

Université de Neuchâtel  
Centre de linguistique de Corpus

Corpus  
Les discours du général de Gaulle  
(1958-1969)

Décembre 2017

[dominique.labbe@umrpacte.fr](mailto:dominique.labbe@umrpacte.fr)

Ensemble des interventions publiques disponibles du Général de Gaulle de mai 1958 à avril 1969 : allocutions, entretiens, conférences de presse, messages...

En décembre 2017, il comporte 459 textes, soit 410 492 mots et un vocabulaire de 9 004 vocables différents.

- les interventions radio-télévisées : textes effectivement prononcés par le général de Gaulle ;

- les autres discours et messages : textes communiqués par les services de la présidence et diffusés par les agences de presse.

Ces documents ont été produits dans le cadre des fonctions de président, diffusés sans restriction par la présidence et largement publiés dans la presse.

Plus de 400 de ces textes ont été reproduits dans : Charles de Gaulle. *Discours et messages* (tomes 3 à 5). Paris : Plon, 1970. Pour ces textes, la référence est donnée dans l'entête du fichier afin de permettre de s'y reporter commodément.

## **Contenu du dossier**

Deux sous-corpus sont distingués grâce à des listes placées en tête du dossier :

A. Ensemble des interventions radiotélévisées (allocutions, entretiens, conférences de presse) ;

## B. Autres interventions.

Le corpus est décrit dans "0 Corpus....xls", placé en tête du dossier et lisible avec n'importe quel tableur.

Les textes sont classés par ordre chronologique (année, mois, jour). Exemple : DeGaulle19580515 est un communiqué émis le 15 mai 1958.

Les trois premières lignes de chaque fichier donnent :

- les indications bibliographiques (auteur, titre, lieu et date)
- la référence du document papier ou électronique auquel on peut se reporter,
- le nom de la personne ayant réalisé les opérations de saisie, de standardisation orthographique, de balisage et de lemmatisation avec la date de ces opérations.

Chaque texte est sous deux formats :

- le texte original (NomdeFichier.txt). La graphie de chaque mot a été contrôlée, notamment celle des noms propres. Tout le paratexte est placé entre <> (références du texte, questions des journalistes...).
- une version lemmatisée (même nom de fichier précédé du radical CN).

Dans le fichier CN, chaque mot du texte apparaît sur une ligne avec sa graphie standard, son entrée de dictionnaire (par exemple, l'infinitif du verbe) et son code grammatical, chaque champ est séparé par une virgule. La liste des codes grammaticaux est donnée dans le fichier "Nomenclature CN.pdf".

- chaque texte est accompagné d'un index (même nom de fichier précédé du radical II). Cet index est utilisé par les différents outils informatiques d'analyse des corpus.

Le vocabulaire du corpus est donné dans :

- index alphabétique (nom de corpus précédé du radical IA), lisible sous excell. Dans les deux premières colonnes, les mots du corpus - transcrits sous leurs graphies standards - sont rangés sous leurs entrées de dictionnaire (par ordre alphabétique). En troisième colonne, l'effectif dans l'ensemble du corpus. Dans les colonnes suivantes, les textes où ils apparaissent avec leurs nombres d'occurrences (effectifs absolus).
- index hiérarchique (nom de corpus précédé du radical IH) : principaux renseignements sur le vocabulaire du corpus avec les vocables les plus utilisés rangés par catégories grammaticales.

## 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 (adresse électronique au début de ce document).

Les corpus étiquetés sont couverts par la licence BSD (<http://www.opensource.org/licenses/bsd-license.html>), Copyright (c) 2017, Dominique Labbé, Denis Monière, Cyril Labbé.

Les outils d'exploitation des fichiers CN sont disponibles auprès de Dominique Labbé.

## Bibliographie

- La première étude de statistique appliquée au discours politique

Cotteret Jean-Marie & Moreau René. *Le vocabulaire du Général de Gaulle*. Paris : Presses de la Fondation des sciences politiques, 1967.

- Les normes de standardisation et de lemmatisation utilisées pour produire le corpus

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>)

- Les discours des présidents de la république française depuis 1958

Edward Arnold, Cyril Labbé, Dominique Labbé & Denis Monière. *Parler pour gouverner : Trois études sur le discours présidentiel français*. Grenoble : Laboratoire d'Informatique de Grenoble, 2016, p. 38-53.

- Ouvrages et articles utilisant le corpus de Gaulle (ordre chronologique inverse)

Cyril Labbé & Dominique Labbé. *La répartition du vocabulaire*. Grenoble : Laboratoire d'informatique de Grenoble, septembre 2017.

Cyril Labbé & Dominique Labbé. Existe-t-il un langage propre à la politique ? In David Banks (Ed). *Aspects linguistiques du texte politique*. Paris : L'Harmattan, 2014, p. 7-28.

- Cyril Labbé & Dominique Labbé. La diachronie dans le discours politique. Le général de Gaulle. In David Banks (Ed). *Aspects diachroniques du texte de spécialité*, Paris, l'Harmattan, 2010, p. 129-148.
- Dominique Labbé. Le général de Gaulle en campagne. In David Banks (Ed.). *Aspects linguistiques du texte de propagande*. Paris : L'Harmattan, 2005, p. 213-233.
- Pierre Hubert, Cyril Labbé & Dominique Labbé. Automatic Segmentation of Texts and Corpora. *Journal of Quantitative Linguistics*, december 2004, 11-3, p. 193-213.
- Dominique Labbé. La richesse du vocabulaire politique : de Gaulle et Mitterrand. In Sylvie Mellet & Marcel Vuillaume. *Mots chiffrés et déchiffrés. Mélanges offerts à Etienne Brunet*. Paris : Champion, 1998, p. 173-186.
- Dominique Labbé. La France chez de Gaulle et Mitterrand. In Pierre Fiala et Pierre Lafon (dir). *Des mots en liberté. Mélanges Maurice Tournier*. Fontenay-aux-Roses : ENS Editions, 1998, p. 183-193.
- Dominique Labbé. Le "nous" du général de Gaulle. *Quaderni di studi linguistici*. 4/5, 1998, p 331-354.
- Pierre Hubert & Dominique Labbé. La structure du vocabulaire du général de Gaulle. In Sergio Bolasco, Ludovic Lebart et André Salem. *III Giornate internazionali di Analisi Statistica dei Dati Testuali*. Rome : Centro d'Informazione e stampa Universitaria, 1995, tome II, p. 165-176.