



• **Le Forum** des droits
sur l'internet

RECOMMANDATION

Internet et développement durable II : Langues et internet

22 décembre 2009

www.foruminternet.org

Le Forum des droits sur l'internet
6. rue Déodat de Séverac 75017 PARIS

Introduction	5
I. – Percevoir la langue comme un enjeu stratégique du numérique	9
A – Le multilinguisme, un enjeu au carrefour de préoccupations multiples	9
1. Un enjeu politique pour le rayonnement d'un pays	9
a. Les sources écrites du droit	9
b. Les documents scientifiques	10
c. Les publications institutionnelles et documents de nature officielle	11
2. Un enjeu commercial pour les acteurs économiques	12
a. Une question commerciale ou marchande	12
b. Une question d'image	12
c. Améliorer la protection du consommateur	12
B – Mettre en place un dispositif public solide au service de la diversité linguistique	13
1. Définir une stratégie linguistique au niveau national	13
a. Améliorer l'accessibilité linguistique des ressources critiques	13
b. Le cas des marchés publics	14
c. Mieux mesurer la diversité linguistique	16
d. Promouvoir la labellisation des sites internet selon des critères linguistiques et culturels	16
2. Valoriser l'effort national de traduction	16
a. Les services publics de la traduction	17
b. Exemple étranger : l'Allemagne	17
c. Les propositions d'évolution	18
3. Conforter le cadre juridique français de la langue	19
a. Le cadre juridique de la langue française	19
b. Proposition complémentaire : un dispositif d'incitation fiscale pour le multilinguisme ?	21
C – Valoriser l'apport des technologies numériques de la langue	21
1. État de l'art des technologies numériques linguistiques et perspectives	22
a. Quelles sont les possibilités techniques en matière d'assistance linguistique aujourd'hui et que peut-on en attendre ?	22
b. Quelles sont les perspectives d'évolution ?	23
2. Améliorer la recherche sur les technologies de la langue et leur usage	24
a. Renforcer la dotation des pouvoirs publics aux programmes de recherche français	24

- b. Améliorer la visibilité de l'action européenne en termes de technologies de la langue et d'aide à la traduction 26
- c. Produire davantage de ressources linguistiques 26
- d. Mesurer les performances linguistiques des technologies de la langue 27
- e. Des technologies utiles pour les personnes en situation de handicap 27
- 3. Les technologies de la langue et l'enseignement 28
 - a. L'apport des technologies de la langue 28
 - b. Enseignement de la langue et internet 29

D – Mieux accompagner l'évolution de la langue sur internet ; le cas du français en France 29

- 1. L'évolution et la vitalité de la langue française 30
 - a. Quelle est la présence de la langue française sur l'internet ? 30
 - b. Internet : un moteur favorisant les échanges écrits mais qui provoque aussi une mutation de la langue 31
 - c. Les technologies de l'information et l'orthographe française 33
- 2. La création et la diffusion des nouveaux termes du français de l'internet en France 34
 - a. Les besoins des acteurs 34
 - b. Le dispositif public existant 34
- 3. L'accessibilité de la langue sur internet 37
 - a. La langue permet l'accès à l'information... 37
 - b. ...mais il est nécessaire d'adapter le niveau d'accessibilité linguistique à son public 38

II. – La mise en œuvre de bonnes pratiques linguistiques sur internet 40

A – Les normes techniques et la normalisation autour de la langue sur internet 40

- 1. La traduction des messages d'erreur des serveurs internet 40
- 2. Uniformisation du répertoire d'étiquetage des langues utilisé sur l'internet 40
- 3. Les noms de domaine en caractères étendus sur les domaines de tête nationaux 41
 - a. Les normes techniques, matériels et logiciels sont-ils prêts ? 41
 - b. Enjeux du nommage en Unicode sur les extensions nationales de premier niveau 42
- 4. Améliorer l'accès aux ressources linguistiques et notamment aux dictionnaires 43
- 5. Le clavier français est-il adapté à la langue française dans un contexte fortement multilingue ? 43
- 6. L'utilisation des normes d'encodage les plus récentes 44

B – Comment gérer et mettre en œuvre le multilinguisme sur internet ?	45
1. Gérer un projet internet multilingue	45
a. Comment organiser un service de traduction ?	45
b. La distinction entre les contenus statiques et les contenus périodiques (actualités)	46
c. Gérer une communauté de développeurs multinationaux	46
d. Quelles langues choisir ?	47
2. Pour une utilisation raisonnée des technologies linguistiques	47
a. Pour la traduction de contenus	47
b. Pour d'autres usages	48
3. Comment utiliser au mieux l'étiquetage et le marquage linguistique des pages internet ?	49
a. Le marquage des pages internet	49
b. Le marquage des liens internet	49
c. La détection automatique de la langue du visiteur	50
4. La matérialisation de la langue sur le site internet	51
a. La représentation de la langue sur les pages du site	51
b. La neutralité de la page d'accueil	52
c. L'encodage de caractères	53
d. Langue, nom de domaine et <i>URL</i>	54
e. Traduire les pages présentant des erreurs techniques	55
ANNEXE	56
Annexe: Composition du groupe de travail	57

INTRODUCTION

Dans la mesure où l'accès à l'internet devient progressivement un droit garanti et opposable à tous les citoyens¹ et qu'il permet désormais l'accès à l'information et à l'emploi, il ne faudrait pas que les évolutions technologiques soient l'occasion d'augmenter les freins à l'insertion, que ce soit à cause du handicap ou des difficultés liées à la maîtrise de la langue.

Le Forum des droits sur l'internet a installé en mai 2008 un groupe de travail sur le thème « internet et développement durable »² afin d'examiner de façon transversale les problématiques de développement durable qui peuvent se poser dans la société de l'information.

Ce cadre reste encore à explorer ; il a été largement ignoré lors du Grenelle de l'environnement³ qui s'est déroulé en France de mai⁴ à octobre 2007, mais a donné lieu récemment à un rapport significatif sur le sujet, « Renforcer la contribution des technologies du numérique au développement durable », qui a été remis à Mme Christine Lagarde, ministre de l'Économie, et à M. Christian Estrosi, ministre chargé de l'Industrie, le 8 septembre 2009.

Dans ce contexte, le Forum des droits sur l'internet a rendu une première Recommandation, adoptée le 25 novembre 2008 par son Conseil d'orientation, invitant les pouvoirs publics à mettre en place un dispositif incitatif pour l'accessibilité numérique⁵.

L'accessibilité des contenus de l'internet à tous passe non seulement par un respect attentif des différentes normes techniques et des recommandations pour l'accessibilité, mais aussi par une gestion linguistique des contenus adaptée aux publics qui sont visés. C'est cette recherche qui a fondé la deuxième partie des réflexions du groupe de travail « internet et développement durable ».

Culture et développement durable ?

Les problématiques de développement durable sont progressivement entrées au cœur de la pensée collective : d'abord cantonnée aux aspects purement « écologiques », c'est-à-dire aux interactions entre les milieux physiques et les organismes vivants⁶, cette réflexion s'est ensuite rapidement portée sur les sphères sociales et économiques et plus récemment culturelles, formant ainsi les quatre « piliers » du développement durable. La création, en France, en 2007, d'un ministère d'État en charge du Développement durable atteste, si besoin en était, de cet intérêt croissant des pouvoirs publics et de la société en général pour ces réflexions.

¹. Sur l'idée d'un « droit opposable » à tous les citoyens en matière d'accès à internet en haut débit, consulter le discours de M. Éric Besson, Secrétaire d'État en charge du Développement de l'économie numérique, prononcé le 20 octobre 2008 à l'Élysée à l'occasion de l'annonce du plan « France Numérique 2012 » :

http://www.elysee.fr/documents/index.php?cat_id=7

². Voir la page de présentation du groupe de travail sur le site du Forum des droits sur l'internet :

http://www.foruminternet.org/spip.php?action=redirect&id_article=2686

³. Voir sur ce sujet le document récapitulatif des tables rondes finales du Grenelle de l'environnement :

http://www.legrenelle-environnement.gouv.fr/grenelle-environnement/IMG/pdf/07-11-23_GE_TR_recap_bis.pdf

⁴. Le Grenelle de l'environnement a été annoncé par Alain Juppé, ministre de l'Écologie, du Développement et de l'Aménagement durable, le 18 mai 2007.

⁵ http://www.foruminternet.org/spip.php?action=redirect&id_article=2809&var_mode=calcul

⁶. Charles J. Krebs, *Ecology*, Benjamin Cummins, 5^e éd., 2001.

Pour beaucoup, la signification de cette expression reste pourtant floue : qu'entend-on exactement par « développement durable » ? Selon la première définition⁷ proposée en 1987 par la Commission mondiale sur l'environnement et le développement durable, il s'agit d'un « *développement qui répond aux besoins des générations du présent sans compromettre la capacité des générations futures à répondre aux leurs* »⁸.

Il s'agit donc d'un concept très vaste, qui trouve une application dans tous les domaines de nos sociétés. Comme il a été dit, la sphère culturelle en est une des premières déclinaisons : dans cette optique, il s'agit de préserver la diversité des solutions locales afin d'assurer la continuité culturelle.

Ainsi, le Sommet mondial sur la société de l'information (SMSI), institué par les Nations unies⁹, a dès le départ placé la culture comme vecteur essentiel du développement durable de la société de l'information. Les problématiques de diversité et d'identité culturelle, et de diversité linguistique notamment, ont été au cœur de la réflexion menée par les participants.

Le considérant 54 de la déclaration de principes de Genève¹⁰ indique ainsi que « *la préservation du patrimoine culturel constitue une composante fondamentale de l'identité et de la compréhension de soi qui relie une communauté à son passé. La société de l'information devrait mettre en valeur et préserver le patrimoine culturel pour les générations futures, par toutes les méthodes appropriées, y compris la numérisation.* »

Le Forum international pour le multilinguisme¹¹, qui s'est tenu à Bamako en janvier 2009, a rappelé la nécessité d'un engagement pour le multilinguisme universel, en améliorant la sensibilisation internationale sur ces sujets, en favorisant l'éducation multilingue, et en œuvrant pour un cyberspace multilingue.

La culture constitue donc un pilier fondamental du développement durable de la société de l'information, dans le but de satisfaire à deux objectifs : la préservation du patrimoine, d'une part, et son enrichissement, d'autre part.

Il est important de souligner que si la culture constitue un enjeu en soi du développement durable, elle est aussi une problématique transversale : l'absence de prise en compte de la dimension culturelle dans le développement durable constituerait un frein évident pour l'obtention de résultats tangibles en matière écologique, économique ou sociale.

Pourquoi s'intéresser à la langue ?

La question de l'environnement culturel étant cruciale pour une politique de développement durable, la langue constitue assurément un sujet de préoccupation prioritaire, puisqu'elle est la première des expressions culturelles – un constat renforcé au sein de l'univers numérique.

⁷. La définition complète est la suivante : « *Un développement qui répond aux besoins des générations du présent sans compromettre la capacité des générations futures à répondre aux leurs. Deux concepts sont inhérents à cette notion : le concept de "besoins", et plus particulièrement des besoins essentiels des plus démunis, à qui il convient d'accorder la plus grande priorité, et l'idée des limitations que l'état de nos techniques et de notre organisation sociale impose sur la capacité de l'environnement à répondre aux besoins actuels et à venir.* »

⁸. Rapport dit « Brundtland » intitulé Notre avenir à tous, soumis à l'Assemblée générale des Nations unies en 1986 par la Commission mondiale sur l'environnement et le développement.

⁹. Résolution 56/183 du 21 décembre 2001 de l'Assemblée générale de l'organisation des Nations unies.

¹⁰. Déclaration de principes de Genève (10-12 décembre 2003), en ligne sur le site des Nations unies : http://www.itu.int/wsis/documents/doc_multi.asp?lang=fr&id=116110

¹¹. Voir le détail du programme du Forum international de Bamako (Mali) de janvier 2009 : <http://www.acalan.org/fr/confeven/forum/forum.php>

Le medium écrit prend en effet avec le numérique un nouvel essor. Alors que les médias apparus progressivement au XX^e siècle (radio, télévision, cinéma) laissaient une faible part à l'écrit, l'internet renverse la donne et consacre la langue écrite comme vecteur fondamental d'échange d'informations entre les individus, et, de par son caractère mondial, entre les nations, les peuples et les cultures.

Ce renouveau de la langue écrite soumet celle-ci à de fortes pressions : tandis que certaines langues, bien maîtrisées par les populations « pionnières » de l'internet, ont connu une forte expansion, d'autres langues risquent au contraire de se retrouver écartées du réseau. Le problème peut être aggravé par des circonstances de nature technique ne facilitant pas l'expression linguistique dans telle ou telle langue : clavier non adapté, choix par des éditeurs de logiciels de normes de codage obsolètes ne permettant pas l'emploi de jeux de caractères différents...

Enfin, les usages particuliers de la langue écrite sur l'internet peuvent avoir un impact significatif sur celle-ci : ainsi, les applications « en temps réel » de l'internet, telles que les messageries instantanées, les réseaux sociaux, les forums et salons de discussion, qui se sont fortement développés ces dernières années, imposent une communication brève et rapide, ce qui est une situation relativement nouvelle. Les usagers ont donc naturellement « adapté » leurs utilisations des mots et des règles de grammaire, afin de pouvoir s'exprimer aussi vite que possible, tout en restant compris par leurs interlocuteurs.

Le défi qui est aujourd'hui posé aux langues, et ainsi à la langue française, est de savoir s'adapter aux technologies, tout en conservant leur richesse et leur intelligibilité.

Il importe par conséquent de s'assurer de l'évolution constante de la langue, en adéquation avec l'évolution des technologies de communication, et de réfléchir au développement pérenne de la langue, afin que les contenus demeurent accessibles par les générations futures.

Par ailleurs, compte tenu du caractère mondial du réseau internet, la diffusion de contenus sur le réseau pose la question de la traduction des documents dans de multiples langues, pour rendre accessibles ces contenus à un public toujours plus étendu.

Le multilinguisme est un important vecteur de communication, d'échange et de dialogue entre les cultures et les peuples. « *On ne saurait souligner assez l'importance des langues pour réaliser la cohésion sociale et lever les obstacles à l'intégration* », déclarait M. Léonard Orban, commissaire européen chargé du multilinguisme, lors de l'inauguration en octobre 2009 d'une plate-forme internet en faveur du multilinguisme¹².

Le Conseil de l'Union européenne, sous la présidence française, a adopté une résolution¹³ le 21 novembre 2008 visant à réfléchir à la mise en place d'un programme spécifique de soutien à la traduction, à la hauteur de ses enjeux culturels, économiques et technologiques, afin de remédier au manque de visibilité et à la faiblesse de l'action communautaire dans le domaine de la traduction.

La Recommandation du Forum des droits sur l'internet

Ce document se limitera aux points les plus sensibles liés aux questions linguistiques dans le cadre de l'internet, **à ses enjeux juridiques et stratégiques, notamment pour les pouvoirs publics (I), et aux bonnes pratiques qui en découlent pour l'ensemble des acteurs de l'internet (II).**

¹². Voir le communiqué de presse en date du 23 octobre 2009 sur le site europa.eu : <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/1574&format=HTML&language=FR>

¹³. Résolution du Conseil de l'Union européenne du 21 novembre 2008 (PDF) : http://www.consilium.europa.eu/uedocs/cms_Data/docs/pressdata/fr/educ/104231.pdf

Le Forum a souhaité proposer **aux pouvoirs publics** des recommandations opérationnelles sur l'organisation du cadre juridique français et des services de l'administration pour une meilleure prise en compte des problématiques linguistiques. Il propose ensuite **aux acteurs privés** une série de bonnes pratiques à suivre en matière de multilinguisme en ligne et de respect des normes techniques de l'internet, à partir des débats et des travaux du groupe de travail « internet et développement durable » qui se sont déroulés au Forum entre février et septembre 2009.

L'approche retenue par le Forum des droits sur l'internet dans ces travaux se veut **résolument universelle**, et non pas focalisée sur le cadre national de la langue, ou sur la défense de la seule langue française. Les problématiques qui y sont abordées et les recommandations qui en découlent ont donc vocation à nourrir la réflexion collective sur la langue bien au-delà de nos frontières, dans le cadre mondial que constitue l'internet.

Cette Recommandation a fait l'objet d'une consultation auprès de l'ensemble des membres du Forum des droits sur l'internet du 1^{er} décembre 2009 au 15 décembre 2009. Elle a été définitivement adoptée par le Conseil d'orientation du Forum le 22 décembre 2009.

I. – PERCEVOIR LA LANGUE COMME UN ENJEU STRATÉGIQUE DU NUMÉRIQUE

La langue écrite connaît, depuis l'ère du numérique et l'essor de l'internet, un renouveau considérable. Elle est redevenue le vecteur de communication et d'échange majeur entre les personnes et les peuples. C'est pourquoi, pour tout acteur présent sur l'internet, la langue est un enjeu au carrefour de préoccupations multiples (A). Dans ce cadre, il importe donc que les pouvoirs publics mettent en place un dispositif solide au service de la diversité linguistique (B). L'apport des technologies numériques a, quant à lui, profondément modifié la façon d'appréhender l'enjeu de la langue ; il doit être pris en compte, mieux utilisé par les décideurs et davantage valorisé (C). Enfin, se pose la question de l'évolution et de l'accompagnement de la langue française dans le contexte de l'internet (D).

A – Le multilinguisme, un enjeu au carrefour de préoccupations multiples

Lorsque l'on aborde le sujet de la langue, il est légitime de s'interroger sur l'intérêt de réaliser des traductions et de communiquer en plusieurs langues.

Traduire des documents en plusieurs langues et mettre en œuvre le multilinguisme répond en premier lieu à un objectif très concret : rendre ses contenus plus **accessibles**, en donnant la possibilité à une population plus importante de les lire.

Traduire, c'est également favoriser la **diversité culturelle** et faciliter les échanges entre les cultures. L'internet permet par exemple de visiter un musée à l'autre bout du monde, de percevoir le regard sur le monde d'un étranger à travers son blog de photographies et de commentaires, ou encore de discuter en ligne avec des personnes du monde entier.

Mais la langue constitue aussi **pour un État** un enjeu politique d'ordre stratégique pour son rayonnement dans le monde (1), et, **pour les entreprises**, un enjeu commercial et d'image significatif (2).

1. Un enjeu politique pour le rayonnement d'un pays

Le multilinguisme peut permettre à un État d'accroître son rayonnement dans le monde¹⁴, et de diffuser à une plus grande diversité de locuteurs son patrimoine scientifique, juridique ou culturel. Au niveau national, plusieurs types de documents apparaissent en effet particulièrement stratégiques, et pourraient bénéficier d'une plus grande ouverture vers des langues étrangères.

a. Les sources écrites du droit

Une grande partie du droit écrit français est accessible directement sur l'internet : le site Légifrance¹⁵ diffuse ainsi l'ensemble des codes législatifs, des textes réglementaires, ainsi qu'une part importante des décisions significatives rendues par les juridictions françaises.

¹⁴. « Le ministre de la culture et de la communication tient à rappeler que le multilinguisme est une dimension clé du rayonnement de la France à l'étranger », question ministérielle n° 46 914 et réponse ministérielle publiée au Journal officiel le 17 novembre 2009 :

<http://questions.assemblee-nationale.fr/q13/13-46914QE.htm>

¹⁵. Voir le site Légifrance, le site public de la diffusion du droit :

<http://www.legifrance.gouv.fr/>

Un travail de traduction a été mené sur les principaux codes publiés sur Légifrance : est disponible la traduction en texte intégral de 11 codes français, et de la loi sur la sous-traitance de 1975, en anglais¹⁶, et en espagnol¹⁷ (castillan). Ces traductions datent cependant de 2006, et n'ont pas été mises à jour depuis lors.

La traduction des normes juridiques écrites est particulièrement stratégique, car elle permet **de diffuser le droit français**, de tradition romano-germanique, dans le monde, et par ce fait d'inspirer parfois largement des systèmes de droit étrangers en cours de refonte.

Elle permet également aux investisseurs d'obtenir davantage d'informations sur le cadre juridique local (par l'intermédiaire d'un conseil juridique généralement), ce qui peut faciliter l'implantation de structures ou de filiales dans le pays.

Le travail de traduction est cependant rendu délicat par la spécificité de certains principes propres au droit écrit romano-germanique, et nécessite des ressources importantes pour parvenir à une traduction la plus proche possible de l'original. Dans certaines situations, il pourra donc être plus pertinent de rédiger un résumé fidèle des principales dispositions normatives, à l'attention des lecteurs étrangers.

En outre, il faut signaler le fait que la norme, même traduite dans les meilleures conditions possibles, n'aura jamais force de loi, et ne pourra être appliquée ou revendiquée comme telle devant une juridiction. C'est là une limite particulière de la traduction de documents normatifs en droit français¹⁸.

Le Forum recommande aux pouvoirs publics d'accentuer leur effort en termes de traduction de normes juridiques, dans une double optique d'attractivité économique et de diffusion du droit continental.

b. Les documents scientifiques

La traduction de publications scientifiques constitue, elle aussi, une activité particulière, et à fort enjeu stratégique et politique. La traduction des résultats issus de la recherche peut leur assurer une meilleure visibilité à l'étranger, ainsi qu'une plus grande notoriété. Elle assure **la promotion de la pensée scientifique nationale à l'étranger**.

L'importance d'une politique de la traduction en matière scientifique est d'autant plus à souligner que de nombreux linguistes mettent en évidence le lien entre créativité et création dans sa propre langue, d'où l'intérêt de produire une recherche originale dans la diversité des langues du monde.

Toutefois, la traduction des publications scientifiques se situe dans une catégorie à part de la traduction : ces documents nécessitent en effet d'importantes ressources humaines pour parvenir à une traduction fidèle à l'original, et la collaboration étroite des traducteurs avec les chercheurs.

En France, la traduction de certains de ces documents devrait être assurée en partie par le financement du Conseil pour la création artistique¹⁹, sous le pilotage de M. Guy

¹⁶. Le droit positif français disponible en anglais :

<http://195.83.177.9/code/index.phtml?lang=uk>

¹⁷. Le droit positif français disponible en espagnol :

<http://195.83.177.9/code/index.phtml?lang=esp>

¹⁸. Dans des pays voisins de la France (la Suisse par exemple), plusieurs langues peuvent être reconnues simultanément pour l'applicabilité des normes, ce qui rend bien sûr d'autant plus cruciale la concordance entre les différentes versions. Pour certaines normes, la Confédération suisse pratique ainsi non pas la traduction mais la co-rédaction.

¹⁹. Le Conseil pour la création artistique a été créé par le décret n° 2009-113 du 30 janvier 2009. Il est présidé par le président de la République et actuellement animé par Marin Karmitz. Sa mission est « d'éclairer les choix des pouvoirs publics en vue d'assurer le développement et l'excellence de la création »

Walter²⁰. On ne sait pas cependant s'il s'agira de traduire en français des résultats de recherches étrangères ou des résultats de recherches françaises publiées en anglais ou encore de traduire en (bon) anglais des recherches françaises afin de leur donner de la visibilité.

c. Les publications institutionnelles et documents de nature officielle

La communication institutionnelle et la documentation officielle des pouvoirs publics répondent à des exigences plus hautes en termes de qualité de la traduction : est en jeu l'image de l'État vis-à-vis de ses administrés, mais aussi vis-à-vis des pays étrangers.

L'absence totale de traduction sur les sites institutionnels de l'État peut aussi renvoyer **une image négative** aux yeux des étrangers, qui souhaiteraient mieux connaître le fonctionnement du pays par l'intermédiaire de ses sites internet officiels. Contrairement à ses voisins européens²¹, les sites officiels de l'exécutif français²² ne sont plus traduits²³.

Le Forum recommande aux pouvoirs publics d'améliorer leur effort de traduction pour les sites internet officiels du pouvoir exécutif : elysee.fr et gouvernement.fr.

En matière de traduction, il faut aussi prendre en compte la qualité. Si une traduction automatique qui n'est pas retravaillée ne peut pas être de qualité, le contraire n'est en revanche pas aussi simple : une traduction faite par des traducteurs n'est pas forcément de bonne qualité. S'il est facile de détecter les erreurs de style, il sera moins évident d'ajuster le fond de la traduction, de sorte qu'elle soit parfaitement en ligne avec le document original.

Il convient donc de rappeler qu'il existe de nombreuses normes internationales définissant la qualité d'un travail de terminologie ou de traduction²⁴, pour la plupart issues du comité technique ISO/TC 37²⁵.

artistique française, de promouvoir sa diffusion la plus large, notamment internationale, et d'arrêter les orientations de nature à permettre leur mise en œuvre ».

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020176855&fastPos=1&fastReqId=2119219138&categorieLien=cid&oldAction=rechTexte>

²⁰. « [Le Conseil] financera la traduction de textes scientifiques français. Guy Walter, patron de la Villa Gillet, à Lyon, pilotera le projet », Les dix travaux de Marin Karmitz pour la culture, Le Monde, édition du 11 septembre 2009, page 22.

²¹. Voir par exemple le site internet du gouvernement allemand, traduit intégralement en français et en anglais :

<http://www.bundesregierung.de/Webs/Breg/FR/Homepage/home.html>

²². Voir notamment le site de la présidence de la République, et celui du gouvernement :

<http://www.elysee.fr>

<http://www.gouvernement.fr/>

²³. Les précédentes versions d'elysee.fr et de premier-ministre.gouv.fr étaient également disponibles l'une et l'autre en allemand, anglais, espagnol ; il s'agissait pour le premier de contenus statiques et pour le second de contenus dynamiques, régulièrement mis à jour. Jusqu'en 2004, premier-ministre.gouv.fr s'affichait en allemand, anglais et espagnol, mais aussi en italien et en portugais, le tout dans des versions dynamiques. Cela représentait une marque de courtoisie et d'estime indéniable, au moins à l'égard des pays européens les plus proches.

L'Association diversum, qui se charge notamment d'observer le comportement linguistique des vingt à vingt-cinq principaux sites publics dans chacun des pays de l'Union européenne et en Suisse, relève pour la France une régression des indicateurs, que ce soit pour la moyenne de versions linguistiques par site (passée de 2,11 au 30/06/2008, à 2,05 au 31/12/2008, puis à 1,95 au 30/06/2009) ou pour le pourcentage de sites multilingues, c'est-à-dire disponibles en au moins trois langues (26,32 % au 30/06/2008, 25 % au 31/12/2008 et 21,05 % au 30/06/2009). La baisse est encore plus marquée si l'on s'en tient au multilinguisme dynamique, c'est-à-dire avec mise à jour fréquente : 15,79 % au 30/06/2008, 15 % au 31/12/2008 et 5,26 % au 30/06/2009). Par comparaison, sur ces mêmes indicateurs, au 30/06/2009, l'Allemagne affiche 2,29 versions linguistiques en moyenne, 35,29 % de sites multilingues et 29,41 % des sites en situation de multilinguisme dynamique.

²⁴. Voir notamment la norme EN 15038 sur les exigences requises pour les prestations de services de traduction :

Le Forum recommande aux pouvoirs publics de ne mettre en ligne que des traductions de qualité sur ses sites officiels, de façon à ne pas dévaloriser l'image de l'État français aux yeux de la communauté internationale. Il est par exemple fortement déconseillé de recourir à des traductions automatiques qui ne seraient pas retravaillées par un service de traduction²⁶.

En outre, notons que l'intérêt de mettre en ligne des traductions réalisées uniquement par le biais d'outils automatiques est faible car les utilisateurs peuvent eux-mêmes, s'ils le souhaitent, recourir à des traductions automatiques au moyen d'outils gratuits²⁷.

2. Un enjeu commercial pour les acteurs économiques

a. Une question commerciale ou marchande

Le multilinguisme constitue, pour les entreprises, **un enjeu commercial** : un site marchand traduit en langue étrangère pourra ainsi toucher une clientèle plus internationale et permettra d'augmenter de façon significative le chiffre d'affaires réalisé.

Le souhait pour un acteur économique de traduire ou non un site internet doit donc être mis en relation avec sa **politique de développement à l'international**.

b. Une question d'image

La traduction et le multilinguisme affiché contribuent aussi, de façon plus indirecte, à améliorer **la notoriété** d'un acteur présent sur l'internet : les visiteurs, même s'ils n'utiliseront pas toutes les langues proposées, apprécieront l'effort de traduction, et cet effort pourra être considéré comme un gage de travail sérieux ou de qualité et d'ouverture sur le reste du monde.

c. Améliorer la protection du consommateur

La traduction est aussi un moyen de **mieux informer le consommateur sur ses droits**, c'est pourquoi la traduction des textes spécifiques à destination du consommateur doit faire l'objet d'une attention toute particulière.

Le Forum des droits sur l'internet souhaite attirer l'attention des exploitants de sites internet commercialisant des biens en France et bénéficiant d'une version française de leur site sur la nécessité de traduire les conditions générales de vente aussi finement que possible, en s'assurant de l'équivalence juridique des stipulations contractuelles.

Il est par ailleurs recommandé de mettre à jour l'ensemble des versions linguistiques des conditions générales de vente dès lors qu'une modification est faite sur le document en langue originale.

http://www.boutique.afnor.fr/NEL5DetailNormeEnLigne.aspx?&nivCtx=NELZNELZ1A10A101A107&aff=461&ts=8410360&CLE_ART=FA137571

Il existe par ailleurs d'autres normes plus spécifiques comme par exemple la norme ISO 12616 relative à la terminographie axée sur la traduction :

<http://www.iso.org/iso/fr/pressrelease.htm?refid=Ref828>

²⁵. Comité technique ISO/TC 37, Terminologie et autres ressources linguistiques.

²⁶. En dehors de certains cas plus spécifiques où les outils automatiques peuvent être suffisants, comme par exemple la traduction des bulletins météorologiques.

²⁷. Voir par exemple *Google Traduction*, qui propose des traductions dans 41 langues et dans 1 640 paires de langues (avril 2009) : http://www.google.fr/language_tools

Il est enfin recommandé aux exploitants de retirer toutes les traductions de conditions générales qui ne seraient plus à jour avec la version originale, celles-ci pouvant induire en erreur les consommateurs.

B – Mettre en place un dispositif public solide au service de la diversité linguistique

La diversité linguistique doit donner lieu, au niveau des pouvoirs publics, à une véritable orientation stratégique (1) et il semble nécessaire d'identifier clairement quels sont les destinataires et les utilisateurs des ressources mises à disposition du public, afin d'améliorer leur accessibilité linguistique. La question de la traduction et de son organisation se pose ensuite, dans la perspective de la mise en œuvre opérationnelle des objectifs stratégiques préétablis (2). Enfin, on doit se demander si le cadre juridique national de la langue est adapté au dispositif souhaité pour la diversité linguistique (3).

1. Définir une stratégie linguistique au niveau national

Au niveau national, il semble indispensable d'accentuer l'effort linguistique sur certaines ressources apparaissant comme particulièrement stratégiques, mais aussi de mieux promouvoir la diversité linguistique.

a. Améliorer l'accessibilité linguistique des ressources critiques

Les normes techniques sont les documents officiels et acceptés comme tels par une communauté de développeurs ou d'utilisateurs pour la réalisation de programmes ou de matériels de nature technique. Elles jouent un rôle fondamental dans le domaine des nouvelles technologies, puisqu'elles assurent le plus large développement des projets. L'accès à ces normes est donc indispensable pour pouvoir participer à des projets internet ou liés aux nouvelles technologies.

Certains organismes diffusent librement²⁸ l'ensemble de leurs normes techniques : c'est le cas notamment du W3C²⁹ (*Web Consortium*), ou encore de l'IETF³⁰ (*Internet Engineering Task Force*). D'autres diffusent les normes techniques librement depuis peu : c'est le cas de l'UIT³¹ (Union internationale des télécommunications).

Enfin, d'autres organismes ne diffusent pas leurs normes techniques librement, c'est le cas notamment de l'ISO³² (Organisation internationale de normalisation). Les normes techniques de cet organisme sont en effet payantes³³, ce qui peut décourager les plus petits acteurs d'utiliser ou de se conformer à celles-ci. Cela est regrettable car l'ISO met

²⁸. La mise en œuvre de certaines normes techniques peut être restreinte par certains brevets. Voir à ce sujet [RIM cache ses brevets à l'IETF](http://www.bortzmeyer.org/rim-cache-ses-brevets.html), Stéphane Bortzmeyer, 23 novembre 2009 : <http://www.bortzmeyer.org/rim-cache-ses-brevets.html>

²⁹. Voir, par exemple, la dernière version des recommandations internationales pour l'accessibilité (*Web Content Accessibility Guidelines* ou WCAG 2.0) : <http://www.w3.org/TR/WCAG/>

Il en va de même pour leur traduction officielle agréée en français :

<http://www.w3.org/Translations/WCAG20-fr/>

³⁰. La page d'accueil du site de l'IETF (en anglais) : <http://www.ietf.org/>

³¹. L'UIT a annoncé officiellement en septembre 2007 la disponibilité libre et gratuite de l'ensemble de leurs normes techniques au format PDF : http://www.itu.int/newsroom/press_releases/2007/21.html

³². Page d'accueil (en français) du site de l'ISO : <http://www.iso.org/iso/fr/home.htm>

³³. Voir par exemple le prix et les modalités de distribution de la norme ISO 639-3 : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39534

pourtant en œuvre, contrairement à d'autres acteurs, une politique de plurilinguisme, le français étant notamment une des langues officielles de l'organisation.

Il existe toutefois une procédure établie permettant de demander, par l'intermédiaire de la commission de normalisation de l'Association française de normalisation (AFNOR), la gratuité des normes techniques relatives aux technologies de l'information et de la communication élaborées dans le cadre joint de l'ISO et de la Commission électrotechnique internationale (CEI)³⁴.

Enfin, au niveau national, certaines normes payantes de l'ISO sont traduites en français par l'AFNOR³⁵ et disponibles librement sur le site de l'organisme français, contrairement aux normes ISO. C'est le cas notamment pour toutes les normes techniques rendues d'application obligatoire en France³⁶.

Le Forum recommande aux pouvoirs publics de renforcer l'effort de traduction des normes techniques internationales en en demandant la traduction, lorsque cela est possible, par les organismes nationaux de normalisation.

Dans certains cas, la traduction n'a pas de valeur officielle, et, en cas de contestation, seule la norme technique officielle peut faire foi. La valeur ajoutée de la traduction de la norme technique est donc plus faible, quoique dans certains cas elle soit l'objet d'importants enjeux, et elle tient en partie à la participation d'experts français et francophones aux commissions élaborant les normes techniques.

Il est recommandé de privilégier la rédaction en plusieurs langues de notes explicatives concises d'une norme technique plutôt que d'en réaliser une traduction intégrale, ce qui devrait permettre de réduire le coût de la traduction de façon significative, tout en permettant aux décideurs de faire référence à la même norme technique.

b. Le cas des marchés publics

Dans le cadre des marchés publics, il est régulièrement fait référence à des normes techniques, notamment lorsque l'appel d'offres touche la sphère des réseaux, des télécommunications, de l'internet.

Cette situation porte préjudice aux plus petits prestataires qui ne disposent pas d'un service de traduction et peuvent difficilement travailler avec des normes techniques qui ne sont pas disponibles librement en français.

Le Forum recommande aux pouvoirs adjudicateurs³⁷ nationaux de veiller à l'emploi de la langue française dans tous les marchés publics, ceci afin de ne pas écarter les entreprises françaises qui ne disposeraient pas des ressources suffisantes pour traduire les documents du marché public en langue française.

³⁴. Page d'accueil de la CEI :
<http://www.iec.ch/about/mission-f.htm>

³⁵. Page d'accueil de l'AFNOR :
<http://www.afnor.org/>

³⁶. Selon l'article 17 du décret n° 2009-697 du 16 juin 2009 relatif à la normalisation et qui entrera en vigueur au 1^{er} janvier 2010, « les normes rendues d'application obligatoire sont consultables gratuitement sur le site internet de l'Association française de normalisation » :
<http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020749979&fastPos=2&fastReqId=622438787&categorieLien=id&oldAction=rechTexte>

³⁷. Au sens de l'article 2 du Code des marchés publics : « Les pouvoirs adjudicateurs soumis au présent code sont : 1° L'État et ses établissements publics autres que ceux ayant un caractère industriel et commercial ; 2° Les collectivités territoriales et les établissements publics locaux. Sauf dispositions contraires, les règles applicables à l'État le sont également à ceux de ses établissements publics auxquels s'appliquent les dispositions du présent code. De même, sauf dispositions contraires, les règles applicables aux collectivités territoriales le sont également aux établissements publics locaux. »

Il est par ailleurs recommandé aux pouvoirs adjudicateurs de faire référence à des normes techniques disponibles librement en français, et d'utiliser le vocabulaire technique français quand cela est possible.

En ce qui concerne les marchés publics européens émis par les différents États membres, il a été constaté un déficit important de traduction des appels à propositions, qui ne sont généralement pas disponibles en langue française, ce qui pénalise fortement les acteurs économiques francophones et peut créer des **distorsions de concurrence**.

L'exemple du site internet CORDIS³⁸ est révélateur des pratiques en la matière : « service communautaire d'information sur la recherche et le développement, source officielle d'information sur les appels de propositions lancés au titre du septième programme-cadre (7e PC) », la page d'accueil du site est traduite en six langues, mais la majorité des autres pages du site sont exclusivement en anglais. Les appels à propositions, même ceux qui visent des technologies de la langue ou des outils de traduction, sont monolingues³⁹.

Il en va de même pour le portail de la direction de la société de l'information⁴⁰ : bien que six langues soient disponibles depuis la barre de navigation, peu de contenus sont disponibles en français. Les appels à propositions qui y figurent sont, quant à eux, presque exclusivement monolingues⁴¹.

Le Forum encourage les institutions de l'Union européenne à améliorer leurs efforts de traduction en matière de marchés publics européens, et de publier, aussi souvent que possible, des traductions multilingues des appels à propositions émis par les différents États membres. Il est recommandé aux pouvoirs publics de relayer cette demande de manière appuyée, de façon à stimuler la croissance économique des acteurs économiques nationaux.

De manière plus générale, en ce qui concerne la normalisation, il semble indispensable d'adapter le processus et la diffusion des normes à l'ère des échanges numériques de l'internet, et de privilégier les organismes de normalisation qui permettent un libre accès de leurs normes aux utilisateurs.

Une réflexion sur la modernisation de la normalisation⁴² a été engagée au niveau européen sous l'impulsion de Günter Verheugen, vice-président de la Commission chargé de la politique des entreprises et des industries ; plusieurs possibilités d'actions sont proposées dans le livre blanc, et notamment la possibilité de reconnaître automatiquement les organismes de normalisation selon certains critères tels que leur ouverture (publication des normes adoptées) plutôt que de recourir à une liste fermée d'organismes.

Cette initiative semble utile pour faciliter l'utilisation des normes par les acteurs et notamment les acteurs publics, garantir l'interopérabilité des produits et services au citoyen et au consommateur, et **stimuler l'innovation dans le monde numérique**.

³⁸. Voir la page d'accueil en français :

http://cordis.europa.eu/home_fr.html

³⁹. Voir par exemple l'appel à propositions relatif aux outils et technologies pour la traduction automatique (budget de 26 millions d'euros) :

http://cordis.europa.eu/fp7/ict/language-technologies/fp7-call4_en.html

⁴⁰. Voir le site de la direction de la société de l'information, sous-partie du site de la Commission européenne :

http://ec.europa.eu/information_society/index_fr.htm

⁴¹. Voir cet exemple d'appel à propositions, disponible exclusivement en anglais :

http://ec.europa.eu/information_society/tools/fwc2010/index_en.htm

⁴². Voir le communiqué de presse de la Commission en date du 3 juillet 2009, sur le site europa.eu :

<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/1085&format=HTML&aged=0&language=FR>

c. Mieux mesurer la diversité linguistique

Il est aujourd'hui délicat d'évaluer avec précision la diversité linguistique sur l'internet et de l'apprécier de façon standardisée à partir d'une grille de comparaison dont les critères exacts restent à définir.

Plusieurs études, parfois annuelles, ont permis dans une certaine mesure de connaître le comportement linguistique de certains acteurs ou communautés d'acteurs. Beaucoup d'obstacles existent cependant, soit pour automatiser l'exercice, soit pour apprécier la qualité ou l'étendue des traductions.

La connaissance des comportements linguistiques sur l'internet et la mesure de la diversité en particulier demeurent des questions importantes, dont les réponses s'avèrent nécessairement utiles aux entreprises comme aux administrations, afin d'adopter des stratégies linguistiques pertinentes.

Le Forum des droits sur l'internet recommande aux pouvoirs publics d'accroître leurs efforts pour mieux mesurer la diversité linguistique sur l'internet, en procédant périodiquement à des audits des pratiques en la matière, notamment en s'appuyant sur les travaux des spécialistes du sujet⁴³.

d. Promouvoir la labellisation des sites internet selon des critères linguistiques et culturels

Face aux enjeux liés à la promotion de la diversité culturelle sur l'internet (pour un monde numérique durable), la labellisation⁴⁴ des sites internet exemplaires, soucieux de la qualité de la langue et d'une ouverture vers les autres locuteurs, permettrait la mise en valeur des bonnes pratiques existantes tout en stimulant les comportements responsables en la matière.

Le Forum recommande aux pouvoirs publics d'encourager des initiatives de labellisation des sites internet permettant de favoriser le développement de la diversité culturelle numérique. La valorisation des comportements linguistiques responsables pourra être encouragée en priorité, compte tenu du rôle prééminent joué par la langue, première des expressions culturelles, et de la primauté de l'écrit sur internet.

2. Valoriser l'effort national de traduction

La mise en œuvre au niveau national des politiques linguistiques doit s'accompagner d'un pilotage stratégique performant au niveau interministériel, et dépendant du Premier ministre.

La traduction des contenus des sites internet publics relevant d'une préoccupation stratégique, il est recommandé aux pouvoirs publics de mettre en place un **pilotage interministériel de la traduction**.

Elle implique aussi une réorganisation des services de traduction des services publics, aujourd'hui sous la compétence de plusieurs ministères.

⁴³. L'Observatoire des pratiques linguistiques, association subventionnée par les pouvoirs publics, l'Association diversum, ou au sein de la DGLFLF.

⁴⁴. À titre d'exemple, citons le label *diversum internet*, géré par l'Association diversum : <http://www.diversum.net/fr/association/aa.php?t=salle-de-presse&g=presse&p=1>

a. Les services publics de la traduction

Il existe aujourd'hui trois services de traduction dans les services publics rassemblant au total 54 personnes :

- le Centre des liaisons européennes et internationales de la sécurité sociale⁴⁵ (CLEISS), qui emploie 23 traducteurs ;
- le service de traduction du ministère de l'Économie et des Finances, qui emploie 19 traducteurs, traducteurs-terminologues et linguistes spécialistes des outils informatiques Métier ;
- le service de traduction du ministère des Affaires étrangères, qui emploie 12 traducteurs.

Ces trois services sont indépendants les uns des autres. Il existe toutefois un embryon d'inter-ministériarité de par l'existence du Groupe de travail interministériel de la traduction (G.I.T.), qui se réunit deux fois par an, et qui a créé un Atelier de terminologie administrative.

Ces services de traduction, qui comptent donc un effectif réduit de linguistes (traducteurs, traducteurs-terminologues, linguistes informaticiens) sont **relativement peu connus des pouvoirs publics** : ces derniers recourent, pour la très grande majorité de leurs besoins, à des prestataires externes (des agences de traduction, le plus souvent). Il apparaît également que l'État délègue à des opérateurs externes des projets de traduction⁴⁶.

Les services ministériels ne disposent pas, par ailleurs, de la capacité juridique qui leur permettrait de travailler pour d'autres administrations, comme les autres ministères ou les universités, ces dernières souffrant au demeurant d'un manque de visibilité dommageable en raison de l'absence de traductions de leurs sites internet.

Enfin, il n'existe pas de cellule spécialisée pour les contenus internet au sein de ces services de traduction, ce qui est regrettable, étant donné la forte exposition de ces contenus spécifiques aux locuteurs étrangers. Les services de traduction de Bercy et du Quai disposent cependant de personnels compétents dans le domaine des technologies et des outils linguistiques.

b. Exemple étranger : l'Allemagne

Il n'est pas possible de comparer le modèle français, pays de tradition monolingue⁴⁷, à la situation de pays plurilingues tels que la Belgique, la Suisse ou encore l'Espagne. En revanche, la comparaison avec l'Allemagne, pays de taille comparable et lui aussi de tradition monolingue de par sa Constitution, est particulièrement éclairante.

Chaque ministère fédéral dispose de son propre service de traduction, qui emploie, en fonction de la structure administrative, de 3 jusqu'à 110 linguistes (au ministère des Affaires étrangères).

Parallèlement, les pouvoirs publics allemands bénéficient d'un service linguistique interministériel – l'Office fédéral des langues – qui compte 310 linguistes (traducteurs, interprètes, etc.) et environ 120 agents supports.

⁴⁵. Voir le site internet du CLEISS :
<http://www.cleiss.fr/>

⁴⁶. Un bon exemple est la volonté récente du Conseil pour la création artistique de confier à la Villa Gillet la traduction de textes scientifiques français. Voir *supra* note 20 page 11.

⁴⁷. Article 2 de la Constitution de 1958 : « La langue de la République est le français » :
<http://www.conseil-constitutionnel.fr/conseil-constitutionnel/francais/la-constitution/la-constitution-du-4-octobre-1958/texte-integral-de-la-constitution-de-1958.5074.html>

Dans cette organisation, très performante, c'est l'Office fédéral des langues (rattaché au ministère de la Défense), qui prend en charge notamment les besoins qui ne peuvent être satisfaits dans les ministères.

Ce dispositif ne concerne par ailleurs que l'échelon fédéral, chaque Land pourvoyant à ses besoins propres.

c. Les propositions d'évolution

L'organisation optimale reproduirait en France, à une échelle plus modeste, le dispositif allemand, qui a amplement fait ses preuves, et ce, en s'appuyant sur l'expérience et les compétences de l'existant :

- un nombre restreint de pôles linguistiques constitueraient le réseau ministériel (notamment : Affaires étrangères, Économie et Finances, Affaires sociales et autres, Défense) ;
- un service interministériel, qui pourrait compter à terme une centaine de linguistes, traiterait de toutes les questions ne pouvant être prises en charge à l'échelon ministériel (interprétariat, intelligence économique, traitement informatique des langues, etc.)

Ce schéma ne représenterait, avec environ deux cents linguistes, encore que 30 % des moyens déployés par l'Allemagne ; il permettrait toutefois à la France de faire par rapport à la situation existante **un bond qualitatif considérable en termes d'orientation stratégique et de mise en œuvre du multilinguisme.**

Par ailleurs, la traduction de contenus sur internet répond à des spécificités propres : accessibilité du document à des visiteurs du monde entier, nécessités liées au bon référencement du contenu, contexte technique plus fortement marqué, utilisation de liens hypertextes pour la navigation ou encore d'aides contextuelles pour faciliter la lecture du texte.

Le Forum des droits sur l'internet recommande aux pouvoirs publics de renforcer le dispositif interministériel existant (Groupe interministériel de la traduction) et d'en confier le pilotage au Premier ministre.

Du fait de la spécificité des projets internet, il est recommandé de mettre en place, au sein de ce service linguistique interministériel, un pôle de compétence spécialisé pour la traduction des contenus publiés sur l'internet, accessible à toutes les administrations centrales.

Au niveau des choix linguistiques, le Forum recommande aux pouvoirs publics de s'ouvrir davantage aux langues à fort potentiel économique et politique (arabe et chinois notamment) tout en s'efforçant de préserver la diversité culturelle en mettant en avant des langues rares.

Il semble par ailleurs indispensable d'accentuer l'effort de traduction des sites internet des universités françaises, afin de **valoriser la recherche et l'enseignement national**, et d'attirer davantage d'étudiants et de chercheurs étrangers.

Le Forum recommande au ministère de la Recherche de davantage prendre en compte les problématiques de diversité linguistique dans les universités, afin d'offrir aux locuteurs étrangers une bonne visibilité des activités et travaux menés dans chaque structure pédagogique.

Il est notamment recommandé au ministère de la Recherche d'améliorer la mutualisation des compétences et des savoir-faire afin de faciliter les efforts de traduction des sites internet des universités, et de piloter la mise en œuvre de plates-formes collaboratives

permettant à tous, et notamment aux enseignants et aux étudiants de participer eux-mêmes aux efforts de traduction des pages internet les concernant sur les sites internet des universités.

Au niveau européen, notons que les trois institutions que sont le Parlement européen, le Conseil européen et la Commission européenne, bénéficient de services de traduction conséquents et d'une solide expérience en matière de traitement de documents officiels.

Le Forum recommande aux pouvoirs publics de se rapprocher davantage des institutions européennes et notamment de la Commission européenne, afin de mettre en œuvre un partenariat stratégique permettant l'échange d'expérience et de bonnes pratiques entre les services publics de la traduction et leurs homologues européens.

3. Conforter le cadre juridique français de la langue

a. Le cadre juridique de la langue française

Le principal texte juridique encadrant l'emploi de la langue française est la loi du 4 août 1994 dite « loi Toubon ». Il concerne à la fois les acteurs publics⁴⁸ et les acteurs privés⁴⁹. Il ressort des réflexions et des échanges avec les acteurs concernés que ce dispositif semble, aujourd'hui encore, satisfaisant. Toutefois, il peut paraître, à certains égards, insuffisant. Une proposition de loi de M. Philippe Marini, sénateur, a été adoptée⁵⁰ par le Sénat le 10 novembre 2005, mais n'a jamais été débattue devant l'assemblée. Elle proposait notamment d'ajouter à l'article 2 de la loi de 1994 la mention « ou par voie électronique » afin de clarifier l'applicabilité de la disposition au contexte numérique.

Le Forum considère que le « cadre » juridique de la langue, constitué notamment par la loi du 4 août 1994, **est dans l'ensemble adapté** aux besoins des acteurs publics et privés dans le contexte national.

Il considère à ce titre que la formulation de l'article 2 de la loi est sans équivoque, et s'applique tout autant à la publicité sur l'internet qu'à celle réalisée sur les autres médias.

Pour davantage de clarté, le Forum recommande cependant aux pouvoirs publics de clarifier la notion de « **mode d'emploi** » telle que définie dans cet article, pour englober

⁴⁸. Dispositions visant les acteurs publics et les acteurs privés investis d'une mission de service public :
Article 4 : « Lorsque des inscriptions ou annonces visées à l'article précédent, apposées ou faites par des personnes morales de droit public ou des personnes privées exerçant une mission de service public font l'objet de traductions, celles-ci sont au moins au nombre de deux. »

Article 5 : « Quels qu'en soient l'objet et les formes, les contrats auxquels une personne morale de droit public ou une personne privée exécutant une mission de service public sont parties sont rédigés en langue française. Ils ne peuvent contenir ni expression ni terme étrangers lorsqu'il existe une expression ou un terme français de même sens approuvés dans les conditions prévues par les dispositions réglementaires relatives à l'enrichissement de la langue française. »

Article 7 : « Les publications, revues et communications diffusées en France et qui émanent d'une personne morale de droit public, d'une personne privée exerçant une mission de service public ou d'une personne privée bénéficiant d'une subvention publique doivent, lorsqu'elles sont rédigées en langue étrangère, comporter au moins un résumé en français. »

⁴⁹. Dispositions visant les acteurs publics ou privés :
Article 2 : « Dans la désignation, l'offre, la présentation, le mode d'emploi ou d'utilisation, la description de l'étendue et des conditions de garantie d'un bien, d'un produit ou d'un service, ainsi que dans les factures et quittances, l'emploi de la langue française est obligatoire. Les mêmes dispositions s'appliquent à toute publicité écrite, parlée ou audiovisuelle. »

Article 3 : « Toute inscription ou annonce apposée ou faite sur la voie publique, dans un lieu ouvert au public ou dans un moyen de transport en commun et destinée à l'information du public doit être formulée en langue française. »

⁵⁰. La proposition de loi adoptée par le Sénat :
<http://www.senat.fr/leg/tas05-029.html>

toutes les notices qui pourraient être délivrées sur l'internet, et non pas seulement celles incluses dans l'emballage du produit.

Enfin, la dynamique d'ouverture linguistique établie par l'article 4 de la loi vis-à-vis du secteur public pourrait être prolongée par les exploitants privés non chargés d'une mission de service public. Il leur est donc recommandé, lorsqu'ils souhaitent traduire leur site internet, de le faire, lorsque cela est possible, en plusieurs langues au lieu d'une seule.

Dans le cadre particulier des logiciels et de certains services internet, certaines traductions peuvent se révéler délicates à mettre en œuvre, pour des raisons techniques. C'est le cas notamment de certains messages et erreurs techniques.

En effet, l'accès à ces messages nécessiterait de revoir des couches logicielles « basses », ce qui impliquerait des coûts financiers significatifs pour les éditeurs et auteurs de logiciels. Techniquement, il paraît également difficile, dans certains cas, d'identifier avec précision la langue de l'utilisateur⁵¹.

Il faut par ailleurs souligner que certains de ces messages ne sont pas visibles de l'utilisateur, ou tout du moins, pas directement. Un exemple familier est celui de l'erreur 404⁵² « *File not found* » renvoyé par un serveur *web* lorsque le document demandé dans l'adresse *URL* par l'internaute n'existe pas sur le serveur. Dans ce cas-là, il reste possible pour le créateur du site d'aménager une page spécifique plus explicative que le simple message d'erreur du serveur⁵³, message dont la traduction par l'éditeur du logiciel faisant office de serveur HTTP serait beaucoup plus complexe.

Le Forum recommande aux pouvoirs publics d'exonérer de l'obligation de traduction les messages ou erreurs techniques des programmes, logiciels ou sites internet nécessitant une implication lourde en termes de développement.

Il est recommandé aux pouvoirs publics, avec l'aide des spécialistes ou organismes de normalisation concernés, de procéder à une analyse fine des messages pouvant être traduits, et de ceux qui ne doivent pas l'être, et, en ce qui concerne les messages ou erreurs les plus usuels, de fournir une traduction séparée accessible aux publics concernés.

Enfin, il paraît utile de mentionner un texte encore en vigueur aujourd'hui et qui impose à toutes les juridictions nationales de motiver leur décision en langue française et donc d'utiliser une terminologie adaptée : il s'agit de l'ordonnance de Villers-Cotterêts d'août 1539⁵⁴.

⁵¹. Cas des services asynchrones comme la messagerie électronique, posant le problème de la traduction des messages d'erreur en cas de non-réception du message, souvent envoyés par le « *Mail Delivery Subsystem* ».

⁵². Voir une explication de cette erreur technique sur : http://fr.wikipedia.org/wiki/Erreur_HTTP_404

⁵³. Pour plus d'explications sur les bonnes pratiques multilingues et la gestion des messages et erreurs techniques des logiciels serveurs *HTTP*, se référer à l'ouvrage de Patrick Andries, Unicode en pratique, chapitres 8 à 10 : <http://hapax.qc.ca/>

⁵⁴. Voir notamment les articles 110 et 111 de l'ordonnance d'août 1539 :
« Art. 110. – Et afin qu'il n'y ait cause de douter sur l'intelligence desdits arrêts, nous voulons et ordonnons qu'ils soient faits et écrits si clairement, qu'il n'y ait ni puisse avoir aucune ambiguïté ou incertitude ne lieu à demander interprétation.
Art. 111. – Et pour ce que telles choses sont souvent advenues sur l'intelligence des mots latins contenus esdits arrêts, nous voulons d'oresnavant que tous arrêts, ensemble toutes autres procédures, soient de nos cours souveraines et autres subalternes et inférieures, soient de registres, enquestes, contrats, commissions, sentences, testaments, et autres quelconques, actes et exploits de justice, ou qui en dépendent, soient prononcés, enregistrés et délivrés aux parties en langage maternel français et non autrement. »

Étant donné que les décisions de justice ne sont pas explicitement visées par la loi de 1994, il est logique que la Cour de cassation préfère viser l'ordonnance de 1539 pour sanctionner l'emploi de citations non traduites en français par les juridictions nationales⁵⁵.

- b. Proposition complémentaire : un dispositif d'incitation fiscale pour le multilinguisme ?

La traduction présente un coût non négligeable. La question de l'aide et du soutien des pouvoirs publics aux travaux de traduction des sites internet des acteurs privés est donc d'actualité.

Les pouvoirs publics ont proposé en 2005 aux PME françaises un dispositif d'aide à l'exportation dans le cadre de la prospection commerciale⁵⁶. La question qui se posait était de savoir si ce dispositif d'aide, qui permet d'obtenir un crédit d'impôt correspondant à 50 % de la dépense, et allant jusqu'à 40 000 euros, pouvait être sollicité pour des traductions d'un site internet utilisé à des fins de prospection commerciale.

La question a été posée au ministre de la Culture et de la Communication par le député Bruno Bourg-Broc et publiée au Journal officiel le 18 avril 2006, et la réponse⁵⁷ publiée le 4 juillet 2006. Il est précisé que le dispositif peut être, en l'état, utilisé pour des opérations de traduction d'un site internet utilisé à des fins de prospection commerciale. Cette réponse a par ailleurs été reprise dans le rapport au Parlement sur l'emploi de la langue française⁵⁸.

La réponse apportée par le ministre de l'Économie étant assez claire dans son principe, il serait dommage que ce dispositif d'aide ne soit pas commercialement utilisé par les traducteurs et les développeurs de sites internet pour inciter leurs clients entreprises à concevoir des sites internet multilingues.

Le Forum recommande aux acteurs privés d'utiliser le dispositif d'aide à l'exportation pour la mise en œuvre du multilinguisme sur les sites internet, et, pour les prestataires, de mieux intégrer cette disposition dans leur offre commerciale.

C – Valoriser l'apport des technologies numériques de la langue

Les technologies de la langue peuvent apporter une contribution précieuse à deux types de besoins :

- d'une part, **la préservation et l'enrichissement des cultures**, grâce aux technologies (dites monolingues), qui permettent par exemple de manipuler et de faciliter les échanges de contenus dans une langue ou de transcrire un contenu audiovisuel en langue écrite pour mieux y accéder depuis un moteur de recherche ;

<http://www.assemblee-nationale.fr/histoire/villers-cotterets.asp>
55. Cass. 1^{re} civ., 25 juin 2009, n° 08-11-226, F-D, Sté Galerie d'art Milmo Penny Fine art Ltd c/ X. JurisData n° 2009-048977.

56. Voir l'article 244 quater H du Code général des impôts, et la note de la Direction générale des impôts : http://www.impots.gouv.fr/portal/dqi/public?paf_dm=popup&paf_gm=content&paf_gear_id=500018&type=Page=cpr02&docOid=documentstandard_4500&espId=-1&temNviPopUp=true

57. Voir la question n° 92 711 et sa réponse sur le site de l'Assemblée nationale : <http://questions.assemblee-nationale.fr/q12/12-92711QE.htm>

58. Voir le rapport au Parlement sur l'emploi de la langue française de l'année 2006 : http://www.dglf.culture.gouv.fr/rapport/2006/rapport_parlement_2006.pdf

- d'autre part, **la communication et l'échange entre les peuples**, grâce aux technologies (dites interlingues), qui facilitent la traduction de contenus d'une langue vers une autre.

Ces technologies présentent des bénéfices multiples dans le cadre linguistique, dont les perspectives d'évolution semblent intéressantes (1). Elles doivent donc être appréhendées avec soin par les décideurs et chefs de projets. Se posent néanmoins les questions du soutien, de la recherche et du développement liés à ces technologies (2), et de leur utilisation dans le cadre de l'enseignement (3).

1. État de l'art des technologies numériques linguistiques et perspectives

- a. Quelles sont les possibilités techniques en matière d'assistance linguistique aujourd'hui et que peut-on en attendre ?

Voici quelques-unes des fonctionnalités que l'on peut attendre des technologies de la langue :

- Traduction automatique de contenus textuels et aide à la traduction de contenus : la qualité du résultat est très variable suivant les couples de langues utilisés, la syntaxe et les champs sémantiques et lexicaux du document original. Dans tous les cas, le résultat nécessitera d'être retravaillé par une équipe de traduction, mais la technologie permettra généralement un gain de temps significatif. Certains outils de ce type sont utilisables gratuitement⁵⁹.
- La translittération permet la conversion d'un texte d'un jeu de caractères dans un autre. Dans le cas où les correspondances sont univoques (on parle alors de bijection plus ou moins totale), les résultats peuvent être excellents, comme pour le passage du serbe (serbo-croate en cyrillique) au croate (serbo-croate en latin), ou encore du moldave en azéri.
- Les technologies monolingues permettent de manipuler le texte, de proposer des outils de correction orthographique ou grammaticale, de générer des listes de mots-clés à partir d'un document volumineux, ou encore un « résumé ». Ces technologies d'indexation de contenu sont largement utilisées dans la plupart des outils et moteurs de recherche, et permettent d'obtenir des résultats correspondant au plus près à la requête de l'utilisateur.
- Les technologies de reconnaissance vocale et de synthèse vocale permettent de transformer du langage parlé en texte écrit, et inversement. Elles permettent par exemple de mettre en œuvre un sous-titrage en temps réel de contenus audiovisuels, mais aussi de générer des résumés de documents audiovisuels, qui deviennent dès lors plus accessibles par l'intermédiaire d'un moteur de recherche sur internet⁶⁰.
- Les technologies vocales sont largement utilisées pour les besoins spécifiques de certaines personnes en situation de handicap : sous-titrage de discours ou de conférences pour des malentendants, synthèse vocale de documents écrits, pages internet, messages électroniques pour les malvoyants. À ces technologies peut s'ajouter une technologie interlingue : transcription d'un discours parlé en anglais en texte français par exemple.
- Enfin, les technologies de la langue trouvent de nombreuses applications en termes de détection automatique de contenus illicites à des fins de lutte contre la

⁵⁹. Voir par exemple *Google Traduction* ou *Systran* :
http://www.google.fr/language_tools
<http://www.systran.fr/>

⁶⁰. Voir par exemple *Voxalead News*, moteur de recherche d'extraits vidéo issus de chaînes de télévision, utilisant un moteur de synthèse pour extraire une série de mots clés pour chaque vidéo :
<http://voxaleadnews.labs.exalead.com/>

contrefaçon, de protection de l'enfance, ou encore de sécurité (reconnaissance automatique d'image ou de voix par exemple), bien que leur fiabilité ne soit pas encore optimale.

Les usages, quant à eux, se généralisent : le moteur de recherche Google, disponible dans 124 langues, propose à ses visiteurs des traductions gratuites dans 2 500 paires de langues des contenus proposés dans ses résultats. Ces outils rencontrent un vif succès, et permettent d'accéder à des informations qui n'étaient pas accessibles jusqu'alors.

Ces outils rendent néanmoins possible la collecte d'informations personnelles ou confidentielles par les éditeurs de ces solutions : un document qui est envoyé en ligne à un outil de traduction automatique doit alors être envoyé sur un serveur de traitement distant pour en réaliser la traduction, et il convient par conséquent d'être prudent lors de l'utilisation de ces outils en ligne.

Le Forum des droits sur l'internet recommande aux utilisateurs de solutions de traitement linguistique ou d'aide à la traduction d'être particulièrement vigilants par rapport à l'utilisation qui peut être faite des données transmises en ligne.

Il est notamment recommandé aux acteurs publics ou privés de ne pas recourir à des outils d'aide à la traduction en ligne pour des contenus sensibles ou confidentiels.

b. Quelles sont les perspectives d'évolution ?

Ces technologies font de plus en plus régulièrement l'objet **d'évaluations quantitatives**⁶¹. Les systèmes sont évalués en les confrontant à des situations représentatives des applications visées. Ces évaluations permettent d'orienter et de mesurer les progrès. Par exemple, la reconnaissance vocale est évaluée en mesurant le taux de mots mal reconnus sur des échantillons de quelques heures de parole.

Des évaluations sont conduites tous les ans depuis plus de vingt ans, et couvrent aujourd'hui plusieurs langues. Les meilleurs systèmes commencent à offrir des taux d'erreur et des traitements qui les rendent utilisables pour indexer de grands volumes de documents sonores. On devrait ainsi voir se démocratiser à brève échéance des moteurs de recherche permettant d'accéder à des documents audiovisuels. Les systèmes de dialogue oral sont également appelés à se répandre.

Les technologies de traduction évoluent rapidement aussi depuis une dizaine d'années. Elles devraient poursuivre leur progression régulière en termes de qualité de traduction et de langues couvertes, et permettre ainsi de faire tomber de plus en plus souvent la barrière de la langue. Combinées aux technologies vocales, elles feront à terme de la traduction vocale une réalité quotidienne.

Ces progrès sont stimulés par l'amélioration des performances tant en termes de calcul que de stockage des technologies numériques. Elles bénéficient d'interfaces de plus en plus modernes et faciles à utiliser, comme par exemple des lunettes projetant une traduction en temps réel sur la rétine de leur utilisateur⁶².

Les nouvelles technologies dites de « réalité augmentée », qui permettent d'afficher des informations en temps réel dans l'environnement visuel immédiat d'une personne, offrent, elles aussi, de nouvelles perspectives pour les systèmes de traduction automatiques.

⁶¹. Une des méthodes d'évaluation la plus connue est le *BLEU Score (Bilingual Evaluation understudy)* : <http://en.wikipedia.org/wiki/BLEU>

⁶². *Au Japon, des lunettes permettent de traduire en direct*, Numerama.com, 14/11/2009 : <http://www.numerama.com/magazine/14489-au-japon-des-lunettes-permettent-de-traduire-en-direct.html>

2. Améliorer la recherche sur les technologies de la langue et leur usage

Le rythme des progrès dépend évidemment de l'intensité des efforts de recherche. Les technologies de la langue font l'objet de plusieurs programmes de recherche en France et dans le monde.

Les principaux projets sont conduits par les États-Unis et financés principalement par la *Defense Advanced Research Project Agency*⁶³ (DARPA). Les recherches couvrent l'ensemble des technologies en se concentrant sur un petit nombre de langues d'intérêt stratégique. On estime les budgets à plus de 50 millions de dollars par an.

En Inde, le multilinguisme constitue l'une des dix priorités du programme national sur l'informatisation de la société⁶⁴ (TDIL). Il vise, pour traiter l'anglais et 18 langues⁶⁵ indiennes, à produire plusieurs outils tels que la traduction automatique, la synthèse à partir du texte, la reconnaissance vocale, les moteurs de recherche, la reconnaissance de caractères, les correcteurs orthographiques et des ressources linguistiques pour chacune des 19 langues.

En Europe, plusieurs appels à projets ont été menés dans le cadre du **7^e programme cadre pour la recherche et le développement technologique (PCRD)**, concernant directement les technologies de la langue. En 2009, l'objectif d'interaction par le langage a donné lieu à un programme de 26 millions d'euros sur la traduction automatique et à un programme de 14 millions d'euros pour le multilinguisme sur internet. Des réseaux d'infrastructures linguistiques se constituent progressivement⁶⁶.

- a. Renforcer la dotation des pouvoirs publics aux programmes de recherche français

En France, de nombreux projets et programmes sont conduits dans différents cadres. Plusieurs projets sont conduits dans le cadre de l'Agence nationale pour la recherche (ANR).

Le programme Technolangue⁶⁷ est un programme significatif qui a donné lieu à des résultats publics. Il s'est déroulé entre 2002 et 2005 et a été subventionné à hauteur de deux millions d'euros par an. Il a produit de nombreuses ressources linguistiques (dictionnaires, lexiques, corpus et bases de données terminologiques). Le corpus produit pour la reconnaissance vocale comprend notamment 1 600 heures de parole et 100 heures transcrites ce qui correspond à peu près à un million de mots. Il a également produit des outils de traduction automatique de l'arabe et de l'anglais vers le français. Il a donné lieu à de nombreuses campagnes d'évaluation qui ont lancé une dynamique forte au niveau national.

Bénéficiant de cette dynamique, **le programme Quaero**⁶⁸ est un programme fédérateur de recherche et d'innovation soutenu par OSEO⁶⁹. Son objectif est d'améliorer le traitement des documents multilingues et multimédias.

⁶³. La page d'accueil du site du DARPA :

<http://www.darpa.mil/>

⁶⁴. Page d'accueil du programme indien sur l'informatisation de la société :

<http://tdil.mit.gov.in/>

⁶⁵. Assamais, bengali, gujarati, hindi, kannada, kashmiri, konkani, malayalam, manipuri, marathi, napali, oriya, panjabi, sanskrit, sindhi, tamoul, télougou, ourdou.

⁶⁶. Voir notamment CLARIN, Infrastructure de Ressources Linguistiques pour les Sciences Humaines et Sociales, et FLareNet, programme sur les ressources et les technologies de la langue :

<http://www.clarin.eu/>

<http://www.flarenet.eu/>

⁶⁷. Le portail internet du projet Technolangue :

<http://www.technolangue.net/>

⁶⁸. Voir la page d'accueil du programme Quaero :

Établi entre 2008 et 2013, il représente un budget total d'environ 200 millions d'euros ; l'État subventionne le programme à hauteur de 100 millions d'euros, par l'intermédiaire d'OSEO.

L'objectif du programme est d'améliorer le traitement des documents multilingues et multimédias. Il est structuré autour du développement de 30 technologies répondant aux demandes de six projets applicatifs. Neuf langues sont traitées dans le cadre du projet.

Un des projets applicatifs, porté par Exalead⁷⁰, vise à développer des moteurs de recherche multimédias permettant la recherche de contenus audiovisuels à partir de l'analyse sémantique qui en est faite.

*VoxaleadNews*⁷¹, par exemple, fait appel aux technologies de reconnaissance vocale pour générer une transcription complète et une liste de mots clés des contenus audiovisuels publiés par les différentes chaînes de télévision dans le monde. L'outil Chromatik⁷², bien que plus anecdotique, illustre également le propos.

Grâce aux nombreux efforts de recherche soutenus par ces projets, les technologies arrivent ainsi à maturité pour un nombre croissant d'applications. Les offres commerciales de produits et de services linguistiques sont encore peu nombreuses, et peu d'acteurs se sont positionnés sur ces secteurs innovants.

Le marché est cependant en train d'évoluer avec l'arrivée des technologies à maturité, et les besoins en termes de recherche se renforcent. Elles ne représentent actuellement qu'une très faible proportion de la recherche consacrée aux technologies de l'information alors qu'elles sont en train de devenir **un facteur majeur d'innovation**.

Le Forum considère qu'il est important que les pouvoirs publics augmentent leur soutien financier aux programmes de recherche sur les technologies de la langue et que les organismes de recherche publics et les universités accroissent leur effort de recherche sur les technologies de la langue en général.

Il est recommandé aux pouvoirs publics d'informer les utilisateurs et les chercheurs sur les résultats obtenus dans les différents programmes de recherche et, dans la mesure où cela est possible⁷³, de publier un maximum de résultats sous forme accessible et librement utilisable par la communauté.

Il leur est par ailleurs recommandé d'améliorer le suivi des projets réalisés en partenariat avec les acteurs privés, en augmentant la part de la participation humaine des pouvoirs publics dans la gestion des programmes de recherche.

Il est enfin recommandé aux pouvoirs publics d'accroître leur effort en matière d'enseignement et de formation dans ce domaine.

<http://www.quaero.org/modules/movie/scenes/home/>

69. Établissement public favorisant l'aide à l'innovation, la garantie des concours bancaires et le financement en partenariat :

<http://www.oseo.fr/>

70. Page d'accueil du moteur de recherche multimédia Exalead :

<http://www.exalead.com/search/>

71. Démonstration du site Voxalead News, cf. *supra* note 60 page 22.

72. Chromatik permet de rechercher des images à partir de leur couleur dominante :

<http://chromatik.labs.exalead.com/>

73. Lorsque l'État ne finance qu'en partie un programme de recherche, conjointement avec des acteurs privés, il n'est pas pleinement propriétaire des matériels et logiciels développés. La restitution d'éléments sous forme accessible et réutilisable est donc un élément de négociation entre pouvoirs publics et acteurs privés. Au contraire, dans le cadre d'une commande publique financée exclusivement par les pouvoirs publics, les résultats appartiennent exclusivement à l'État, les droits sur les logiciels ainsi développés étant gérés par l'Agence pour la propriété immatérielle de l'État (APIE).

- b. Améliorer la visibilité de l'action européenne en termes de technologies de la langue et d'aide à la traduction

Le 7^e PCRD⁷⁴ (Programme cadre de recherche et de développement technologique) est le principal instrument de financement et de recherche à l'échelon européen. Il consacre, entre 2007 et 2013, 50,5 milliards d'euros à la recherche, dont 9 milliards sont spécifiquement alloués aux technologies de l'information et de la communication.

Parmi les différentes aides, celle consacrée aux « interactions basées sur le langage » ne s'élève qu'à une hauteur de 26 millions d'euros, soit 0,29 % du budget consacré aux technologies de l'information et de la communication seulement, ce qui semble nettement insuffisant compte tenu des objectifs ambitieux fixés par l'Union européenne⁷⁵.

Le Forum considère que le soutien financier aux technologies de la langue par l'Union européenne est insuffisant pour atteindre les objectifs en termes d'échanges inter-culturels et d'efforts de traduction fixés par les États membres, et qu'il serait par conséquent opportun d'allouer davantage de moyens aux interactions basées sur le langage (technologies de traitement automatique des langues individuelles et de traduction de l'une à l'autre) dès la deuxième phase du 7^e PCRD et dans le cadre du 8^e PCRD.

Cette action devrait se faire de manière coordonnée avec les programmes de recherche des États membres concernant ce même domaine, afin d'assurer une couverture aussi large que possible des langues traitées, tant individuellement que pour la traduction de l'une à l'autre.

- c. Produire davantage de ressources linguistiques

Les technologies linguistiques nécessitent des ressources pour fonctionner correctement : dictionnaires, corpus de traductions notamment. Il implique donc **d'accroître l'effort de production de ces ressources**, ce qui permettra d'améliorer la qualité des outils d'aide au traitement de la langue au final.

En outre, le cadre juridique pourrait être adapté pour faciliter le développement et l'utilisation des technologies de la langue, dans le respect des droits d'auteur des ayants droit. Il est notamment apparu nécessaire de faciliter l'accès des chercheurs aux corpus de traduction existants au sein des différents organismes publics ou privés effectuant des traductions.

Aux États-Unis, la réutilisation des corpus de traduction est facilitée par le principe du *Fair Use*⁷⁶, qui permet aux organismes de recherche de réutiliser librement certains contenus protégés par le droit d'auteur à des fins de recherche ou d'enseignement. Un tel principe n'existe pas en Europe.

Dans ce contexte, plusieurs voies ont été envisagées :

- renforcer la coopération entre le monde de la recherche et les éditeurs publics ou privés propriétaires d'importants corpus de traduction ainsi que des principaux projets multilingues (Europeana⁷⁷, Gutenberg⁷⁸, Google Books⁷⁹) ;

⁷⁴. La page d'information sur le 7^e PCRD européen (2007-2013), en ligne sur le site welcomeurope.com : <http://www.welcomeurope.com/default.asp?id=1900&MyPays=fr>

⁷⁵. Voir, *supra* note 13, page 7.

⁷⁶. Pour plus de détails sur le *Fair Use* aux États-Unis, voir notamment : http://fr.wikipedia.org/wiki/Fair_use

⁷⁷. La bibliothèque numérique Europeana contient 4,6 millions de documents dans 26 langues : <http://www.europeana.eu/portal/>

⁷⁸. Le projet Gutenberg, mettant à disposition du public des livres du domaine public : http://www.gutenberg.org/wiki/Main_Page

- améliorer le partage des ressources écrites des principaux organismes internationaux tels que l'Organisation des Nations unies (ONU), le Fonds monétaire international (FMI), l'Union européenne ou l'UNESCO ;
- ajouter une nouvelle exception au droit d'auteur à l'article L. 122-5 du Code de la propriété intellectuelle⁸⁰, permettant aux chercheurs de technologies de la langue d'avoir accès à des corpus de traduction.

d. Mesurer les performances linguistiques des technologies de la langue

Il est aujourd'hui complexe d'évaluer la qualité et le comportement des différentes technologies existantes d'aide au traitement de la langue. Cet aléa rend difficile l'évaluation et la comparaison qualitative des différentes offres disponibles pour l'utilisateur, ce qui ne favorise pas le développement du marché.

Il existe cependant un nombre croissant d'évaluations disponibles publiquement⁸¹.

Au niveau international, la Réunion internationale annuelle sur la traduction et la terminologie assistées par ordinateur⁸² (JIAMCATT), sous l'égide des Nations unies, a pour objet d'offrir à ses partenaires un cadre de discussion, d'échanges et de coopération dans les domaines de la *terminotique*, de la *traductique*, de l'interprétation et de la recherche documentaire.

Le Forum des droits sur l'internet recommande aux pouvoirs publics de mettre en place au sein du service interministériel de la traduction⁸³ un pôle de compétence dont la mission serait d'évaluer et de mesurer les performances des technologies linguistiques, à partir d'un référentiel composé de différents critères à déterminer.

e. Des technologies utiles pour les personnes en situation de handicap

En termes d'accessibilité, les technologies de la langue sont très **prometteuses pour les personnes en situation de handicap**, notamment les technologies inter-médias, qui permettent de transformer automatiquement des contenus d'un médium dans un autre (passage de contenus visuels en parole et réciproquement passage de la parole en contenu visuel pour faciliter la communication des personnes malvoyantes, passage de contenus audio en texte ou en langue des signes produite par des agents artificiels et réciproquement synthèse vocale à partir de textes ou de langues des signes pour faciliter la communication des personnes malentendantes).

Les dispositions légales sur l'accessibilité numérique⁸⁴, qui obligent les éditeurs publics de services sur l'internet à diffuser des contenus accessibles à tous, conformément aux

⁷⁹. Google Books donne accès à 7 millions de documents en 44 langues :

<http://books.google.com/books>

⁸⁰. Voir l'article sur le site Légifrance :

<http://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006278917>

⁸¹. Voir notamment, coté américain, les travaux du *National Institute of Standards and Technology* (NIST) :

<http://www.itl.nist.gov/iad/mig/tests/>

Au niveau européen, les travaux du projet *Euromatrix* :

<http://www.euromatrix.net/>

Et ceux de l'*European Language Resource Association* :

<http://www.hlt-evaluation.org/>

En France, le programme Technolangue a également donné lieu à des évaluations publiques :

<http://www.technolangue.net/>

⁸². Description du JIAMCATT sur le wiki de la Réunion :

<http://jiamcatt.ourwiki.net/tiki-index.php?page=JIAMCATT-fr>

⁸³. Sur les propositions relatives à un service interministériel de la traduction, voir *supra* le paragraphe « Valoriser l'effort national de traduction », page 16.

⁸⁴. Voir notamment l'article 47 de la loi du 11 février 2005 :

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000809647>

recommandations internationales pour l'accessibilité⁸⁵, sont généralement peu connues des spécialistes de la langue œuvrant sur les technologies de traitement automatique.

Le Forum recommande aux pouvoirs publics, et notamment au Secrétariat d'État chargé des Personnes handicapées, à la Direction générale de la cohésion sociale, au Comité interministériel au handicap et aux associations du milieu du handicap, de se rapprocher des éditeurs de technologies linguistiques, ces dernières pouvant constituer un atout précieux pour rendre accessibles certains contenus aux personnes en situation de handicap, notamment en ce qui concerne les médias audiovisuels.

3. Les technologies de la langue et l'enseignement

Les technologies de l'information et de la communication ouvrent de nouvelles portes en termes d'enseignement de la langue, et notamment les technologies de la langue, et l'internet.

a. L'apport des technologies de la langue

Les technologies de la langue permettent de recourir à des correcteurs orthographiques et grammaticaux, qui peuvent constituer une aide précieuse pour la rédaction des documents écrits dans un cadre scolaire. Leur fonctionnement et leur utilisation nécessitent cependant d'être expliqués aux élèves, faute de quoi ces outils risquent d'être mal ou peu utilisés.

Il ne s'agit pas seulement d'expliquer le fonctionnement technique des correcteurs orthographiques, ce qui entre déjà dans le cadre du Brevet informatique et internet⁸⁶ (B2i), mais plutôt de proposer un cours de culture générale sur l'informatique dans lequel sont expliqués **les limites et le recul nécessaire à l'utilisation des outils informatiques** et, en particulier, les outils de correction d'orthographe et de grammaire, qui ne permettent pas de s'affranchir d'une relecture orthographique.

Il est recommandé au ministère de l'Éducation de poursuivre la réflexion sur l'apprentissage scolaire et sur l'utilisation des outils de correction orthographique comme complément de la relecture des documents écrits, et notamment d'intégrer dans le B2i un enseignement sur l'utilisation pratique et les limites des outils de correction orthographiques et grammaticaux.

Ces outils constituent par ailleurs **une aide significative pour les élèves dyslexiques**, qui représentent environ dix pour cent⁸⁷ des écoliers français. Pour ces élèves, qui font l'objet d'une « prise en charge » spécifique⁸⁸, le passage à l'écrit est souvent difficile. Les outils de correction d'orthographe et de grammaire leur permettent de compenser plus facilement leurs difficultés et de rentrer dans une situation d'apprentissage.

⁸⁵. Voir notamment les recommandations internationales pour l'accessibilité (WCAG 2.0) publiées par le Web Consortium (W3C) et traduites en français : <http://www.w3.org/Translations/WCAG20-fr/>

⁸⁶. Présentation du Brevet informatique et internet (B2i) sur le site EducNet : <http://www.educnet.education.fr/textes/reglementaires/competences/b2i/brevet-informatique-et-internet>

⁸⁷. État de la recherche sur la dyslexie, Sonia Krifi, Université Lyon II, 2004 : <http://www.apedys.org/dyslexie/article.php?sid=32>

⁸⁸. Voir par exemple les programmes personnalisés de réussite éducative (PPRE), les projets d'accueil individualisés (PAI), ou encore les projets personnalisés de scolarisation : <http://eduscol.education.fr/cid45633/les-programmes-personnalises-de-reussite-educative-ppre.html>
<http://www.afpssu.com/Le-Projet-d-Accueil-Individualise--179.html>
<http://www.education.gouv.fr/bo/2006/32/MENE0602187C.htm>

Il est recommandé au ministère de l'Éducation de favoriser l'utilisation des outils de correction orthographiques et grammaticaux dans le cadre de l'apprentissage spécifique des élèves dyslexiques.

b. Enseignement de la langue et internet

Le vecteur de l'internet constitue un nouveau levier pour améliorer l'enseignement de la langue française aux jeunes. Ce levier semble aujourd'hui sous-utilisé, et il paraît souhaitable d'encourager les initiatives utilisant les technologies de l'information pour faciliter l'apprentissage et la connaissance de la langue française ou des autres langues.

Dans le cadre du volet numérique de l'appel à projets du plan de relance des pouvoirs publics⁸⁹, deux projets de « *serious game* » ou jeux sérieux ont été retenus⁹⁰ : « *Kaisha* », jeu permettant de faire découvrir aux jeunes le vocabulaire et les fonctionnements dans une entreprise, et « *Star Teachers* », une plate-forme d'apprentissage des langues fondée sur des scènes de cinéma.

Le Forum des droits sur l'internet recommande aux pouvoirs publics et notamment aux ministères de l'Éducation et de la Recherche d'utiliser davantage les technologies de l'information et de la communication dans l'apprentissage des langues aux plus jeunes.

D – Mieux accompagner l'évolution de la langue sur internet ; le cas du français en France

L'internet et les technologies de l'information sont en évolution constante, à un rythme qui ne cesse de s'accélérer avec la multiplication des moyens d'échanges logiciels entre les utilisateurs, et l'amélioration continue des matériels et réseaux de diffusion. Face aux enjeux que posent ces modifications dans la façon de chacun de s'exprimer et de communiquer, quelle doit être la place des pouvoirs publics et des différents acteurs concernés dans l'accompagnement de la langue ?

Au niveau international, en ce qui concerne les langues latines, les situations sont très variées. La langue italienne ne bénéficie pas, *a priori*, d'un accompagnement particulier des pouvoirs publics. Elle est donc actuellement très ouverte aux mots d'origine étrangère, à tel point que, pour un exemple éloquent en matière informatique, il serait possible en italien de traduire « la souris de l'ordinateur » par « *Il mouse del computer* ».

D'un pays à l'autre, les modèles de « régulation » varient fortement. En Catalogne par exemple, le catalan fait l'objet d'une régulation très forte par les pouvoirs publics régionaux ainsi que d'une promotion active sur les médias traditionnels ou numériques. De nombreuses ressources terminologiques et dictionnaires en catalan sont disponibles en ligne⁹¹.

Au Québec, région francophone dans un pays majoritairement anglophone, l'utilisation du français fait l'objet de dispositions légales étendues⁹², obligeant un double affichage en

⁸⁹. Voir la description du volet numérique du plan de relance : <http://www.telecom.gouv.fr/rubriques-menu/soutiens-financements/programmes-nationaux/volet-numerique-du-plan-relance/397.html>

⁹⁰. Voir la liste des projets de « *serious game* » (jeux sérieux) retenus dans le cadre du plan de relance : <http://www.telecom.gouv.fr/rubriques-menu/soutiens-financements/programmes-nationaux/volet-numerique-du-plan-relance/resultats-deux-appels-projets/liste-projets-retenus-serious-game-2217.html>

⁹¹. Voir par exemple le site du centre de terminologie catalan, TermCat : <http://www.termcat.cat/>

⁹². Voir notamment la charte de la langue française, ou « Loi 101 » :

anglais et en français sur la plupart des documents et enseignes, ou encore dans le monde du travail. Les pouvoirs publics ne cachent pas leur interventionnisme.

La terminologie au Québec dispose de moyens conséquents ; l'Office québécois de la langue française (OQLF) a l'habitude de proposer des équivalents en français pour la quasi-totalité du vocable technique. La tendance semble pourtant s'infléchir⁹³, puisque l'emploi de ces terminologies francophones en lieu et place de leur équivalent étranger ne donne plus lieu à une officialisation, mais simplement à plusieurs propositions terminologiques.

À l'opposé, parmi les langues les plus imperméables aux mots étrangers, citons l'exemple de la langue islandaise, qui ne comporte presque aucun terme d'origine étrangère.

Il convient donc de mesurer comment la langue évolue dans le contexte particulier de l'internet (1), de réfléchir au dispositif qui peut accompagner ces évolutions (2) pour améliorer les pratiques et favoriser une meilleure accessibilité linguistique des documents sur internet (3).

1. L'évolution et la vitalité de la langue française

Quelle est aujourd'hui la présence de la langue française sur l'internet ? Et quels sont les impacts des technologies sur la langue et sur les usages des francophones ?

a. Quelle est la présence de la langue française sur l'internet ?

Selon une étude⁹⁴ de l'Union Latine réalisée en novembre 2007⁹⁵, le français représente 9,80 % des pages internet par rapport à celles qui sont en langue anglaise, et 4,41 % du total des pages internet. L'évolution de la présence francophone est globalement positive, puisqu'elle n'était que de 4 % en 1998. **Le français est la première langue latine de l'internet**, et elle se situe à un niveau proche de la langue allemande, qui est estimé à 13,12 % des pages internet par rapport à celles qui sont en langue anglaise, selon l'étude précitée.

Selon l'étude, on compte 58,4 millions d'internautes francophones, ce qui représente 5,1 % du total des internautes. Le rapport entre nombre d'internautes francophones et nombre de pages francophones est de 0,87 %, ce qui correspond à la deuxième place parmi les langues latines les plus dynamiques, après l'italien.

La langue française se distingue aussi par une très forte présence dans certaines communautés : le livre et les bibliothèques numériques bénéficient notamment d'une grande quantité de documents en français⁹⁶ ; les *blogs* francophones sont très nombreux sur la toile, la France étant un des pays comptant la communauté de *blogueurs* la plus active⁹⁷ ; enfin, le taux de participation des internautes français sur les réseaux sociaux est en très forte croissance⁹⁸.

<http://www.olf.gouv.qc.ca/charte/charte/index.html>

⁹³. À ce sujet, lire par exemple La Cour suprême charcute la loi 101 en qualifiant la loi 104 d'inconstitutionnelle, AmériQuébec.net, 23 octobre 2009 :

<http://www.ameriquebec.net/actualites/2009/10/23-la-cour-supreme-charcute-la-loi-101-en-qualifiant-la-loi-104-dinconstitutionnelle.qc>

⁹⁴. Étude Langues et culture sur la toile de la Direction terminologie et industries de la langue (DTIL) de l'Union Latine réalisée en novembre 2007 par FUNREDES :

http://dtil.unilat.org/LI/2007/fr/resultados_fr.htm

⁹⁵. Un changement intervenu sur les techniques de référencement des principaux moteurs de recherche a rendu la comparaison avec des résultats plus récents impossible.

⁹⁶. Cf. *supra* notes 77, 78, 79 page 27.

⁹⁷. Selon le rapport Les médias et le numérique de Danièle Giuzzi, publié en septembre 2008, la France comptait en 2008 environ 9 millions de *blogs*, dont 2.5 millions de *blogs* actifs, et se situerait au 4^e rang mondial en nombre de bloc-notes (*blogs*) par internaute :

Au niveau de la terminologie employée, une étude⁹⁹ conduite par l'Association diversum et l'Observatoire des éditions numériques montre que les mots francophones de l'internet tendent à remplacer leurs équivalents anglo-saxons sur les pages des sites internet publics français : le terme *Home*, équivalent anglais d'Accueil, a ainsi complètement disparu ; on parlera d'actualités plutôt que de *news*, de webmestre (62,15 %) plutôt que de *webmaster* (22,95 %), de lettre d'information (45,40 %) plutôt que de *newsletter* (20,66 %), ou encore de courriel (48,09 %) plutôt que de *mail* (13,79 %).

Il faut par ailleurs noter que le média internet a permis l'apparition de communautés linguistiques promulguant l'emploi de **langues régionales ou locales**. Souvent, ces communautés sont réduites en termes de personnes, mais sont en revanche très actives et apportent un grand nombre de pages sur les espaces collaboratifs.

Sur l'encyclopédie en ligne Wikipédia, on compte ainsi plus de 30 000 pages sur le Wikipédia en breton¹⁰⁰, près de 3 500 pages en haut-normand¹⁰¹, plus de 5 000 pages en corse¹⁰², ou encore près de 50 000 pages en basque¹⁰³.

L'internet, et notamment les espaces collaboratifs, constituent donc une chance pour la diversité linguistique et la sauvegarde du patrimoine culturel et linguistique. Toutefois, il faut noter que les langues régionales ou minoritaires ne sont pas du tout utilisées dans les échanges commerciaux sur internet.

Une étude¹⁰⁴ conduite par le *Language Observatory Project* japonais explique ainsi qu'en Afrique, les langues majoritaires (anglais, français, portugais et arabe) ont pris le pas sur les langues minoritaires ou locales. En effet, malgré la constitution de communautés autour de ces langues locales, c'est la langue dominante utilisée pour les échanges commerciaux qui s'est imposée sur la toile.

Le Forum des droits sur l'internet considère que l'internet, et notamment les espaces collaboratifs, sont une chance pour la diversité linguistique, l'échange culturel, et la préservation du patrimoine linguistique.

Il considère que l'internet assure le maintien et l'essor des langues locales ou régionales de France, sans que cela constitue pour autant un danger pour la langue française.

- b. Internet : un moteur favorisant les échanges écrits mais qui provoque aussi une mutation de la langue

L'internet a provoqué une évolution importante des langues écrites. Le caractère mondial du réseau a incité les internautes à s'exprimer dans un langage différent de leur langue maternelle, afin de pouvoir communiquer avec le plus grand nombre, même si ce

-
- <http://danielegiazzi.typepad.fr/ump/2008/09/rapport-les-mdi.html>
- ⁹⁸. Voir notamment l'étude ComScore *Social Networking Has Banner Year in France, Growing 45 Percent*, 17 février 2009 : http://www.comscore.com/Press_Events/Press_Releases/2009/2/Social_Networking_France
- ⁹⁹. Étude *Le comportement linguistique de l'État sur l'internet* réalisée par l'Association diversum et l'Observatoire des éditions numériques, avec le soutien de la Délégation générale à la langue française et aux langues de France (DGLFLF), 12 mars 2009 : http://www.diversum.net/fichiers/File/com_12-03-2009.pdf
- ¹⁰⁰. Wikipédia en breton : <http://br.wikipedia.org>
- ¹⁰¹. Wikipédia en haut-normand : <http://nrm.wikipedia.org>
- ¹⁰². Wikipédia en corse : <http://co.wikipedia.org/>
- ¹⁰³. Wikipédia en basque : <http://eu.wikipedia.org>
- ¹⁰⁴. Voir l'enquête liée aux langues dans le cyberspace africain sur le site du *Language Observatory* : <http://gii2.nagaokaut.ac.jp/gii/blog/lopdiary.php?catid=154&blogid=8>

langage n'est pas parfaitement maîtrisé. L'anglais, qui était, aux origines des réseaux *arpanet* et de l'internet, la seule langue utilisée, demeure aujourd'hui encore la langue de référence sur le réseau mondial.

Il est cependant utilisé par de nombreuses personnes qui ne le maîtrisent pas couramment ; le souhait d'être intelligible par le plus grand nombre a pu conduire à un appauvrissement du vocabulaire utilisé en anglais sur internet. L'apparition du site Wikipédia en *Simple English*¹⁰⁵, une version simplifiée de la langue anglaise, illustre ce phénomène.

Des déclinaisons simplifiées du langage écrit¹⁰⁶ se sont répandues dans diverses communautés. Le meilleur exemple est celui de la « *langue texto* », historiquement utilisée par les utilisateurs de téléphones mobiles pour rédiger des messages texto ou SMS. En raison du faible nombre de touches, de la taille limitée de l'écran et de la limite maximale du nombre de caractères par message¹⁰⁷, le langage écrit vise à comporter un nombre de caractères minimal tout en restant intelligible.

Le langage SMS est utilisé, dans certaines communautés¹⁰⁸ sur internet, le plus souvent par des jeunes ou des adolescents. Il est davantage utilisé à mesure que le besoin d'instantanéité est prégnant : discussion en temps réel dans des salons de discussion (*chats*) ou sur les messageries instantanées, certains forums de discussion, certains réseaux sociaux. Il faut par ailleurs noter que dans certaines communautés, l'utilisation du langage SMS dans des forums de discussion peut être très mal perçue, ce qui pourra entraîner la suppression des messages rédigés sous cette forme, voire l'exclusion temporaire ou permanente de l'utilisateur s'étant exprimé ainsi.

Le langage SMS est en revanche beaucoup moins fréquemment utilisé lorsque la conversation est privée, restreinte à une ou plusieurs personnes : courrier électronique et listes de diffusion électroniques notamment.

L'internet a donc favorisé l'émergence d'une multiplicité de nouveaux modes de communication, dérivés de la langue écrite traditionnelle, et plus adaptés aux supports de communication utilisés.

Pour autant, celle-ci ne semble pas menacée¹⁰⁹ : elle reste en effet l'apanage des conversations personnelles bilatérales et professionnelles, et de la grande majorité des textes à vocation journalistique (médias traditionnels, bloc-notes ou *blogs*), **signe de sa**

¹⁰⁵. Le Wikipédia en *Simple English*, qui comporte plus de 50 000 entrées :
<http://simple.wikipedia.org/>

¹⁰⁶. Le phénomène ne se limite pas à la seule langue française : voir par exemple, en ce qui concerne la langue espagnole (castillan), l'article (en espagnol) [El lenguaje SMS ¿amenaza el español?](http://www.elcastellano.org/noticia.php?id=514), 24 janvier 2008, [elcastellano.org](http://www.elcastellano.org) :

¹⁰⁷. La limite technique est de 160 caractères, mais certains terminaux mobiles permettent l'envoi d'un message plus long, qui est en réalité envoyé en plusieurs messages (et peut donc être facturé en conséquence par l'opérateur téléphonique).

¹⁰⁸. Un bon exemple de plate-forme accueillant de nombreux férus de la « langue texto », Sky Blogs :
<http://www.skyrock.com/blog/>
Pour autant, le phénomène ne laisse pas indifférent, et les voix s'élevant contre cet usage de la langue écrite sont de plus en plus nombreuses, comme par exemple ici :
<http://vivelefrançais.skyrock.com/>
<http://sms.informatiquefrance.com/index.htm>

¹⁰⁹. [Le discours SMiste, un nouveau corpus ?](http://www.cahiers-pedagogiques.com/spip.php?article2166), Sabine Pétilon, 4 février 2006, Cahiers pédagogiques :
<http://www.cahiers-pedagogiques.com/spip.php?article2166>
[Le langage SMS : révélateur d'1compétence](http://www.sms4science.org/userfiles/le%20langage%20SMS%20r%C3%A9v%C3%A9lateur%20d%271c%20omp%C3%A9tence.pdf), Cédric Fairon, Jean René Klein, Sébastien Paumier, Les Cahiers du CENTAL :
<http://www.sms4science.org/userfiles/le%20langage%20SMS%20r%C3%A9v%C3%A9lateur%20d%271c%20omp%C3%A9tence.pdf>
[Les « SMS » ou « textos » signent-ils l'arrêt de mort de l'orthographe ?](http://www.curiosphere.tv/ressource/12485-comment-lorthographe-vient-aux-enfants/77066-les-sms-ou-textos-signent-ils-larret-de-mort-de-lorthographe) Curiosphere.tv :
<http://www.curiosphere.tv/ressource/12485-comment-lorthographe-vient-aux-enfants/77066-les-sms-ou-textos-signent-ils-larret-de-mort-de-lorthographe>

vivacité et de son adaptabilité face aux nouveaux usages liés aux nouvelles technologies.

En revanche, l'apparition et le développement de plusieurs formes de langage écrit sur l'internet risquent de favoriser les phénomènes communautaires, de cloisonner ou de stigmatiser certaines catégories de la population et de générer davantage d'exclusion quant à l'utilisation de l'internet et des technologies de l'information et de la communication.

Le Forum des droits sur l'internet recommande aux pouvoirs publics de se saisir de la question de l'utilisation de la langue sur les nouveaux supports de communication tels que l'internet et les réseaux mobiles et d'ouvrir ce débat avec notamment les utilisateurs et les spécialistes des technologies de la langue.

Le Forum recommande aux exploitants de sites internet privés permettant l'expression de leurs visiteurs au moyen notamment de forums de discussion d'adopter une position neutre et modérée quant aux différents usages de la langue par leurs membres, en s'efforçant de **limiter les phénomènes communautaires ou d'exclusion** liés à l'usage de la langue et de l'orthographe.

c. Les technologies de l'information et l'orthographe française

La question de l'orthographe de la langue française est un sujet d'actualité. Le livre polémique de François de Closets¹¹⁰, « Zéro faute », a ré-ouvert le débat en septembre 2009. L'auteur, partant du constat que l'orthographe est de moins en moins respectée de nos jours, prône un assouplissement des règles orthographiques, et insiste sur la complexité du vocabulaire, qui pourrait « léser les personnes dépourvues d'une bonne mémoire ».

Pourtant, avec le renouveau de la langue écrite, l'écriture est désormais assistée par l'ordinateur : des technologies de correction orthographique sont présentes dans les différents programmes informatiques, et notamment dans les suites bureautiques¹¹¹ et, depuis peu, dans les navigateurs internet¹¹². Mais il y a toujours autant de fautes d'orthographe dans les correspondances écrites¹¹³ ou dans la presse papier ou internet.

Il semble donc que les utilisateurs portent une confiance trop grande à l'égard de leurs outils de correction d'orthographe. Pourtant, ces technologies ne peuvent se substituer à la relecture. Elles ne peuvent ni détecter ni corriger toutes les fautes, et notamment celles de grammaire. Il faut donc insister sur le fait que **ces outils ne sont que des compléments** permettant de gagner du temps lors de la relecture du document, mais absolument pas de s'en affranchir.

Si l'on souhaite améliorer le niveau de qualité orthographique française, il y a donc plusieurs actions à mener auprès des différents acteurs.

¹¹⁰. Zéro faute - L'orthographe, une passion française, François de Closets, Mille et Une Nuits, septembre 2009, ISBN 978-2755501360.

¹¹¹. À titre d'exemple, le correcteur orthographique de *Microsoft Word* est disponible dans 126 langues et 233 variantes régionales, et le correcteur grammatical est disponible dans 6 langues et 61 variantes régionales. La suite *OpenOffice.org* intègre quant à elle 96 langues et variantes régionales.

¹¹². Le navigateur internet *Mozilla Firefox* inclut, depuis sa version 3.0, un correcteur orthographique actif lors de la saisie d'informations dans les champs des formulaires des pages internet. 18 langues de dictionnaires sont incluses dans la distribution publique, d'autres dictionnaires peuvent-être ajoutés sous forme d'extensions à partir de la page suivante : <https://addons.mozilla.org/fr/firefox/browse/type:3>

¹¹³. Les fautes d'orthographe et les langages abrégés ne sont pas une nouveauté : voir à ce titre l'article d'Alain Rey, J'ai déniché une trace de langage SMS en 1840, paru dans *Regards* n° 45, novembre 2007 : <http://www.regards.fr/article/?id=2994>

Le Forum recommande aux pouvoirs publics de renforcer leur collaboration avec les éditeurs d'outils de correction orthographique, de façon à veiller à l'amélioration constante de ces technologies, en intensifiant les recherches en analyse syntaxique automatisée et sur les correcteurs grammaticaux.

Il est par ailleurs recommandé à la société civile et notamment aux groupes de presse et aux autres éditeurs d'informations sur internet de ne pas s'appuyer exclusivement sur les outils de correction orthographique. Ces outils ne doivent être qu'un complément de la relecture humaine, qui demeure indispensable et qui doit être assortie de moyens conséquents proportionnés à l'importance de l'activité d'édition de contenus.

2. La création et la diffusion des nouveaux termes du français de l'internet en France

L'aménagement de la langue, en la dotant de termes capables de désigner les réalités nouvelles, s'inscrit dans une logique de développement durable, afin de préserver la diversité des langues et des cultures. Lorsqu'une langue ne dispose plus d'un vocabulaire de spécialité, c'est une partie de cette langue qui ne vit plus. Pour permettre à une langue de continuer à décrire la modernité, il convient donc de l'équiper au fur et à mesure de la nouveauté.

a. Les besoins des acteurs

La spécificité de l'internet (dans une plus grande mesure encore que pour certaines autres spécialités – comme l'économie ou la gestion) réside dans le fait que ce vocabulaire de spécialité est désormais revendiqué par le plus grand nombre et à rythme soutenu.

Beaucoup d'acteurs sont très sensibles aux problématiques d'aménagement de la langue. Un espace de débat et de travail collaboratif permettrait une large mobilisation des internautes et des industries de l'internet autour de ces questions.

Le Forum des droits sur l'internet appelle de ses vœux l'émergence d'une **initiative privée** autour d'un projet de site internet de discussion et de travail sur la langue, s'appuyant sur les bases terminologiques préexistantes¹¹⁴.

Idéalement, cet espace de discussion, opéré par un ou plusieurs acteurs privés, pourrait se structurer autour d'un wiki ou d'un forum de discussion. Il permettrait par ailleurs d'alimenter utilement la réflexion des pouvoirs publics et des commissions spécialisées sur les terminologies à prendre en compte.

b. Le dispositif public existant

La France est dotée d'un dispositif public de terminologie et de néologie, placé sous la tutelle du Premier ministre. Les origines de ce dispositif remontent à un décret de 1972¹¹⁵ dans lequel il est prévu des « commissions de terminologie » placées sous toute administration centrale qui en fait la demande.

Le dispositif a peu à peu évolué, puis a été entièrement revu par le décret du 3 juillet 1996 relatif à l'enrichissement de la langue française¹¹⁶, qui institue une commission générale coordonnant l'action des différentes commissions spécialisées (créées par arrêté

¹¹⁴. Cf. *infra* note 121 page 35.

¹¹⁵. Décret n° 72-19 du 7 janvier 1972 relatif à l'enrichissement de la langue française : http://www.legifrance.gouv.fr/jo_pdf.do?cidTexte=JORFTEXT00000879206

¹¹⁶. Décret n° 96-602 du 3 juillet 1996 relatif à l'enrichissement de la langue française : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000005621310&dateTexte=20091116>

ministériel) et introduit la nécessité de passer par un avis de l'Académie française avant toute publication.

Le dispositif¹¹⁷ est composé d'une vingtaine de commissions spécialisées, chacune composée de spécialistes bénévoles et d'un haut représentant de l'État. Ces commissions spécialisées se réunissent périodiquement et transmettent des propositions de termes à la Commission générale de terminologie et de néologie. Celle-ci examine les propositions sur la forme et sur le fond¹¹⁸ et peut proposer des modifications aux commissions spécialisées.

Les projets de termes sont transmis à l'Académie française, qui peut valider, amender, ou rejeter les propositions. Lorsque le terme reçoit l'accord de l'ensemble des partenaires, il est ensuite transmis pour accord au ministre compétent¹¹⁹, en fonction du domaine de la spécialité. La Commission générale publie au Journal officiel la liste des nouveaux termes validés au cours de ce long processus.

Il existe également une voie plus rapide, qui permet à l'administration, lorsqu'elle doit s'exprimer sur un sujet particulier, de solliciter directement la Commission générale d'une demande de traduction d'un terme existant dans une langue étrangère.

Les termes retenus et publiés au Journal officiel **s'imposent à l'administration en lieu et place de termes étrangers dans un esprit d'exemplarité**¹²⁰. Ce dispositif, mobilisant de nombreux acteurs de la société civile mais d'abord tourné vers les besoins de l'État, vise à lui permettre de s'exprimer en français, en toutes circonstances, comme cela est prévu par le droit positif.

Le dispositif, de par l'ampleur des spécialistes qui lui sont associés, constitue en France une ressource inégalée. Les termes produits pour l'administration sont utiles à la société civile en général, aux côtés d'autres initiatives terminologiques, comme celles des éditeurs de dictionnaires¹²¹.

Par ailleurs, toute réflexion concernant ce dispositif (pour améliorer les processus d'identification des besoins et de diffusion des travaux) doit être menée en appréciant **le cout d'opportunité** – en termes de cohésion sociale et d'incidence sur la valeur patrimoniale du français – que pourrait représenter une « perte de domaine » dans un champ lexical aussi stratégique que celui des sciences de l'information.

La base terminologique du dispositif est accessible en ligne sur le site FranceTerme¹²², mais reste peu connue du grand public, ce qui est regrettable. Les conditions juridiques d'utilisation de cette base de données sont inconnues et mériteraient d'être précisées.

Comme nous l'avons dit, le processus ne correspond pas seulement à un besoin de l'administration, mais à **une œuvre d'intérêt général**, en contribuant à

¹¹⁷. Pour en savoir plus sur le dispositif de terminologie et de néologie français, consulter : <http://franceterme.culture.fr/FranceTerme/terminologie.html>

¹¹⁸. La Commission générale comprend quatre scientifiques et est à même de faire des propositions sur le vocabulaire scientifique, qui constitue une part importante des propositions de termes.

¹¹⁹. À défaut de réponse du ministère, le terme est considéré comme validé passé un délai de un mois sans réponse.

¹²⁰. L'article 11 du décret n° 96-602 dispose que « *les termes et expressions publiés au Journal officiel sont obligatoirement utilisés à la place des termes et expressions équivalents en langues étrangères* », mais les termes sont publiés à la rubrique « avis divers » du Journal officiel et non plus sous forme d'arrêté. D'autre part, aucune sanction n'est prévue si les termes ne sont pas utilisés. Enfin, en pratique, de nombreux agents publics ignorent l'existence même du dispositif.

¹²¹. Parmi les éditeurs proposant des bases terminologiques libres d'utilisation, citons notamment Trésor de la langue française informatisé (TLFI), et le projet *Wiktionnaire* de Wikipédia : <http://atilf.atilf.fr/tlf.htm>
<http://fr.wiktionary.org/>

¹²². Voir en ligne le site FranceTerme : <http://franceterme.culture.fr/FranceTerme/>

l'enrichissement de la langue. Il apparaît important, par conséquent, d'assurer au site FranceTerme un référencement et une visibilité optimaux, moyennant sans doute des partenariats à tous les stades du processus de création des mots.

Il est recommandé d'améliorer le fonctionnement technique du site « FranceTerme » en procédant à des ajustements sur les points suivants :

- fournir un format structuré de données¹²³ directement réutilisable et permettant l'automatisation des processus d'exploitation de la base terminologique ;
- améliorer l'accessibilité du site conformément aux recommandations internationales pour l'accessibilité ;
- améliorer le fonctionnement des flux de syndication (flux *RSS*) en rendant le flux chronologique et détectable par les navigateurs internet ;
- permettre un accès direct à un mot et à sa définition au moyen de son *URL*.

Il est également recommandé de préciser clairement le régime juridique d'utilisation de la base de données terminologique FranceTerme.

Par ailleurs, afin d'améliorer la visibilité des termes validés par l'administration, le Forum recommande aux pouvoirs publics de renforcer le suivi des termes lorsqu'ils sont validés, en développant des partenariats avec des éditeurs et relais de diffusion.

Le Forum attire l'attention du Conseil Supérieur de l'Audiovisuel (CSA) sur le rôle prescripteur des journalistes de la radio et de la télévision en matière de langue et sur l'intérêt que puissent être utilisés des mots francophones dans les médias, ce qui répond à un besoin de cohésion sociale important.

D'autres ajustements peuvent être envisagés : la boîte à idées¹²⁴, par exemple, est une simple boîte de contact par courrier électronique et ne donne pas lieu à un suivi public sur le site FranceTerme, faute de ressources. Elle permet de recueillir environ 1 200 messages d'internautes par an, dont 300 environ constituent une proposition intéressante, soit parce que le terme est déjà en cours d'étude dans le dispositif d'enrichissement, ce qui confirme son intérêt, soit parce qu'il peut être transmis aux commissions spécialisées. Il n'y a pas de suivi public sur les propositions déposées par les internautes.

Le Forum recommande aux pouvoirs publics de mettre en place une « boîte à idées » publique, ouverte aux commentaires et permettant aux utilisateurs ou aux spécialistes de déposer des demandes de nouveaux termes en langue française.

En outre, il est souvent impossible de faire des propositions directement auprès des commissions spécialisées, qui, pour la plupart, n'ont pas d'existence sur la toile. La Commission spécialisée de terminologie et de néologie de l'informatique et des composants électroniques (CSTIC), qui est en partie compétente pour les notions et principes de l'internet, dispose d'un site internet¹²⁵, mais celui-ci n'est plus actualisé, et ne donne aucun moyen à l'internaute d'apporter sa contribution¹²⁶. Il présente les actualités de la Commission par ordre chronologique, à l'inverse de ce qui se fait habituellement.

¹²³. Comme par exemple le format *XML*.

¹²⁴. La boîte à idées du site FranceTerme :
<http://franceterme.culture.fr/FranceTerme/boiteidees.html>

¹²⁵. Le site de la CSTIC :
<http://ensmp.net/cstic/>

¹²⁶. Le lien « Participez aux études préparatoires » est inopérant :
<http://ensmp.net/cstic/etudes/participez.html>

Le Forum recommande aux commissions spécialisées de terminologie et de néologie de s'efforcer de mettre en œuvre un contact permanent avec les administrés, *via* une plateforme internet qui pourrait être opérée et mutualisée par la Commission générale.

L'adhésion des usagers de la langue aux choix retenus dans le cadre du dispositif auquel l'État a recours est évidemment à rechercher puisque l'intérêt du processus, tout en contribuant à l'enrichissement de la langue, est de faciliter l'accessibilité des notions nouvelles par le public. Plusieurs exemples récents montrent, en l'occurrence, que l'appropriation d'un vocabulaire francophone par les internautes est tout à fait possible.

3. L'accessibilité de la langue sur internet

La langue permet l'accès à l'information, il est donc nécessaire de l'adapter afin d'optimiser, pour le lecteur, l'accessibilité de l'information mise en ligne.

a. La langue permet l'accès à l'information...

Réaliser de multiples traductions d'un document et le rendre disponible dans plusieurs langues, c'est avant tout améliorer son accessibilité. Toutefois, contrairement à une diffusion « papier » traditionnelle, les lecteurs de l'information sont plus difficiles à identifier : potentiellement, le monde entier peut accéder à une même page internet.

Notons cependant que, contrairement à l'accessibilité numérique, qui impose notamment aux pouvoirs publics de rendre leurs diffusions audiovisuelles accessibles à tous en suivant les recommandations internationales pour l'accessibilité¹²⁷, le défaut de traduction ne pourra pas être sanctionné, et ne constitue pas une discrimination. L'intelligibilité de la langue n'est donc pas une obligation, mais n'en demeure pas moins fortement recommandée : **la langue est une composante majeure de l'accessibilité.**

À titre d'exemple, un document en anglais soutenu ne sera lisible que par des personnes maîtrisant bien cette langue (environ 500 millions), tandis que s'il est écrit avec un vocabulaire et une grammaire restreints, il pourra être potentiellement compris par plus de deux milliards de personnes. Ainsi, *Voice Of America*¹²⁸, une chaîne de radio américaine, diffuse des bulletins d'informations en anglais simplifié, pour s'adresser à un public international le plus large possible.

Le bulletin d'informations quotidien en français facile¹²⁹, diffusé tous les jours à 21 heures temps universel par Radio France International, est un autre exemple de programme mettant en avant son accessibilité linguistique pour toucher un public le plus large possible.

En outre, d'autres obstacles peuvent se poser à la compréhension d'un texte : il s'agit de la technicité et de la spécialisation du vocabulaire employé. Une page d'aide du site des impôts relative à l'utilisation d'un certificat électronique devra, pour remplir son objectif, utiliser une terminologie très « grand public », malgré la technicité du sujet, pour être comprise par tous les utilisateurs du service.

¹²⁷. Voir la Recommandation du Forum des droits sur l'internet « Internet et développement durable I : l'accessibilité des services de communication publique en ligne du secteur public » adoptée le 25 novembre 2008 :

http://www.foruminternet.org/spip.php?action=redirect&id_article=2808

¹²⁸. Voir le site de la chaîne de radio *Voice Of America* :

<http://www.voanews.com/>

¹²⁹. Description du programme sur le site de Radio France International (RFI) :

http://www.rfi.fr/lffr/statiques/accueil_apprendre.asp

La rédaction d'un tel article ne pourrait donc pas être confiée exclusivement à un technicien mais devra être relue et améliorée par un responsable éditorial conscient des enjeux de l'accessibilité linguistique et du public visé.

- b. ...mais il est nécessaire d'adapter le niveau d'accessibilité linguistique à son public

L'utilisation de termes trop techniques peut donc rendre l'information inaccessible aux internautes non techniciens. C'est malheureusement un cas trop fréquent sur l'internet et une opportunité manquée de s'ouvrir à de nouveaux utilisateurs ou clients.

L'expérience linguistique de Debian France¹³⁰ est ici particulièrement enrichissante : l'association a pour buts le soutien et la promotion du Projet Debian en France, un système d'exploitation qui est essentiellement utilisé par une communauté d'utilisateurs très technicienne, puisque le logiciel a fait ses preuves en matière de serveurs internet.

Cependant, la volonté stratégique de l'association est de rester ouverte au plus grand nombre, et notamment aux utilisateurs qui ne sont pas forcément des spécialistes de l'informatique ou des réseaux.

Au-delà de l'effort de traduction considérable qui a été mené par les bénévoles de l'association, et qui leur permet de proposer aujourd'hui à la communauté un disque d'installation de leur logiciel en 41 langues, représentant 55 % des locuteurs de la planète, une attention toute particulière a été portée sur la documentation du logiciel : tout en restant très technique, les idiomes et terminologies utilisés dans celle-ci sont simples d'accès pour le plus grand nombre. Elle peut être comprise par la majorité des utilisateurs et ne requiert pas de compétences techniques particulières.

Le problème se posera chaque fois que le rédacteur du contenu n'a pas le même niveau de spécialité que le lecteur : ce sera le cas lorsqu'un adulte rédige un contenu à destination des plus jeunes ou encore lorsqu'un juriste écrit un article à destination d'un public non-juriste. Faute d'identifier clairement le public visé, et la terminologie la plus adaptée à ce public, le document ne sera pas accessible.

La relecture par un spécialiste de la langue est donc conseillée lorsque l'on souhaite toucher un public particulièrement bien identifié, comme par exemple, les adolescents. Le travail de mise en accessibilité linguistique s'apparente alors à celui d'une traduction.

Le Forum recommande aux exploitants de sites internet de mener une réflexion en amont de tout projet internet sur l'accessibilité du vocabulaire employé et sur les publics visés par les contenus publiés en ligne.

Il est recommandé d'adapter le vocabulaire du projet au public visé, en prenant soin d'éviter l'emploi d'un vocabulaire technique ou fortement spécialisé lorsque c'est le grand public qui est visé.

Il est fortement déconseillé aux exploitants de sites internet de confier tout ou partie des tâches rédactionnelles aux techniciens, spécialistes ou créatifs en charge du projet internet. Le Forum souhaite insister sur l'importance d'une équipe spécifique et autonome en charge des aspects rédactionnels pour tout projet internet.

La question de l'utilité d'un glossaire peut se poser lorsqu'il y a une abondance de termes techniques. Utilisé modérément et à bon escient, il peut s'avérer utile, mais il ne devrait être nécessaire qu'en cas d'impossibilité manifeste de trouver un terme plus adapté, car tous les visiteurs ne feront pas la démarche d'aller le consulter.

¹³⁰. *Debian* est un système d'exploitation libre apprécié des équipes techniques en charge des serveurs internet, du fait de sa robustesse : <http://www.debian.org/>

L'utilisation à ce titre d'*infobulles*¹³¹ est la bienvenue, car elle permet au visiteur d'avoir accès à une définition ou à une précision sans avoir à sortir du contexte du terme qui pose problème, c'est-à-dire sans quitter la page en cours pour charger une page séparée spécifique au glossaire. Il faudra cependant veiller à utiliser un mécanisme d'affichage accessible à tous¹³².

Au niveau des sites internet des services publics, un **effort particulier de simplification** est mené sur les formulaires administratifs, notamment dans le cadre de la mise en ligne progressive de ces documents sur le nouveau portail personnalisé de l'administration, mon.service-public.fr¹³³. Les formulaires sont revus par la Direction générale de la modernisation de l'État (DGME), qui est chargée du projet.

Il existait jusqu'à récemment¹³⁴ un conseil spécialisé pour cette tâche, le Comité d'orientation de simplification du langage administratif¹³⁵ (COSLA), placé sous la double tutelle du ministre chargé de la réforme de l'État et du ministère de la Culture. Il proposait une aide aux administrations pour l'amélioration de l'accessibilité linguistique des documents et formulaires administratifs, de façon à les rendre plus compréhensibles du grand public.

Aujourd'hui, la Délégation générale à la langue française et aux langues de France¹³⁶ (DGLFLF) n'est plus en mesure d'intervenir sur les aspects linguistiques des documents administratifs pour en améliorer l'accessibilité.

Le Forum des droits sur l'internet recommande aux pouvoirs publics une collaboration plus étroite entre la DGME en particulier et les services de l'administration en général, avec les services de la DGLFLF, dans le cadre de l'élaboration et de l'amélioration linguistique des documents administratifs et notamment des formulaires des services publics mis en ligne sur internet.

¹³¹. Encart apparaissant automatiquement dans la fenêtre de navigation en cours lorsque l'utilisateur survole avec le curseur de sa souris un mot spécifique du texte.

¹³². Les balises de programmation des pages internet *HTML* prévoient la possibilité d'afficher simplement un texte bref permettant de préciser un acronyme ou une abréviation : balise `<Acronym>` ou `<abbr>`. Ces balises sont compatibles avec toutes les recommandations pour l'accessibilité. D'une manière générale, on proscriera le recours à tout langage de script tel que *Javascript* pour automatiser l'affichage de ces *infobulles*.

¹³³. Voir le site internet mon.service-public.fr (identification nécessaire) :

<https://mon.service-public.fr/portail/>

¹³⁴. Le mandat du COSLA n'a pas été renouvelé et il a disparu juridiquement avec l'application du décret de 2006.

¹³⁵. Voir une présentation succincte du COSLA et de ses activités :

http://www.fonction-publique.gouv.fr/IMG/pdf/cosla_151203.pdf

¹³⁶. Organe de réflexion, dévaluation et d'action, la DGLFLF, rattachée au ministère de la Culture, élabore la politique linguistique du Gouvernement en liaison avec les autres départements ministériels, anime et coordonne l'action des pouvoirs publics pour la promotion et l'emploi du français et veille à favoriser son utilisation comme langue de communication internationale. Elle s'efforce de valoriser les langues de France et de développer le plurilinguisme.

<http://www.dglf.culture.gouv.fr/>

II. – LA MISE EN ŒUVRE DE BONNES PRATIQUES LINGUISTIQUES SUR INTERNET

Cette partie vise à présenter aux acteurs une série de bonnes pratiques afin d'améliorer le multilinguisme et l'accessibilité linguistique sur internet. Le respect des normes techniques existantes est un préalable (A), mais cela ne suffit pas : la gestion d'un projet multilingue peut être complexe et certains conseils semblent utiles à rappeler à l'égard des exploitants de sites internet (B).

A – Les normes techniques et la normalisation autour de la langue sur internet

1. La traduction des messages d'erreur des serveurs internet

La question de la traduction des messages d'erreur rencontrés sur internet a déjà été posée¹³⁷. Elle implique un travail d'identification des messages pouvant poser problème aux utilisateurs et l'élaboration d'une réponse à ce besoin. L'AFNOR a été sollicitée par les pouvoirs publics sur ce sujet.

Le Forum appelle de ses vœux une participation la plus large possible de la communauté d'utilisateurs aux travaux de l'AFNOR sur le sujet de la traduction des messages d'erreur rencontrés sur internet¹³⁸.

2. Uniformisation du répertoire d'étiquetage des langues utilisé sur l'internet

L'étiquetage d'une langue permet de désigner celle-ci par un acronyme court. La liste de ces différents codes constitue un répertoire. Il existe plusieurs répertoires d'étiquetage des langues, tels que la norme ISO 639-3, qui comprend plus de 7 000 langues, mais qui n'est pas « stable », une langue pouvant être retirée de façon unilatérale, ce qui rend obsolète le codage déjà réalisé sur des pages internet. Par ailleurs, cette norme ne permet pas de spécifier les dialectes, ni les alphabets de caractères¹³⁹.

D'autres normes techniques sont déjà utilisées dans les protocoles XML, HTML, et HTTP¹⁴⁰, qui reposent tous sur la norme technique IETF RFC 5646¹⁴¹.

Le Forum recommande aux exploitants de sites internet et aux éditeurs de logiciels d'utiliser le répertoire d'étiquetage de langue basé sur la norme IETF RFC 5646, déjà utilisé couramment dans les protocoles et formats de page de l'internet.

¹³⁷. Cf. *supra* « Le cadre juridique de la langue française », page 19.

¹³⁸. Les travaux de l'AFNOR se déroulent généralement en deux phases : l'enquête publique, qui est ouverte à tous les participants, et la discussion en commission, qui nécessite l'adhésion de la structure participante auprès de l'AFNOR. L'article 14 du décret n° 2009-697 du 16 juin 2009 relatif à la normalisation et qui entrera en vigueur au 1^{er} janvier 2010 exonère certaines associations et entreprises de moins de 250 salariés des frais d'adhésion : <http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020749979&fastPos=2&fastReqId=622438787&categorieLien=id&oldAction=rechTexte>

¹³⁹. Une même langue peut s'écrire en plusieurs alphabets très différents, comme par exemple, l'azéri.

¹⁴⁰. Protocoles *eXtensible Markup Language* (langage de balisage extensible), *Hypertext Markup Language* (langage de balisage hypertexte) et *Hypertext Transfer Protocol* (protocole de transfert hypertexte).

¹⁴¹. Voir le standard IETF RFC 5646 (en anglais) : <http://tools.ietf.org/html/rfc5646>

3. Les noms de domaine en caractères étendus sur les domaines de tête nationaux

La situation au niveau mondial est à distinguer de la situation particulière française.

a. Les normes techniques, matériels et logiciels sont-ils prêts ?

La plupart des offices gérant des registres locaux¹⁴² (ccTLD) ou génériques (dont *.com*, *.net*, *.org*, *.info* sont les plus connus) permettent aux déposants d'enregistrer un nom de domaine composé de « caractères étendus », c'est-à-dire tous les caractères qui ne sont pas compris dans la table de 128 caractères¹⁴³ de la norme ASCII¹⁴⁴ datant de 1961, comme par exemple les caractères accentués, les symboles ou les unités de monnaie, et certains signes de ponctuation. Le domaine *.eu a*, quant à lui, ouvert le dépôt de noms de domaine en caractères étendus le 10 décembre¹⁴⁵.

La norme Unicode¹⁴⁶, qui permet l'écriture dans la majorité des langues et des jeux de caractères présents sur la planète, peut être utilisée dans le domaine du nommage¹⁴⁷ : il est dès lors possible d'enregistrer et d'utiliser¹⁴⁸ un nom de domaine comprenant des caractères accentués, ou encore d'autres jeux de caractères tels que le chinois ou l'arabe.

Cette faculté revêt une importance particulière dans les pays dont la langue s'écrit dans un jeu de caractères différent de l'alphabet latin. Les populations utilisant ces jeux de caractères ne maîtrisent que très peu ou mal l'alphabet latin et ne peuvent donc utiliser facilement des adresses internet composées exclusivement de caractères ASCII.

Un obstacle vient d'être levé au niveau international : l'utilisation d'extensions de premier niveau en jeux de caractères étendus. Aujourd'hui encore, même s'il est possible d'utiliser un nom de domaine composé de caractères en alphabet chinois par exemple, l'internaute devra saisir l'extension « *.cn* » correspondant au registre chinois en caractères latins, ce qui constitue une barrière linguistique et technique.

Selon une annonce récente¹⁴⁹ de l'ICANN (*Internet Corporation for Assigned Names and Numbers*), le déploiement des extensions de premier niveau en caractères étendus « IDN » (*International Domain Names*) est prévu prochainement. Il devrait être facilité par le plan *FastTrack* de l'ICANN, qui est ouvert depuis le 15 novembre 2009¹⁵⁰, et qui permet aux pays utilisant des jeux de caractères non-latins de faire une demande prioritaire d'une extension de tête en caractères étendus.

¹⁴². Voir par exemple les registres locaux *.ch* ou encore *.de*, qui acceptent et utilisent des noms de domaines en jeux de caractères étendus.

¹⁴³. Sur ces 128 caractères, il n'y a en fait que 95 caractères affichables : les lettres de l'alphabet latin en majuscule et en minuscule, les chiffres arabes, les signes de ponctuation et d'arithmétique basiques et quelques symboles tels que « @ », « # », « & », « \$ », « / », « \ », « | » ou « _ ».

¹⁴⁴. Pour plus d'informations sur la norme ASCII, voir :

http://fr.wikipedia.org/wiki/American_Standard_Code_for_Information_Interchange

¹⁴⁵. Voir le communiqué de presse de l'EURid et la liste des nouveaux caractères autorisés :

<http://www.eurid.eu/fr/noms-de-domaine-eu/les-idns-pour-le-eu>

¹⁴⁶. Pour plus d'informations sur la norme UNICODE, voir :

<http://fr.wikipedia.org/wiki/Unicode>

¹⁴⁷. Pour plus de détails, consulter le standard IETF RFC 3490 :

<http://tools.ietf.org/html/rfc3490>

¹⁴⁸. On pourra par exemple utiliser l'URL www.nestlé.com depuis un navigateur récent pour accéder au site du groupe.

¹⁴⁹. Voir le communiqué de presse du 30 septembre 2009 relatif au plan *FastTrack* IDN sur le site de l'ICANN :

<http://www.icann.org/fr/annoncements/announcement-2-30sep09-fr.htm>

¹⁵⁰. Formulaire de dépôt d'une demande du programme *FastTrack* :

<http://www.icann.org/en/topics/idn/fast-track/>

Au niveau des logiciels, la majorité des navigateurs internet récents permettent d'utiliser des caractères Unicode dans la barre d'adresse pour accéder directement à un site internet.

Le Forum considère que le bon fonctionnement et l'utilisation de noms de domaine et d'extensions de premier niveau en caractères étendus (Unicode) sont une condition nécessaire pour assurer l'accessibilité de toutes les populations à l'internet et leur intégration dans le monde numérique. Il est par conséquent urgent de lever les derniers obstacles techniques empêchant l'utilisation généralisée de ces jeux de caractères étendus dans le système international du nommage.

b. Enjeux du nommage en Unicode sur les extensions nationales de premier niveau

La langue française, qui repose sur l'alphabet latin, utilise néanmoins un certain nombre de caractères dits diacritiques, c'est-à-dire qu'ils sont composés d'une lettre et d'un signe d'accentuation : accent grave, aigu ou circonflexe ou encore tréma ou cédille. Il existe par ailleurs certaines ligatures utilisées en français qui ne sont pas présentes dans l'alphabet latin, comme le æ (d'origine médiévale et appliquée en français à certaines expressions latines, par exemple curriculum vitæ ou Lætitia) ou le œ (œuf).

Ces caractères, fréquents en français, et qui ne sont pas représentés dans la table de caractères ASCII, ne sont aujourd'hui pas utilisables dans les noms de domaine des registres nationaux¹⁵¹, alors que tous les obstacles techniques semblent être levés. Le problème n'a pas été évoqué lors de la récente consultation publique sur le nommage¹⁵² et n'a d'ailleurs pas suscité de réactions de la part des participants.

La volonté d'ouvrir le dépôt des noms de domaine en caractères étendus se heurte en effet à la volonté de certains acteurs économiques, arguant notamment que cette libéralisation provoquerait un risque accru de contrefaçon de marque sur internet, ainsi que de cybersquat et de typosquat de noms de domaine, l'internaute pouvant soi-disant plus difficilement identifier la différence entre un nom de domaine frauduleux en caractères accentués et le nom de domaine original en caractères ASCII.

Les acteurs économiques avancent par ailleurs le fait que cette ouverture du nommage présenterait un coût non négligeable pour les entreprises, qui devraient multiplier par deux ou trois le budget annuel consacré à leur parc de noms de domaine.

Pourtant, il paraît opportun de pouvoir utiliser la langue française dans son ensemble en matière de nommage, ce qui permettrait de simplifier l'utilisation des noms de domaine pour les internautes non avertis, qui utilisent souvent des caractères étendus dans la barre d'adresse de leur navigateur. Il y a par ailleurs un risque que les accents français disparaissent de l'internet s'ils ne peuvent être utilisés dans les noms de domaine nationaux.

Le Forum recommande aux pouvoirs publics et à l'office d'enregistrement compétent de permettre le dépôt et l'utilisation de noms de domaine en caractères étendus sur les registres nationaux (.fr, .re).

Afin de limiter les tentatives de cybersquat, de typosquat ou de contrefaçon, il pourrait être envisagé de limiter le dépôt de noms de domaine comprenant des caractères

¹⁵¹. Voir la seule déclaration publique de l'AFNIC à ce sujet, Avertissement de l'AFNIC sur : les noms de domaine internationalisés (IDN), février 2003 : <http://www.afnic.fr/actu/nouvelles/nommage/NN20030217>

¹⁵². Résultats de la consultation publique sur les domaines internet .fr et ultramarins : <http://www.telecom.gouv.fr/rubriques-menu/organisation-du-secteur/dossiers-sectoriels/noms-domaine/domaines-internet--dot-fr-ultramarins-resultats-consultation-1888.html>

accentués aux seuls titulaires des noms de domaine équivalents en caractères ASCII. Ces questions devront donc être étudiées par l'office d'enregistrement compétent lors de l'ouverture du dépôt de noms de domaine en caractères étendus.

Exemple : Seule la marque titulaire du nom de domaine nestle.fr pourrait déposer le nom de domaine nestlé.fr.

La mise en place des extensions de premier niveau (IDN) en jeux de caractères étendus n'a, quant à elle, que peu d'impact pour la France, seule l'extension .re (île de la Réunion) pouvant poser problème et être doublée d'un .ré.

4. Améliorer l'accès aux ressources linguistiques et notamment aux dictionnaires

Les dictionnaires en format ouvert sont une ressource critique car ils permettent le développement de nouveaux outils de la langue (traduction notamment) et l'intégration de nouvelles langues dans les modules linguistiques des différents logiciels, comme par exemple une suite bureautique qui disposerait d'une fonction de correction d'orthographe.

Le Forum des droits sur l'internet encourage les pouvoirs publics et les éditeurs à améliorer l'accès aux ressources linguistiques en ligne et notamment aux dictionnaires, *via* un protocole ouvert et documenté dans une norme publiquement disponible, et déjà disponible dans un logiciel libre, tel que le protocole DICT¹⁵³ issu de la norme technique IETF RFC 2229¹⁵⁴.

Le protocole doit pouvoir permettre l'interrogation et l'accès aux ressources du dictionnaire par des programmes, et pas seulement de façon interactive, par un humain.

5. Le clavier français est-il adapté à la langue française dans un contexte fortement multilingue ?

Comme il a été expliqué précédemment¹⁵⁵, le français écrit comporte un certain nombre de caractères accentués, qu'il n'est pas toujours possible d'utiliser facilement avec un clavier français.

Citons l'exemple des majuscules accentuées, dont l'emploi est prescrit par l'Académie, et qui ne sont pas présentes sur la majorité des claviers français, les ligatures æ et œ, les guillemets français à « double chevrons », les apostrophes courbes françaises, ou encore les trémas sur certaines voyelles comme le Ÿ. Pour l'ensemble de ces caractères, il est généralement impossible de les saisir directement au clavier, et il est nécessaire de recourir à un logiciel adapté¹⁵⁶ pour les obtenir.

L'utilisation dans des conditions normales des caractères étendus et spécifiques à la langue française répond à une demande sociale¹⁵⁷, elle est d'ailleurs particulièrement

¹⁵³. Pour plus d'informations sur le protocole DICT, consulter l'article [RFC 2229 : A Dictionary Server Protocol](#), Stéphane Bortzmeyer, 1er mai 2007 : <http://www.bortzmeyer.org/2229.html>

¹⁵⁴. Le standard IETF RFC 2229 : <http://tools.ietf.org/html/rfc2229>

¹⁵⁵. Cf. *supra* « Les noms de domaine en caractères étendus sur les domaines de tête nationaux », page 41

¹⁵⁶. Les modules de correction orthographiques de certains logiciels de traitement de texte ajoutent par exemple automatiquement les ligatures ou les accents sur les majuscules.

¹⁵⁷. Voir par exemple le site de l'association « Mon nom accentué » : <http://accentuez.mon.nom.free.fr/>

forte en ce qui concerne les noms patronymiques, qui ne sont généralement pas accentués¹⁵⁸.

Au-delà des difficultés pour obtenir certains caractères de la langue écrite en français, il n'existe aucune norme technique ou spécification officielle du clavier français¹⁵⁹, bien qu'il y ait eu des tentatives de normalisation¹⁶⁰. En fonction du fabricant du matériel, les touches seront disposées différemment sur le clavier¹⁶¹. La position de l'arobase « @ », par exemple, varie d'un fabricant à l'autre.

Pourtant, nombreux sont les pays à disposer d'une norme de clavier, ce qui garantit l'homogénéité du placement des touches sur les claviers vendus dans le pays. Citons pour l'exemple le Canada, qui dispose d'une norme de clavier multilingue (et dont les premières touches sont disposées en QWERTY).

Le Forum recommande aux pouvoirs publics, aux fabricants de matériels et aux éditeurs et auteurs de logiciels de solliciter les organismes de normalisation pour réaliser une norme française des claviers dont le respect serait obligatoire (norme de type NF) et dont les objectifs seraient :

- d'unifier le positionnement de certaines touches sur les claviers vendus en France ;
- de permettre la saisie de l'ensemble des caractères étendus français et notamment des majuscules accentuées sans qu'il soit nécessaire de recourir à un dispositif logiciel particulier.

6. L'utilisation des normes d'encodage les plus récentes

L'internet a connu une grande diversité de normes techniques permettant l'utilisation de caractères de plus en plus variés. Chaque caractère doit en effet être « codé » dans le langage binaire propre à l'informatique pour pouvoir être utilisé et affiché sur des matériels informatiques et il est donc nécessaire d'utiliser un format d'encodage de ces caractères.

Aujourd'hui, les différentes normes d'encodage de caractères de type UTF¹⁶² (UTF-8¹⁶³, UTF-16¹⁶⁴ notamment) permettent l'utilisation sur l'internet et les réseaux de communication d'une grande variété de jeux de caractères et d'alphabets tels que décrits dans la norme Unicode.

Il demeure que ces normes d'encodage ne sont pas systématiquement utilisées par les éditeurs de logiciels et de sites internet¹⁶⁵.

¹⁵⁸. Les caractères pourront être pris en compte avec leurs signes diacritiques dans le Répertoire national d'identification des personnes physiques (RNIPP) à partir du 1er juillet 2010. La prise en compte des nouvelles entrées (essentiellement des naissances) dépend d'une évolution de la prise en charge des signes diacritiques sur les claviers et dans les logiciels des maternités jusqu'aux officiers d'état-civil. Les conditions d'un rattrapage du stock de noms patronymiques ne sont pas encore connues, les signes diacritiques des prénoms devraient être rétablis de façon automatique.

¹⁵⁹. Mais il existe des propositions alternatives de claviers, non officielles : voir par exemple le clavier bépo : <http://accentuez.mon.nom.free.fr/Clavier-bepo.html>

¹⁶⁰. *Le clavier bureautique et informatique*, Yves Neuville. Cedic/Nathan, 1985

¹⁶¹. La norme internationale ISO/IEC 9995 définit cependant la structure de base et les différentes « zones » des claviers : http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=51644

¹⁶². UTF est l'abréviation de *Universal Multiple-Octet Coded Character Set Transformation Format*.

¹⁶³. Présentation du format d'encodage UTF-8 sur Wikipédia : <http://fr.wikipedia.org/wiki/UTF-8>

¹⁶⁴. Présentation du format d'encodage UTF-16 sur Wikipédia : <http://fr.wikipedia.org/wiki/UTF-16>

¹⁶⁵. Pour l'exemple, citons le cas des caractères accentués qui ne fonctionnent pas sur la version pour mobile du site communautaire *twitter*.

En ce qui concerne la messagerie électronique, qui est un cas particulier du fait de l'ancienneté des protocoles utilisés¹⁶⁶, il est depuis longtemps possible d'utiliser des caractères Unicode dans les messages eux-mêmes¹⁶⁷. En revanche, la faculté d'utiliser des adresses électroniques comportant des caractères accentués ou non-latins est seulement très récente¹⁶⁸.

Le Forum recommande aux exploitants de sites internet, de réseaux sociaux, et aux éditeurs et auteurs de logiciels, d'utiliser systématiquement les normes d'encodage les plus récentes afin de s'assurer de la portabilité, à l'avenir, de leur applicatif dans différents langages en jeux de caractères Unicode, quel que soit le terminal utilisé.

B – Comment gérer et mettre en œuvre le multilinguisme sur internet ?

Cette partie a pour but d'aider les décideurs et les chefs de projet internet dans la prise en compte de l'enjeu linguistique de leurs projets en proposant une série de « bonnes pratiques ».

1. Gérer un projet internet multilingue

a. Comment organiser un service de traduction ?

La traduction peut être réalisée par « couches successives » comme c'est souvent le cas sur la bibliothèque communautaire Wikipédia : dans cette hypothèse, la première passe de traduction est effectuée par un outil d'aide à la traduction. Dans la deuxième passe, la traduction « littérale » est effectuée par un traducteur non connaisseur du sujet. La troisième passe correspond au « raffinement » de la traduction par un expert du sujet, qui y consacrera peu de temps.

Le Forum recommande aux exploitants de sites internet de faire intervenir différents types de personnels (traducteurs, spécialistes) sur les traductions des contenus pour optimiser les coûts et garantir un bon niveau de traduction.

La distinction entre les contenus et les éléments de l'interface est importante : faut-il distinguer les deux éléments dans l'organisation de la traduction, et proposer d'une part des traductions de l'interface de navigation, et, d'autre part, des traductions du contenu ?

Le Forum recommande aux exploitants de sites internet de privilégier la traduction des éléments de contenu, la seule traduction des éléments d'interface n'apportant pas de valeur ajoutée significative au visiteur.

À quel niveau du projet intervient le service de traduction ? S'il intervient trop en amont, les travaux de traduction nécessiteront de multiples révisions au fur et à mesure de l'évolution du projet. S'il intervient trop en aval du projet, les traductions pourront nécessiter de nouveaux développements imprévus (dans le cas où l'applicatif est mal adapté au contexte multilingue, ou encore que les mots sont trop longs pour être intégrés dans une interface graphique spécifique).

¹⁶⁶. Les origines du protocole de messagerie électronique remontent à 1965 : http://fr.wikipedia.org/wiki/Courrier_électronique

¹⁶⁷. Grâce au standard MIME (*Multipurpose Internet Mail Extensions*) : <http://fr.wikipedia.org/wiki/MIME>

¹⁶⁸. Voir à ce propos l'explication technique et les normes concernées : <http://www.bortzmeyer.org/courrier-entierement-internationalise.html>

Le Forum recommande aux exploitants de sites internet de **prendre en compte l'enjeu linguistique dès le départ de tout projet internet**, mais d'effectuer la réalisation des traductions au fil du projet.

b. La distinction entre les contenus statiques et les contenus périodiques (actualités)

La traduction des seuls éléments statiques d'un site internet a-t-elle une valeur ajoutée ? La réponse dépend de la nature du site. S'il s'agit d'un site « vitrine », la traduction des contenus fixes sera utile et le cout dans le temps de la traduction sera faible.

Si les éléments d'actualité constituent l'attrait principal du site pour les visiteurs, il faut se poser la question de l'intérêt de traduire les éléments fixes du site internet. En effet, un site dont les contenus périodiques ne seront pas traduits n'accueillera pas de visiteurs étrangers. La traduction des éléments de contenu fixes ne présentera qu'une faible valeur ajoutée.

Le Forum invite les exploitants de sites internet dont le principal intérêt réside dans l'actualisation régulière et périodique des contenus à se poser la question de l'utilité de la traduction de certains contenus fixes pris isolément.

Le cout de la traduction des interfaces et des éléments statiques est relativement faible comparé à la traduction des contenus périodiques, qui nécessite une organisation particulière de la chaîne de publication et une équipe permanente dédiée à la traduction. La possible externalisation de la traduction des contenus périodiques est plus délicate à mettre en œuvre, elle nécessitera une bonne coopération entre les ressources disponibles en interne, et le prestataire.

Le Forum tient à souligner que la traduction de contenus actualisés périodiquement et régulièrement peut exiger d'importantes ressources financières ou humaines.

En matière de traduction sur un site ouvert au public, il est recommandé de privilégier la qualité ; une traduction de mauvaise qualité donnera en effet **une image négative** de l'organisme à ses visiteurs. Lorsque les volumes sont trop importants, il paraît donc plus opportun de préférer une traduction plus synthétique mais de qualité.

Le Forum recommande aux exploitants de sites internet de préférer le recours à une traduction synthétique mais de qualité (résumés) plutôt que de réaliser une traduction intégrale de moins bonne qualité.

c. Gérer une communauté de développeurs multinationaux

Il est de plus en plus fréquent de recourir à des développeurs de différentes nationalités dans des projets de développement internet (*outsourcing*, délocalisation, projets à code source ouvert basés sur le bénévolat). Si le recours à une langue minimum pour les échanges semble s'imposer, il faut s'efforcer de permettre à chacun de contribuer.

En revanche, si le projet est suffisamment large (milliers de contributeurs), il sera possible d'utiliser des langues intermédiaires pour faciliter la participation de développeurs de différents pays dans leur langue maternelle.

Enfin, il importe de distinguer les contributions de traduction des contributions techniques. Dans le second cas, le passage par la langue de référence n'est pas nécessaire.

Pour tout projet de développement, il est recommandé de **choisir une langue de référence** applicable à tous les échanges techniques entre les développeurs. Pour les projets les plus importants, et s'il existe des communautés locales dynamiques, il peut être envisagé de créer des sous-groupes utilisant une langue différente.

Afin de faciliter les échanges entre les développeurs, il peut être opportun **d'établir un référentiel terminologique** des mots techniques utilisés dans la langue de référence.

Il semble par ailleurs nécessaire d'imposer l'usage d'une langue unique pour les variables de développement, mais de veiller à rendre ces variables intelligibles par tous en proscrivant les expressions trop complexes.

Il est recommandé de choisir une seule langue pour les variables utilisées dans les routines d'un programme ou applicatif ainsi que pour les commentaires et de retenir des termes ou expressions les plus simples possibles pour qu'ils demeurent compréhensibles par le plus grand nombre.

En ce qui concerne les commentaires insérés dans le code des applicatifs, il faudra veiller au respect de la langue de référence et à leur accessibilité linguistique. Ce point est d'autant plus important que les commentaires pourront être lus par des prestataires différents, pour qui les commentaires constituent la base d'une bonne compréhension de la logique de programmation retenue par les développeurs initiaux.

d. Quelles langues choisir ?

C'est un enjeu stratégique à identifier dès le départ en termes de **conquête de nouveaux marchés** pour une entreprise ou d'influence **politique et diplomatique** pour un État.

Le niveau d'intérêt pour une langue n'est absolument pas lié au nombre de locuteurs présents sur l'internet, mais au niveau de sensibilité des visiteurs pour la langue. La sensibilité du public est par exemple très forte sur les langues régionales (basque, catalan, etc.).

Les langues les plus sensibles sont les langues maternelles, et les langues régionales lorsqu'elles sont utilisées par les populations y résidant. Plus le niveau de sensibilité de la langue est élevé, plus la traduction dans cette langue aura un effet positif sur le visiteur et présentera une forte valeur ajoutée pour l'exploitant du site.

Le Forum recommande aux acteurs privés d'appréhender les choix linguistiques en termes de **retour sur investissement potentiel**, notamment par rapport aux marchés potentiels existants ou à venir. La sensibilité particulière d'une population vis-à-vis d'une langue, qu'elle utilise par exemple comme langue maternelle, ne devra pas être négligée.

2. Pour une utilisation raisonnée des technologies linguistiques

a. Pour la traduction de contenus

La qualité des outils d'aide à la traduction est en net progrès¹⁶⁹ mais la qualité de la traduction reste très aléatoire et dépend des situations et de l'accès à des corpus très étendus¹⁷⁰. Un travail de relecture et de correction reste indispensable après l'utilisation de ces outils. Il est par conséquent largement déconseillé de recourir aux outils d'aide à

¹⁶⁹. Cf. *supra* « État de l'art des technologies numériques linguistiques et perspectives », page 22.

¹⁷⁰. Les traductions automatiques effectuées au niveau de la Commission européenne fonctionnent bien, car elles bénéficient d'un grand nombre de textes de référence, que le vocable utilisé est sensiblement le même d'un texte à l'autre, et enfin, car la structure des documents est très marquée.

la traduction sans effectuer de relecture et d'amélioration « humaine » pour traduire des contenus destinés à être publiés.

Le Forum recommande aux exploitants de sites internet qui souhaitent faire des traductions de leurs contenus d'affecter ou de former des personnels spécifiques pour ces tâches, même si des outils d'aide à la traduction sont utilisés, l'intervention humaine étant nécessaire pour la relecture et l'amélioration du document avant sa publication sur internet.

Il est recommandé de laisser l'appréciation de l'usage ou non de ces technologies aux personnels compétents affectés aux tâches de traduction.

Ces outils peuvent néanmoins constituer une aide à la traduction, dont l'apport est variable selon la langue, ou les ressources humaines disponibles.

Le Forum recommande aux exploitants de sites internet souhaitant mettre en œuvre le multilinguisme d'envisager l'utilisation des technologies d'aide à la traduction automatique en complément des ressources humaines qui sont affectées aux tâches de traduction.

b. Pour d'autres usages

Les technologies de la langue ne sont pas limitées à la traduction de contenus et peuvent présenter un intérêt significatif dans de nombreux autres domaines. Pour la **veille concurrentielle ou stratégique**, ces outils constituent un outil de choix pour les entreprises, qui pourront s'informer à moindre coût de la situation sur des marchés étrangers en traduisant des contenus en langue étrangère vers la langue maternelle.

Le Forum considère que les technologies d'aide à la traduction constituent un atout majeur et particulièrement opportun pour les entreprises dans le cadre de la veille stratégique ou concurrentielle, ces outils leur permettant d'avoir accès à un moindre coût à des informations sur des marchés potentiels.

Il recommande par conséquent l'utilisation de ces technologies aux entreprises désireuses de conquérir de nouveaux marchés pour lesquels la langue est un obstacle.

Pour le grand public, ces technologies, intégrées dans les **moteurs de recherche**, facilitent les échanges culturels et l'accès à des informations autrefois inaccessibles. Leur usage doit être mentionné à l'internaute ; il doit être possible de les désactiver dans les préférences du moteur de recherche ; les contenus en version originale doivent enfin être prioritaires.

Le Forum considère que la présence de résultats traduits de façon automatique dans les moteurs de recherche n'est pas de nature à nuire à l'utilisateur, mais au contraire à **enrichir** l'étendue de sa recherche à des sources auxquelles il n'aurait pas eu accès sans ces technologies de traduction.

Il recommande néanmoins aux éditeurs de ces technologies et aux moteurs de recherche de laisser la possibilité à l'utilisateur de désactiver l'affichage des contenus traduits de façon automatique.

Le Forum recommande à ces éditeurs d'indiquer le plus clairement possible, lorsque cela est le cas, qu'un contenu a été traduit de façon automatique.

Dans tous les cas, il doit être possible pour l'utilisateur d'accéder facilement à la version originale du document, au moyen d'un lien présent dans un bandeau statique situé en haut de la page par exemple.

3. Comment utiliser au mieux l'étiquetage et le marquage linguistique des pages internet ?

a. Le marquage des pages internet

Il est possible d'insérer dans le code des pages internet une balise¹⁷¹ indiquant quelle est la langue utilisée sur la page en cours, comme par exemple :

```
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr" dir="ltr">
```

Les codes utilisés pour décrire une langue ont été standardisés sur la base de la norme ISO 639, qui a donné lieu à plusieurs évolutions, et qui a également fait l'objet de normes techniques issues de l'IETF¹⁷².

S'il y a plusieurs langues sur la page, il est conseillé de marquer la langue principale du contenu et non celle de l'interface¹⁷³.

Le Forum recommande aux exploitants de sites internet de veiller au bon étiquetage de langue des pages internet de leur site, l'étiquette retenue devant correspondre à la langue principale utilisée sur la page.

Cependant, les résultats des moteurs de recherche ne prennent généralement pas en compte les étiquettes utilisées par les webmasters pour leurs sites internet, ce qui est regrettable. Elles sont sans effet sur le référencement.

Les moteurs de recherche utilisent en effet des technologies dites heuristiques pour détecter de façon automatique la langue d'une page internet et l'identifier comme telle. Cela fonctionne correctement pour certaines langues et notamment pour le français, mais pose des difficultés avec des langues telles que le danois et le norvégien, qui ne sont pas distinguées correctement.

Le Forum recommande aux éditeurs de moteurs de recherche de davantage tenir compte des étiquettes de langues utilisées par les éditeurs de sites internet, ce qui permettra de limiter les dysfonctionnements techniques rencontrés sur certaines langues.

b. Le marquage des liens internet

Les liens hypertextes permettent de pointer vers un document d'un autre site internet ou d'une autre langue. Il peut être très décevant, pour le visiteur, de passer d'une langue à l'autre à partir d'un lien hypertexte. Il est par conséquent conseillé d'indiquer la langue d'un lien sortant si celle-ci est différente de la page de départ.

Le Forum recommande aux exploitants de sites internet, et, de manière plus générale, à tous les internautes amenés à contribuer sur des sites ou portails collaboratifs, ou dans les messages électroniques, d'indiquer de façon explicite la langue de la page vers laquelle pointe un lien hypertexte, si celle-ci est différente de celle de la page d'origine.

Il est recommandé d'employer la syntaxe suivante :

<http://www.monlien.fr/english.html> [anglais]

¹⁷¹. Pour plus d'informations techniques sur les balises de langue, il est conseillé de consulter ce site de référence (en anglais) :

<http://www.langtag.net/>

¹⁷². Voir *supra* note 141 page 40.

¹⁷³. Pour plus d'informations sur l'utilisation des marqueurs de langue dans une page internet, consulter [Les language tags](#), Stéphane Bortzmeyer, 10 octobre 2006 : <http://www.bortzmeyer.org/files/langtags-PRINT.pdf>

Dans l'hypothèse où une langue peut s'écrire de plusieurs façons, ce qui est le cas du chinois qui peut s'écrire en alphabet simplifié ou traditionnel, il est également recommandé d'indiquer le jeu de caractères utilisé sur la page de destination lorsqu'il est différent du jeu de caractères utilisé sur la page d'origine :

<http://www.monlien.fr/chinese.html> [chinois (simplifié)]

Techniquement, il est également possible de signaler *via* une balise *HTML* spécifique au navigateur internet qui parcourt la page que le lien en question renvoie vers un contenu dans une autre langue. Malheureusement, cette balise ne provoque pas toujours d'effet direct au niveau de l'affichage¹⁷⁴.

Le codage *HTML* du lien précédemment cité comme exemple pourra se faire de la façon suivante :

```
<a href="http://www.monlien.fr/english.html" hreflang="en"> monlien.fr  
[<span lang="en"> anglais </span>]</a>
```

Il est néanmoins recommandé aux exploitants de sites internet, d'utiliser, dans la mesure du possible, l'attribut *hreflang* dans les liens hypertextes de leurs sites internet.

La prise en charge de cette balise par les **logiciels de synthèse** utilisés par les personnes en situation de handicap devrait en revanche permettre de changer la langue du synthétiseur de façon automatique.

On pourra utilement se référer aux répertoires de codage des différentes langues pour trouver l'extension correspondant à une langue en particulier.

c. La détection automatique de la langue du visiteur

Il est prévu un système de **détection automatique de la langue d'un visiteur**, basé sur le réglage de son navigateur internet. Dans ce dernier, il est en effet possible de choisir plusieurs langues, par ordre de préférence.

Malheureusement, peu d'utilisateurs vont modifier ce réglage, si bien qu'il sera dans certains cas mal configuré (il est en général réglé en anglais par défaut).

Il n'est par ailleurs pas possible pour l'utilisateur de choisir un réglage qui correspondrait à la « version originale » directement depuis son navigateur (bien que l'absence de préférence linguistique puisse avoir le même effet).

Les sites internet peuvent ensuite « lire » le réglage de langue effectué par l'utilisateur et s'en servir pour le rediriger de manière totalement transparente vers la version qui correspond le mieux à ses préférences linguistiques.

Ce système de détection automatique devra être utilisé à bon escient et uniquement lorsque les contenus traduits sont d'aussi bonne qualité que les contenus originaux.

Lorsque les contenus sont différents en fonction de la langue de traduction (par exemple, un article sur Wikipédia), il sera préférable de ne pas recourir au système de détection automatique de la langue.

Si le système de détection automatique de la langue est finalement activé et utilisé, il faudra prévoir une voie d'accès manuelle aux autres langues proposées sur le site, dans le cas où le réglage du navigateur ne correspond pas à sa préférence linguistique.

¹⁷⁴. Pour plus de détails sur cette question, voir notamment :
<http://www.bortzmeyer.org/attribut-lang-liens.html>

Le Forum recommande aux exploitants de sites internet de recourir aux technologies de détection automatique de la langue par les navigateurs internet uniquement lorsque les contenus traduits sont disponibles dans la même quantité et dans la même qualité que les contenus originaux.

Si ces technologies sont utilisées sur un site internet, son exploitant devra veiller à proposer à l'utilisateur une interface accessible depuis n'importe quelle page du site et permettant de choisir une langue différente de celle détectée automatiquement.

Dès lors que les versions traduites s'écarteraient en qualité ou en quantité de contenus traduits de la version originale, il est recommandé aux exploitants de sites internet de ne pas utiliser les technologies de détection automatique de la langue des visiteurs et de proposer une interface permettant à chacun de choisir la langue de son choix manuellement.

4. La matérialisation de la langue sur le site internet

a. La représentation de la langue sur les pages du site

Lorsque l'exploitant choisit de mettre en œuvre le multilinguisme sur son site internet, il est confronté à la délicate question de la matérialisation des langues sur l'interface du site.

Il conviendra donc de déterminer **un emplacement fixe**, commun à toutes les pages du site, pour l'interface permettant la sélection de la langue. Rappelons à ce titre que la zone située en haut à gauche d'une page internet est généralement la plus visible et la plus utilisée par le visiteur.

Au niveau de la matérialisation de la langue en elle-même, il convient de **proscrire à tout prix l'utilisation de pictogrammes** (et notamment de drapeaux nationaux), souvent polémiques, et ne couvrant que très rarement une zone de locuteurs¹⁷⁵ : certains pays sont multilingues et certaines langues sont parlées dans plusieurs pays.

Le Forum recommande aux exploitants de sites internet de ne pas utiliser d'icônes ou d'éléments graphiques représentant des drapeaux ou tout autre symbole officiel d'un État pour mettre en œuvre une option de choix de la langue.

De la même manière, pour les raisons précitées, on ne demandera pas au visiteur quel est son pays d'origine pour déterminer la langue qui sera utilisée par la suite sur le site internet. Ce choix ne devrait logiquement intervenir que lorsque l'adresse du visiteur est nécessaire (dans les échanges transactionnels tels que le commerce électronique par exemple).

De la même manière, il n'est pas recommandé de présenter une liste de pays dans une option de choix de langue, de nombreux pays étant multilingues.

¹⁷⁵. Un des rares contre-exemples est le japonais, qui est la seule langue parlée au Japon, et qui est uniquement parlée au Japon.

L'option la plus simple et la plus accessible demeure finalement de présenter les différentes **langues endonymes**¹⁷⁶ disponibles sur le site **en toutes lettres** dans la langue visée, en bas de chaque page du site internet, comme par exemple :

Català | Français | English | Español | हिन्दी | Italiano | 日本語

Pour faciliter l'accessibilité aux différentes traductions d'une page ou d'un contenu, il est recommandé aux exploitants de sites internet de matérialiser en toutes lettres le nom de la langue dans sa langue d'origine en bas de chaque page du site.

Pour les pages intérieures du site et dans l'hypothèse où le nombre de langues proposées serait important, il sera acceptable d'utiliser les abréviations d'un répertoire de langues tel qu'ISO 639-1, à condition que l'exploitant ait indiqué en page d'accueil l'équivalence entre le nom de la langue et son abréviation.

La question de **l'ordre** dans lequel doivent être indiquées les langues peut poser problème : s'il s'agit uniquement de langues utilisant le même jeu de caractères, il n'y a pas de difficultés et il est recommandé de les présenter par ordre alphabétique.

Dans le cas contraire, il est impossible de classer les langues par ordre alphabétique, les caractères étant différents. Il demeure possible de les classer par ordre alphabétique de leur exonyme¹⁷⁷ anglais (par exemple, le chinois, *Chinese* en anglais, sera classé à « C »), solution pratiquée par la plupart des exploitants, mais qui n'est pas la plus élégante.

Le choix de **l'interface** à utiliser pour présenter à l'utilisateur un nombre important de langues peut aussi poser des difficultés. S'il y a quelques dizaines de langues proposées, le menu déroulant sera l'interface la plus simple à utiliser, mais si l'on présente à l'utilisateur plusieurs centaines de langues¹⁷⁸, celui-ci deviendra très fastidieux à manipuler.

Si le nombre de langues disponibles est trop élevé, il est recommandé de recourir à un menu de sélection des langues, sous réserve de certains problèmes d'ergonomie restant à résoudre.

Dans l'hypothèse où un menu déroulant doit être prévu du fait du nombre important de langues proposées, un icône générique pourrait être utilisé pour indiquer l'accès à l'interface de choix de la langue¹⁷⁹.

b. La neutralité de la page d'accueil

Lorsque l'exploitant décide de faire un site multilingue, la question de la page d'accueil est particulièrement délicate : faut-il choisir une langue de référence (l'anglais) et laisser

¹⁷⁶. L'endonymie désigne le fait qu'un groupe de personnes emploie régulièrement et couramment un certain nom pour se désigner lui-même. Ce nom est l'endonyme. Par extension, toute dénomination d'un groupe d'individus dans sa propre langue, qu'elle corresponde ou non au nom régulier, est un endonyme : <http://fr.wikipedia.org/wiki/Endonymie>

¹⁷⁷. L'exonymie est le fait qu'un groupe de personnes dénomme un autre groupe de personnes par un nom distinct du nom régulier employé par l'autre groupe pour se désigner lui-même. Ce nom est l'exonyme. Par extension, toute dénomination d'un groupe d'individus dans une langue étrangère, si elle ne correspond pas phonétiquement au nom régulier, est un exonyme. <http://fr.wikipedia.org/wiki/Exonymie>

¹⁷⁸. Voir par exemple l'installateur du logiciel *OpenOffice.org* : <http://extensions.services.openoffice.org/dictionary>

¹⁷⁹. Tentative de standardiser un icône de choix de la langue, « *Language Icon* » : <http://languageicon.org/>

les visiteurs trouver la version qu'ils préfèrent ? Faut-il imposer un écran de sélection de la langue, nécessitant un clic supplémentaire pour accéder aux contenus du site ?

De ces deux possibilités, la seconde semble préférable. Il semble en effet important de préserver autant que possible la **neutralité linguistique de la page d'accueil**, c'est-à-dire en rendant la page aussi lisible pour chaque langue utilisée sur le site.

En fonction de la quantité de langues proposées, il sera donc possible de mettre en ligne une quantité plus ou moins grande de texte de présentation, mais ces textes seront généralement très peu détaillés¹⁸⁰.

En revanche, il sera plus impératif d'utiliser une page d'accueil neutre si l'exploitant du site propose des versions de son site dans des langues utilisant des jeux de caractères différents, ce qui permettra au visiteur japonais, indien ou arabe de choisir plus facilement sa langue que sur une page remplie de caractères latins.

La neutralité de la page d'accueil devrait enfin permettre un meilleur référencement du site sur les différentes versions linguistiques des moteurs de recherche.

Le Forum recommande aux exploitants de sites internet visant un public international de préserver la neutralité linguistique de la page d'accueil de leur site et d'offrir directement aux visiteurs la possibilité de choisir leur langue dès cette page. La neutralité de la page d'accueil sera d'autant plus importante si les visiteurs attendus utilisent des jeux de caractères différents (cas d'un site internet recevant des visites fréquentes de populations européennes, russes, chinoises ou japonaises par exemple).

c. L'encodage de caractères

Il est bien évidemment indispensable d'utiliser un format d'encodage compatible avec l'ensemble des langues utilisées ou pouvant ultérieurement être utilisées sur le site internet.

La norme d'encodage de caractères ISO 8859-1¹⁸¹, encore largement répandue sur internet, n'est pourtant pas parfaitement adaptée au français, certains caractères tels que le Ÿ, la ligature œ ou le symbole euro (€) ne pouvant être utilisés¹⁸².

En raison des difficultés techniques liées au changement de l'encodage d'un site internet, il n'est cependant pas conseillé de migrer les données existantes.

En revanche, lors d'un nouveau projet, compte tenu de l'absence de certains caractères français et des potentialités de recourir à des traductions mêmes partielles dans des langues de plus en plus utilisées sur l'internet comme le chinois, son usage est fortement déconseillé.

Le Forum recommande aux exploitants de sites internet et aux prestataires techniques, sauf contrainte technique majeure, de restreindre au maximum le nombre de codages de

¹⁸⁰. Voir à ce titre quelques exemples de pages d'accueil « neutres » :

Le Parlement européen :

<http://www.europarl.europa.eu/>

Le portail de l'Union européenne :

<http://europa.eu/>

La Cour de justice des communautés européennes (CJCE) :

<http://curia.europa.eu>

¹⁸¹. Voir le descriptif de la norme ISO 8859-1 et sa table de caractères :

http://fr.wikipedia.org/wiki/ISO_8859-1

¹⁸². La norme ISO 8859-15 remédie à ce problème, mais n'est que très rarement utilisée en Europe occidentale, Unicode étant le choix de référence quand il s'agit de mettre à jour l'encodage d'un système.

http://fr.wikipedia.org/wiki/ISO_8859-15

caractères utilisés sur un même site internet et de retenir le codage acceptant la plus grande diversité de jeux de caractères (UTF-8 ou UTF-16 notamment¹⁸³).

Il leur est recommandé de choisir les encodages précités pour tout nouveau projet de développement internet.

d. Langue, nom de domaine et URL

Doit-on utiliser un nom de domaine différent pour chaque langue d'un site internet ?

Le Forum recommande aux exploitants de sites internet de n'utiliser, lorsque cela est possible, qu'un seul nom de domaine de premier niveau pour chaque site, indépendamment des traductions mises en œuvre sur ce site.

Étant donné qu'il existe des langues multi-pays, et des pays multilingues, il est notamment fortement déconseillé d'utiliser l'extension du nom de domaine de premier niveau (.fr, .com, .eu) pour distinguer plusieurs déclinaisons linguistiques du même site internet.

Il est au contraire recommandé d'utiliser une extension de nom de domaine géographique (.fr, .de...) pour désigner des **filiales ou antennes locales de l'organisme** ou entreprise responsable du site internet.

Que doit-on faire lorsque le nom de la marque, de l'organisme ou de l'institution est décliné dans plusieurs langues ? Le problème de la neutralité linguistique du nom de domaine doit également être posé et ses solutions envisagées au regard d'exemples.

Certains sites d'organismes internationaux ont été confrontés au problème : la banque mondiale constitue un bon exemple. Le site utilise un nom de domaine différent pour certaines déclinaisons linguistiques : *banquemondiale.org*, *worldbank.org*, *bancomundial.org*, mais le même nom de domaine pour d'autres langues : *worldbank.org/eca/russian/*, ou encore *worldbank.org.cn/Chinese/*. Le choix n'est finalement pas très clair.

Dans le cas où le site s'adresse à une grande diversité de locuteurs, il pourra être préférable de retenir par exemple **un seul nom de domaine**, qui peut être compris par un groupe le plus large possible de locuteurs, tel qu'un acronyme ou encore un mot en latin pour un site s'adressant à des locuteurs utilisant l'alphabet latin.

Cependant, dans le cas le plus fréquent, qui concerne notamment les **pays multilingues**, qui ne reconnaissent généralement pas plus de deux ou trois langues officielles¹⁸⁴, il pourra être raisonnable de réserver les variantes du nom de domaine dans chacune de ces langues.

Il est recommandé de faire apparaître le choix de la langue dans l'URL utilisée sur le site en utilisant les étiquettes normalisées¹⁸⁵. Deux possibilités peuvent alors être envisagées :

- utiliser un sous-domaine : fr.site.ext/ ;
- utiliser une arborescence au premier niveau : www.site.ext/fr/.

Ces deux choix techniques permettent de conserver la même structure d'URL pour un même contenu, quelle que soit la langue. L'utilisateur averti pourra détecter avant de

¹⁸³. Pour davantage de précisions sur l'emploi du codage Unicode sur un site internet, cf. *supra* note 53 page 20.

¹⁸⁴. Par exemple, la Suisse (3 langues), la Belgique (2 langues), le Canada (2 langues).

¹⁸⁵. Cf. *supra* « Uniformisation du répertoire d'étiquetage des langues utilisé sur l'internet », page 40.

cliquer sur le lien la langue de la page et essayer d'en choisir une autre en modifiant lui-même l'*URL* du document.

Il est recommandé d'indiquer la langue de façon explicite dans l'*URL* des pages du site, et, dans la mesure du possible, aussi haut que possible dans l'arborescence du site internet, comme par exemple : <http://www.monsite.fr/en/article1.html>.

Enfin, les différents moteurs de recherche proposent généralement une option permettant d'indiquer la langue utilisée sur le site ou les différentes « zones » linguistiques présentes sur un site multilingue. Ce paramètre sera important pour le bon référencement du site dans ses différentes versions linguistiques.

Le Forum recommande aux exploitants de sites internet de prêter une attention particulière à la définition des différentes zones linguistiques de leur site internet au moment du référencement de leurs pages sur un moteur de recherche.

e. Traduire les pages présentant des erreurs techniques

Un logiciel mettant en œuvre des pages internet, plus communément appelé serveur *web*¹⁸⁶ ou serveur *HTTP*, génère automatiquement certaines erreurs lorsque le visiteur du site internet demande des pages qui n'existent pas « *404 Page not found* » ou encore quand le visiteur n'a pas les privilèges nécessaires pour afficher la page « *403 Access forbidden* ».

Bien que ces messages d'erreur ne soient pas traduits, il est pourtant facile pour le responsable du site internet de créer des pages personnalisées dans la langue du site et permettant d'afficher un message un peu plus explicite au visiteur, en lui offrant la possibilité d'effectuer une recherche sur les pages du site.

Il est recommandé aux exploitants de sites internet de mettre en œuvre **des pages d'erreur personnalisées** dans la langue principale du site ou, le cas échéant, dans les principales langues utilisées sur le site internet pour remplacer les messages d'erreur des logiciels serveurs et expliquant au visiteur les raisons de l'erreur technique, tout en lui proposant des solutions pour trouver autrement le contenu qu'il recherche (moteur de recherche notamment).

Notons que dans cette situation, il est impossible de détecter la langue du visiteur et de délivrer une page spécifique en fonction du réglage particulier de sa langue.

¹⁸⁶. Exemples de logiciels serveur *HTTP* : *Apache*, *Microsoft IIS*.

ANNEXE

ANNEXE : COMPOSITION DU GROUPE DE TRAVAIL

Représentants des acteurs économiques :

- **Google France**
 - Olivier Esper, relations institutionnelles
- **Idéose**
 - Pierre Guillou, dirigeant
- **Livre-rare-book**
 - Pascal Chartier, gérant
- **Microsoft France**
 - Isabelle Leung-Tack, responsable des affaires publiques en charge de la citoyenneté
- **Renaissance Numérique**
 - Éric Von Rompay, délégué général de l'association
- **Société Générale**
 - Aude de Corbiac, Chargée de Mission au sein de l'équipe des Relations Interbancaires

Représentants des utilisateurs :

- **Association française pour le nommage internet en coopération (AFNIC)**
 - Stéphane Bortzmeyer, ingénieur R&D, spécialiste des techniques multilingues
- **Association française de normalisation (AFNOR)**
 - Jean-François Legendre, responsable développement
 - Philippe Magnabosco, secrétaire JTC1/SC35
- **Association pour la promotion et la recherche en informatique libre (APRIL)**
 - Frédéric Couchet, délégué général
- **La Documentation Française**
 - Bénédicte Roullier, responsable du site portail service-public.fr
 - Olivier Roumieux, administrateur des sites, département des produits internet
- **Association diversum**
 - Jérôme Gouadain, président
- **Union Latine**
 - Daniel Prado, directeur de la direction Terminologie et industries de la langue
 - Annelore Lemoulinier, direction Terminologie et industries de la langue

Observateurs des pouvoirs publics :

- **Centre de traduction des ministères économique et financier**
 - Alain Repaux, chef du centre de traduction des ministères économique et financier
- **Conseil général des nouvelles technologies de l'information (CGTI), ministère de l'Économie, de l'Industrie et de l'Emploi**
 - Robert Picard, ingénieur général
- **Délégation générale pour l'armement (DGA)**
 - Edouard Geoffrois, directeur technique du programme Quaero
- **Délégation générale à la langue française et aux langues de France (DGLFLF), ministère de la Culture**
 - Luc Bouiller, chef de projet multimédia auprès de la mission sensibilisation et développement des publics
 - Odile Canale, chef de la mission emploi et diffusion de la langue française
 - Florence Desmoulière, chargée de mission auprès de la mission développement et modernisation de la langue française
 - Bénédicte Madinier, chef de la mission développement et modernisation de la langue française
 - Xavier North, Délégué général à la langue française et aux langues de France

Spécialistes :

- **Institut national des télécommunications (INT)**
 - Fabrice Flipo, ingénieur, docteur en Philosophie des Sciences et Techniques, maître de conférences au département langues et sciences humaines
- **Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) / Centre National de la Recherche Scientifique (CNRS)**
 - Joseph Mariani, chercheur au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI, CNRS) et directeur de l'Institut des technologies Multilingues et Multimédias de l'Information (IMMI)

Auditions :

- **Association Debian France**
 - Christian Perrier, responsable langues et internationalisation
- **Centre de traduction des ministères économique et financier**
 - Alain Repaux, chef du centre de traduction des ministères économique et financier
- **Commission Spécialisée de Terminologie et de Néologie de l'Informatique et des Composants Électroniques**
 - Jérôme Trollet, président

- **Délégation générale à la langue française et aux langues de France (DGLFLF)**
 - Etienne Quillot, chargé de mission Commissions de terminologie
- **Délégation générale pour l'armement (DGA)**
 - Edouard Geoffrois, directeur technique du programme Quaero
- **Google France**
 - Olivier Esper, relations institutionnelles
 - Jean-François Wassong, Directeur Technologies
- **Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) / Centre National de la Recherche Scientifique (CNRS)**
 - Joseph Mariani, chercheur au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI, CNRS) et directeur de l'Institut des technologies Multilingues et Multimédias de l'Information (IMMI)
- **Société française de terminologie**
 - Loïc Depecker, président
- **Association Wikimedia France**
 - Julien Fayolle, vice-président

La coordination des travaux est assurée par **Thibault GROUAS**, assisté de **Stéphane GRÉGOIRE**, chargés de mission au Forum des droits sur l'internet.