

Le numérique au service de la langue française et des langues de France

Les enjeux du numérique sont progressivement devenus un sujet de préoccupation et de négociation majeur pour les États et les entreprises à travers le globe. Alors qu'il y a désormais 2 milliards d'utilisateurs d'internet dans le monde, les Français sont devenus férus des réseaux numériques et, notamment, des réseaux sociaux. Dans un contexte de libéralisation des données, les usages numériques se sont fortement diversifiés et sont de plus en plus mobiles : en 2011, un téléphone mobile sur trois est un terminal intelligent (*smartphone*).

Ces technologies de l'information et de la communication concourent grandement aux objectifs de la politique de la langue : préserver la langue française et les langues de France pour répondre aux besoins d'expression et de communication des citoyens et des institutions d'une part, favoriser le multilinguisme en permettant le passage entre le français et les autres langues d'autre part.

Repères

Délégation générale à la langue française et aux langues de France

La langue au cœur des réseaux numériques

Les réseaux et les technologies numériques ont révolutionné la façon de s'informer, de communiquer et d'enseigner. L'information en temps réel passe de plus en plus par l'écrit plutôt que par l'oralité et la publication sur internet permet à la fois l'instantanéité et l'universalité. En outre, les réseaux sociaux bouleversent la façon dont est diffusée l'information : contrairement aux médias traditionnels, chaque destinataire d'une information peut devenir à son tour diffuseur, au moyen d'un simple clic. Chaque internaute a ainsi la possibilité de diffuser une information à la planète entière et, potentiellement, de toucher autant de lecteurs qu'un média professionnel, au moyen d'une simple connexion au réseau.

Les internautes francophones ont très tôt montré leur engouement pour ces nouvelles possibilités d'expression, utilisant spontanément le français pour créer leurs premiers sites personnels, sur un réseau pourtant très anglophone à ses débuts. Mais c'est surtout le succès de l'internet participatif en France, réputé pour son grand nombre de blogueurs, qui a contribué à développer l'usage de la langue française sur

l'internet. L'implication massive des internautes dans des projets contributifs phares tels que l'encyclopédie en ligne *Wikipedia* permet de promouvoir activement la langue française.

Face à ces mutations, comment mettre ces technologies au service de la promotion du français dans le monde ? Comment assurer et renforcer la présence des médias traditionnels sur la toile ? Peut-on s'appuyer sur une communauté de contributeurs actifs mais décentralisée, et de quelle façon ? Quelles perspectives présentent les technologies du numérique en matière d'enseignement, de formation à distance ? De quelle façon les pouvoirs publics peuvent-ils prolonger les initiatives privées en ce qui concerne la mise en ligne de contenus numériques ? Quelle est la place de la langue française dans les bibliothèques numériques ou encore dans les projets de portails de données publiques « *open data* » tels qu'*Etalab* ? En matière de *web sémantique*, comment promouvoir la constitution d'ontologies francophones dans les différentes filières et secteurs d'activité ?



Délégation générale à la langue française et aux langues de France

6 rue des Pyramides
75001 Paris
Téléphone : 01 40 15 73 00
Télécopie : 01 40 15 36 76
Courriel : dglff@culture.gouv.fr
Internet : www.dglf.culture.gouv.fr
ISSN en ligne : 2260-2887

Le web de données

Le développement en français d'outils de création, de transmission et de diffusion culturelle permet de conforter le rôle historique de la langue française comme langue internationale de diffusion et de production des savoirs.

Grâce aux nouvelles technologies du *web* sémantique, l'internet est en passe de devenir une base de connaissances mondiale. Ce que l'on appelle le « *web*

de données » consiste en effet à interconnecter d'immenses référentiel (terminologies, catalogues d'œuvres, etc.) ouvrant la voie à des usages radicalement nouveaux des données numériques. En effet, les données sont ici structurées et reliées entre elles, ce qui permet de constituer un réseau global d'informations ouvrant des possibilités nouvelles pour le multilinguisme.

Construire le web de données culturelles

D'abord confiné au domaine de la recherche, le *web* de données s'ouvre progressivement à des contenus intéressant le plus grand nombre, notamment ce qui concerne les données culturelles. L'Institut national de recherche en informatique et en automatique (INRIA) a proposé au ministère de la Culture et de la Communication un projet inédit de constitution d'une version entièrement sémantisée de l'encyclopédie *Wikipédia* francophone. Ce projet de développement, propre à assurer la présence

de la langue française dans le *web* de données, va dans le sens de la politique développée par le ministère de la Culture en valorisant le maillage culturel des territoires français et de la francophonie et la prise en compte des expressions artistiques émergentes. Il va par ailleurs dans le sens de la politique de l'État, qui est de développer l'accès aux données ouvertes « *Open data* » et de s'impliquer davantage dans les réseaux de données structurées ou « *Linking Open Data (LOD)* ».

Le rôle de Wikipédia

Wikipédia, projet collaboratif d'encyclopédie multilingue mondialement connu, est un objet d'étude particulièrement enrichissant pour qui s'intéresse à la présence des langues sur l'internet. L'encyclopédie excelle en effet dans le domaine du multilinguisme et compte 38 déclinaisons linguistiques comportant plus de 100 000 pages. On y trouve aussi des centaines de déclinaisons dans des langues peu fréquentes sur l'internet, ou encore dans des langues régionales comme le breton, le

basque ou le tahitien. La langue française, quant à elle, se place au troisième rang des versions linguistiques de l'encyclopédie, après l'anglais et l'allemand, et à la seconde place dans le *Wiktionnaire*, un dictionnaire collaboratif. Ces deux outils sont une formidable opportunité de développement et de rayonnement pour la langue et la culture françaises, mais aussi pour les langues de France.

Wikipédia est un projet d'encyclopédie multilingue sur internet, rédigée par les internautes et publiée sous licence libre. Consultée par plus de 16 millions de Français différents chaque mois ou 1,8 million de Français différents chaque jour, *Wikipédia*

est le 8^e site le plus visité en France. Créée en 2001, la version francophone compte actuellement 1 200 000 articles, ce qui la place au 3^e rang derrière l'anglophone et la germanophone.

Glossaire

Web sémantique

Le terme de *Web* sémantique a été proposé par Tim Berners Lee en 2001 pour désigner une évolution de l'internet qui permettrait aux données d'être plus facilement utilisables et interprétables par des logiciels afin que ceux-ci puissent opérer des raisonnements élaborés. Les technologies du *Web* sémantique permettent de créer de telles données, de constituer des vocabulaires et les règles qui les décrivent, et de bâtir des systèmes capables de les manipuler dans de bonnes conditions d'interopérabilité.

Web de données

Le *Web* de données (*Linked Data*) vise à favoriser la publication de données structurées sur la toile, non pas sous la forme de banques de données isolées les unes des autres, mais en reliant les données entre elles pour constituer un réseau global d'informations.

Ontologie

C'est l'ensemble structuré des termes et concepts que constitue un champ d'informations. Elle constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Son objectif premier est de modéliser un ensemble de connaissances dans un domaine donné, comme, par exemple, le nucléaire.

Internet collaboratif

Données, pages ou sites internet issus d'un travail collaboratif, c'est-à-dire fondé sur un nouveau mode de collaboration et de mise en réseau des personnes grâce aux technologies de l'information et de la communication. Les outils nouveaux permettent ainsi de maximiser la créativité et l'efficacité d'un groupe associé à des projets d'envergure même s'il est très dispersé dans l'espace et le temps.

Technologies de la langue

D'après le programme européen MetaNET, le terme « technologies de la langue » désigne l'ensemble des technologies informatiques spécialisées dans le traitement du langage humain. D'après cette définition les technologies sont classées en fonction des contenus traités (texte, voix, image...).

Domaine de premier niveau

Dans le système de noms de domaine, un domaine de premier niveau ou un domaine de tête (en anglais *top-level domain*) est un domaine qui est au sommet de la hiérarchie des domaines. Dans un nom de domaine, il est indiqué par le label suivant le dernier point dans le nom du domaine (exemple : dans *dglf.culture.gouv.fr*, le domaine de premier niveau est *fr*).

Sources :

Wikipédia, Bibliothèque nationale de France.

De nombreuses initiatives en faveur de la diversité linguistique

Forte de ces constats, la Délégation générale à la langue française et aux langues de France soutient plusieurs actions menées par des associations œuvrant pour la diversité linguistique dans le cyberspace :

> **Wikimédia France**, qui encourage la constitution d'encyclopédies et de dictionnaires en langues de France, notamment dans l'outre-mer, dans le but de mieux valoriser le patrimoine inestimable que peuvent constituer ces langues. L'association a publié récemment deux rapports détaillés sur le français et sur les langues de France dans les projets *Wikipédia*, dont les principaux résultats figurent dans le *Rapport au Parlement sur l'emploi de la langue française* 2011.

> **Le réseau Maaya**, réseau mondial pour la diversité linguistique, qui sert de plateforme d'échanges et de partage sur ces questions et assure la modération du sous-thème « diversité linguistique » du plan d'action du Sommet mondial pour la société de l'information (SMSI).

> **L'association Diversum**, qui œuvre pour la prise en compte de l'environnement culturel dans les politiques de développement durable et qui a notamment su développer une expertise et une méthodologie propres en termes d'évaluation du degré de plurilinguisme des sites internet. L'association a réalisé en 2011 un rapport exhaustif consacré au multilinguisme sur les sites internet de l'État en France et en Europe, figurant lui aussi dans le *Rapport au Parlement* 2011.

> **L'association française pour le nommage internet** (AFNIC) assure la gestion des noms de domaine de premier niveau français *.fr* et de certaines extensions de l'outre-mer telles que *.re* et *.pm*. L'AFNIC a entrepris d'ouvrir le dépôt de noms de domaines sur ces extensions à tous les caractères utilisés par le français et les langues de France : il sera dès lors bientôt possible d'utiliser des caractères accentués dans un nom de domaine en *.fr*.

Les technologies de la langue et les ressources linguistiques

Outre les réseaux numériques, un autre facteur d'influence est apparu en même temps que le passage au numérique et se révèle d'une importance capitale pour la langue et le multilinguisme : les technologies de la langue. On entend par technologies de la langue toutes les techniques permettant de manipuler, transformer, synthétiser la langue, qu'elle soit écrite, orale, ou audiovisuelle.

Les technologies les plus connues permettent notamment la traduction dite « automatique » et la synthèse vocale, mais il y a bien d'autres applications : aide à la traduction, sous-titrage automatique, aide à la rédaction (correcteurs orthographiques et grammaticaux), reconnaissance vocale, indexation automatique de documents...

Ces nouvelles possibilités offertes par le numérique contribuent à l'amélioration de la vie quotidienne de nos concitoyens, au développement de notre économie et au renforcement de nos échanges. Elles permettent notamment une meilleure intercompréhension entre les populations parlant différentes langues et elles peuvent, à ce titre, jouer un rôle fondamental au sein de l'Union européenne, ou encore dans l'outre-mer

français, qui compte un grand nombre de langues maternelles.

Surtout, elles constituent un enjeu culturel et linguistique considérable : en effet, plus une langue est facilement accessible, traduisible et manipulable au moyen de ces technologies, plus elle sera couramment utilisée sur les réseaux numériques. Il importe donc de s'assurer de la bonne place du français et des langues de France dans le développement de ces technologies, en soutenant la recherche, en favorisant les partenariats public-privé, en participant aux travaux et aux études financés par la Commission européenne.

Enfin, les technologies de la langue posent de nombreuses questions de société, car elles influencent profondément les usages de la langue. Quel est l'avenir de la traduction « humaine » face à la traduction automatique ? Et comment vont évoluer les métiers liés à la traduction ? Qu'en est-il pour le sous-titrage ? Les technologies de la langue nous affranchiront-elles de choisir des mots clés, de faire des résumés, des ontologies à l'avenir ? La reconnaissance vocale nous affranchira-t-elle de l'écrit ?

Sitographie

culture.fr

Le site de la Culture.

dismoidixmots.culture.fr

Le site de la *Semaine de la langue française et de la Francophonie* et de l'opération « Dis-moi dix mots qui te racontent ».

wikilf.culture.fr

Le site participatif pour l'enrichissement de la langue française.

cblog.culture.fr

Le blog de la culture et du numérique.

etalab.gouv.fr

Le blog de la mission Étalab.

data.gouv.fr

La plateforme française d'ouverture des données publiques.

afnic.fr

Site internet de l'Association française pour le nommage internet.

fr.wikipedia.org

Encyclopédie en ligne *Wikipédia*.

fr.wiktionary.org

Dictionnaire contributif en langue française.

diversum.net

Le site de l'Association diversum.

zigzag-francophonie.eu

Une fenêtre ouverte sur le monde de la francophonie.

maaya.org

Le site du réseau mondial pour la diversité linguistique, maaya

technolangue.net

Le portail des technologies de la langue.

Retrouvez le *Rapport au Parlement sur l'emploi de la langue française* en ligne sur le site de la Délégation générale à la langue française et aux langues de France : dglf.culture.gouv.fr



wikiLF

Participez à l'enrichissement de la langue française

Le ministère de la Culture et de la Communication a lancé en 2011 un outil collaboratif pour associer les internautes à l'évolution de la langue française : le wikiLF.

wikilf.culture.fr

Ce site internet permet à tous les internautes de donner leur avis sur le choix d'un nouveau terme pour désigner une notion en français, de proposer des nouveaux termes avec leurs définitions, ou encore de faire des commentaires sur les contributions, les avis et les suggestions des autres participants.