Wordscore

La dernière méthode pour situer la position des partis politiques est le programme informatique Wordscore développé récemment par Michael Laver et ses collègues (Laver 2003). Contrairement au CMP et aux méthodes s'appuyant sur un dictionnaire qui traitent les textes comme des discours qui doivent être compris et interpréter pour en comprendre la signification, soit par un codeur humain ou un ordinateur, la technique Wordscore traite les textes (plus précisément les mots contenus dans ces textes) comme des données qui contiennent de l'information à propos de la position des auteurs du texte sur une dimension politique prédéfinie.

À partir d'un ensemble de textes de référence dont les positions politiques sont déterminées a priori, la technique permet d'extraire des données de ces textes de référence sous la forme de fréquences de mots et utilise cette information pour estimer la position politique de textes vierges à propos desquels on ne sait rien. Pour paraphraser Laver, Benoit et Garry (2003) ce que l'on sait d'un texte vierge est limité aux mots qu'il contient. Le programme Wordscore compare les mots avec ceux contenus dans les textes de référence qui sont utilisés pour représenter les positions extrêmes sur un axe défini a priori (l'axe gauche-droite, par exemple, où le texte de référence de gauche a un score de 0 et celui de droite, 10). Un programme d'analyse de données génère un premier score (appelé *wordscore*) pour chaque mot selon la fréquence relative de son apparition dans le texte de référence, puis un deuxième score pour la fréquence relative des tous les scores dans un texte vierge (appelé *textscore*). L'étape finale consiste à transformer les *textscores* des textes vierges pour les ramener à l'échelle métrique originale. C'est ce calcul qui permet de situer les textes dans l'espace politique.

Le principal avantage de la méthode Wordscore est qu'elle produit une distribution des scores autour d'un score moyen estimé. Cela permet d'obtenir une mesure d'erreur-type et donc, d'établir un intervalle de confiance autour du score moyen estimé. La méthode Wordscore a l'avantage unique par rapport aux autres méthodes examinées ici de fournir une mesure statistique pour établir à quel point deux textes vierges sont différents l'un de l'autre dans leur

vocabulaire. En fait, deux textes sont statistiquement distincts si leurs intervalles de confiance ne se chevauchent pas. Bien entendu, les scores sont davantage valides si on a confiance dans le choix des textes de références et dans la mesure utilisée pour décider quels sont leurs positions données sur une échelle ou un axe prédéfini. Laver, Benoit et Garry (2003) font plusieurs recommandations importantes au sujet du choix des textes de référence et de l'attribution de cotes à ceux-ci, une d'entre elle étant que les textes vierges et les textes de référence doivent partager un cadre de référence commun.
La méthode Wordscore convient bien à la comparaison entre textes de même nature (deux programmes de partis lors d'une même élection par exemple) mais elle échoue à comparer des textes politiques de nature différente (un programme électoral et un discours du trône par exemple).

Bibliographie

Laver, Michael, Kenneth Benoit et John Garry (2003), « Extracting Policy Positions from Political Texts Using Words as Data », *American Political Science Review*, vol. 97, n° 2, p. 311-331.