



# ASSEMBLÉE NATIONALE

14ème législature

## équipements

Question écrite n° 62697

### Texte de la question

M. Thierry Lazaro attire l'attention de Mme la ministre de la culture et de la communication sur la mise en œuvre par la Commission européenne de plusieurs lignes directrices pour des financements destinés à améliorer les technologies de traduction automatique, d'analyse des contenus et de lecture. Il souhaite connaître les intentions du Gouvernement en la matière.

### Texte de la réponse

La ministre de la culture et de la communication veille, en liaison avec les autres départements ministériels concernés, à ce que la Commission européenne donne à la langue française la place qui lui revient dans les programmes visant à améliorer les technologies linguistiques. Ainsi, sur le volet recherche et innovation du programme H2020, figure un appel à projets intitulé « Cracking the language barrier » qui accompagnera quatre projets de développement dans le domaine de la traduction automatique, de la traduction médicale, et de la traduction des cours en ligne ouverts et massifs (CLOM). Le projet « QT21 », notamment, vise à améliorer la qualité de la traduction automatique depuis et vers l'anglais pour les langues de l'Europe centrale et de l'Est, et l'on peut regretter que le français ne soit pas traité dans ce projet majeur. La France souhaite que ce projet puisse être étendu à d'autres langues représentant la diversité linguistique de l'Union européenne et notamment le français. Le ministère de la culture et de la communication conduit, par ailleurs, une action volontariste pour que les technologies de traitement automatisé du langage soient mises au service de la langue française et des langues de France. Ces technologies représentent en effet un enjeu culturel majeur : plus une langue est accessible, traduisible et utilisable au moyen de ces technologies, plus elle est couramment utilisée sur les réseaux numériques. Si les problèmes d'ingénierie linguistique sont loin d'être résolus et nécessiteront encore longtemps des efforts importants de recherche, de nombreuses applications rendent d'ores et déjà de grands services au public et génèrent une forte activité économique réelle : traduction automatique, reconnaissance et synthèse vocales... Les 19 et 20 février derniers, à l'initiative de la DGLFLF, un colloque a permis de réunir des acteurs tant institutionnels (au niveau de l'État comme des collectivités locales), qu'académiques, scientifiques (Centre national de la recherche scientifique) et associatifs pour réfléchir à l'apport des technologies du langage au français et aux langues régionales de France. Le ministère de la culture et de la communication poursuit, pour sa part, son effort dans trois directions. D'une part, il soutient tout d'abord la numérisation de nombreux documents en langues de France. D'autre part, le programme Corpus de la parole a pour objet de constituer, conserver, mettre à disposition et valoriser des corpus oraux enregistrés dans toutes les langues de France. Des centaines d'heures sont désormais disponibles 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. Le troisième axe de travail concerne le développement du web sémantique et du web de données avec, par exemple, la contribution à la mise en place du service DBpédia en français, qui rend librement accessible à tous les créateurs de sites et services innovants, au format du web sémantique, l'ensemble des pages (en français et dans plusieurs autres langues) de l'encyclopédie Wikipédia.

Cet outil, réalisé par l'établissement public de recherche Inria dans le cadre d'une convention de partenariat « Sémanticpédia » établie avec le ministère de la culture et de la communication et l'association Wikimedia France, est aujourd'hui utilisé par plusieurs dizaines de sites Internet. Par ailleurs, le projet « JocondeLab » permet désormais d'accéder en quatorze langues, dont quatre langues régionales de France, à près de 300 000 notices décrivant des oeuvres des musées de France. Ce projet a été couronné en mars 2014 par le prix « Data Access » des « Data Intelligence Awards », qui récompense un projet améliorant significativement l'accès à l'information publique par le champ des données offertes et la pertinence des résultats de recherche. Le ministère de la culture et de la communication entend poursuivre et amplifier à l'avenir son action selon ces différents axes de travail, qui s'inscrivent dans sa politique en faveur de la diversité culturelle et linguistique.

## Données clés

**Auteur :** [M. Thierry Lazaro](#)

**Circonscription :** Nord (6<sup>e</sup> circonscription) - Les Républicains

**Type de question :** Question écrite

**Numéro de la question :** 62697

**Rubrique :** Politiques communautaires

**Ministère interrogé :** Culture et communication

**Ministère attributaire :** Culture et communication

## Date(s) clé(s)

**Question publiée au JO le :** [5 août 2014](#), page 6575

**Réponse publiée au JO le :** [28 avril 2015](#), page 3221