

# AT2 : ANALYSE TEXTUELLE

Le développement de la **communication électronique** produit de plus en plus de grands corpus de textes, notamment ceux qui sont associés à des débats publics. On se souvient ainsi du rôle joué par les forums en ligne à propos du projet de référendum sur la constitution européenne. Plus récemment, le « grand débat sur l'identité nationale » restera sans doute comme l'une des polémiques majeures du quinquennat en cours.

L'analyse des corpus montre **des positionnements spécifiques des médias et des internautes dans la structuration même de la référence**. Certains objets traités de façon homogène par les médias, sont différenciés par les internautes et, inversement, les médias établissent des cloisonnements dans des objets que les internautes traitent de façon unifiée.

Pour étudier l'articulation entre cadres médiatiques et cadres personnels dans la construction du débat public, il faut considérer que **les cadres sont des facettes du monde social, dont l'éclairage est pris en charge par des groupes sociaux ou institutions** constitués en sources et se livrant une compétition pour l'accès aux médias.

C'est par la définition ou l'éclairage donné à certaines thématiques, le champ de la connaissance dans lequel elles sont développées, que se posent les termes du débat relatif à tout phénomène social. Sur un même objet, mobilisant globalement le même vocabulaire, on a pu **mettre en évidence des structurations lexicales différentes**. Cela signifie bien que ce n'est pas la présence ou l'absence, ni le nombre d'occurrence de telle forme lexicale ("compter des mots") qui importe. Ce qui crée la signification, c'est **l'organisation de ces formes lexicales et les relations** qu'elles entretiennent pour fournir une représentation cohérente. Il s'agit donc, non seulement de décrire la structure de la discussion sur la toile et celle de son traitement médiatique, mais également d'identifier leurs possibles relations.

S'agissant de corpus textuels de plus en plus importants, cet objectif n'est réalisable qu'à condition de **disposer d'outils de mesures lexicales adaptés**.

Le développement du logiciel Iramuteq est particulièrement destiné à répondre à ces nouvelles exigences.

Le premier objectif que s'était fixé le développement de cet outil était de **reproduire de façon assez fidèle des analyses textuelles classiques**. Le développement actuel s'oriente dans cinq directions : la production d'outils d'aide à l'interprétation des analyses, l'implémentation de nouvelles méthodes (application des analyses de similitude sur les corpus découpés en segments de texte), l'adaptation aux corpus de grande taille, le perfectionnement de la phase de lemmatisation et l'internationalisation (les corpus anglais et allemand sont en phase pré-expérimentale ; l'italien, l'espagnol et le portugais devront être rapidement ajoutés). Plusieurs de ces objectifs sont déjà atteints (outils d'aide à l'interprétation, analyse de similitude), les autres sont à l'étude ou en cours de réalisation.

L'ensemble de logiciels libres sur lesquels repose cet outil (notamment R) permettent également d'**envisager l'incorporation rapide de nouvelles méthodologies d'analyses**. S'ils font, d'ores et déjà, d'Iramuteq l'un des logiciels d'analyse de discours les plus complets du domaine - et rendent également son utilisation très simple dans les recherches et dans les formations (le logiciel fonctionne sous Windows, Mac OS X et linux et est totalement gratuit) - une phase de développement et d'expansion de l'outil est appelée à démarrer.