séquences *il (le) sait pas* et *on (le) sait pas* n'apparaissent, quant à elles, que 7 fois et 38 fois, respectivement ;



*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales*

- de même, la probabilité est grande pour que *pas* suive immédiatement *je (le) sais*. À ce propos, le corpus renferme 900 occurrences de *je (le) sais*, ce qui revient à dire que *pas* figure à sa droite dans 89 % des cas<sup>6</sup>.

En somme, ce n'est pas seulement la fréquence élevée de *je (le) sais pas* qui frappe dans l'ensemble des séquences complexes utilisées dans notre corpus ; c'est aussi – et peut-être même davantage – la forte attirance exercée entre *je*, *(le)*, *sais* et *pas*. De là, on peut supposer que le degré d'« entrenchment » (c'est-à-dire d'enracinement) relatif à la suite considérée est grand, qu'elle est mémorisée en tant que séquence morphologiquement complexe ayant atteint, dans certains contextes, un degré élevé de figement (entre autres, Langacker 1987 : 59-60). Parmi ces contextes se trouvent ceux où elle agit à titre de MD<sup>7</sup>.

## **2.2 Je sais pas : séquence verbale**

*Savoir* a un potentiel pour la factivité : le locuteur en use, dans certains contextes, en présupposant la véracité de la proposition complément (Kiparsky & Kiparsky, 1970). Nous disons que ce verbe a *un potentiel* pour la factivité, parce qu'il n'est pas systématiquement employé de manière factive, ce qui en fait un verbe semi-factif (Levinson, 1983 ; Korzen, 2001). Il sera factif notamment dans les contextes positifs comme celui relevé en (1) où la proposition P renvoie à l'idée, présentée comme vraie par J.-M., selon laquelle la prière effectuée par une tierce personne (en l'occurrence, une étudiante prénommée *Hasmina*) diffère quelque peu de celle faite par les catholiques.

- (1) M. : non mais [ <all<t'sais>> elle se décrit comme une bonne musulmane là je veux dire elle respecte ses paRENTS elle respecte t'sais <dim<la plupart des des>> (.) pis elle fait sa

<sup>6</sup> Les pourcentages sont établis sur la base des calculs suivants :

- dans le premier cas, 802 occurrences de *je (le) sais pas* sur 1 050 occurrences de *(le) sais pas* équivaut à 76 % ;
- dans le second cas, 802 occurrences de *je (le) sais pas* sur 900 occurrences de *je (le) sais* correspond à 89 %.

<sup>7</sup> Dans ce qui suit, nous faisons l'économie du *le* entre parenthèses pour des raisons de simplicité, étant entendu que les deux prononciations relevées plus haut et transcrites respectivement *je sais pas* et *je le sais pas* sont possibles.

prière sauf que elle elle nous l'a bien dit que genre la prière  
 pi :s peu importe les ablutions tout ce qui va avec (.) euh (.)  
 t'sais pour elle c'est comme (.) dans le Coran son inter-  
 prétation à elle en tout cas c- elle ce qu'elle nous disait c'est  
 que

[...]

J.-M. : mais (.) la prière du vendredi **je sais que** c'est une  
 prière :re qui est comme un peu différente des autres là  
 (CFPQ, sous-corpus 10, segment 1, p. 3, ligne 5)

*Savoir* perd sa factivité dans certains contextes négatifs<sup>8</sup>, dont celui présenté en (2) : nous y trouvons alors la séquence qui nous intéresse particulièrement, c'est-à-dire *je sais pas*. Dans cet exemple, *savoir* n'est pas factif puisque J affirme précisément ne pas savoir à quelle heure elle s'est levée, en réponse à la question de K. Notons qu'il refuse la construction complétive *\*je ne sais pas que P* pour des raisons sémantiques évidentes : il serait pour le moins curieux d'affirmer qu'on ne sait pas si ce qu'on présente comme vrai est vrai<sup>9</sup>. La construction en *si P* serait en revanche acceptable puisqu'elle donnerait une valeur hypothétique à la subordonnée conformément au sémantisme du verbe lorsqu'il est nié à la première personne du présent de l'indicatif, ce que montre (3).

- (2) K : [tu t'es levée à quelle heure toi à matin/  
 J : [(inaud.)  
 J : à quelle heure/  
 K : hum hum  
 J : **je sais pas**  
 C : (RIRE) tard  
 (CFPQ, sous-corpus 17, segment 1, p. 11, ligne 7 ; 8 min, 4 s)

<sup>8</sup> Nous écrivons que *savoir* perd sa factivité dans certains contextes négatifs, car celle-ci peut réapparaître à des temps autres que l'indicatif présent, comme à l'imparfait (p. ex. : *je savais pas que P*) ou au plus-que-parfait (p. ex. : *j'avais pas su que P*).

<sup>9</sup> L'interrogative indirecte du type *je ne sais pas à quelle heure (que) je me suis levé* est toutefois acceptable, parce que l'affirmation d'un non-savoir porte cette fois sur le syntagme nominal qui suit *savoir* (cf. à *quelle heure*) et non sur la complétive. Dans cet exemple, la complétive introduit une idée présentée comme vraie (cf. *je me suis levé*) car elle n'est pas directement sous la portée de *savoir*.

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familières*

- (3) VE : **je sais pas si** ça dérange si on est cinq / (RIRE)  
(CFPQ, sous-corpus 19, segment 6, p. 54, ligne 26 ; 3 min 18 s)

À côté de l'emploi où la séquence *je sais pas* sert au locuteur à indiquer son ignorance vis-à-vis de la proposition P, comme en (2), on en trouve un second, où elle lui permet cette fois d'indiquer un doute ou une incertitude face à la véracité de cette proposition, comme en (4). Cet emploi n'est donc pas sans rappeler celui illustré en (3), sauf qu'ici, il y a économie de la proposition *si P*.

- (4) J : pis le Planétarium  
K : ouin non tu feras je pense pas que tu vas pouvoir y aller  
(*en souriant et en hochant légèrement la tête négativement*)  
C : <p<**je sais pas**>> (*en souriant légèrement*)  
J : <pp<euh je pense que tu peux>> (*en hochant la tête affirmativement*) mais si tu te fais refuser (*en haussant les épaules comme pour signifier « tant pis »*) euh ben tu demanderas à ton prof (*en inclinant la tête vers la droite comme pour désigner la personne dont elle parle*)  
(CFPQ, sous-corpus 17, segment 4, p. 43, ligne 15 ; 2 min 40 s)

Ainsi, en (4) comme en (2), *je sais pas* renvoie à l'expression d'un savoir déficient, ce que souligne Diani 2004 pour *I don't know*. Par exemple, en réaction à l'affirmation de l'interlocutrice K, selon laquelle il ne sera pas permis à C d'aller au Planétarium, cette dernière utilise *je sais pas* en (4) pour signaler une incertitude vis-à-vis de ce qui est dit (et non une négation catégorique). *Je sais pas* est alors proche de formules comportant un modalisateur épistémique du type *je suis pas certain* ou *je suis pas sûr*. Dans l'exemple considéré, tout porte à croire que l'incertitude exprimée par C quant à la véracité de P tient au fait qu'elle cherche à ne pas contrer K ouvertement. De ce point de vue, l'usage de la séquence examinée relève d'une stratégie de politesse qui consiste à ne pas afficher de but en blanc son désaccord avec l'interlocuteur afin de ne pas heurter sa face positive (Brown & Levinson, 1987 ; Diani, 2004 ; Aijmer, 2014).

De même, en (5), le locuteur S utilise *je sais pas* pour exprimer une incertitude face à l'affirmation de R et non pour

s'y opposer. L'intonation montante, bien notée dans cet exemple par le transcripateur, paraît assez typique de l'emploi discuté. Celle-ci n'induit pas une valeur de question, en ce sens que le locuteur ne se demande pas « s'il sait que P ». Au contraire, l'emploi sous examen paraît davantage correspondre à ce que Kerbrat-Orecchioni (1991) appelle une « quassertion » ou une « semi-question », c'est-à-dire à un acte illocutoire intermédiaire entre l'assertion pure et simple et la question en bonne et due forme. *Je sais pas* joue ici un double rôle : en premier lieu, il permet au locuteur d'exprimer une incertitude face à ce qui est dit et, en second lieu, il constitue un moyen, peu contraignant pour l'interlocuteur, de tenter de vérifier, sans en avoir l'air, l'exactitude de ses propos, un peu comme un *ah bon*↑, un *ah oui*↑ ou un *vraiment*↑. Nous disons qu'il s'agit d'un moyen *peu contraignant* d'agir de la sorte, parce que l'interlocuteur peut décider de ne pas réagir à une semi-question, alors qu'il lui serait plus difficile, sur le plan interactionnel, d'ignorer une véritable question sans risque de paraître un tantinet malpoli. Ainsi, R laisse en suspens l'assertion-question de S en (5) sans qu'aucun malaise entre les interactants semble s'installer.

- (5) [En parlant d'un enfant turbulent dont on est en train de raconter les méfaits, le locuteur dit :]  
 R : il est (.) il est dur pour son corps pis il est dur sur le corps des autres  
 J : oui (*dit en riant*)  
 (RIRE GÉNÉRAL)  
 B : sur le corps pis sur le moral  
 S : mais que tu le connaisse tu vas comprendre tout qu'est-ce qu'on te dit/ (.) seigneur  
 [...]  
 R : les oreilles [doivent lui siller lui  
 J : [mais il comprendra [pas (.)  
 S : [je sais pas↑ (*en réaction à ce que dit Robert*)  
 J : il comprendra pas (.) pas parce qu'il est pas fin ↓ mais parce que tout le monde qui le rencontre il trouve qu'il est juste charmant  
 (CFPQ, sous-corpus 15, segment 8, p. 138, ligne 12 ; 4 min 55 s)

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales*

Dans les exemples précédents, la séquence *je sais pas* a incontestablement un statut verbal. À ce titre, elle n'est pas figée à la première personne du présent de l'indicatif, si bien qu'on rencontre des emplois tels *je savais pas* et *on sait pas*, comme en (6) et (7). En (6), *je savais pas* nie un savoir, en écho à l'exemple (2) ; en (7), *on sait pas* indique une possibilité, parallèlement aux exemples (4) et (5).

- (6) É : ouin ouin il y a juste le : le coussin là (*en désignant le coussin sur lequel elle est assise*) [le rembourrage là que j'ai fait faire  
I : [ah ouin/ ah <p<<ouin ça je je>> [ah/ ouais ah/ **je savais pas**\  
É : [mai :s <f<il a du TALENT le frerot↑ hein↑>>  
(CFPQ, sous-corpus 16, segment 1, p. 2, ligne 19 ; 1 min)
- (7) A : oui apparemment qui q- apparemment qu'ils avaient été averTIS (*en pointant son crayon dans les airs comme pour insister sur ses propos*) (.) que c'était (.)  
R : ben c'est possible qu'ils aient pu faire quelques éléments de risques additionnels  
AN : ah ça on peut **on sait pas** hein /  
(CFPQ, sous-corpus 20, segment 7, p. 78, ligne 15 ; 1 min 32 s)

### **2.3 Je sais pas : marqueur discursif**

Comme cela était à prévoir, le MD *je sais pas* ne joue pas un rôle au sein de l'énoncé, si bien que sa suppression ne rend pas agrammaticale la répartie du locuteur qui en use<sup>10</sup>. Ainsi, le segment textuel *pis c- me semble ça nous CALME ça fait du bien* serait syntaxiquement recevable en (8).

- (8) M : on s'éVADE là-dedans (.) c'est pas croyable comme on s'évade dans ça (.) des fois là euh : ben là ça fait euh : quasiment six mois que j'ai pas touché à mon piano là mais

---

<sup>10</sup> On le sait, le fait qu'une séquence puisse être omise n'est pas une caractéristique exclusive des MD (par exemple, les adverbes d'énonciation sont également optionnels). Le caractère optionnel d'une séquence n'est donc pas une condition suffisante pour qu'il y ait MD. Il faut plutôt voir une condition nécessaire (sauf, encore là, dans les cas connus de mots-phrases tels *ok* ou *coudon* lorsqu'ils constituent par eux-mêmes une intervention).

euh quand ça me le disait là je m'en allais au piano pis j'improvisais là pis t'sais euh t'sais euh •vas-y vas-y° là pis c-  
**je sais pas** me semble ça nous CALME ça fait du bien (.)  
ouin (*en hochant la tête affirmativement*)  
(CFPQ sous-corpus 11, segment 5, p. 60)

Dans l'emploi considéré, la séquence *je sais pas* correspond à ce que Schneider 2007a et 2007b appelle une « clause parenthétique réduite » à valeur modale. Elle présente, grosso modo, les particularités sémantiques les plus communes dégagées par Urmson 1952, dans l'extrait suivant, pour ce qu'il nomme « verbes parenthétiques ».

They [= *parenthetical verbs*] themselves have not, in such a use, any descriptive sense but rather function as signals guiding the hearer to a proper appreciation of the statement in its context, social, logical, or evidential. They are not part of the statement made, or additional statements, but function with regard to a statement made rather as 'READ WITH CARE' functions in relation to a subjoined notice, or as the foot stamping and saluting can function in the Army to make clear that one is making an official report. Perhaps they can be compared to such stage-directions as 'said in a mournful (confident) tone' with reference to the lines of the play. They help the understanding and assessment of what is said rather than being a part of what is said. (Urmson, 1952 : 495-496)

Dans ce cadre, *je sais pas* se loge dans la troisième des 4 classes (concernant les clauses parenthétiques réduites) identifiées par Schneider 2007a et 2007b. Celle-ci réunit des marqueurs mettant en jeu des verbes de croyance qui ont *grosso modo* comme fonction, selon l'auteur, « d'atténuer directement l'obligation communicative du locuteur ». Cela étant, il reste maintenant à préciser comment s'opère le passage du verbe *savoir* au *je sais pas* discursif à portée atténuante.

Dans l'un de ses emplois verbaux, nous l'avons vu, la séquence *je sais pas* permet au locuteur d'indiquer que, pour lui, la proposition P est indéterminée. Lorsque celle-ci fait office de MD, il y a déplacement de l'indétermination : elle se

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familières*

situé non plus vis-à-vis d'une proposition P et donc du dit, mais plutôt vis-à-vis du dire. De ce fait, le locuteur signale, en usant du marqueur, qu'il est engagé dans un processus cognitif relatif à la mise en mots de sa pensée. Il hésite et se demande à voix haute que dire et/ou dans quels termes. En guise d'exemple, la locutrice VE manifeste explicitement, par *je sais pas* en (9), qu'elle est en mode réflexif, qu'elle cherche comment poursuivre sa narration. Cela ressort d'autant mieux, ici, que la séquence sous examen est précédée d'une marque d'hésitation explicite (cf. *euh*) et d'une pause relativement longue (de près de 3 secondes). De plus, les premiers mots qui suivent immédiatement *je sais pas* ne forment pas un texte fluide. On y retrouve ce que Blanche-Benveniste (1997 : 21) appelle un « entassement paradigmatique » : *c'était si* est en concurrence, au plan paradigmatique, avec *c'est* qui est finalement l'option retenue pour l'enchaînement narratif (cf. *c'est important de faire des beaux partys de Noël*). Ce court extrait montre donc que *je sais pas* est une trace, parmi d'autres, qui traduit, dans le texte, l'effort cognitif que fournit le locuteur au moment même où il met en mots sa pensée.

- (9) VE : c'est cool des partys de Noël sérieux c'est comme euh (2,8") **je sais pas** je trouve que c'était si c'est important de faire des beaux partys de Noël  
(CFPQ, sous-corpus 19, segment 6 p. 52, ligne 2 ; 11 s)

En affichant ouvertement, grâce à *je sais pas*, qu'il est en mode réflexif et qu'il est hésitant, le locuteur pose ouvertement un regard critique sur les limites de son savoir : son dit a pour cette raison une allure plus ou moins assurée. Ce faisant, celui-ci collabore honnêtement à l'échange. Il respecte en tous points les maximes conversationnelles du type formulées par Grice 1979 et plus spécifiquement la sous-maxime de qualité que l'auteur énonce comme suit : « N'affirmez pas ce que vous croyez être faux ; n'affirmez pas ce pour quoi vous manquez de preuve » (p. 61). *Je sais pas* est donc un marqueur foncièrement interactionnel qui s'inscrit dans une démarche collaborative. Par exemple en (10), sa présence illustre le fait que H, sans y avoir réfléchi au préalable, cherche dans le vif de la discussion, une

solution au dilemme apparent d'un tiers absent<sup>11</sup> : écouter une émission de télévision ou aller à une fête à l'occasion d'un mariage. La solution avancée est qu'il n'y a pas forcément à choisir. Les deux activités sont possibles, si l'émission de télévision est enregistrée ou écoutée lors d'une reprise le lendemain.

- (10) H : enregistre-le **je sais pas** ou écoute-le demain là c'est un party qu'il y a à soir là •ah je connais pas ben ben Domingue là° (.) pis ça crime (.) c'est un mariage là (.) en plus c'est la c'était la fin  
(CFPQ, sous-corpus 14, segment 8, p. 86, ligne 6)

En somme, si *je sais pas* discursif n'est pas nécessaire au plan référentiel, il l'est au plan pragmatique. Voici un dernier exemple qui, à nouveau, le montre bien. En (11), R (un homme dans la soixantaine) rend explicite, par le biais du marqueur, l'effort cognitif qu'il déploie au moment de parole pour exprimer son point de vue sur un sujet délicat relativement aux groupes de jeunes gens dont le pouvoir collectif exerce trop souvent, selon lui, une pression excessive sur les choix individuels. Aussi, en plus d'être un marqueur collaboratif, *je sais pas* en est-il un foncièrement égotiste. En effet, grâce à lui, le locuteur se prémunit contre d'éventuelles représailles résultant d'une prise de parole qui, autrement, aurait pu être perçue, par l'interlocuteur, comme étant peu réfléchie.

- (11) R : à un certain moment donné en tout cas **je sais pas** je vois peut-être ça mal mais moi je me dis là à un certain moment donné une gang ensemble [...] une gang ensemble tu sais pas quoi faire  
[...]  
R : à un certain moment donné il y en a un qui sort quelque chose il sort quelque chose (*en claquant des doigts comme pour représenter la rapidité de l'événement dont il parle*)  
[...]  
R : et go pis tout le monde est ensemble pis aïe je peux pas dire non/ écoute je vas avoir l'air d'un maudit pissou pissou  
(CFPQ, sous-corpus 20, segment 3, p. 28, ligne 19 ; 4 min 32 s)

---

<sup>11</sup> Dans l'exemple considéré, H fait allusion à un problème rencontré par un ami qui ne participe pas à l'échange.



*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales*

Pour terminer, remarquons que le glissement de *savoir* à *je sais pas* discursif a sans doute été facilité par le fait que, dans son emploi verbal, celui-ci renvoie, si l'on en croit Rémi-Giraud 1986, à un savoir où « le sujet reste dans la réalité intérieure d'un processus mental » (p. 250). Or c'est également ce qui se passe, on l'a vu, avec *je sais pas* discursif. Dans les termes de Rémi-Giraud, *savoir*<sup>1</sup> signifie « avoir dans l'esprit l'acte mental qui permet de former et d'affirmer une phrase (exprimée dans le C.O.D.) » (p. 254). Par comparaison, son synonyme proche, *connaître*<sup>1</sup>, renvoie à un « savoir d'expérience lié à l'espace des réalités non linguistiques » ; il signifie « avoir dans l'esprit l'image qui correspond à un objet extérieur (exprimé dans le C.O.D.) » (p. 254). Dans cette veine, pourquoi, à côté de *je sais pas* discursif, n'a-t-on pas, par exemple, *je connais pas* discursif ? Une hypothèse plausible est que *connaître* n'est pas foncièrement tourné vers la pensée mais vers le monde, contrairement au verbe « assertif fort », *savoir*, tout entier centré sur « le cours ou l'aboutissement [d'un] processus cognitif » (Borillo 1982 : 35) ; cela est aussi le propre de *je sais pas* discursif<sup>12</sup>.

### 3. Conclusion

Cette étude avait comme premier objectif de présenter le contexte général ayant conduit à la confection d'un corpus de langue orale usitée au Québec, en contexte informel, dans les années 2000, à savoir le CFPQ. Elle visait également à exposer les grands principes méthodologiques ayant présidé à la constitution du corpus, ainsi que ses principales caractéristiques (allant de l'enregistrement des données sur support audiovisuel à la mise en ligne des transcriptions et à leur interrogation).

Le second objectif poursuivi était d'illustrer en quoi un corpus de langue familière comme le CFPQ pouvait s'avérer utile pour choisir et analyser finement nombre d'expressions fréquentes, caractéristiques de la variété diatopique/diaphasique

---

<sup>12</sup> Sur le plan syntaxique, on pourrait encore ajouter que les constructions *\*connaître que P / \*connaître si P* n'existent pas. Ce fait, signalé par les chercheuses précitées comme étant une conséquence de surface distinguant *savoir* et *connaître*, serait donc également en lien avec l'absence de *je ne connais pas* dans la zone discursive.

de langue ciblée. De telles expressions sont généralement absentes des dictionnaires, y compris des plus récents.

Pour atteindre ce deuxième objectif, l'intérêt s'est porté du côté de la séquence la plus fréquente du corpus composée de 3 unités graphiques, c'est-à-dire *je sais pas*. Cette séquence est tantôt verbe, tantôt MD. Lorsque *je sais pas* est verbe, soit il nie catégoriquement un savoir, soit il laisse en suspens sa véracité. Dans ce cas, la séquence considérée n'est pas figée à la première personne du présent de l'indicatif et elle participe au contenu propositionnel de l'énoncé. Le statut de *je sais pas* se modifie lorsque l'indétermination face à un contenu propositionnel se transpose au plan métadiscursif. La séquence, qui ne joue alors plus un rôle au plan propositionnel ou référentiel, devient syntaxiquement optionnelle. Elle traduit en pareil cas l'effort déployé par un locuteur engagé dans un processus cognitif complexe consistant à mettre en mots sa pensée. Celui-ci hésite : il cherche que dire et/ou dans quels termes s'exprimer. Dans ce cas-ci, *je sais pas* est employé comme MD.

Pour bien faire, il faudrait désormais confronter les résultats obtenus ici pour *je sais pas* à des données prélevées dans d'autres corpus oraux du français, tels le CFPP2000, ESLO et OFROM. La voie est tracée : nous aurons certainement l'occasion de nous y engager dans une prochaine étude.

### **Bibliographie**

- Aijmer K. (2014). « *I don't know* as a marker of youth language », in K. Helgesson *et al.* (éd.) *Fint språk/Good Language. Festskrift till Lars-Gunnar Andersson*. Göteborgs : Utgiven i serie vid Göteborgs universitet, 1-14.
- Blanche-Benveniste Cl. (1997). *Approches de la langue parlée en français*. Paris : Ophrys.
- Borillo A. (1982). « Deux aspects de la modalité assertive : croire et savoir », *Langages* 67 : 33-53.
- Brown P. et S. C. Levinson (1987). *Politeness. Some Universals in Language Usage*. Cambridge : Cambridge University Press.

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familières*

- De Sève S. (2005). « Quand la morphologie devient une ressource interactive », in D. Banks (éd.) *Les marqueurs linguistiques de la présence de l'auteur*. Paris : L'Harmattan, 243-254.
- Diani G. (2004). « The Discourse Functions of *I Don't Know* in English Conversation », in K. Aijmer et A.-Br. Stenström (éd.) *Discourse Patterns in Spoken and Written Corpora*. Amsterdam : Benjamins, 157-171.
- Gauvin K. (1999). *Une approche énonciative et interactive de je sais et je sais pas*, mémoire de maîtrise. Moncton : Université de Moncton.
- Grant L. E. (2010). « A Corpus Comparison of the Use of *I Don't Know* by British and New Zealand Speakers », *Journal of Pragmatics* 42, 8 : 2282-2296.
- Grice H.-P. (1979). « Logique et conversation », *Communications* 30 : 57-72.
- Kiparsky P. et C. Kiparsky. (1970). « Fact », in M. Bierwisch (éd.) *Progress in Linguistics. A collection of Papers*. The Hague : Mouton, 143-173.
- Kerbrat-Orecchioni C. (1991). « L'acte de question et l'acte d'assertion : opposition discrète ou continuum ? », in C. Kerbrat-Orecchioni (éd.) *La Question*. Lyon : Presses universitaires de Lyon, 87-111.
- Korzen H. (2001). « Factivité, semi-factivité et assertion. Le cas des verbes *savoir, ignorer, oublier et cacher* », in H. Kronning et al. (éd.) *Langage et référence*. Acta Universitatis Upsaliensis : Uppsala, 323-333.
- Langacker R. W. (1987). *Foundations of Cognitive Grammar. Theoretical Prerequisites*, vol. 1. Sandford : Stanford University Press.
- Levinson St. C. (1983). *Pragmatics*. Cambridge : Cambridge University Press.
- Poirier Cl. (2015). « Un pas en avant, un pas en arrière. Analyse du dictionnaire de l'Équipe FRANQUS », *Cahiers de lexicologie* 105, 1 : 21-53.

- Pop L. (2009). « Quelles informations se pragmatisent ? Le cas des verbes plus ou moins marqueurs », *Revue roumaine de linguistique* 54, 1-2 : 61-172.
- Récanati Fr. (1984). « Remarques sur les verbes parenthétiques », in P. Attal et Cl. Muller (éd.) *De la syntaxe à la pragmatique*. Amsterdam : Benjamins, 319-352.
- Rémi-Giraud S. (1986). « Étude comparée du fonctionnement sémantique et syntaxique des verbes *savoir* et *connaître* », in S. Rémi-Giraud et M. Le Guern (éd.) *Sur le verbe*. Lyon : Presses Universitaires de Lyon, 169-306.
- Schneider St. (2007a). *Reduced Parenthetical Clauses as Mitigators. A Corpus Study of Spoken French, Italian and Spanish*. Amsterdam et Philadelphia : Benjamins.
- Schneider St. (2007b). « Les clauses parenthétiques réduites en français, italien et espagnol. Une analyse pragmatique fondée sur des corpus de la langue parlée », in D. Trotter (éd.) *Actes du XXIV<sup>e</sup> Congrès international de linguistique et de philologie romanes. Aberystwyth, 2-5 août 2004*, vol. 3. Tübingen : Niemeyer, 423-432.
- Schneider St. (2013). « Clauses parenthétiques réduites et type d'interaction verbale. Quelques considérations », in H. Chuquet (éd.) *Des sentiments au point de vue : études de linguistique contrastive*. Rennes : Presses universitaires de Rennes, 175-194.
- Urmson J. O. (1952). *Parenthetical Verbs*, *Mind* 61, 244 : 480-496.
- Usito 2012-, Équipe Franqus. Sous la direction éditoriale de H. Cajolet-Laganière et P. Martel. Sherbrooke : Éditions Delisme inc.
- En ligne : <https://www.usito.com>. (Consulté le 5 mars 2015).

### Corpus

- CFPP2000 (Corpus de français parlé parisien des années 2000), Université Paris 3 – Sorbonne nouvelle.  
Site : <http://recherche.flsh.usherbrooke.ca/cfpq/> (Consulté le 21 septembre 2014).

*Le Corpus de français parlé au Québec (CFPQ)  
et la langue des conversations familiales*

CFPQ (Corpus de français parlé au Québec), CATIFQ,  
Université de Sherbrooke.

Site : <http://recherche.flsh.usherbrooke.ca/cfpq/>

Sous-corpus 1 à 21 exploités dans le cadre de la présente  
étude (= environ 31 h ½ d'enregistrement). (Consulté le  
21 septembre 2014)

*Corpus lexicaux québécois*, Secrétariat à la politique  
linguistique, Gouvernement du Québec, 1997-2007. Site :  
<http://www.spl.gouv.qc.ca/languefrancaise/corpuslexicaux/>  
(Consulté le 5 mars 2015).

ESLO (Enquêtes sociolinguistiques à Orléans, Université  
d'Orléans.

Site : <http://eslo.huma-num.fr/> (Consulté le 21 septembre  
2014)

OFROM (Corpus oral de français parlé en Suisse romande),  
Université de Neuchâtel.

Site : <http://www11.unine.ch/> (Consulté en 21 septembre  
2014).