

Délégation générale à la langue française et aux langues de France

LE NUMERIQUE AU SERVICE DE LA POLITIQUE LINGUISTIQUE

I. Problématique et enjeux

Au début de l'année 2013, la population utilisant internet s'élevait à **environ 2,2 milliards**. Plus de **80 % des Français sont abonnés à internet**, 92 % d'entre eux s'y connectent tous les jours, et le recours aux connexions mobiles est de plus en plus fréquent : en 2013, **près de 80 % des téléphones vendus devraient être des téléphones intelligents** « *smartphones* » permettant d'accéder à internet.

L'avenir de nos langues se joue donc aujourd'hui essentiellement sur les réseaux et l'observation des usages numériques montre combien il est important d'assurer la présence du français et des langues de France dans cet univers qui est **une chance pour la diversité linguistique et le multilinguisme** : on compte à ce jour 127 déclinaisons de Wikipédia comprenant plus de 10 000 articles, alors que l'encyclopédie libre n'était disponible qu'en anglais en 2001.

Le **traitement automatique des langues** offre quant à lui des perspectives nouvelles tant en termes économiques que culturels. Certaines technologies comme la **reconnaissance vocale**, devenue le premier argument de vente du téléphone *iPhone* (Siri), deviennent incontournables, et modifient radicalement notre façon de communiquer.

L'observation des pratiques est donc indispensable, car elle permet de mettre en œuvre **une politique de la langue adaptée à un monde en rapide évolution** et dans lequel les frontières traditionnelles ont peu d'importance. **Écrire en français sur la toile, c'est pouvoir être lu par plus de 200 millions de francophones à travers le monde**. Il importe donc d'être particulièrement vigilant quant à la diffusion de la culture et des langues sur la toile, et trois enjeux paraissent plus particulièrement stratégiques à ce titre :

- **le web des données** ou « web sémantique », qui est en passe de devenir le moyen incontournable pour accéder à l'information et à la culture sur l'internet et dès lors **assurer la visibilité d'une langue et d'une culture** dans un contexte fortement multilingue ;
- **la mise à disposition de données « ouvertes » et librement réutilisables**, ceci afin d'encourager leur dissémination sur les réseaux et notamment sur les portails collaboratifs tels que Wikipédia, dont près de la moitié des pages sont à caractère culturel ;
- **le développement des ressources et technologies pour le traitement automatisé de la langue** (reconnaissance vocale, synthèse vocale, traduction assistée ou automatique) pour le français mais aussi pour les langues de France.

II. État des lieux/bilan des actions entreprises

Ce changement rapide et radical des usages a nécessité **la mise en place d'une nouvelle mission** en charge des questions numériques à la DGLFLF l'année

dernière et **l'adaptation de certaines méthodes de travail et de diffusion aux nouveaux médias**. Deux exemples nous semblent particulièrement pertinents :

- d'une part, l'ouverture du dispositif d'enrichissement de la langue à la **participation des internautes**, via une plate-forme collaborative inaugurée en septembre 2011, le « WikiLF » : wikilf.culture.fr
- d'autre part, la mise en ligne en janvier 2014 d'un site expérimental « JocondeLab » s'appuyant sur le web de données et permettant **la diffusion de près de 300 000 notices illustrées des œuvres des musées de France en 14 langues** pleinement fonctionnelles, dont 4 langues régionales (breton, basque, occitan, catalan), projet lauréat des *Data Intelligence Awards* 2014.

III. Chantiers à conduire

– à court terme (année 2014)

Le chantier prioritaire est le **second projet innovant mené dans le cadre du programme de sémantisation de ce ministère et qui s'appuie sur le partenariat stratégique** Sémanticpédia avec Wikimedia France et Inria. Celui-ci vise à exposer en plusieurs langues étrangères des extraits sonores des différentes langues de France, mises à disposition sur le site Corpus de la parole.

Le deuxième chantier à court terme est l'organisation à Paris, avec l'appui du CNRS et du ministère de la Recherche, d'une **journée d'ateliers sur la thématique du développement des technologies de la langue pour les langues de France** de façon à poser les bases d'un grand programme national d'aide au développement d'outils pour ces langues.

– à moyen et long terme

À plus long terme, nous souhaitons **pérenniser la plate-forme de recherche Semanticpedia** mise en place par le MCC avec l'INRIA et Wikimedia France en l'élargissant aux établissements publics culturels, pour faire aboutir d'autres projets stratégiques, par exemple avec le **Centre des monuments nationaux**. Cette démarche déjà entamée dans le cadre d'un groupe de travail adhoc avec le Secrétariat général pourrait se concrétiser par un la signature d'un partenariat « **Sémanticpédia V2** » renforcé notamment en termes financiers.

Sur le volet des technologies de la langue, nous souhaitons lancer un programme de développement de ces technologies, qui pourrait s'intituler « **Technolangue II** », s'inspirant du programme « **Technolangue** » initié en 2003 sur ce même thème. Ce programme, dont l'utilité est soulignée dans le rapport sur la francophonie remis par Jacques Attali en août 2014, pourrait être cofinancé dans le cadre du **Programme d'Investissements d'Avenir (PIA)**.

Il faut par ailleurs renforcer nos partenariats avec **l'Union Européenne**, notamment en ce qui concerne le multilinguisme en Europe, l'utilisation des **technologies de la langue**, et, bien sûr **Europeana** ainsi qu'avec **l'Institut Français**, dont l'action sur le numérique menée à travers le monde apparaît comme **complémentaire** de celle de la DGLFLF menée principalement en France.