

La politique linguistique de l'Union européenne repose sur le principe du multilinguisme. En effet, le multilinguisme est une exigence légale: que la législation européenne soit directement applicable dans l'ensemble des États membres ou qu'elle doive d'abord être transposée en droit national, il est essentiel que les textes législatifs soient traduits et publiés dans les différentes langues officielles de l'UE. De plus, le multilinguisme garantit que les institutions européennes sont accessibles à l'ensemble des citoyens de l'UE.

Différentes mesures sont toutefois nécessaires, si l'on veut que ce principe soit respecté. Parmi celles-ci figurent le recours à des traducteurs professionnels, la formation linguistique des administrateurs et la fourniture d'outils linguistiques, comme la Traduction Automatique (TA), capables d'offrir des moyens rapides et efficaces pour obtenir une traduction brute de rapports, de comptes rendus et de courriers électroniques.

La Commission européenne gère un service de TA depuis de nombreuses années: les résultats sont utilisés par les administrateurs dans leur travail quotidien et par les traducteurs, qui s'en servent comme base pour réaliser des traductions professionnelles. Le service est accessible à tout le personnel des institutions européennes ainsi qu'aux administrations publiques des États membres.

## ≡ Qu'est-ce que la traduction automatique?

Un système de TA peut être défini comme un outil de traduction informatique qui fonctionne en décomposant des phrases ou d'autres segments de texte dans une langue source donnée, en les analysant en contexte et en essayant ensuite de recréer leur sens dans une langue cible donnée, en tenant compte de l'inflexion, du registre et de l'ordre des mots.

Il existe différents types de systèmes de TA, comme des systèmes «à base de règles», «à mémoire syntagmatique» et «à base de statistiques». Le premier emploie des règles linguistiques complexes combinées à des dictionnaires pour analyser un texte, puis générer une traduction. Les deux autres utilisent des traductions existantes stockées dans des corpus bilingues - de vastes dépôts de textes source alignés avec leurs traductions humaines officielles - et s'appuient largement sur l'expérience passée.

Les systèmes de TA peuvent également être «automatiques» (les utilisateurs n'interviennent

pas dans le processus de traduction) ou «interactifs» (si l'ordinateur a un problème de terminologie, par exemple, il peut demander aux utilisateurs quelles sont leurs préférences).

Par ailleurs, certains systèmes peuvent préférer une «langue contrôlée», où les textes doivent être écrits d'une manière clairement définie, avec un vocabulaire restreint, alors que d'autres accepteront sans difficulté toutes sortes de textes (système en plein texte).

### **Le service de traduction automatique de la Commission européenne**

Le service de traduction automatique de la Commission européenne est construit autour de EC Systran, version spécifique du système SYSTRAN complété initialement par le World Translation Center (États-Unis). Depuis 1976, SYSTRAN a été complété et adapté par la Commission européenne à des fins internes.

EC Systran est une technologie à base de règles, en plein texte et entièrement automatique qui tourne sur des plates-formes UNIX/Linux au Data Centre de la Commission à Luxembourg. Des centaines de milliers de pages sont soumises chaque année pour traduction automatique: près de 80 % d'entre elles proviennent de la Commission elle-même. Parmi les États membres, les autorités espagnoles et allemandes (au niveau national et régional) sont les principaux utilisateurs, mais la demande progresse également en France.

### **Combinaisons linguistiques**

Le choix des combinaisons linguistiques, ou couples, est influencé par différents facteurs, comme les besoins internes de la Commission, la qualité de traduction attendue pour les langues combinées et la volonté des États membres à participer au cofinancement de projets comme le programme Société de l'information multilingue (MLIS).

La qualité des résultats de la TA varie en fonction du couple sélectionné et du contenu du texte source.

Le service de TA offre actuellement 28 couples de langues:

<b>Langue source (traduction à partir de)</b>	<b>Langue cible (traduction vers)</b>
<i>Danois</i>	Anglais*
<i>Néerlandais</i>	Anglais*, Français*
<i>Anglais</i>	Néerlandais, Français, Allemand, Grec, Italien, Portugais, Espagnol
<i>Français</i>	Néerlandais, Anglais, Allemand, Grec*, Italien, Portugais, Espagnol
<i>Allemand</i>	Anglais, Français
<i>Italien</i>	Anglais*, Français*
<i>Portugais</i>	Anglais, Français
<i>Espagnol</i>	Anglais*, Français
<i>Grec</i>	Anglais*, Français
<i>Suédois</i>	Anglais*

\* = *prototype en phase d'essai*

Par exemple, le service de TA est plus adapté aux rapports et comptes rendus qu'aux discours ou aux textes littéraires. Des phrases courtes sont plus faciles à analyser que des phrases longues. De plus, la machine ne peut pas «deviner» ce que devrait signifier un mot mal orthographié, il est donc utile de vérifier l'orthographe des textes source avant de les envoyer.

En général, le résultat brut est suffisant pour prendre connaissance rapidement des informations, mais si le texte est destiné à la diffusion, la traduction doit être révisée.

### **Utilisation de la TA**

À l'heure actuelle, le personnel des institutions européennes demande une TA à l'aide d'une interface Web interne, alors que les administrateurs dans les États membres doivent utiliser leur messagerie électronique habituelle, une méthode donc moins conviviale. Dans les deux cas, l'utilisateur doit spécifier la langue source et la ou les langues cible, puis joindre un fichier ou copier/taper le texte à traduire. Le résultat de la

TA est renvoyé dans la boîte aux lettres de l'utilisateur, normalement dans les 20 minutes.

L'une des missions du projet IDA(BC) TA a été de développer une procédure d'accès commune qui sera introduite dans un avenir proche.

## IDA(BC) TA

IDA(BC) TA est destiné à fournir un accès efficace et convivial au service de TA de la Commission. L'initiative repose sur le principe selon lequel un accès plus large à un tel outil aidera les administrations publiques européennes à surmonter les barrières linguistiques lors de l'échange de données et à réduire les frais généraux associés à la communication multilingue.

Le projet a débuté en 2002 avec une étude de faisabilité comprenant une analyse du service de traduction automatique, une enquête sur les besoins en TA des administrations publiques européennes et des recommandations pour répondre à ces besoins.

L'étude a mis en évidence un certain nombre de faiblesses dans le service de TA de la Commission, comme le manque de convivialité et l'accès de mauvaise qualité pour les utilisateurs externes. En effet, de nombreux administrateurs nationaux ignoraient qu'une telle technologie était à leur disposition. Parmi les autres questions soulevées figuraient l'amélioration de la qualité de traduction, l'intégration de la terminologie des États membres et l'étendue de la couverture linguistique.

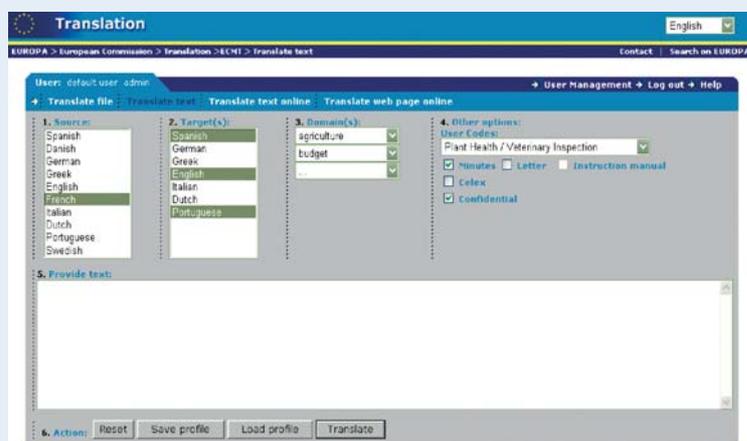
Les traductions réalisées par une machine sont assurément moins fiables que les traductions humaines, mais, dans un environnement de travail multilingue, elles peuvent être d'un grand secours lorsque les administrateurs doivent comprendre l'essentiel de textes rédigés dans des langues qu'ils ne maîtrisent pas ou qu'ils connaissent mal.

Suite à cette étude, la terminologie fournie par les administrations nationales a été ajoutée aux dictionnaires de TA, sur une base d'essai, pour la traduction entre le français, l'allemand et l'anglais (2003).

Des travaux ont également été engagés pour mettre à niveau l'architecture de la TA. Les améliorations ont été apportées en deux étapes (2003/2004). Tout d'abord, une procédure d'accès commune a été développée, de manière à ce que tous les utilisateurs puissent demander une traduction automatique via le Web, qu'ils travaillent pour les institutions européennes ou les administrations des États membres. Dans le même temps, l'interface utilisateur a été complètement remaniée et une nouvelle fonction a été ajoutée: la traduction en temps réel. Cela signifie que le résultat de la TA peut désormais être renvoyé directement sur l'écran de l'utilisateur, ainsi qu'à sa boîte aux lettres électronique.

Dans un deuxième temps, l'interface a été placée sur une plate-forme de services Web, de façon à ce que les administrations nationales puissent l'intégrer dans leurs propres sites Web et l'adapter à l'apparence de leur environnement informatique.

Ce nouveau service sera accessible à toutes les administrations européennes par le biais du réseau TESTA et d'Internet (pour les administrations non connectées à TESTA) au cours du premier semestre 2005. Les utilisateurs découvriront que le service leur offre une grande diversité d'options, dont certaines pourront influencer sur le résultat final de la traduction:



La nouvelle interface de TA

- *Domaines*: il s'agit de dictionnaires thématiques, qui, lorsqu'ils sont sélectionnés, sont prioritaires sur les dictionnaires généraux. Dans le domaine nucléaire, par exemple, le mot français "cœur" est traduit en anglais par "core" plutôt que par le terme plus général "heart". De même, "arrêt" est traduit par "shut-down" plutôt que par "stop". 36 domaines sont disponibles au total et l'utilisateur peut en choisir jusqu'à trois. La profondeur de la terminologie du domaine varie en fonction du couple de langues utilisé.
- *Code utilisateur*: code spécial donné à un utilisateur ou à un groupe d'utilisateurs pour accéder à un ensemble terminologique personnalisé. Un Code utilisateur est prioritaire à la fois sur les dictionnaires généraux et les domaines.
- *Types de texte*: option limitée à certains couples de langues. La sélection d'un type de texte («Lettre», «Compte-rendu» ou «Manuel d'utilisation») indique au système de traduire un texte en utilisant des conventions stylistiques spécifiques. Par exemple, les comptes-rendus rédigés au passé en anglais seront rendus au présent en français, espagnol et italien; la traduction du français vers l'anglais offre le même service, mais en sens inverse.
- *Confidentiel*: cette mention garantit que le document à traduire restera sur le serveur uniquement durant le temps nécessaire à son traitement, qu'il n'est pas archivé et qu'il n'est pas accessible aux gestionnaires du système.
- *Extraction des titres dans Celex*: si cette option est sélectionnée, toutes les références textuelles à la législation communautaire seront directement extraites et recherchées dans la base de données législative CELEX. Le titre intégral du texte législatif pertinent sera ensuite récupéré pour les langues source et cible et envoyé sous forme de pièce jointe séparée avec le résultat de la traduction automatique.

Enfin, IDA TA a demandé une étude sur les produits de TA existant pour les 10 langues

officielles de l'UE qui ne sont pas encore couvertes par ce service de la Commission: finnois, tchèque, estonien, hongrois, letton, lituanien, maltais, polonais, slovaque et slovène. Dans la perspective des futures adhésions, le bulgare et le roumain ont également été inclus. En plus d'une étude de marché, l'étude comporte une évaluation linguistique et technique des produits (qualité à des fins de compréhension, adaptabilité, options utilisateur, etc.) et une analyse coûts/bénéfices. Les résultats de cette étude devraient être disponibles d'ici fin 2005.

## Comment participer?

L'adresse de la nouvelle interface de TA sera diffusée sur le site Web IDABC et ailleurs, dès qu'elle sera accessible en ligne courant 2005. Dans l'intervalle, les administrateurs nationaux peuvent s'inscrire pour obtenir un accès par courrier électronique en contactant le Help Desk TA de la Commission: [systran-helpdesk@cec.eu.int](mailto:systran-helpdesk@cec.eu.int). Les utilisateurs sont invités à signaler les erreurs de traduction qu'ils décèlent.

La traduction automatique (TA) n'est pas destinée à remplacer les traducteurs humains. En effet, les ordinateurs ne peuvent pas entrer en concurrence avec ceux-ci, car ils n'ont pas la connaissance et l'expérience du monde acquises par les êtres humains. La traduction automatique est simplement un outil supplémentaire qui peut offrir une traduction brute et rapide lorsque le temps presse.

## Qui sommes-nous?

Le service de traduction automatique de la Commission européenne est géré par la Direction générale de la Traduction (DGT). IDA TA a été développé et financé dans le cadre du programme IDA. Il sera poursuivi en tant que service d'infrastructure dans le cadre des mesures horizontales IDABC.

Vous obtiendrez de plus amples informations sur IDABC et IDA TA, en consultant la page <http://europa.eu.int/idabc> ou en envoyant un courrier électronique à l'adresse suivante: [idabc@cec.eu.int](mailto:idabc@cec.eu.int).

