

Université de Neuchâtel
Centre de linguistique de Corpus

Corpus « Le discours des Premiers ministres canadiens »

Novembre 2017

dominique.labbe@umpacte.fr

denis.moniere@umontreal.ca

Depuis 1867, la confédération canadienne est dotée d'institutions parlementaires. Ces institutions ont le français comme langue officielle.

Le Premier ministre s'exprime en chambre où il présente le programme législatif, conduit la majorité et répond aux questions des députés. Cet aspect est décrit dans le corpus « discours gouvernemental » (sous-corpus « CTrone »). Mais la majeure partie de sa communication a lieu en dehors de la chambre où il s'adresse à ses concitoyens, soit par les médias audiovisuels ou écrits, soit directement devant divers auditoires. Cette communication fait l'objet du corpus « Premiers ministres canadiens ».

Les textes ont été rassemblés par Denis Monière (Université de Montréal). La collecte s'est étendue de 1995 à 2010, périodes pendant lesquelles, les services du Premier ministre ont retranscrit ses principales interventions hors de la chambre et les ont placés sur le site internet du gouvernement fédéral où elles demeuraient en ligne (<https://pm.gc.ca/fra>). A partir du printemps 2010, les discours anciens sont effacés, la plupart des interventions hors chambre font l'objet d'un simple communiqué et celles qui sont retranscrites ne demeurent en ligne que peu de temps.

Les archives fédérales d'Ottawa n'ayant pas souhaité numériser ces discours, il a été impossible de compléter le corpus pour les années antérieures à 1995 et postérieures à 2010.

Le corpus se compose de :

Jean Chrétien (1995-2003) : 270 textes, 442 789 mots, 9 210 vocables différents

Paul Martin (2004-2005) : 49 textes, 116 972 mots, 5 611 vocables différents

Stephen Harper (2006-2010) : 310 textes, 400 710 mots, 10 615 vocables différents.

Total : 629 textes, 960 471 mots 14 260 vocables différents.

Contenu du dossier

L'ensemble est au format windows.

Les textes sont classés par ordre chronologique.

Les titres des fichiers correspondants aux textes sont :

- Nom du premier ministre
- Année, mois, jour

Le corpus se compose de trois sous-dossiers (un par Premier ministre).

• En tête de chaque dossier :

- la description du corpus (format excel) : nom de fichier, auteur, nature du texte et date, nombre de mots, de formes graphiques et de vocables ;

- l'index alphabétique (nom du corpus précédé du préfixe IA, format excel) : vocables classés selon l'ordre alphabétique avec les formes graphiques sous lesquelles ils apparaissent dans chaque texte et/ou sous-corpus (en colonne) et leur nombre d'occurrences (effectifs absolus) ;

- l'index hiérarchique (nom du corpus précédé du préfixe IH, format txt windows) : densités des catégories grammaticales et vocables les plus utilisés.

• Le texte de chaque discours (txt windows) est accompagné du texte lemmatisé (même nom de fichier précédé du préfixe CN) et de son index électronique (préfixe II).

La première ligne indique l'auteur, sa fonction, la nature du texte, son titre, la date et le lieu.

Dans les fichiers lemmatisés (préfixe CN), chaque mot occupe une ligne :

- première position : la forme graphique standard correspondant au mot,
- deuxième position : l'entrée de dictionnaire correspondant à la forme (par exemple, infinitif du verbe),

- troisième position : le code grammatical (pour la signification de ces codes, se reporter au document « NomenclatureCN » dans ce même dossier).

Règles d'usage :

Ces fichiers ne peuvent faire l'objet d'aucun usage commercial.

Les premières lignes ne doivent pas être enlevées.

Toute publication réalisée avec ces fichiers indiquera la source et sera envoyée à Dominique Labbé et Denis Monière (adresses électroniques au début de ce document).

Bibliographie

- Les normes de standardisation et de lemmatisation utilisées pour produire ces fichiers

Dominique Labbé. *Normes de saisie et de dépouillement des textes politiques. Cahier du CERAT n° 7*. Grenoble : CERAT-IEP, avril 1990 (ouvrage consultable en ligne : <https://www.researchgate.net/publication/40270233>)

- La lexicométrie

Cyril Labbé & Dominique Labbé. Lexicométrie : quels outils pour les sciences humaines et sociales ? *Communication aux Journées d'étude Usages de la lexicométrie en sociologie*. Université de Versailles, 12-13 juin 2013 (<https://hal.archives-ouvertes.fr/hal-00834039>)

- Ouvrages et articles utilisant le corpus des discours des premiers ministres canadiens :

Labbé Dominique & Monière Denis. Quelle est la spécificité des discours électoraux? Le cas de Stephen Harper. *Canadian Journal of Political Science / Revue canadienne de science politique*, 43:1, (March/ mars 2010), p. 69–86.

Labbé Dominique & Monière Denis. Les particularités d'un discours politique : les gouvernements minoritaires de Pierre Trudeau et de Paul Martin au Canada. *Corpus*, 4, 2005, p. 79-104.