

Ressources et outils pour la recherche en acquisition du français langue maternelle au Canada

Gisèle Chevalier, CRLA, Université de Moncton (Canada)

chevalg@umoncton.ca

Ce cours présente l'état de lieux au Canada, de la recherche en acquisition du français langue maternelle qui prend appuie sur la linguistique de corpus. Il ne sera donc pas question de l'acquisition du français comme langue seconde par les anglophones ou les allophones (citoyens parlant une langue autre que les deux langues officielles : l'anglais et le français), ni de l'acquisition des langues autochtones. Nous verrons que dans un pays où les communautés francophones vivent en contact intense avec l'anglais, la définition de la langue maternelle ou première pose problème. On comprendra vite à mon discours que je dirige ma lunette sur les recherches en acquisition tardive aux niveaux morphologique, morpho-syntaxique, lexical et pragmatique (Voir Y. Rose pour la phonologie), et que j'adhère à une perspective sociolinguistique de type variationniste en raison de la situation sociolinguistique du français dans l'ensemble du Canada.

Les corpus repérés à ce jour

Je présenterai dans un premier temps les maigres résultats d'une cueillette d'information sur les corpus et les outils existants pour traiter le français écrit et parlé dans les différentes régions du Canada. L'intérêt est là, si j'en juge au nombre de réponses que j'ai reçues, on manque de moyens et, peut-être, de direction. J'illustrerai la diversité des français du Canada en présentant des échantillons tirés des corpus qui ont été portés à mon attention par des publications ou lors de ma petite enquête : Allard 2002; Chiasson-Albert 2005; Godard 1991; Saint-Pierre 1991, Losier 2005. Je rappellerai brièvement les raisons historiques et socio-politiques à la source de la variation par un bref survol de l'évolution des français au Canada suivant (Poirier 2005; Leclerc 2000).

Les outils de traitement automatique des corpus

En ce qui a trait aux outils, plusieurs logiciels développés pour le traitement automatique du français aux différents niveaux d'analyse ont été mis en marché ces dernières années ou sont téléchargeables (Voir les outils présentés à l'École de linguistique de corpus tenue à Caen (2003). Ces logiciels traitent les textes français à des niveaux de sophistication impensables il y a quelques années à peine, mais la matière à étudier dans les corpus de productions écrites ou les transcriptions d'oral chez les enfants, passent par les mailles du filet des dictionnaires et grammaires standard. Ces formes « non reconnues » : formes orales familières non consignées dans les ressources lexicales informatisées, emplois régionaux, orthographe déviant des normes orthographiques, emprunts, constituent notre matière première. Ils apparaissent de façon récurrente, et atteignent des hautes fréquences d'emploi, s'agissant de la morphologie verbale et de mots grammaticaux (pronoms, conjonctions, connecteurs).

Outil de traitement automatique du français oral

Constant l'absence d'outils performants pour le traitement des corpus oraux transcrits, une équipe de linguistes et d'assistants à la recherche, assistée du concepteur de l'environnement linguistique NooJ, v.1.21 (Silberztein 2004) s'est donné comme objectif d'adapter les fonctionnalités du logiciel NooJ pour le traitement efficace de corpus acadiens écrits ou transcrits de l'oral (Chevalier, Kasparian et Silberztein 2004).

Efficace signifiant ici la description morphologique des occurrences, la reconnaissance du lien des occurrences avec un lemme, la mise en relation de variantes linguistiques et orthographiques avec un super lemme (chevals, chevaux, cheval, jeval : CHEVAL; maintenant, astheur, asteur, asteure, now : MAINTENANT) description des propriétés syntaxico-sémantiques des verbes n'en est qu'à ses débuts. À titre d'exemple, en fr. can., le verbe *échapper* entre dans une construction transitive : Echapper, (agent, thème) *J'ai échappé le savon*. En Acadie, la structure sémantique du verbe *appartenir* inverse l'ordre des arguments : elle requiert un sujet [+Hum] : *mon oncle appartient le restaurant sur la [rue] Church*.

Je ferai une démonstration du type d'analyse que permet à ce stade-ci de son développement le module NooJ Acadien, qui sera mis en ligne le 1^{er} juin, J'argumenterai qu'ayant construit des ressources lexicales pour une variété conservatrice (maintien des formes obsolètes en français de référence) en contact intense avec l'anglais, et fortement marquée quant à l'oral, l'utilité du logiciel s'étend au-delà des limites de l'Acadie du Nouveau-Brunswick. Je ferai enfin état des ajustements à faire au logiciel développé à partir de corpus de productions par des adolescents et des adultes, pour le traitement des productions enfantines répertoriées.

Pour conclure, je tracerai les lignes directrices pour la constitution d'un corpus pan-canadien de productions orales d'enfants francophones.

Les corpus de productions enfantines

- 1) Allard, Suzanne 2002. (Université Sainte-Anne, Nouvelle-Écosse). Non diffusé.
- 2) Chiasson Albert, Nathalie 2005 (Campus de Shippagan, Université de Moncton). Récit de 24 enfants de 4 à 12 ans élicité par une séquence d'images. Non diffusé.
- 3) Godard, Lucie 1991 (UQAM, Québec) : Récits écrits et oraux inventés par des de 5 à 12 ans. Non diffusé.
- 4) Losier, Line 2005 (Université de Moncton, Nouveau-Brunswick). Conversations filmées de groupes d'enfant de 2 à 5 ans pendant une session de jeux libre. Non diffusé.
- 5) Saint-Pierre, Madeleine 1989 (UQAM, Québec) : enregistrements audio-visuels d'enfants de la maternelle (5-6ans) en situation de jeu de rôle. Non diffusé.

Références

- CANAC-MARQUIS, Steve, et POIRIER, Claude (2005). « Origine commune des français d'Amérique du Nord: le témoignage du lexique », dans Albert Valdman, Julie Auger et Deborah Pison-Halten (sous la dir. de), *Le français en Amérique du Nord. État présent*, [Sainte-Foy], Les Presses de l'Université Laval, coll. Langue française en Amérique du Nord, p. 517-538.
- CHEVALIER, Gisèle., KASPARIAN, Sylvia. et SILBERZTEIN, Max (2004). « Éléments de solution pour le traitement automatique d'un français oral régional ». *Revue française en traitement automatique du langage (TAL)* Vol 2 N°45 : La linguistique de corpus sous la direction de J. Véronis, Paris, Éditions Lavoisier, pp. 41-62.
- Leclerc, Jacques (2000). *L'aménagement linguistique dans le monde*. <http://www.tlfq.ulaval.ca/axl/amnord/canada.htm>
- MOUGEON, R. & BENIAK, É. (1991). *Linguistic consequences of language contact and restriction: The case of French in Ontario*, Oxford: Oxford University Press.
- MOUGEON, Raymond et Edouard BENIAK (1989). *Le Français canadien parlé hors Québec : aperçu sociolinguistique*. Éditeur: Sainte-Foy : Presses de l'Université Laval.
- SILBERZTEIN, Max (2004). Logiciel NooJ v1.21 <http://perso.wanadoo.fr/rosavram/>