



# ASSEMBLÉE NATIONALE

14ème législature

## langues régionales

Question écrite n° 24596

### Texte de la question

M. Pierre Morel-A-L'Huissier attire l'attention de Mme la ministre de la culture et de la communication sur l'ancrage des langues régionales dans la modernité à travers l'internet et les nouvelles techniques. Il lui demande de bien vouloir lui faire part de ses intentions en la matière.

### Texte de la réponse

Le ministère de la culture et de la communication veille à ce que l'émergence des nouvelles techniques de l'information soit bénéfique aux langues de France comme à la langue française. Il a collaboré activement avec l'Association française pour le nommage Internet en coopération (AFNIC), chargée par l'État d'enregistrer les noms de domaine en . fr et dans les extensions suivantes, correspondant à des zones où sont en usage des langues régionales : . re (la Réunion), . wf (Wallis et Futuna), . yt (Mayotte) pour rendre possible, depuis 2012, l'enregistrement de noms de domaines comportant des signes diacritiques propres au français et aux langues de France (accents, tréma, cédille). Sur ce thème du nommage, le ministère se réjouit que la société pour l'attribution des noms de domaine et des numéros sur Internet (ICANN), de droit américain, ait récemment rendu un avis positif lors de l'évaluation du projet de nouvelle extension de premier niveau . bzh porté par le conseil régional de Bretagne. Le ministère soutient également la numérisation de nombreux documents en langues de France, en liaison avec des organismes tels que la Fédération des associations de musiques et danses traditionnelles (FAMDT), les Francophonies en Limousin, la maison de la culture yiddish, l'association Bigun Kakol (langue basque), le Centre interrégional de développement de l'occitan (CIRDOC), l'Institut d'études occitanes ou encore l'association Nosauts de Bigorra. Par ailleurs, la ministre de la culture et de la communication a lancé le 19 novembre 2012 un partenariat avec l'établissement public de recherche Inria et l'association Wikimedia France dans le but de sémantiser, c'est-à-dire de rendre plus accessibles des corpus encyclopédiques présents sur le site Internet Wikipédia, de façon à permettre leur réutilisation sur d'autres sites et notamment ceux mis en oeuvre par le ministère. Ces jeux de données disponibles en de nombreuses langues de France (400 000 pages disponibles en catalan, 150 000 en basque, 72 000 en occitan ou encore près de 50 000 en breton), désormais liés aux pages en français facilitent la mise en ligne de collections de ressources dans les langues régionales, en plus du français. Plusieurs projets expérimentaux de ce type seront menés par la délégation générale à la langue française et aux langues de France (DGLFLF) entre 2013 et 2015. Depuis plusieurs années, la DGLFLF travaille par ailleurs avec le Centre national de la recherche scientifique au programme Corpus de la parole, qui a pour objet de constituer, de conserver, de mettre à disposition et de valoriser des corpus oraux enregistrés dans toutes les langues de France. Des centaines d'heures sont disponibles sur le site Corpus de la parole, avec pour objectif non seulement le développement d'une base de données sur l'oral, mais aussi d'outils de traitement automatique des langues et d'ingénierie linguistique. Indispensables à la recherche, les données ainsi assemblées acquièrent en outre, avec le temps, un caractère patrimonial et de ressources langagières. Enfin, la DGLFLF conduit cette année plusieurs projets intéressant directement les langues de France dans le contexte numérique : un projet de cartographie et de géocodage des différentes langues parlées en France permettant d'alimenter une future norme représentant les langues du

monde sur le globe terrestre (projet Linguasphère) ; un projet mené avec le laboratoire européen ELRA établissant un inventaire exhaustif des ressources multilingues et des sources orales et écrites en langues de France, dans le but de développer un outillage propre à certaines de ces langues (outils d'aide à la traduction, de synthèse ou de reconnaissance vocale par exemple) ; un projet mené avec le Réseau Maaya qui permettra de mieux appréhender la présence exacte des langues de France sur la toile ; un « wikicamp » ou projet d'atelier d'écriture collaborative en langues d'outre-mer pour encourager la présence de ces langues, aujourd'hui sous-représentées sur la toile. D'une manière générale, les langues de France sont systématiquement intégrées aux projets de développement du numérique au service des langues.

## Données clés

**Auteur :** [M. Pierre Morel-A-L'Huissier](#)

**Circonscription :** Lozère (1<sup>re</sup> circonscription) - Les Républicains

**Type de question :** Question écrite

**Numéro de la question :** 24596

**Rubrique :** Culture

**Ministère interrogé :** Culture et communication

**Ministère attributaire :** Culture et communication

## Date(s) clé(s)

**Question publiée au JO le :** [23 avril 2013](#), page 4315

**Réponse publiée au JO le :** [17 septembre 2013](#), page 9651