



# LES TECHNOLOGIES DU LANGAGE HUMAIN POUR L'EUROPE

Travail commissionné par :



Trente, Italie

Réalisé avec la collaboration de :

ACCIPIO  
CONSULTING  
Aix-la-Chapelle, Allemagne

*Ce travail a été financé par l'Union Européenne dans le cadre du projet intégré TC-STAR – Technologies et Corpus pour la traduction parole-parole. Chef de projet: Gianni Lazzari – IST-2002-FP6-506738, Web Site : <http://www.tc-star.org>*

Avril 2006

Note légale : Le contenu de ce document ne reflète que le point de vue de l'auteur. La Commission européenne et le Consortium TC-STAR ne sont pas responsables de l'usage qui peut être fait des renseignements contenus ci-après.

## Préface

La richesse des langues est considérée par beaucoup comme le couronnement de l'évolution humaine. Le langage nous fournit le moyen de communiquer idées, émotions et savoir et d'exprimer notre identité culturelle. Toutes les réalisations humaines – dans le domaine des sciences et de la technologie, de la philosophie, de l'art et de la culture – sont possibles grâce au langage.

En Europe la diversité linguistique est un fait acquis. L'Union européenne la considère comme une composante inaliénable de notre patrimoine culturel; le principe de l'égalité des langues est donc présent dans les traités institutionnels de la Communauté européenne. La sauvegarde de la diversité linguistique est depuis toujours au centre de la politique communautaire. La diversité linguistique requiert, toutefois, un investissement constant et conséquent, et les institutions européennes allouent une partie considérable de leurs budgets opérationnels aux services de traduction et d'interprétariat.

Pour l'industrie et le commerce européens, la diversité linguistique est à la fois un défi et un point fort, comme l'a récemment fait remarquer la Commission dans sa première communication concernant un nouveau cadre stratégique pour le multilinguisme<sup>1</sup>. Il est certain que des investissements, faits en temps utile dans les technologies de la communication multilingue, peuvent fournir un accès rapide à de nouveaux marchés émergents partout dans le monde, ce qui constitue un facteur vital pour le succès à long terme de l'Europe.

La capacité d'accéder aux informations et de les utiliser dans les différentes langues est d'une importance vitale pour les citoyens, pour les gouvernements et le commerce, et les technologies du langage peuvent, par conséquent, jouer un rôle important pour faciliter la communication entre les personnes, les administrations et les entreprises. En collaboration avec les états membres, l'Union européenne a parrainé, au cours des vingt dernières années, plusieurs actions de recherche et de développement qui ont contribué à créer des compétences, des ressources et une infrastructure linguistique paneuropéenne.

<sup>1</sup> [http://europa.eu.int/comm/education/policies/lang/key/legislation\\_en.html](http://europa.eu.int/comm/education/policies/lang/key/legislation_en.html)

L'Europe est actuellement l'un des marchés les plus avancés en ce qui concerne les technologies du langage et de la traduction automatique. L'Union européenne s'engage à faire en sorte que tous les outils et les ressources nécessaires soient mis à disposition de toutes les langues de l'UE et des principales langues utilisées dans les échanges commerciaux, ouvrant ainsi la voie à une société de l'information multilingue touchant toute l'Europe. En fournissant des produits et des services multilingues, tels que les systèmes de traduction automatique et la recherche d'information interlinguistique, la Commission européenne entend réaliser le but ambitieux de généraliser l'accès à l'information à tous les citoyens européens – ce qui est l'un des objectifs clés de l'initiative i2010.

Le multilinguisme est désormais une politique au niveau européen et il sera soutenu par une série d'actions dans le cadre des programmes de l'éducation, de la formation et de la recherche, tels que les programmes pour l'apprentissage des langues, la recherche sur la diversité linguistique et sur les technologies du langage humain, ainsi que les programmes sur le contenu digital.

Ce document est un rapport précieux sur l'état de l'art, sur les défis et les possibilités qui s'affichent pour l'Europe dans cet important domaine de la recherche; il représente également une source d'inspiration pour les chercheurs, pour l'industrie et pour les responsables politiques, et il contribuera certainement à rendre l'Europe plus multilingue encore.



*Viviane Reding*  
*Membre de la Commission européenne*  
*Responsable de la Société*  
*de l'Information et des Médias*



*Ján Figel'*  
*Membre de la Commission européenne*  
*Responsable de l'Éducation, Formation*  
*Culture et Multilinguisme*

## Sommaire

Pour l'Europe multilingue, la communication et l'échange d'informations entre langues différentes revêtent une importance capitale. Vingt langues européennes officielles, soit 190 couples de langues ou 380 directions de traduction, sont un coût considérable pour toute activité impliquant des contenus linguistiques, au gouvernement, dans les activités économiques et dans les communautés. Même si cet effort est relativement petit pour certains types de transactions et de communications, il représente toutefois une barrière insurmontable pour la plupart des communications et des transactions. À l'avenir cet effort pourra considérablement diminuer suite à la capacité de traduction fournie par les systèmes automatiques, qui ne seront pas aussi parfaits que les traducteurs humains professionnels, mais qui seront plus économiques, plus rapides, disponibles et utiles à bon nombre d'usages. Alors que l'on s'attend à ce que ces systèmes permettent d'améliorer l'efficacité des traductions humaines traditionnelles, le majeur impact de ces technologies se vérifiera dans les applications automatiques. La traduction automatique du langage parlé et écrit commencera par être efficace dans certaines niches de marché, mais elle s'étendra rapidement à d'autres secteurs, tout-à-fait indépendamment des services commerciaux de traduction existant actuellement. Ces technologies agiront comme des technologies catalysantes, en stimulant le commerce et l'économie de l'Europe. C'est en effet une nécessité stratégique pour l'Europe que de disposer de technologies du langage, pour faciliter le plus possible la communication et l'échange d'informations entre langues différentes.

Ce document commence en illustrant l'importance des technologies du langage, pour l'Europe en particulier, et il en décrit l'état actuel. Puis il examine la perspective européenne dans un contexte global, avec une attention particulière pour les États-Unis, l'Inde et l'Asie orientale. Viennent la description de l'état de l'art dans le domaine de la recherche et des activités économiques, et la définition des attentes pour l'essor futur des marchés. Et enfin quelques interviews de décideurs et de spécialistes dans le domaine de la recherche et de l'économie permettent d'approfondir le sujet et d'élargir la perspective.



## Index

|   |    |
|---|----|
| 1. Que sont les technologies du langage?  | 9  |
| 2. L'importance des technologies du langage pour l'Europe   | 11 |
| 2.1. Un obstacle déterminant pour le marché intérieur européen  | 11 |
| 2.2. Les langues dans le monde  | 12 |
| 2.2.1. Les principales langues du monde   | 12 |
| 2.2.2. Langues en voie d'extinction   | 14 |
| 2.2.3. L'importance de la diffusion : langues primaires, secondaires et tertiaires et forces du marché                              | 15 |
| 2.3. Un regard de près sur l'Union Européenne   | 16 |
| <i>Interview</i> de Karl-Johan Lönnroth   | 17 |
| 2.4. La communication   | 19 |
| 2.5. Le prochain pas de l'industrialisation: des machines élaborant le texte et la parole.  | 20 |
| 2.6. Le marché  | 21 |
| 2.6.1. La traduction comme facteur de coût  | 21 |
| 2.6.2. Les marchés consolidés de la localisation et de la traduction  | 22 |
| <i>Interview</i> de Renato Beninato   | 25 |
| 2.6.3. De la traduction humaine à la traduction automatique : extraordinaire réduction des coûts et amélioration de l'accessibilité | 27 |
| <i>Interview</i> de Michael Anobile   | 28 |
| 2.6.4. Innovations technologiques incrémentielles et de rupture   | 31 |
| 3. Asie, Europe et États-Unis: analogies et différences   | 34 |
| 3.1. L'union Européenne   | 34 |
| 3.2. Les États-Unis   | 35 |
| 3.2.1. Le rôle stratégique des TLH  | 35 |
| 3.2.2. Les programmes de recherche  | 35 |
| <i>Interview</i> de Joseph Olive  | 37 |
| 3.3. L'Asie orientale   | 40 |
| 3.3.1. L'anglais comme langue franche en Asie orientale   | 40 |
| 3.3.2. Les couples linguistiques asiatiques et l'importance croissante du chinois   | 40 |
| 3.3.3. Les programmes de recherche  | 41 |
| <i>Interview</i> de Jun-ichi Tsujii   | 42 |
| 3.4. L'Inde   | 45 |
| 3.5. Les conditions économiques contextuelles   | 45 |

|  |    |
|--|----|
| 3.6. Une stratégie pour l'Europe   | 47 |
| <i>Interview</i> de Joseph Mariani   | 48 |
| 4. Où nous en sommes aujourd'hui   | 50 |
| 4.1. Le travail de traduction aujourd'hui  | 50 |
| 4.2. Les technologies utilisées dans la traduction professionnelle                 | 50 |
| <i>Interview</i> de Kevin Bolen  | 52 |
| 4.3. La recherche sur les technologies à la base de la traduction<br>parole-parole | 55 |
| 4.4. Le projet TC-STAR   | 58 |
| 5. Le pouvoir d'une technologie catalysante  | 60 |
| 5.1. Les besoins insatiables del l'être humain                                     | 60 |
| <i>Interview</i> de Deimitris Sabatakakis  | 63 |
| 6. Conclusions   | 65 |
| 7. Bibliographies et autres renseignements complémentaires                         | 66 |
| 7.1. Bibliographie   | 66 |
| 7.2. Autres lectures   | 67 |
| 7.3. Tableaux  | 67 |
| 7.4. Figures   | 68 |
| 7.5. Liste des acronymes   | 69 |

## 1. Que sont les technologies du langage?

Nous savons depuis longtemps que la technologie peut radicalement changer le monde qui nous entoure ainsi que notre façon de vivre. Par rapport au passé, le monde a subi des changements extraordinaires grâce à l'informatique et aux télécommunications. Et vu que chacune de ces révolutions s'est imposée graduellement au cours des dernières années, le changement, bien que continu, n'a pas été particulièrement traumatique.

Nous sommes maintenant conscients qu'il existe une autre technologie informatique ayant un potentiel énorme, et nous devons admettre d'être en présence d'une autre révolution. Son évolution a été plutôt lente, et nombre des thèmes de recherche qui la caractérisent ont été explorés pendant tant d'années, que certains ont abandonné tout espoir. Il est très difficile d'apprendre aux ordinateurs à traiter le langage humain – que soit dans sa forme écrite ou parlée – comme réussissent à le faire les personnes: parler de façon naturelle, comprendre ce qui a été dit (et le sens), synthétiser un document ou une conversation, trouver un enregistrement audio selon son contenu, traduire d'une langue à l'autre. Nous aimerions être en mesure de nous interfacer avec les machines à travers la voix et le langage, car ce sont les moyens de communication

que nous utilisons, et nous aimerions que les ordinateurs soient en mesure de traiter les informations selon les modalités que nous estimons utiles. L'ensemble des technologies qui s'occupent de ces arguments est connu sous le nom de technologies du langage (Technologies du langage humain, TLH). L'identification automatique de la voix, la traduction automatique et la synthèse vocale représentent les technologies de pointe, mais il y en a beaucoup d'autres. Comme cela est déjà arrivé pour l'informatique et les télécommunications, les TLH ont le potentiel nécessaire pour modifier radicalement notre façon de nous confronter et de travailler avec les informations, parce qu'elles accèdent et élaborent les informations codifiées du langage d'une façon tout à fait nouvelle. Ce document se concentre sur un seul aspect des TLH, la capacité d'aller outre la barrière créée par une autre langue, que ce soit dans la communication entre personnes ou dans l'élaboration des informations contenues dans des textes écrits non structurés.

Une importante série d'applications se basant sur les technologies de la traduction automatique et de la traduction du langage parlé<sup>1</sup> compose ce que l'on appelle *cross-lingual information retrieval, summarization, and data assimilation* (récupération, synthèse et assimila-

<sup>1</sup> Pour plus de détails sur ces technologies cfr. Paragraphe 4.3.

*tion d'informations entre langues différentes*). Une deuxième série, tout aussi importante, comprend la communication entre personnes parlant des langues différentes. Dans ce cas – selon le type d'input ou d'output, voix ou texte écrit – l'on parle de traduction automatique du texte écrit, traduction

du langage parlé ou traduction de parlé à parlé. Ces technologies sont tellement récentes qu'aucun terme n'a encore été inventé pour les définir – ce pourrait être *cross-lingual information and communication technology* (*Technologies de l'information et de la communication entre langues différentes*)?

## 2. L'importance des technologies du langage pour l'Europe

### 2.1. Un obstacle déterminant pour le marché intérieur européen

Quatre types fondamentaux de liberté sont prévus par le traité de la Communauté européenne<sup>2</sup>, la *libre circulation des biens, des personnes, des services et des capitaux*. Le marché intérieur, principal résultat de l'intégration européenne, fut réalisé fin 1992 comme une aire sans frontières intérieures pour les biens et les services. Pour ce qui est du commerce au sein des pays de l'Union Européenne, il n'y a pas de taxe d'importation et les systèmes nationaux de taxation doivent respecter ces quatre libertés fondamentales.

En ce qui concerne les processus d'harmonisation des règlements et des exigences juridiques nationales, beaucoup a été fait pour soutenir le marché intérieur. La libre circulation des biens est assurée par le *principe de reconnaissance mutuelle* sur chaque marché, ce qui élimine la nécessité d'une harmonisation complète de la législation nationale dans chaque Etat membre : la vente de biens produits légalement dans l'un des états membres de l'UE ne peut

être interdite sur le territoire d'un autre état membre, même si ce dernier réalise des produits ayant des caractéristiques techniques ou de qualité différentes<sup>3</sup>. Ce même principe vaut pour les services.

**Les barrières linguistiques sont le dernier obstacle au commerce des services informatiques en Europe.**

Avec la réglementation de la taxation et l'harmonisation des lois, deux importantes barrières commerciales ont pratiquement été éliminées. Sur le versant logistique, la distribution des produits est encore un problème. Pour ce qui est de l'information, la venue d'Internet a apporté une révolution : sa distribution est devenue tellement rapide et économique, qu'elle est perçue comme étant gratuite et immédiatement disponible sur demande. Par conséquent, alors que les trois précédentes barrières du commerce entre les pays ont été abattues, il en reste une quatrième : la barrière du langage<sup>4</sup>. L'effort requis pour présenter un produit dans la langue du client final peut varier de façon remarquable et dépend du type de produit pris en considération. Gé-

<sup>2</sup> Article 14 du Traité de la CE.

<sup>3</sup> Seule exception permise – en cas d'empiètement sur l'intérêt général, telle que la protection de la santé, du consommateur ou de l'environnement – des conditions rigoureuses sont imposées.

<sup>4</sup> Cette simplification ignore le fait qu'il existe – outre les problèmes liés à la Langue, des différences culturelles qui doivent être prises en considération lorsque l'on introduit un bien ou un service sur un marché local. Mais bien souvent, cela représente un problème mineur face à l'effort requis pour la traduction.

néralement, cet effort est proportionnel aux services informatifs demandés. Ces considérations mènent au tableau suivant :

|              | Biens | Information |
|--------------|-------|-------------|
| Taxation     | 😊     | 😊           |
| Conformité   | 😊     | 😊           |
| Distribution | 😐     | 😊           |
| Langue       | 😐     | 😞           |

Tableau 1 : conditions commerciales pour le commerce international au sein de l'U.E, pour les biens et l'information. Pour les services de l'information, la barrière de la langue représente un obstacle de taille à franchir.

Des quatre barrières prises en considération, la langue demeure le majeur obstacle, spécialement pour la fourniture de services informatifs dans différents pays. En utilisant les technologies du langage qui font l'objet de ce document, cet obstacle considérable, qui représente l'ultime et la plus grande différence entre notre marché commun européen et un grand marché intérieur tel que celui des États-Unis, sera éliminé – avec des avantages économiques significatifs.

## 2.2. Les langues dans le monde

Sur une échelle globale, il existe 6.912 langues connues et vivantes<sup>5</sup>. Nombre de ces langues se trouvent en Asie/Pacifique et en Afrique (cf fig. 2).

<sup>5</sup> Source : [Gor]. – Il faut pour le moins mentionner le fait qu'il existe un problème d'identification de la langue. – Une langue vivante est telle s'il existe au moins un locuteur natif.

<sup>6</sup> [Cry] et autres sources. L'on remarquera que les chiffres cités pour la littérature sont inconsistants, et qu'ils dépendent beaucoup du niveau de compétence présumé dans la langue étrangère.

| Pos. | Langue native       | Millions de personnes |          |
|------|---------------------|-----------------------|----------|
|      |                     | source A              | source B |
| 1    | Chinois             | 1.113                 | 1.123    |
| 2    | Anglais             | 372                   | 322      |
| 3    | Indien/Pakistanaïis | 316                   | 236      |
| 4    | Espagnol            | 304                   | 266      |
| 5    | Arabe               | 201                   | 202      |
| 6    | Portugais           | 165                   | 170      |
| 7    | Russe               | 155                   | 288      |
| 8    | Bengali             | 125                   | 189      |
| 9    | Japonais            | 123                   | 125      |
| 10   | Allemand            | 102                   | 98       |
| 11   | Français            | 70                    | 72       |
| 12   | Italien             | 57                    | 63       |
| 13   | Malais              | 47                    | 47       |

Tableau 2 : Les principales langues du monde en millions de locuteurs natifs selon deux différentes sources, (A) The English Company's engco model [Gra] et (B) chiffres comparatifs tirés de Ethnologue ([Gri]; cf [Gra]).

### 2.2.1. Les principales langues du monde

L'anglais est probablement la langue la plus importante, mais en terme de locuteurs natifs, elle est dépassée par le chinois. Toutefois si l'on ajoute aux locuteurs natifs anglais (375 millions) les personnes parlant anglais comme seconde langue (375 autres millions) ainsi que celles qui le parlent comme langue étrangère (750 millions), l'on atteint le résultat significatif de 1,5 milliard de personnes<sup>6</sup> en mesure de parler anglais.

|    | <b>Langue</b>     | <b>Influence</b> |
|----|-------------------|------------------|
| 1  | Anglais           | 100              |
| 2  | Allemand          | 42               |
| 3  | Français          | 33               |
| 4  | Japonais          | 32               |
| 5  | Espagnol          | 31               |
| 6  | Chinois           | 22               |
| 7  | Arabe             | 8                |
| 8  | Portugais         | 5                |
| 9  | Malais            | 4                |
| 10 | Russe             | 3                |
| 11 | Indien/Pakistanaï | 0,4              |
| 12 | Bengali           | 0,09             |

Tableau 3 : 'Influence globale' des 12 langues principales selon le modèle engco (cf tableau 2). Le classement 100 représente la position de l'anglais en 1995 [Gra].

Le tableau 2, tiré de la littérature existante, donne la liste complète des premières langues, classées en terme de locuteurs natifs. Il est intéressant de noter que ce tableau présente les 13 premières langues<sup>7</sup>, et non pas les 10 ou 20 premières; un numéro choisi, du moins selon les apparences, pour pouvoir inclure le français et l'italien. Quand il s'agit de la langue, même dans les arides statistiques, entrent en jeu le sentiment national, les questions culturelles et les perspectives nationales (ou européennes). La langue est très liée à la culture et au sentiment d'appartenance à une communauté.

Il est plutôt évident que l'importance que nous attribuons intuitivement à une langue n'est pas en relation directe avec le nombre de personnes qui la

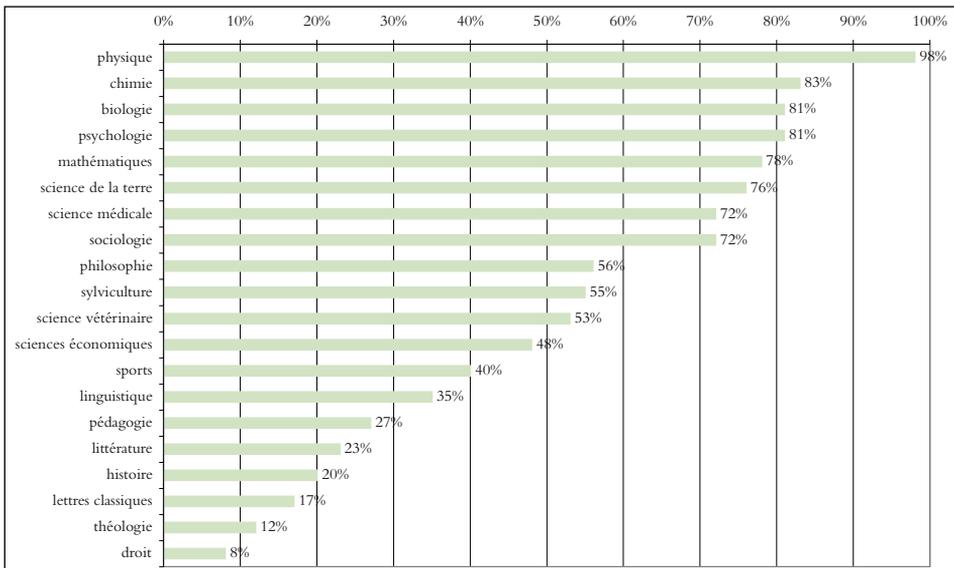


Tableau 4 : Disciplines dans lesquelles les universitaires allemands déclarent que l'anglais est leur langue de travail [Gra].

<sup>7</sup> Ce tableau est tiré de [Gra] qui indique les 13 premières langues.

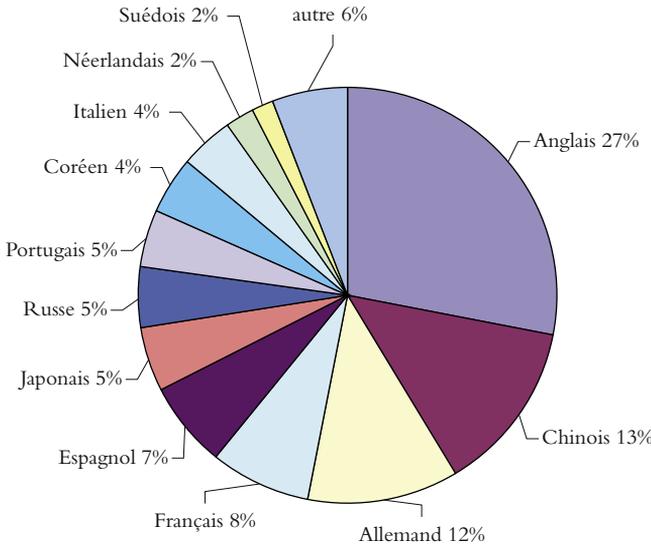


Fig. 1 : Proportion des livres publiés chaque année dans le monde dans chaque langue. L'anglais est la langue la plus utilisée pour la publication des livres: plus de 60 pays publient leurs titres en anglais [Gra].

parent. Par exemple, le français est perçu comme une langue importante et il s'agit de la deuxième langue plus enseignée dans les écoles européennes après l'anglais. L'importance (perçue) d'une langue dépend aussi d'autres facteurs, telle que l'importance économique et politique (outre l'héritage culturel). Le *tableau 3* indique l'importance globale de certaines langues par rapport à l'an-

glais. L'on remarquera que l'importance d'une langue varie selon la discipline ou l'aire d'utilisation. Un exemple particulier est représenté au *tableau 4*.

### 2.2.2. Langues en voie d'extinction

Le danger d'extinction des langues est lié à deux facteurs : le nombre de locuteurs et le nombre de fonctions pour lesquelles la langue est utilisée. Générale-

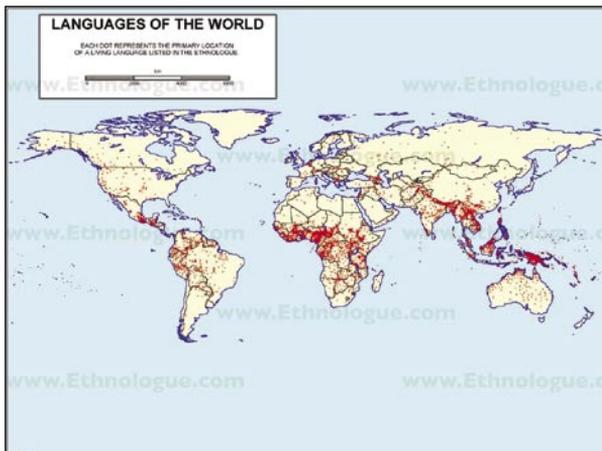


Fig. 2 : Les langues du monde. Chaque point représente la principale localisation des langues vivantes citées dans Ethnologue.

ment, les personnes bilingues commencent par utiliser seulement la deuxième langue avec leurs enfants, ou bien ils utilisent leur langue maternelle de moins en moins fréquemment. Selon Ethnologue<sup>8</sup>, 500 langues environ sont considérées comme étant presque éteintes<sup>9</sup>. Il existe une certaine préoccupation pour ce risque d'extinction, car la langue est étroitement liée à la culture; ce lien est tel, que l'extinction de la langue est presque toujours accompagnée par la disgrégation sociale et culturelle. (À la grande surprise de l'auteur de ce rapport, une autre préoccupation est la perte d'importance de la communauté universitaire qui étudie ces langues!)

### **2.2.3. L'importance de la diffusion : langues primaires, secondaires et tertiaires et forces de marché**

Il n'est pas nécessaire qu'une langue court le risque de s'éteindre pour être désavantagée. Prenez une société ayant l'intention de développer ses affaires du marché local à un marché international. Étant donné que le coût de la localisation<sup>10</sup> pour une nouvelle langue ne dépend pas du nombre des locuteurs, mais qu'il est plus ou moins fixe, il existe sûrement des langues primaires – pour lesquelles il est nécessaire d'avoir une offre de marché – des langues secondaires et, peut-être même des langues tertiaires ayant moins d'importance commercia-

le. Qu'une langue soit considérée secondaire ou tertiaire, cela dépend des circonstances, mais l'on peut affirmer que les forces du marché pénalisent certaines langues, généralement celles ayant une population peu nombreuse parlant cette langue; ou celles qui sont associées à des économies faibles.

L'on retrouve la même situation quand les personnes commencent à étudier une langue étrangère. Un allemand étudierait-il le néerlandais, qui ressemble à l'allemand et qui demanderait par conséquent un moindre effort, ou étudierait-il l'espagnol, étant donné que la population mondiale parlant espagnol est extrêmement plus nombreuse que celle parlant néerlandais? (Cf *tableau 2* pour les chiffres typiques) Et pourquoi étudier le néerlandais quand la plupart des néerlandais parlent anglais?<sup>11</sup>. Nombreuses sont les considérations qui influencent le choix. La langue est parlée dans un pays voisin ou très lointain? Existe-t-il dans ce pays une seconde langue pouvant être utilisée pour communiquer?

### **De nombreuses langues sont sous pression à cause des forces du marché.**

L'attrait d'une langue augmente avec la probabilité que l'on ait besoin de la parler. Naturellement, les langues parlées par une petite communauté sont encore plus désavantagées.

<sup>8</sup> Ethnologue [Gor] est un catalogue de plus de 6.700 langues parlées dans 228 pays.

<sup>9</sup> *Presque éteintes*: ainsi sont définies les langues pour lesquelles la population qui parle cette langue est composée de moins de 50 personnes ou d'une fraction très petite du groupe ethnique.

<sup>10</sup> Per *localisation* l'on entend l'adaptation d'un produit ou d'un service au marché local; explications au chapitre 2.6.2. : *Les marchés consolidés de la localisation et de la traduction*.

<sup>11</sup> 91% de la population néerlandaise gère une conversation dans une autre langue au moins (EB5).

### 2.3. Un regard de près sur l'Union Européenne

La langue est étroitement liée à la culture. Dans certains cas, c'est une composante essentielle de l'identité nationale. Il n'est donc pas étonnant que nous, européens, en créant

|             |    |                 |
|-------------|----|-----------------|
| Allemand    | DE | Deutsch         |
| Anglais     | EN | English         |
| Danois      | DA | Dansk           |
| Espagnol    | ES | Español         |
| Estonien    | ET | Eesti           |
| Finnois     | FI | Suomi           |
| Français    | FR | Français        |
| Grec        | EL | Elinika         |
| Hongrois    | HU | Magyar          |
| Italien     | IT | Italiano        |
| Letton      | LV | Latviesu valoda |
| Lithuanien  | LT | Lietuviu kalba  |
| Maltais     | MT | Malti           |
| Néerlandais | NL | Nederlands      |
| Polonais    | PL | Polski          |
| Portugais   | PT | Português       |
| Slovaque    | SK | Slovenčina      |
| Slovène     | SL | Slovenščina     |
| Suédois     | SV | Svenska         |
| Tchèque     | CS | Čeština         |

Tableau 5 : Les 20 langues officielles de l'Union Européenne et leurs abbréviations [ELPI]. L'Irlandais sera la 21e langue officielle de l'UE à partir du premier janvier 2007.

l'Union Européenne, ayons pris la décision, en connaissance de cause, de ne pas introduire une langue primaire mais de conserver les différentes langues et de leur attribuer les mêmes droits. Bien qu'il ne faille pas sous-estimer la grande importance de l'anglais en tant que langue franche, l'on peut tranquillement affirmer que nous vivons dans un monde multilingue où la langue est importante. L'Union Européenne est née multilingue : au point qu'un Commissaire à l'Education, la Formation, la Culture et au Multilinguisme a été nommé.

Les différentes langues ont la même importance, et certains documents, en particulier les lois et les débats parlementaires, sont disponibles (c'est-à-dire qu'ils ont été traduits) dans toutes les langues officielles. Sur l'autre versant, pour des raisons d'efficacité et d'économie dans l'opérativité quotidienne trois langues sont utilisées : l'anglais, le français et l'allemand.

En résumé, nous avons de bonnes raisons de protéger notre héritage culturel, mais le maintien de plusieurs langues a un coût<sup>12</sup>, en termes économiques (par exemple les coûts de localisation d'un produit) et en termes d'engagement (l'effort demandé à un individu pour apprendre une autre langue). Une quelconque technologie qui serait en mesure de réduire ces coûts représenterait un soutien valable pour notre héritage culturel.

<sup>12</sup> Ne pas fournir de traduction n'est pas une option: les citoyens ont besoin de comprendre les lois, les produits doivent être localisés dans les différentes langues pour être vendus. Nous sommes convaincus que tout frais, même bas, constitue un obstacle, et que toute réduction en terme de coût et d'amélioration à l'accessibilité aura un impact positif.

## Interview de Karl-Johan Lönnroth, Directeur Général, Direction Générale de la Traduction (DGT)

Depuis janvier 2004, Karl-Johan Lönnroth est Directeur Général de la DGT. Il était auparavant: 2000-2003, Vice Directeur Général, Direction pour l'Emploi et les Affaires Sociales (CEC, Commission européenne). 1996-2000, Directeur de la Stratégie pour l'Emploi et les Fonds sociaux européens, Direction Générale pour l'Emploi et les Affaires sociales (CEC). 1991-1996, Directeur du Département pour l'emploi, International Labour Office, Genève. 1971-1991, diverses fonctions au Ministère finlandais du travail en tant que chercheur, responsable de la planification, vice directeur des services pour le marché du travail, et special advisor. 1973-1977, Secrétariat du Conseil nordique des ministres, fonctionnaire responsable à l'emploi, à l'immigration et à la coopération tripartite. Conseil pour l'OECD, le Conseil nordique des ministres et l'Ambassade finlandaise à Stockholm.



**Karl-Johan Lönnroth**  
Director-General, Direction Générale  
de la Traduction (DGT)  
Commission européenne  
Luxembourg, Luxembourg et Bru-  
xelles, Belgique

Formation: Master of Political Sciences, University of Helsinki (1970), Master of Arts, University of Wisconsin, USA (1972), Ecole Nationale d'Administration (ENA) (1983). Aptitude linguistique: finnois, suédois, anglais, français et allemand; connaissance élémentaire du russe et de l'espagnol.

Karl-Johan Lönnroth a une vaste expérience dans la coopération: Pays Nordiques, OECD, UN, ILO; et dans la coopération bilatérale, par exemple Europe orientale. il a publié plus de 40 articles, publications etc, relatifs au travail, à l'emploi, à l'émigration, aux questions sociales et politiques et aux thèmes sociaux.

### *La DGT est une organisation plutôt grande, n'est-ce pas?*

Je suppose que nous sommes la plus grande au monde. Nous travaillons avec 21 langues. Le chiffre global pour tout le service linguistique de l'Union Européenne s'élève à 1.1 milliards d'€uros, soit environ 1% de tout le budget de l'UE. Ceci comprend l'interprétariat et la traduction, et concerne non seulement la Commission mais aussi toutes les institutions tel que le Parlement européen. Cela semble beaucoup, mais en tout cela représente 2,55 € par citoyen.

### *Le grand nombre de langues n'est-il pas un poids pour la société et l'économie européennes?*

Cette question a déjà une réponse implicite. L'on présuppose que ce soit un poids et un coût plutôt qu'un avantage. Le multilinguisme appartient vraiment à un modèle social européen; nous avons cette diversité culturelle. Il faut la considérer comme un point fort et un élément dynamique plutôt que comme une entra-

ve, et je peux affirmer qu'il est évident qu'avoir plusieurs langues ouvre les marchés. La politique du multilinguisme favorise notre société dynamique. En Europe, nous respectons les libertés fondamentales et la diversité culturelle, et à travers cette diversité, l'on peut obtenir des idées nouvelles et donner un nouvel élan à l'économie. Ne connaître qu'une langue est une limitation au sein du marché unique européen du travail. C'est une des raisons pour lesquelles l'Union essaie de promouvoir la connaissance et l'apprentissage des langues.

### **Comment la DGT a-t-elle réagi au défi de l'élargissement de l'UE?**

Durant cette dernière année, nous avons embauché plus de 500 personnes dans notre staff. Je pense que nous affrontons plutôt bien ce nouveau défi. Il nous est difficile de développer et de forger de nouveaux termes et d'améliorer la qualité des traductions, car certains termes de politique européenne n'existent pas dans les nouvelles langues. C'est le défi principal. Nous traduisons généralement un texte législatif en partant de zéro et pour la première fois, et quelquefois ceci est assez compliqué. L'intégration de ces nouvelles langues a assez bien marché.

### **Comment pensez-vous que la traduction automatique se développera?**

La traduction automatique, en tant qu'instrument de traduction avec la *mémoire de traduction* est quelque chose qui est encore en train de se développer, et, selon moi, elle a un brillant avenir. Mais cela ne veut pas dire que l'élément humain diminuera. Le développement va plutôt dans la direction d'une combinaison que je qualifierais de traduction intelligente, le traducteur assurant donc la qualité, la cohérence, la terminologie correcte, et la traduction automatique aidant à comprendre et à améliorer la productivité.

### **Comment voyez-vous la profession de traducteur?**

L'industrie des technologies du langage est l'industrie qui croît le plus vite au monde. La globalisation et la multitude des langues accroît le besoin de services multilingues, ce qui signifie également que cette profession doit être mieux reconnue, car elle joue un rôle important. Vu les grands défis, la figure du traducteur changera elle aussi; ce ne sera plus un simple traducteur de textes, ce sera un conseiller linguistique, un editor. Ainsi continuera la diversification de la profession.

Nous sommes en contact avec des universités et leurs associations pour développer les curricula et la formation, et nous essayons de réaliser ce que nous appelons le *Master européen en traduction*, qui devrait offrir une qualification standard, reconnue par tous, facilitant ainsi les embauches. Ceci devrait mener à une meilleure professionnalisation.

## 2.4. La communication

Nombre d'européens parlent deux ou plusieurs langues. Mais environ la moitié des citoyens de l'Union Européenne ne parlent pas de langue autre que la leur.

Ne serait-ce pas un progrès significatif si les Européens qui ne parlent pas la même langue pouvaient se parler facilement? Une quelconque forme de communication, même à un niveau réduit, serait un progrès par rapport à la situation actuelle. La traduction ne doit pas nécessairement être parfaite. Les qualités requises sont très différentes de celles exigées pour une traduction professionnelle. Comme nous le verrons par la suite, ceci a un effet fondamental sur les forces du marché et spécialement sur les perspectives des nouveaux marchés.

Pouvons-nous imaginer ce que voudrait dire pour nous, aujourd'hui, pouvoir utiliser de façon naturelle et à bas prix des systèmes de traduction? Nous

avons constaté combien peut faire la technologie pour améliorer la communication entre les personnes en réduisant les distances de temps et d'espace. L'une des premières et des plus significatives conquêtes de la culture humaine a été l'essor de l'écriture, qui a permis de transmettre le savoir aux futures générations. Pour ce qui est de l'espace, il suffit de penser à la transmission des informations le long des frontières des empires romain et chinois en utilisant les signaux visuels. Autres pas de géant ont été l'introduction du service postal, du téléphone, de la ligne de télécommunication transatlantique, du téléphone cellulaire et d'Internet.

Nous affrontons actuellement un autre grand défi : la disponibilité instantanée et économique de la traduction, permettant ainsi la communication entre les cultures au-delà du temps et de l'espace. Ce défi remportera-t-il un succès lors des vingt prochaines années? Il nous est difficile d'imaginer un avenir où chacun pourrait communiquer, à

| Langue      | parlée comme langue maternelle | parlée non comme langue maternelle | total locuteurs |
|-------------|--------------------------------|------------------------------------|-----------------|
| Anglais     | 13%                            | 34%                                | 47%             |
| Allemand    | 18%                            | 12%                                | 30%             |
| Français    | 12%                            | 11%                                | 23%             |
| Italien     | 13%                            | 2%                                 | 15%             |
| Espagnol    | 9%                             | 5%                                 | 14%             |
| Polonais    | 9%                             | 1%                                 | 10%             |
| Néerlandais | 5%                             | 1%                                 | 6%              |
| Russe       | 1%                             | 5%                                 | 6%              |

Tableau 6 : Capacité de parler des langues étrangères dans l'UE : pourcentage de citoyens européens parlant leurs langues respectives dans l'UE (comme langue maternelle / comme deuxième langue ou langue étrangère / les deux). L'enquête a été menée dans l'Union Européenne des 25 États membres et, de surcroît, dans les pays entrant (Bulgarie et Roumanie), dans les pays candidats (Croatie et Turquie) et au sein de la communauté turco-chypriote. - Source : [EB5]

des coûts acceptables, avec des personnes parlant des langues différentes en utilisant des systèmes de traduction instantanée du parlé et de l'écrit, tout simplement parce que actuellement ce type de communication n'existe pas. Il est toutefois raisonnable de penser que la disponibilité répandue d'un tel instrument aurait un impact aussi significatif que celui qui a fait suite à l'introduction du téléphone ou d'Internet.

**Après avoir rapproché le temps et l'espace, la liaison entre langues et cultures est le prochain grand défi.**

## **2.5. Le prochain pas dans l'industrialisation : des machines élaborant le texte et la parole.**

Comme nous l'avons à peine fait remarquer, la disponibilité de la traduction à bas prix en tout lieu conduira à des changements fondamentaux. Nous sommes en face d'une véritable révolution : un processus qui, auparavant, ne pouvait être exécuté que par des personnes, peut désormais être exécuté par des machines, et de façon très efficace. Dans les années 60, la technologie de l'information est passée des applications gouvernementales, scientifiques et militaires au domaine commercial, d'abord dans le secteur des assurances et des banques et, plus tard, dans pratiquement tous les secteurs de la vie humaine. La plupart du travail qui, auparavant, était exécuté par des personnes compétentes et expertes, est à présent effectué par des machines plus rapides, économiques et efficaces. La capacité d'élaborer l'information était cependant limitée et

l'élaboration du langage, exception faite pour les données structurées tels que les noms et les adresses, était pratiquement inexistante. Alors que la production de textes écrits, tels que les quotidiens, les livres et le matériel imprimé en général était fortement soutenue par la TI (technologie de l'information), l'analyse et l'élaboration de l'information demeurait très limitée.

La venue d'Internet a porté de nombreux changements. La nécessité d'élaborer le langage s'est accrue énormément pour effectuer des recherches, des synthèses, des traductions et des classifications. L'introduction à grande échelle des libres-services sur Internet a mené à une croissance considérable de la communication avec les usagers finaux, que ce soit pour les entreprises ou pour l'administration. En même temps, les perspectives des usagers se sont accrues, et la plupart d'entre eux donnent pour acquis qu'un e-mail aura une réponse dans les 24 heures. La recherche de documents sur Internet - en se basant sur des mots-clés - effectuée principalement en passant par Google, a pratiquement réussi à entrer dans toutes les maisons. Mais, immédiatement, deux problèmes sont apparus. En premier lieu, la recherche de paroles n'est qu'un substitut de la recherche d'informations, et l'on préférerait avoir à disposition un web sémantique (*semantic web*) et la possibilité de chercher des concepts plutôt que des paroles. En deuxième lieu, étant donné qu'Internet est utilisé par une grande partie de la population, il y a un besoin croissant de localisation, spécialement en ce qui concerne les langues. Ceci crée une énorme demande de tra-

duction sur une grande échelle et/ou en temps réel, spécialement pour la recherche de documents dans des langues différentes et pour la production de sites web multilingues, qui sont devenus l'un des majeurs volants de l'industrie mondiale de la traduction tels que IBM et SAP, Google et Yahoo.

**La traduction est guidée, sur une grande échelle, par la globalisation et par internet.**

## 2.6. Le marché

Si l'on considère la traduction d'un point de vue économique, il n'est pas suffisant de regarder le marché actuel de la traduction, il faut considérer d'autres aspects importants. Pour l'Europe, les frais de traduction et leur effet sur l'économie sont extrêmement importants si l'on considère que, outre les frais directs pour traduire, les barrières linguistiques peuvent être un obstacle à l'exploitation de nouvelles opportunités.

### 2.6.1. La traduction comme facteur de coût

D'une part, si la richesse culturelle de l'Europe, avec ses multiples langues, nous tient à cœur, d'autre part les barrières linguistiques pèsent sur notre économie, vu qu'elles ajoutent des coûts de transaction à toute activité traversant les frontières linguistiques. Dans de nombreux cas typiques, ce coût n'est qu'une petite fraction du coût du produit, à titre indicatif il pourrait être compris entre une valeur minimale de 0,25% et une valeur maximale de 2%.

Certains de ces coûts sont directs (bien qu'il ne soit pas facile de les déterminer) et d'autres sont indirects en ce sens qu'ils empêchent la création de valeur.

Il est plutôt difficile d'évaluer les coûts directs de façon crédible. Dans une grande société, il peut, par exemple, y avoir un budget pour la réalisation d'un site web. Mais il est difficile de comptabiliser toutes les petites activités de traduction menées dans les différents secteurs qui participent à son entretien; des activités qui ne sont pas couvertes par des chiffres définis et qui pourraient facilement faire lever le budget établi. D'autre part même en comptabilisant le salaire de tous les traducteurs, cela serait improbable, vu que nombre d'entre eux travaillent à mi-temps ou comme professionnels. Des difficultés de quantification se manifestent si une organisation paie une société de localisation utilisant un sous-contractant qui, à son tour, paie un professionnel : il faudrait comptabiliser le coût au long de toutes les phases ou seulement dans la première phase? Par conséquent, les chiffres à disposition changent selon la méthode utilisée<sup>13</sup>. Il semble que l'on peut affirmer que le marché global de la traduction oscille entre 8 et 30 milliards d'Euros. Suivent deux données sur les frais du gouvernement de l'U.E. :

- Chaque année, le Parlement européen dépense 300 millions d'Euros, soit 30% de son budget, pour la traduction de tous les débats parlementaires et des documents de l'U.E. dans les 20 langues européennes officielles.

<sup>13</sup> Vous pourrez en savoir plus sur ce sujet en lisant l'interview de Renato Beninato.

- L'Union Européenne dépense 1,1 milliard d'euros par an, soit 1% de son budget, pour tous les services de traduction et d'interprétariat.

### **L'UE dépense 1,1 milliard d'euros par an pour les services de traduction et d'interprétariat.**

#### **2.6.2. Les marchés consolidés de la localisation et de la traduction**

L'opinion publique a au moins une idée approximative du travail exécuté par les traducteurs, qui traduisent des livres ou des documents, mais généralement on ne sait pas grand-chose de l'industrie qui s'occupe de la localisation du software et des pages internet, et qui représente la plus grande partie de l'industrie des technologies du langage. Il est important de comprendre que les sociétés et les institutions concernées ont réellement besoin de la « localisation » et non seulement de la traduction. L'aspect de la traduction mis à part, la localisation du software garantit le bon fonctionnement de ce dernier dans la nouvelle langue : cela signifie aussi intervenir sur les aides et sur la documentation online, sur le fait que le matériel écrit soit formaté de façon appropriée. Il faut, par exemple, prendre en considération la différente longueur des textes traduits, que ce soit en nombre de paroles qu'en nombre de caractères, et il ne faut pas sous-estimer les styles d'écriture. Outre les langues qui s'écrivent de gauche à droite ou de droite à gauche, il faut prendre en considé-

ration le cas des textes bidirectionnels : si l'on inclut des mots latins dans un texte arabe, il faut supporter les deux directions d'écriture! Il existe de nombreux ensembles de caractères qui doivent être supportés, ainsi que des langues avec des caractères représentés par deux bytes tels que le japonais, le chinois et le coréen qui requièrent un support software particulier (le système Unicode est une solution à ces problèmes, mais il se pourrait qu'on ne puisse l'appliquer sur de vieux documents électroniques). En ligne de principe, le software que l'on entend localiser devrait être conçu comme tel dès le début, afin d'éviter des coûts non nécessaires par la suite. En outre, les différences culturelles jouent un rôle important dans la localisation. Pour transmettre un concept dans une langue ou une culture différente et faire en sorte qu'il reste compréhensible, la créativité et l'empathie sont nécessaires.

La méthode habituelle pour la localisation consiste à partager le produit entre ses composantes textuelles et celles de l'interface usager. L'on traduit le texte, et l'interface usager ainsi que la documentation sont modifiées de façon à ce qu'elles fonctionnent de manière adéquate dans la langue traduite.

### **Les principaux segments du marché de la localisation sont la localisation du software et des sites web.**

Selon une source<sup>14</sup>, l'industrie des technologies du langage a produit un re-

<sup>14</sup> Common Sense Advisory [CSA]. D'autres sources estiment que le chiffre global est beaucoup plus élevé, de l'ordre de 30 milliards d'euros. Dans ce document, ce n'est pas le chiffre exact qui compte mais le fait que l'industrie du langage est une industrie qui a bien démarré et qui est en expansion, et si ses dimensions sont inférieures à d'autres industries, elle agit indirectement sur des activités commerciales aux volumes beaucoup plus importants.

venu de 8,8 milliards de US dollars en 2005. Ceci comprend aussi bien la traduction humaine que la traduction faite en utilisant des instruments. Deux importants segments de marché guident actuellement la croissance : le traitement des sites multilingues et la localisation du software. Les sociétés qui contribuent à ce marché sont au nombre de 5.000 dans le monde entier, en ne comptant que celles ayant au moins cinq employés.

Les majeures technologies utilisées

actuellement pour la traduction sont : les mémoires de traduction (MT), les ressources terminologiques, les software et les instruments pour traiter les sites web multilingues et pour faire de la localisation software. Les MT sont une technologie de support bien affirmée (même si elle n'est pas universellement utilisée) qui abat les coûts et améliore la qualité, en particulier la cohérence de la traduction, importante dans les domaines techniques et juridiques, par exem-

|    | Société                     | Pays des Sièges Sociaux | Revenus en US\$ M | Employés | Bu-reaux | Statut |
|----|-----------------------------|-------------------------|-------------------|----------|----------|--------|
| 1  | Lionbridge Technologies     | US                      | 377,1             | 4.000    | 50       | Public |
| 2  | Titan Corp.                 | US                      | 285,4             | --       | --       | Public |
| 3  | SDL International           | UK                      | 146,0             | 1.400    | 36       | Public |
| 4  | STAR AG                     | CH                      | 96,0              | 750      | 33       | Privé  |
| 5  | RWS Group                   | UK                      | 63,4              | 350      | 7        | Public |
| 6  | SDI Media Group             | US                      | 60,3              | 200      | 20       | Privé  |
| 7  | Xerox Global Services       | UK                      | 60,0              | 200      | 4        | Public |
| 8  | Euroscript S.à.r.l.         | LU                      | 54,5              | 600      | 9        | Privé  |
| 9  | Transperfect/Translations   | US                      | 50,2              | 325      | 29       | Privé  |
| 10 | CLS Communication           | CH                      | 36,0              | 260      | 11       | Privé  |
| 11 | Logos Group                 | IT                      | 36,0              | 150      | 17       | Privé  |
| 12 | LCJ EEIG                    | DE/IT/BE/SP             | 21,6              | 140      | 9        | Privé  |
| 13 | Thebigword                  | UK                      | 20,0              | 122      | 7        | Privé  |
| 14 | Hewlett-Packard ACG         | FR                      | 20,0              | 65       | 6        | Public |
| 15 | Moravia                     | CZ                      | 19,0              | 350      | 11       | Privé  |
| 16 | TOIN                        | JP                      | 19,0              | 105      | 5        | Privé  |
| 17 | Merrill Brink International | US                      | 18,5              | 120      | 4        | Privé  |
| 18 | VistaTEC                    | IE                      | 18,2              | 123      | 3        | Privé  |
| 19 | Transware                   | IE                      | 18,0              | 160      | 8        | Privé  |
| 20 | McNeil Multilingual         | US                      | 17,2              | 105      | 9        | Privé  |

Tableau 7 : Placement des 20 premiers fournisseurs de services linguistiques – revenus 2004. (Ces chiffres prennent en considération les deux importantes acquisitions de 2005, l'acquisition de Bowne Global Solutions par Lionbridge et l'acquisition de TRADOS par SDL.)

ple. De même, l'on peut s'attendre à ce que la traduction automatique ou la traduction du langage parlé, utilisées comme un instrument, aient un effet bénéfique sur les prix et sur la qualité de la

traduction humaine. Comme illustré au chapitre 2.6.4., qui définit la terminologie utilisée, la traduction automatique, dans ce contexte, est une innovation de support (sustained).

## Interview de Renato Beninatto, COO, Common Sense Advisory

Beninatto a plus de 20 ans d'expérience au niveau directorial dans l'industrie de la localisation. Il a fait partie des dirigeants de quelques-unes des plus importantes sociétés, plus récemment en qualité, respectivement, de vice-président de Alpnet Inc et de directeur de Berlitz GlobalNET. Il s'est concentré sur les stratégies qui guident la croissance à une échelle globale. C'est un expert pour conduire une société au succès sur le marché global et pour implanter un business au niveau international. Il est actuellement partenaire et analyste responsable de la recherche auprès de Common Sense Advisory, Inc. une société-conseil et de recherche de marché spécialisée dans la traduction et l'industrie de la localisation, avec des clients sur tous les continents.



**Renato Beninatto**  
Chief Operations Officer et  
VP de Consulting Practice  
Common Sense Advisory, Inc.  
Boston, USA

### *J'ai vu des chiffres sur le volume du marché global de la traduction qui ne fournissent pas un cadre cohérent. Pourquoi est-il si difficile d'arriver à ces chiffres?*

Établir les dimensions d'un marché est un exercice complexe qui implique les mêmes doses de logique que de scepticisme. Le but est d'atteindre une approximation crédible du marché. D'une part, mon expérience précédente dans la vente de services de traduction pour deux sociétés publiques et, d'autre part, mes contacts personnels avec les grands usagers de la traduction ont amené la Common Sense Advisory à considérer l'industrie de la traduction. Lorsque l'on parle de marché de la traduction aux outsiders, ceux-ci voient inmanquablement d'énormes possibilités, du moment qu'ils s'aperçoivent du décalage existant entre ce qui est publié dans toutes les langues et ce qui est publié dans leur langue. Mais lorsque l'on arrive au fait, c'est-à-dire lorsque l'on essaie de vendre ces services, la réalité est beaucoup plus dure. La traduction n'est pas stratégique et, par conséquent, elle ne rend pas. Nous avons proposé de comparer le prix de la traduction à celui du papier hygiénique pour décrire combien la traduction est présente dans les budgets des sociétés. Le pire défaut des méthodologies de dimensionnement du marché conduites par les *outsiders* est qu'elles se basent sur les informations des acteurs du marché. L'une des caractéristiques clés du marché de la traduction est le processus des sous-contractants allant jusqu'à quatre niveaux. Par exemple, un vendeur multilingue confie les langues de l'Europe de l'Est dont il a besoin à une société en Hongrie, qui à son tour les passe à d'autres sociétés dans la République Tchèque, en Pologne et en Bulgarie, qui à leur tour s'appuient sur des traducteurs pratiquant la libre profession. Chez Common Sense Advisory, nous ne comptons que le premier outsourcing, qui représente l'argent réellement déboursé pour ce travail par l'utilisateur final.

### **Quelle méthodologie avez-vous suivie?**

Dans "Beggars at the Globalization Banquet"\* un rapport que nous avons publié en novembre 2002, nous avons relevé que les sociétés, selon l'industrie et la taille – dépensent entre 0,25% et 2% de leurs entrées provenant de l'étranger. Ceci représente beaucoup moins des 3% des entrées totales que les organisations assument, généralement, comme hypothèses.

Pour arriver à ces chiffres, nous avons contrôlé d'autres approches au dimensionnement du marché pour voir si nos chiffres étaient cohérents. Nous avons ainsi pris en considération le nombre de sociétés de traduction sur le marché et leurs profits, le nombre de traducteurs dans le monde, et les frais documentés des gouvernements. La comparaison croisée de ces données a confirmé nos chiffres.

### **Qu'en ressort-il d'important?**

Nous menons constamment une recherche de façon analytique sur toute l'industrie des technologies du langage. Il en ressort que la traduction est quelque chose qui s'effectue à un niveau de mid-management et qu'elle n'atteint la salle du Conseil d'Administration que lorsque quelque chose va vraiment de travers. Nous avons en outre relevé qu'avec 9,5 milliards de US\$ en 2006, ce marché a presque la même taille que celle du marché mondial de la bicyclette, et que bien que l'on parle beaucoup de la technologie de la traduction, le revenu fini de tous les acteurs ne s'élève qu'à 100 millions de US\$ environ.

### **Vous dites donc que la traduction et la localisation n'ont pas une grande importance commerciale?**

Non. Au contraire. Même quand la traduction est très économique, elle permet aux sociétés de pénétrer sur de nouveaux marchés et de multiplier leurs profits. En réduire le coût facilitera la réelle globalisation des marchés.

### **Pouvez-vous fournir quelques exemples en chiffres pour les grands usagers de la traduction ?**

Comme vous le savez, la DGT a dépensé 1.1 milliard d'euros en 2004 avec l'entrée des 10 nouveaux pays, mais maintenant les chiffres sont retombés aux alentours de 800 millions d'euros. Le bureau de traduction du Canada dépense moins de 200 millions de US\$. Parmi les sociétés privées, Microsoft et Oracle sont celles qui dépensent le plus. Microsoft dépense environ 300 millions US\$ par an et Oracle environ 200 millions de US\$ par an. Les sociétés du secteur automobile dépensent entre 10 et 35 millions de US\$ par an. Les sociétés qui dépensent plus de 3 millions de US\$ par an sur le marché sont peu nombreuses, et ce sont les plus recherchées.

\* Revenus provenant de l'étranger par rapport au pays d'origine.

### **2.6.3. De la traduction humaine à la traduction automatique : extraordinaire réduction des coûts et amélioration de l'accessibilité**

En laissant de côté pour un instant le problème de la qualité, deux parmi les plus grandes variables conduisant le marché encouragent l'utilisation de la traduction automatique : le coût et l'accessibilité. Baisser les coûts d'un ordre de grandeur alimente sûrement l'utilisation de la traduction, vu que la demande est latente et, pour l'instant, loin d'être satisfaite. Une forte croissance de la demande est très probable et plausible, comme on a pu l'observer dans d'autres cas, tels que les transports aériens à bas coûts et l'industrie des télécommunications avec la venue du VOIP. Il faut souligner que ce n'est pas seulement la traduction en tant que telle qui contribue à l'effet économique de la technologie : une part importante est représentée par de nouveaux types de transactions rendus possibles grâce au développement technologique. L'accessibilité est la deuxième variable. Actuellement, la plupart des traductions ne sont pas produites en temps réel, mais sur un support en papier dans un bureau éloigné. Le matériel à traduire est généralement envoyé autre part, et la traduction est livrée après des heures ou des jours. La traduction en temps réel-

le existe, mais elle est coûteuse. Si on la compare à des délais de livraison en heures, la disponibilité instantanée d'un texte traduit ouvre une série de possibilités à de nouvelles applications, qui seraient inaccessibles autrement.

### **Deux principaux facteurs de marché soutiennent la traduction automatique : Coût et accessibilité**

En termes de qualité, la traduction automatique restera, pendant de nombreuses années, inférieure à la traduction humaine. Par conséquent, les différents segments de marché seront dominés ou par la traduction humaine ou par la traduction automatique. Quel produit peut prévaloir dans un segment de marché donné dépend de la particularité et des requêtes spécifiques de ce segment. La traduction humaine prévaudra dans tous les secteurs où la haute qualité représente une nécessité absolue. Par contre, la traduction automatique aura le dernier mot sur les segments les plus bas du marché<sup>15</sup>, et elle dominera les nouveaux marchés ou les nouveaux segments de marché qui émergeront suite à la disponibilité d'une technologie de traduction à bas prix. À l'avenir, avec des améliorations de qualité et de prestations, la traduction automatique augmentera sa quote de marché.

<sup>15</sup> Basso se réfère à la qualité de la traduction. En termes de temps requis et d'accès la traduction automatique est clairement dans le segment élevé du marché. L'on s'attend à ce qu'elle produise des volumes élevés.

## Interview de Michael Anobile, Managing Director de LISA

Michael Anobile a plus de 25 ans d'expérience dans le secteur TI au niveau international. Il s'est diplômé en Sciences de la Communication à l'Université de Syracuse, et il a participé au programme de Master in political communications de l'Université du Maryland. En 1980, il s'établit en Suisse avec sa famille et devient European Training Manager pour Exxon Office Systems, il remplit ensuite une série de fonctions en tant que senior management européen et suisse dans le secteur de la TI et dans les industries du langage et de la technologie, se focalisant sur le développement d'un business global et sur le marketing.



**Michael Anobile**  
Managing Director  
The Localization Industry Standards  
Association (LISA)  
Romainmôtier, Switzerland.

Membre fondateur de LISA (*The Localization Industry Standards Association*), et Managing Director dès le début, il est responsable de la gestion ordinaire de l'association, y compris le conseil aux organisations développant des standard (par exemple ISO, Unicode, Open18N, W3C, OASIS, etc.) et aux agences gouvernementales (par exemple US Department of Commerce, the DoD, FBI, NVTC, NSA et plusieurs institutions asiatiques, canadiennes et européennes qui s'occupent de standard dans les technologies du langage). Il est également responsable de forum internationaux, programmes de training et de projets de marketing industriel & relations publiques.

### *De quoi s'occupe LISA?*

Nous sommes une organisation dirigée par ses propres membres, qui se focalise sur les sociétés, les gouvernements et les ONG. Outre avec le secteur des PME, nous travaillons à un niveau institutionnel – avec des groupes tels que World Bank, McDonald's, Coca-Cola, IBM, European Directorate-General for Translation, Industry Canada ou le Bureau canadien de la traduction. Certains de ces groupes comprennent plus de 1.000 traducteurs. Nous les aidons à comprendre la perspective globale que la globalisation offre en termes de business, technologie, et flux de travail. Ceci comprend l'internationalisation, la traduction et comment concevoir les produits et les services pour le marché global.

### *Quelle est votre principale recommandation pour faire de la localisation de façon correcte?*

La meilleure pratique requiert qu'un produit soit conçu pour un usage international. Par conséquent, l'effort d'internationalisation, le cycle de vie du produit ou du service, les marchés et la distribution sont des thèmes cruciaux. Cette approche "holistique" suppose que l'information soit traduite. La façon dont le processus de

traduction est automatisé est donc fondamental pour la qualité, pour le consensus de l'utilisateur final et pour le coût.

### ***Pourriez-vous évaluer la demande de traduction, si elle était fournie gratuitement?***

Je n'ai pas idée. L'expérience prouve cependant que lorsque l'on demande à un client s'il achèterait ou non une traduction faite de façon automatique, il est très probable qu'il réponde : non. Mais, si une traduction est faite avec un système automatique, puis est offerte gratuitement au client, il est plus probable qu'il la télécharge et utilise ce document. Le dilemme est évident: il y a une certaine résistance envers ce qui était précédemment appelé "technologie de rupture".

### ***En quelle mesure la technologie a-t-elle contribué à la traduction ?***

La Mémoire de Traduction est l'un des principaux apports de l'industrie à la traduction. C'est, en outre, une technologie qui remonte à il y a vingt ans. Et voyez le temps qu'elle a mis pour trouver son chemin. Je ne puis que confirmer l'efficacité de TRADOS, SDL et autres développeurs de MT tels qu'Atril et Logos, car ils ont vraiment travaillé dur pour implémenter les Mémoires de Traduction à grande échelle, dans de nombreux secteurs commerciaux. MT adopte le concept fondamental de l'editing: si vous avez tapé quelque chose une fois, vous ne devrez plus le taper. Ainsi lorsque vous traduisez quelque chose dont vous savez qu'il sera encore utilisé, vous n'avez plus besoin de traduire cette phrase par la suite. Vous devez, naturellement, contrôler dans quel contexte vous opérez, et vous assurer du réemploi à travers différentes plateformes et instruments. C'est de cette façon que la MT standard de LISA, TMX©, contribue à l'industrie.

### ***Quel est votre point de vue sur la traduction automatique?***

La traduction automatique (TA) est très efficace et valable en terme de coût en cas d'applications de traduction bien définies tels que les bases de données de la connaissance, les call centers, et la documentation technique. La meilleure pratique de l'industrie est de contrôler les processus de gestion de la terminologie et d'authoring. Comprendre comment l'utilisateur final interagira avec l'application, comprendre le niveau d'information et la qualité requise, rendra possible la rationalisation du processus de traduction, en définissant, construisant et maintenant des terminologies spécifiques. L'on aura, par conséquent, une traduction plus soignée et une majeure diffusion. TA travaille très bien lorsque les attentes sont satisfaites et que les ressources adéquates sont employées.

### ***L'une des intentions de ce rapport est de faire croître la conscience envers les technologies de traduction automatique, mais aussi envers la traduction elle-même. Quel est votre message?***

Il est très important que nos leaders politiques et économiques comprennent l'importance des langues. La langue facilite, elle n'est pas un obstacle, elle accroît

la compréhension et la coopération entre les personnes et les cultures. La langue fournit un vaste accès aux informations sociales et politiques les plus importantes. Je suis d'accord avec l'objectif d'augmenter la conscience publique, car il s'agit de quelque chose de plus qu'une simple traduction. Il s'agit de communiquer et de donner accès à des données importantes pouvant aider les personnes à être plus protagonistes d'un point de vue économique, social et politique. Un bon exemple nous est donné par la politique sur le langage de la Direction de la Traduction de l'Union Européenne pour soutenir le développement des marchés de l'Europe orientale et pour les nouveaux pays membres.

### 2.6.4. Innovations technologiques incrémentielles et de rupture

N'est-ce pas étrange que de grandes entreprises, leaders sur le marché et bien gérées, ne réussissent pas à exploiter les innovations émergentes? Bon, nous comprenons désormais beaucoup mieux qu'il y a dix ans les dynamiques de l'innovation. Ce sont certaines caractéristiques innovatrices qui déterminent en grande partie le succès et la possibilité de créer un leader tout à fait nouveau sur le marché; ces caractéristiques sont utilisées pour classifier les innovations en *innovations de soutien* ou *innovations de rupture*, selon la définition qui suit<sup>16</sup>.

La plupart des innovations technologiques améliorent les prestations d'un produit. Ce type d'innovations est appelé *de soutien* (*sustained*). Qu'elles soient incrémentielles ou discontinues,

elles ont en commun l'amélioration des prestations de certains produits selon des critères historiquement consolidés sur les plus grands marchés.

Par contre, il arrive que les innovations empirent les prestations d'un produit, du moins à court terme, mais elles possèdent d'autres caractéristiques qui ont de la valeur pour de nouveaux clients. Ces innovations sont appelées *innovations de rupture* (*disruptive*).

Les innovations de soutien ne changent fondamentalement pas le panorama du marché, c'est-à-dire que le leader du marché les développe (ou les achète) et maintient sa propre leadership. Vu que, par contre, les innovations de rupture offrent des prestations inférieures par rapport à l'état de l'art ou à l'offre de produits existants, celles-ci n'ont aucune valeur pour le leader du marché à ce moment précis.

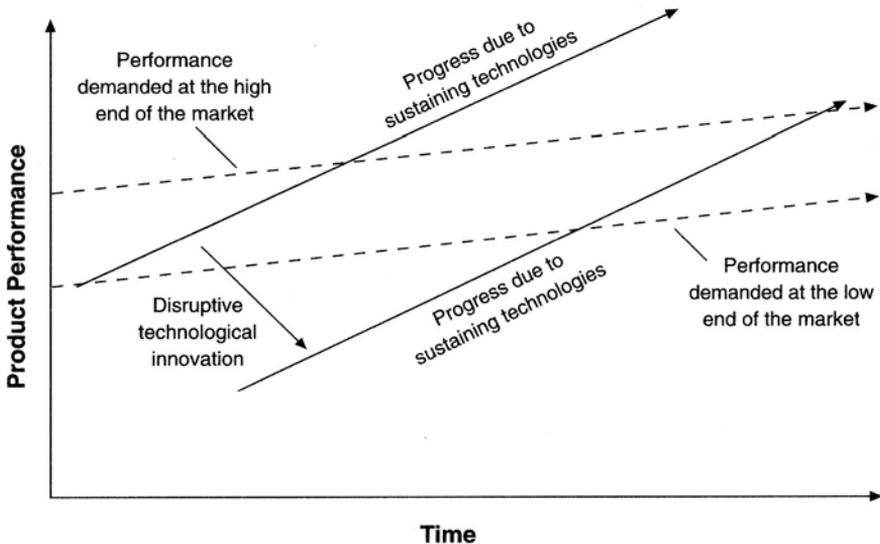


Fig. 3 : Le changement technologique déterminé par les innovations de soutien et de rupture. Source : [CRa].

<sup>16</sup> Les deux paragraphes suivants sont proches de [CRa].

| <b>Technologies affirmées</b>             | <b>Technologie de rupture</b>             |
|---|---|
| Mini-ordinateurs                          | PC  |
| Ordinateurs portables                     | “appareils numériques tenus dans la main” |
| Film photographique à halogénure d'argent | Photographie numérique                    |
| Téléphonie par câble                      | Téléphonie mobile                         |
| Microsoft OS et Office                    | Linux et Open Office                      |
| Commandes de disque dur                   | Bâtons de mémoire                         |

Tableau 8 : Exemples d'innovations de rupture, et technologies correspondantes réalisées. – Source : [Chr], [CRa].

Dans le cas de certains mécanismes plausibles et quantitativement vérifiables, ceci mène à une situation où de nouveaux acteurs entrent en scène et créent une petite niche, qui présente peu d'intérêt pour le leader du marché, jusqu'à ce qu'elle ne grandisse et prenne des dimensions importantes.

Il est intéressant de voir que le leader du marché, suite aux indications de bon management, décide de ne pas poursuivre la nouvelle option.

L'histoire en témoigne, avec les innovations de rupture, le leader du marché est généralement incapable de conserver sa propre position et il est remplacé par un autre acteur du marché. Une demande intéressante se pose pour les acteurs du marché affirmés et nouveaux, mais également pour les politiques : la traduction automatique et la traduction du parlé sont-elles des innovations de soutien ou de rupture ?

Dans le cas de la traduction du langage parlé, l'innovation est de rupture : il n'est pas facile de voir comment cette technologie peut être de support aux interprètes humains et, en outre, si on la compare à la traduction humaine, ses prestations, bien qu'elles soient en voie d'amélioration, demeurent toute-

fois très inférieures. L'on s'attend donc à ce que cette technologie prenne pied sur des marchés très différents des marchés existants et qu'à l'avenir elle étende sa quote de marché. D'autre part, ces marchés cibles ne sont pas couverts par des traducteurs humains (nous pensons à la traduction en direct des émissions télévisées).

La traduction automatique peut être considérée aussi bien comme une innovation de soutien que de rupture, selon le type d'application concerné et le marché sur lequel elle est introduite. Les techniques de mémoire de traduction, par exemple, sont en train de devenir un instrument de préparation et de soutien à la traduction humaine, en facilitant la traduction et en améliorant la qualité. Dans ce contexte, la traduction automatique peut être considérée comme une innovation de soutien : elle améliore ce qui existe déjà.

Tout à fait à l'opposé, la plupart des autres applications de la traduction automatique peuvent être considérées de rupture. Tout comme la traduction du langage parlé, les services de traduction on line des sites web, les chat on line et les self services pour la

recherche de documents en différentes langues donnent des prestations inférieures à la traduction humaine en termes de qualité. L'avantage en termes de coût consent, cependant, à la croissance de ces marchés, qui présentent peu d'attrait pour les services de

traduction humaine à cause des faibles marges de gain.

**TA et TLP sont essentiellement des innovations de rupture, tels que le PC ou la photographie numérique : elles commencent par une niche et se renforcent par la suite.**

### 3. Asie, Europe et États-Unis : analogies et différences

**Le 11 septembre et l'élargissement de l'UE sont les deux événements qui ont guidé la technologie de la traduction ces dernières années.**

#### 3.1. L'Union Européenne

La situation concernant les langues et le besoin de technologies ont été largement illustrés au paragraphe 2.3. : *Un regard de près sur l'Union Européenne*. Si l'on récapitule les principaux points :

- L'Europe a une société multilingue et elle est multilingue par définition.
- Toutes les langues officielles européennes ont les mêmes droits.
- La traduction entre ces vingt langues officielles forme 190 couples linguistiques et requiert un effort considérable.
- Toute mesure apte à faciliter la traduction alimenterait la communication entre les citoyens européens et ferait croître le commerce entre les pays de l'UE.
- Toute société traitant avec le marché de l'UE doit localiser ses pro-

pres produits par rapport aux marchés régionaux en les présentant dans la langue locale.

La Commission européenne est engagée dans la recherche sur les TLH et elle a financé certains projets de recherche sur la traduction automatique. L'on peut accéder rapidement à ces informations sur le site Web de l'IST (Information Society Technologies) en passant par la section *project search*<sup>17</sup>. La liste, qui n'est pas exhaustive, comprend LC-STAR<sup>18</sup>, MATCHPAD<sup>19</sup>, METIS et METIS-II<sup>20</sup>, NESPOLE!<sup>21</sup>, TC-STAR\_p et TC-STAR<sup>22</sup>, TQPRO<sup>23</sup> et TransType2<sup>24</sup>.

La Commission européenne a joué un rôle décisif dans la production des ressources linguistiques nécessaires, en finançant de nombreux projets dans ce domaine. Au sein du VI Programme Cadre, l'Union européenne a dépensé 135 millions d'euros sur les interfaces multimodales et sur les technologies du langage, soit environ 15 millions d'euros par an sur les technologies du langage.

La communauté des chercheurs européens dans le domaine du langage a toutes les caractéristiques nécessaires

<sup>17</sup> IST project search : <http://www.cordis.lu/ist/projects/projects.htm>

<sup>18</sup> LC-STAR : <http://www.lc-star.com/>

<sup>19</sup> MATCHPAD : <http://www.systransoft.com/R&D/Matchpad/index.html>

<sup>20</sup> METIS-II : <http://www.ilsp.gr/metis/>

<sup>21</sup> NESPOLE! : <http://nespole.itc.it>

<sup>22</sup> TC-STAR : <http://www.tc-star.org/>

<sup>23</sup> via la page IST : <http://www.cordis.lu/ist/projects/projects.htm>

<sup>24</sup> TransType2 : <http://tt2.atosorigin.es/>

pour pouvoir affronter une tâche aussi complexe. L'Europe a de nombreux instituts publics de recherche de renommée internationale, qui lui permettent d'affronter les différents secteurs de recherche et d'entreprendre une variété d'approches. Les concepts de multilinguisme et de multiculturalisme sont généralement bien compris et partagés, que ce soit au niveau académique ou économique et ils sont fortement soutenus au niveau politique. En outre, il existe depuis une dizaine d'années une infrastructure pour la préparation et la distribution des ressources linguistiques, l'Association européenne pour les ressources linguistiques (ELRA).

## 3.2. Les États-Unis

### 3.2.1. Le rôle stratégique des TLH

Aux États-Unis le rôle stratégique des TLH est profondément différent du rôle européen. Les États-Unis ont un vaste marché intérieur avec, en substance, une seule langue de base. Ceci, malgré la présence d'une forte population hispanique parlant l'hispano-américain. Il n'y a cependant pas de forte nécessité économique de servir ce groupe économiquement faible, et il n'existe pas non plus de conditions juridiques pour supporter cette langue. Les États-Unis sont essentiellement un marché unique parlant une seule langue.

La suprématie internationale de l'anglais comme langue étrangère est, sous différents aspects, un avantage pour les États-Unis, mais c'est aussi un grave inconvénient quant à la sûreté nationale. Ceci attribue à la traduction, en

particulier des autres langues vers l'anglais, un rôle crucial pour la récolte de renseignements venant du monde et concernant le monde, que ce soit d'intérêt général ou relatif aux opérations de surveillance.

**L'intérêt des USA dans les TLH est lié en grande partie aux aspects de la sûreté nationale.**

### 3.2.2. Les programmes de recherche

Après les attaques du 11 septembre 2001 au World Trade Center, il est évident que la nécessité d'avoir à disposition des traducteurs compétents, quelle que soit la langue, peut augmenter brusquement et dans les plus brefs délais. C'est pour cette raison que les États-Unis ont lancé un programme de recherche sur la récolte de renseignements en différentes langues et de sources multiples, y compris la traduction automatique. Le financement s'élève à plus de 50 millions de dollars par an.

Ces financements – qui sont un multiple de la somme dépensée par la Commission Européenne – et le programme lui-même auront un impact sur la communauté des chercheurs. Alors que certains de ces progrès technologiques et scientifiques seront dans la ligne des besoins spécifiques de l'Europe, d'autres ne le seront pas – la récolte de renseignements à des fins militaires est, sous plusieurs aspects, différente de l'action de traduire pour une communauté multilingue. En outre, la structure essentiellement monolingue du marché intérieur des États-Unis fait naître de sérieux doutes sur le fait que le marché de la technologie du langage déve-

loppée aux USA puisse offrir ce dont l'Europe a besoin.

Sans entrer dans les détails du programme de recherche, l'on peut affirmer qu'en règle générale les États-Unis tendent à financer un petit nombre de grands projets, alors que l'UE finance de grands programmes comprenant de petits projets. L'actuel fleuron est

le projet GALE (Global Autonomous Language Exploitation), qui se focalise sur l'extraction de renseignements à partir d'un texte multilingue et de documents audio au sein d'un domaine illimité (par exemple ce projet affronte aussi la traduction du parlé). Il existe d'autres projets en cours qui s'occupent de traduction, tels que TransTac, STR-DUST et ACTD.

UNCLASSIFIED

| RDT&E BUDGET ITEM JUSTIFICATION SHEET (R-2 Exhibit) |         |         |         |   |         |         | DATE          |         |
|---|---------|---------|---------|---|---------|---------|---------------|---------|
|   |         |         |         |   |         |         | February 2005 |         |
| APPROPRIATION/BUDGET ACTIVITY                       |         |         |         | R.I ITEM NOMENCLATURE   |         |         |               |         |
| RDT&E, Defense-wide<br>BA2 Applied Research         |         |         |         | Information and Communications Technology<br>PE 0602303E, Project IT-04 |         |         |               |         |
| COST (In Millions)                                  | FY 2004 | FY 2005 | FY 2006 | FY 2007   | FY 2008 | FY 2009 | FY 2010       | FY 2011 |
| Language Translation IT-04                          | 0.000   | 57.389  | 65.744  | 69.687  | 75.221  | 75.593  | 65.593        | 60.593  |

(U) **Mission Description:**

(U) This project will develop and test powerful new technology for processing human languages that will provide critical capabilities for a wide range of national security needs. This technology will enable systems to (a) automatically exploit large volumes of speech and text in multiple languages; (b) revolutionize human-computer interaction via spoken and written English and foreign languages; (c) perform computing and decision-making tasks in stressful, time-sensitive situations; and (d) autonomously collate, filter, synthesize and present relevant information in timely and relevant forms. This program element and project were created in accordance with congressional intent in the FY 2005 DoD appropriations bill. Prior year funding was budgeted in PE 0602301E, Project ST-29, and is noted as a memo entry in each program below.

(U) **Program Accomplishments/Planned Programs:**

|  |          |         |         |         |
|--|----------|---------|---------|---------|
|  | FY 2004  | FY 2005 | FY 2006 | FY 2007 |
| Situation Presentation and Interaction | (10.870) | 11.500  | 11.616  | 14.387  |

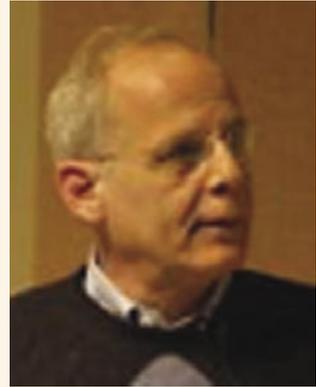
(U) There are two programs involving direct speech-to-speech translation:

- The Compact Aids for Speech Translation (CAST) program is providing the tactical warfighter with real-time, face-to-face speech translation during combat and humanitarian operations in foreign territories. The program addresses domain-specific translation accuracy and response time. Early CAST prototypes relied on simple dictionaries and phrases. The CAST program resulted primarily in quickly making one-way translation systems (from English to multiple foreign languages) available to warfighters in the field. The DARPA Phrasclator is the key prototype system in use today. The system was deployed in Operation Iraqi Freedom and Operation Enduring Freedom. Future versions will offer a more sophisticated, flexible and fluid translation and paraphrasing capability that is robust and conducive to normal human conversations.

Fig. 4 : Budget du DoD (Département de la Défense) des États-Unis pour les technologies du langage concernant la traduction. Budget prévisionnel 2005 pour la recherche, le développement, le test et l'évaluation (RDT&E), défense globale, chez DARPA (Defense Advanced Research Projects Agency). – Source : DARPA.

## Interview de Joseph Olive, Program Manager, DARPA

Joseph Olive est le program manager de l'Information Processing Technology Office chez DARPA. Son actuel portfolio comprend un vaste programme, GALE (Global Autonomous Language Exploitation). Il a plus de trente ans d'expérience dans la recherche et le développement chez Bell-Labs. Il a été le leader mondial dans la recherche de la synthèse de texte à langage parlé et a dirigé un groupe international dans les systèmes de dialogue sur ordinateur et de communication homme machine. Dans son rôle de Directeur de la recherche sur le langage parlé et de CTO chez Lucent Speech Solutions, il a supervisé la production des technologies du langage parlé des Bell-Labs: l'identification, la synthèse et la vérification du locuteur.



**Joseph Olive**

Program Manager  
IPTO (Information Processing Technology Office)  
DARPA  
Arlington, VA, USA.

Joseph Olive s'est diplômé à l'Université de Chicago avec une maîtrise en Physique (physique atomique computationnelle) et un M.A. en composition musicale. Après l'université, il a combiné son intérêt pour la computation et pour la musique et il a entrepris la recherche en acoustique et dans l'élaboration du signal.

Joseph Olive a remporté le prix de la Fondazione Nazionale per le Arti en 1974 pour avoir écrit une œuvre sur l'ordinateur. En 1984 il a également reçu une reconnaissance des Bell-Labs en qualité de Distinguished Member du Technical Staff.

### *Quelles sont les activités de recherche et les objectifs du programme GALE?*

Global Autonomous Language Exploitation (GALE) est un programme qui fournira des renseignements concis qui pourront être utilisés par nos militaires. Du moment que la source peut être dans une langue étrangère, GALE requiert une traduction et une synthèse. L'input peut être texte ou langage parlé et l'output peut être une traduction complète ou une réponse filtrée à une question (pas nécessairement en langage naturel).

Le but de GALE est d'atteindre un soin dans la traduction et la sélection tel que le système pourra être utilisé par les militaires. GALE est la suite de deux programmes précédents de DARPA en TLH: EARS – un programme de transcription et TIDES – un programme pour la traduction, la recherche, l'extraction et la synthèse de l'information. Les précédents programmes n'avaient pas été conçus pour devenir un système unique de traduction, mais ils ont permis d'obtenir de grands progrès dans les TLH.

### ***Quelle est votre position quant à la coopération transatlantique?***

Quelques groupes européens sont impliqués dans GALE. Ce sont des sous-contractants des grands teams créés par les “Principle Investigators” (PI) de GALE et ils collaborent au sein de leurs teams. En règle générale, je pense que la collaboration entre les chercheurs des États-Unis et les chercheurs européens/asiatiques est bonne. Ces relations sont à différents niveaux, au sein du gouvernement, des organismes et de la recherche universitaire.

### ***Quels sont les financements publics aux États-Unis pour les technologies du langage telle que la traduction automatique?***

Le programme total des TLH chez DARPA s'élève à environ 50 millions de US\$ pour 2006. Je ne pense pas qu'il sera réduit, car les TLH sont (ou devraient être) reliées et intégrées pour atteindre le succès dans ce domaine. Je n'ai pas de renseignements sur les dépenses des autres organismes gouvernementaux concernant les TLH, mais je suis sûr que DARPA n'est pas la seule organisation s'intéressant à ce secteur.

### ***Pour les USA quelle est l'importance des TLH, in particulier de la TA et des technologies qui lui sont reliées?***

Personnellement je vois que les TLH sont extrêmement importantes pour le gouvernement des USA et pour le secteur militaire. Il est nécessaire de comprendre et de communiquer avec nos alliés et nos ennemis, et nombre d'entre eux ne parlent pas anglais. À cause de l'énorme quantité de renseignements, il est crucial de chercher le renseignement important comme l'on cherche une “aiguille dans une botte de foin”.

### ***Aux États-Unis, comment le financement public pour la recherche dans les TLH est-il organisé?***

Je n'ai pas vu de plan général pour la recherche dans les TLH aux États-Unis. Il existe un investissement de la part de plusieurs agences nationales et de la part de l'industrie. Il y a une grande collaboration entre la recherche parrainée par le gouvernement, les organismes privés et la recherche universitaire, mais elle est plus informelle. En ce qui concerne DARPA, nous avons une vision et des objectifs très ambitieux pour les TLH. Nous avons des évaluations périodiques pour garantir les progrès. Nous travaillons aussi bien à travers la coopération qu'à travers la compétition. Nos équipes ont pour but de vaincre les évaluations. En outre, nous avons lié nos objectifs et les évaluations à l'utilité de la technologie plutôt qu'à la seule recherche de la perfection. Les TLH ont à faire aux langues et le premier et principal objectif est de préserver le sens de la langue dans les documents (dans leur forme écrite et parlée).

***Quel est l'état de l'art et quels sont les prochains défis scientifiques qui seront traités dans les technologies du langage, dans le domaine de la traduction en particulier?***

La technologie de la traduction automatique s'est beaucoup améliorée ces deux ou trois dernières années. Une grande partie de cette amélioration est due aux paradigmes statistiques associés à des procédures d'optimisation utilisant BLEU. Je crains cependant que ce paradigme n'atteigne rapidement ses limites, s'il ne l'a déjà fait. Je pense que la traduction automatique n'a fait que la moitié du chemin qu'elle doit faire. Je voudrais voir de nouvelles approches au problème, même si cela signifie avoir des résultats non compétitifs au début. Que ce soit pour les technologies de TA ou de ASR, il est nécessaire d'adopter une approche incorporant d'autres techniques de TLN, telles que la RI, les analyses linguistiques, l'extraction des informations etc. Bien qu'il s'agisse aussi de techniques statistiques, elles ne travaillent pas de la même façon, et par conséquent un système combiné utilisant toutes ces technologies pourrait améliorer les résultats et offrir une solution au problème. Il est extrêmement important d'intégrer plus étroitement l'ASR et la TA, en n'utilisant pas seulement le résultat 1-best de l'ASR pour guider la composante de TA.

### 3.3. L'Asie orientale

#### 3.3.1. L'anglais comme langue franche en Asie orientale

Si l'on considère la variété de ses peuples, de ses cultures et de ses langues, l'Asie ressemble plus à l'Europe qu'aux États-Unis. Mais alors que l'Europe a commencé à se considérer comme une entité unique, il n'existe encore, en Asie, ni une politique cohérente ni une identité d'unité autonome. Entre eux les pays utilisent essentiellement l'anglais, et l'ample diffusion en Asie de l'emploi de l'anglais comme *langue franche* implique une demande considérablement inférieure de traduction directe entre les couples linguistiques tels que le thaïlandais-japonais. L'avantage est que chaque pays doit garantir en premier lieu la traduction de sa propre langue de et vers l'anglais, et la restriction à un unique couple de langues réduit la complexité.

#### 3.3.2. Les couples linguistiques asiatiques et l'importance croissante du chinois.

Il existe aussi un désavantage lié à la dépendance de l'Asie à l'anglais comme *langue franche*. Considérons, par exemple, la traduction entre le japonais et le chinois : ces langues ayant une racine commune, l'étymologie des paroles est la même. Il serait naturel de tirer un avantage de ce fait, ainsi que des ressemblances culturelles. L'emploi de l'anglais comme intermédiaire est indirect et mène à des complications inutiles. Selon une nouvelle tendance, les sociétés japo-

naises s'intéressent actuellement de plus en plus à la traduction directe entre le japonais et le chinois. Ceci est lié à la florissante économie chinoise et au fait que de nombreuses sociétés japonaises déplacent leur production manufacturière en Chine. L'intérêt pour la traduction du chinois et vers le chinois est croissant au niveau global ainsi qu'au Japon, et les relations entre la Chine et le Japon sont plus étroites que par le passé.

#### L'anglais est une langue franche en Asie. Le chinois acquiert de l'importance.

Économiquement fort et avec une langue ayant ses racines dans le chinois ancien, la Corée est un pays ayant des liens étroits aussi bien avec la Chine qu'avec le Japon. Les paroles japonaises comme les paroles coréennes ont souvent une origine chinoise, c'est-à-dire qu'elles partagent une même étymologie, même si le coréen n'utilise pas de caractères chinois. Au début, les termes technologiques sont souvent utilisés en japonais, en employant des caractères chinois et, ensuite, ils sont absorbés par le chinois. Cette ressemblance lexicale aide beaucoup la traduction. Ces trois langues ont cependant des prononciations différentes et elles sont assez différentes d'un point de vue linguistique. Si l'on considère ce bagage culturel commun et le fait que le commerce et l'interaction entre ces trois pays sont plus forts qu'avec les autres pays asiatiques, l'on peut penser que pour ces trois langues, chinois, japonais et coréen, des traducteurs directs seront probablement développés dans un proche avenir, tandis que pour

les autres langues asiatiques les traducteurs utiliseront l'anglais comme langue intermédiaire.

**Pour le chinois, le japonais et le coréen des traducteurs directs seront développés dans un proche avenir.**

### **3.3.3. Les programmes de recherche**

Au Japon, la recherche pour les technologies du langage humain avait atteint son apogée dans les années 80 grâce à des financements publics élevés et à l'intérêt des sociétés envers les projets de traduction automatique, tel que le projet *EDR, Electronic Dictionary Project*. Mais, vu que les attentes pour la traduction automatique ne furent pas satisfaites, comme pour le projet européen *EUROTRA*, l'intérêt et les financements publics commencèrent à se réduire. Aujourd'hui, l'on remarque une inversion de tendance.

Étant donné le rôle prédominant de l'anglais, l'attention de la recher-

che pour la traduction concerne la traduction de l'anglais vers la langue maternelle et vice-versa et, en substance, ceci mène à des programmes de recherche sur le plan national. Il serait difficile d'exécuter des projets asiatiques transnationaux, du moment qu'il n'existe pas d'organisations asiatiques appropriées. De toute façon, certains aspects du travail de recherche requièrent une coopération internationale qui, heureusement, est fournie et appuyée par les associations asiatiques, telle que *l'Asian Federation of Natural Language Processing*<sup>25</sup> qui a été récemment constituée.

Afin d'alimenter la recherche entre les langues asiatiques, il serait important de construire des ressources linguistiques pour ces couples linguistiques. Vu l'effort requis et la nécessité de partager les ressources, la coopération internationale est la façon la plus naturelle d'atteindre ces objectifs. La constitution d'une agence d'évaluation serait un autre pas important.

<sup>25</sup> <http://afnlp.org>

## Interview de Jun-ichi Tsujii, Directeur du National Centre for Text Mining à Manchester et Professeur à Manchester et à Tokyo



**Jun-ichi Tsujii**

Directeur, National Centre for Text Mining, Manchester, UK  
Professeur, School of Informatics, University of Manchester, UK  
Professeur, Department of Computer Science, University of Tokyo, Japon.

Le Professeur Jun-ichi Tsujii est, depuis juillet 2005, directeur du *National Centre for Text Mining* et Professeur en Text Mining à la *School of Informatics*, Université de Manchester, Royaume Uni. Il est également Professeur en Natural Language Processing au *Department of Computer Science*, Université di Tokyo, Japon. Dès 1976 il a travaillé dans le domaine du TLN. Il a débuté avec la traduction automatique, il a étendu sa recherche aux formalismes grammaticaux pour les applications pratiques TLN, analyse syntaxique HPSG, extraction de l'information et système de réponses-questions intelligentes. Son équipe de recherche sur TLN à l'université de Tokyo a récemment réussi à appliquer avec succès un analyseur linguistique précis pour produire des représentations sémantiques de tous les abstracts de Medline (1,4 milliard de mots).

Il est reconnu comme l'une des figures de pointe dans le domaine du bio-text mining, de la traduction automatique et du TLN multilingue, et il a été un pionnier passionné de la coopération dans toute l'Asie. Il a été invité pour des *tutorial*, *talk e keynote speeches* lors de nombreuses et importantes conférences, en bio-informatique et TLN. Depuis 1992 il est membre permanent de l'ICCL (International Committee of Computational Linguistics), vice-président (2005) et président (2006) de l'ACL (Association of Computational Linguistics), président (2003-2005) de l' IAMT (International Association of Machine Translation), et vice-président de l' AFNLP (Asian Federation for Natural Language Processing).

### **Dans les années 80, la traduction automatique a été le thème de grands programmes de recherche, mais les résultats n'ont pas été à la hauteur de l'attente. Comment voyez-vous ces efforts aujourd'hui?**

Il est vrai que les efforts des années 80 n'ont pas satisfait les attentes d'un vaste marché potentiel. Pour une personne qui, comme moi, a participé à certains de ces projets, je pourrais dire que ce fut une déception. Cependant, grâce à ces efforts, une solide base pour la recherche et le développement de systèmes TA a été établie. Au Japon, au moins six ou sept producteurs de TA sont encore actifs sur le marché. Je pense, en outre, que la vision que nous avions était alors en avance sur les temps, car les technologies nécessaires pour atteindre nos objectifs n'étaient pas disponibles. Je crois qu'actuellement, vu qu'aujourd'hui nous avons les technologies qu'il nous manquait à l'époque, nous sommes en mesure de revitaliser ce secteur. Par exemple, l'action même d'effectuer l'analyse syntaxique d'une phrase à

travers des formalismes grammaticaux sophistiqués, qui exigeaient des heures de calcul, requiert à présent moins d'une seconde avec notre programme à l'Université de Tokyo.

***Les technologies du langage, y compris la traduction automatique, devraient être importantes pour l'Asie. Le perçoit-on au niveau national?***

Malheureusement, non. Contrairement à ce qui se passe en Europe, l'anglais a longtemps été considéré comme la seule langue internationale pour la communication et, à cause de cela, les personnes ne pensent pas qu'il soit important de prendre en considération les « langues locales » telles que le chinois, le coréen, le japonais etc. C'est une erreur, naturellement, et les personnes commencent à le comprendre. Avec Internet, nous nous sommes aperçus qu'il y a une énorme demande d'élaboration des langues "locales", ou pour mieux dire, que nos langues locales ne sont pas vraiment locales.

***Quelles sont les majeures différences entre les langues asiatiques et les langues européennes?***

Tout dépend de la définition de langues européennes et asiatiques. Les Européens peuvent déclarer qu'en Europe les langues sont différentes, et je suis d'accord. Je pense, toutefois, que la différence entre les langues asiatiques est énorme et qu'elle va bien au-delà de la différence existant entre les langues européennes. De nombreuses langues en Inde, au Moyen-Orient, en Malaisie etc. n'ont aucun rapport avec les langues d'Extrême-Orient, le chinois, le coréen et le japonais. En outre, bien qu'ils partagent certains vocables grâce à une longue histoire d'échanges culturels, le chinois et le japonais/coréen appartiennent à des familles linguistiques complètement différentes.

***Comment pourriez-vous décrire la situation dans les pays asiatiques en général?***

Je ne suis pas la bonne personne pour répondre à cette question, mais la situation dans les pays asiatiques est assez diversifiée. Certaines langues n'ont même pas codifié un système commun de transcription, et encore moins des codes standard pour représenter les caractères. Mais du moment que, dans de nombreux pays tels que l'Inde, la Thaïlande, la Chine etc., les progrès technologiques ont été accélérés, les niveaux technologiques de ces pays sont plus ou moins semblables. Nous remarquons que nombre d'articles de recherche intéressants sont publiés par les chercheurs dans ces pays.

***Existe-t-il en Asie des activités internationales coordonnant la recherche dans ce secteur, ou même une collaboration au niveau global?***

Il y a deux ans nous avons fondé une association académique, l'AFNLP (Asian Federation for Natural Language Processing) afin de promouvoir la coopération et de coordonner les activités. Il y a d'autres activités naturellement. Mais nous sommes encore bien loin du niveau européen de coordination.

### ***Quels sont les prochains pas qu'il faudra faire pour promouvoir les TLH et la TA en Asie?***

Nous avons besoin d'une majeure implication gouvernementale. L'UE a joué un rôle significatif dans la promotion de la recherche et du développement des TLH et de la TA. Jusqu'à la fin des années 80, le Japon était à l'avant-garde non seulement de la technologie mais des financements à la coopération régionale dans ce secteur. Depuis, la situation a changé de façon extraordinaire. Nous voyons que certains pays peuvent contribuer financièrement et intellectuellement à ce secteur. Nous n'avons cependant pas d'organisations gouvernementales panasiatiques pour coordonner les activités. La communauté académique est prête à coopérer, mais nous avons besoin, par exemple, d'organismes de financement qui appuient les grands projets en Asie.

### 3.4. L'Inde

Est-il possible d'imaginer un pays avec autant de diversité linguistique que l'Europe unie? Bon, l'Inde<sup>26</sup> et ses 35 états a 22 langues officielles, c'est-à-dire approuvées par la constitution. Et comme si cette variété n'était pas suffisante, chaque langue a une vingtaine de dialectes environ. Ces langues appartiennent à trois différentes familles linguistiques, et il existe également différents systèmes d'écriture. L'hindi est la langue officielle, mais la constitution indienne affirme que l'anglais peut, lui aussi, être utilisé à des fins officielles. Dans les villes, les personnes connaissent souvent trois langues : la langue nationale (par exemple hindi, bengali, etc.), l'hindi (la langue officielle de l'Inde), et l'anglais. À peu près 30% de la population parle hindi, et environ 5% de la population est à son aise avec l'anglais.

Dans les capitales des états, tous les documents doivent être rédigés en trois langues (anglais, hindi et la langue de l'état). Les directions de traduction le plus fréquemment employées sont de l'anglais à l'hindi et de l'anglais ou hindi dans les respectives langues de l'état. Comme dans d'autres pays cependant, la plupart des traductions est effectuée par des personnes et il n'y a pas assez de traducteurs pour répondre à la demande. Pour ce qui est des instruments techniques, la mémoire de traduction

n'est pas très utilisée, mais la traduction est souvent supportée par des dictionnaires électroniques et par des instruments tels que les analyseurs morphologiques.

La recherche sur la traduction est financée par le gouvernement et par des sociétés internationales telles que IBM, Microsoft, Google et Yahoo, bien qu'elles ne traitent pas encore toutes les langues indiennes. Du moment que ces sociétés utilisent généralement une approche statistique, il existe de nombreux corpora parallèles<sup>27</sup> mais ils ne sont pas accessibles au public. À partir de 2006, ce fossé sera comblé par l'agence LDC-IL<sup>28</sup> (*Linguistic Data Consortium for Indian Languages*), parrainée par le nouveau gouvernement.

### 3.5. Les conditions économiques contextuelles

En Europe, les conditions économiques contextuelles relatives aux technologies du langage sont différentes de celles des États-Unis. En fait, aux États-Unis et dans l'UE, les conditions du marché pour la reconnaissance automatique du langage parlé et pour la traduction automatique sont diamétralement opposées.

Donnons d'abord un coup d'œil à la reconnaissance automatique de la voix. Actuellement d'un point de vue

<sup>26</sup> Nous remercions le Professeur Dr. Pushpak Bhattacharyya (Indian Institute of Technology, Mumbai) pour le bref sur la situation en Inde.

<sup>27</sup> Un *corpus paralelo* est une collection de textes en deux versions de langues, avec l'information que les phrases sont associées l'une à l'autre. (Par exemple, les deux premières phrases de la langue A peuvent correspondre aux trois premières phrases de la langue B, etc.)

<sup>28</sup> <http://www.ciilcorpora.net/ldcil.htm>

économique, l'emploi le plus intéressant de la reconnaissance automatique et des technologies de dialogue est relatif au secteur des systèmes automatiques pour les clients ou, autrement dit, à l'automation des services de *call center*. En ce qui concerne ce marché et cette application, l'Europe diffère des États-Unis pour deux aspects. D'abord, aux USA les systèmes appelés IVR<sup>29</sup> ont eu une large diffusion, alors que ce n'est pas le cas pour l'Europe. Le public américain a utilisé pendant une longue période les systèmes IVR, souvent ennuyeux, et il a généralement été très content de passer des tons DTMF (Dual Tone Multi-Frequency) et des menus rigides aux dialogues parlés.<sup>30</sup> En Europe, où dans certains pays les services de *call center* ont été mis à disposition gratuitement et ont fourni un service plutôt de bon niveau, l'introduction des systèmes automatiques a souvent été perçue comme un pas en arrière par les usagers. Les États-Unis représentent, en outre, un vaste marché assez homogène ayant une seule langue, de dimensions beaucoup plus vastes que n'importe quel marché européen. Par conséquent, il existe une meilleure économie d'échelle aux USA, vu que l'essor d'un dialogue parlé requiert un effort significatif. Pour cette raison, il

ne faut pas s'étonner si l'on découvre que la reconnaissance du langage parlé a été adoptée de façon plus incisive aux USA que dans l'UE.

Les conditions économiques pour la traduction automatique sont exactement à l'opposé : difficiles pour les USA, mais très favorables pour l'Europe que ce soit pour la traduction du langage parlé ou pour la traduction automatique. Ceci est dû au fait qu'il existe de nombreuses langues parlées en Europe, alors que n'importe quel système utilisé pour le vaste marché américain ne couvrirait qu'une petite fraction de toutes les activités. En Europe, faire des affaires signifie être multilingue.

### **Conditions économiques favorables pour les sociétés européennes en tant que fournisseurs de services de TA.**

Il existe également d'importantes forces économiques qui soutiennent l'industrie de la traduction en Asie, mais à cause du rôle dominant de l'anglais, le marché et les perspectives pour une industrie des technologies du langage sont inférieures à celles européennes. Vu la situation actuelle et la situation des sociétés, l'on peut supposer qu'en circonstances normales, l'Europe assumera un rôle phare.

<sup>29</sup> IVR signifie *interactive voice response*. Les systèmes classiques IVR répondent à la pression des touches par des messages pré-enregistrés.

<sup>30</sup> Un rapport cité fréquemment – Nuance Communications – Market Research : “Nuance Speech User Scorecard”, May 2000 – affirme que pour la reconnaissance du langage parlé la satisfaction générale était élevée (environ l'87% des répondants) et de façon significative plus élevée par rapport à celle mise en évidence par les systèmes DTMF. Ce résultat est raisonnable et cohérent par rapport à l'expérience de l'auteur, mais quelque précaution est nécessaire vu que le rapport original n'est plus disponible sur le site de Nuance ou sur internet en cherchant sur Google.

### 3.6. Une stratégie pour l'Europe

Le rôle stratégique des TLH et la situation dans les trois grandes régions peuvent être résumés comme suit :

**Europe** : la technologie du langage est une nécessité économique, politique et culturelle. En franchissant la barrière du langage, l'économie et la communication seront encouragées. Les TLH sont déjà au centre d'un considérable effort européen de recherche, mais l'importance stratégique de cette technologie justifie, pour l'Europe, une très haute priorité dans l'agenda de la recherche.

**USA** : l'emploi des TLH est dominé par des considérations d'ordre militaire et par la lutte contre le terrorisme. Des fonds importants sont actuellement

alloués à la recherche et à la technologie relative aux TLH.

**Asie** : la traduction de et vers l'anglais a actuellement la priorité. Le manque d'identité politique et d'infrastructure commune rend difficile le fait que l'Asie puisse assumer un rôle phare. Il existe une considérable demande de traduction en Inde.

Pour ce qui est de l'essor et de l'exploitation commerciale, l'Europe se trouve simultanément dans une situation d'extrême besoin et dans une position privilégiée. Vu la nature générique de la technologie, elle offre la possibilité d'être commercialisée dans d'autres régions du monde. D'autre part, l'on ne peut s'attendre à ce que nos besoins soient satisfaits par autrui.

## Interview de Joseph Mariani, Directeur, Ministère Français de la Recherche

Joseph Mariani s'occupe de recherche dans les secteurs des technologies du langage, de la communication multimodale homme-machine, de la reconnaissance de la parole, des ressources linguistiques et de l'évaluation.

Il a été président de l'Association Européenne pour les Ressources Linguistiques (ELRA), président de l'Association Européenne (désormais Internationale) pour la Communication Parlée (ISCA), membre du conseil de l'European Network on Language & Speech (ELSNET), et coordinateur du réseau francophone FRANCIL network.

Joseph Mariani a été directeur du LIMSI et chef du département *Communication Homme-Machine* (1989-2001), membre du Conseil Scientifique du CNRS, chair *del'Information Science and Technology Advisory Committee* du CNRS et membre du Comité d'évaluation d'INRIA.

Depuis 2001 il est directeur du département ICT du *Ministère français de la Recherche*, où il est responsable des programmes de recherche dans les télécommunications, les technologies software, les nanotechnologies et les multimedia, y compris les technologies du langage.



**Joseph Mariani**  
Directeur, Directeur, Ministère  
Français de la Recherche,  
Paris, France

### **Par rapport à l'Europe, aux États-Unis et à l'Asie: où sont les ressemblances, et où les différences?**

Avec leur vaste marché intérieur les États-Unis voient le multilinguisme surtout d'un point de vue militaire et de la sûreté : tout le monde comprend l'anglais mais ils ont du mal à comprendre les langues étrangères! L'Asie, tout comme l'Europe, utilise de nombreuses langues, il existe donc un besoin de nature commerciale pour le multilinguisme, même s'il ne s'agit pas d'un marché commun telle que l'UE. Avec plus de 20 langues parlées dans les 25 États membres de l'UE, le multilinguisme est d'une extrême importance pour l'Europe. Outre la dimension économique, qui fait du multilinguisme une véritable nécessité, il existe aussi des dimensions politiques, culturelles et sociales. Ceci rend notre situation très spéciale.

### **Quelle est votre inclination envers la coopération avec les États-Unis ou l'Asie?**

La coopération internationale et une saine compétition font du bien au progrès technologique et scientifique, et j'apprécie et donne mon soutien quand nos pays unissent leurs forces dans la science et les infrastructures pour affronter les défis dans les TLH. Cependant, la coopération internationale ne peut pas remplacer notre agenda et répondre à nos besoins spécifiques. En ce qui concerne le multilinguisme européen, l'UE doit assumer le rôle phare, car personne ne peut ou veut le faire à sa place.

### ***Quels développements sont-ils nécessaires pour améliorer la position de l'Europe?***

Malgré les efforts considérables du passé, je pense que le niveau du financement de la recherche et de la technologie n'est pas encore compatible avec la dimension du défi. Bien que le Commissaire chargé de l'instruction, la culture et le multilinguisme, Ján Figel, ait récemment souligné l'importance du multilinguisme pour l'Europe, les technologies du langage ne sont qu'une petite partie du programme préliminaire du VII programme-Cadre, citées seulement à la fin de la composante technologique "Simulation, visualisation, interaction et réalité mixte". Et les thèmes actuellement sélectionnés pour des actions possibles dans le cadre de l'article 169 sont "Recherche dans la Mer Baltique", "Assistance aux personnes âgées" ou "Météorologie", qui sont des thèmes très respectables, mais, selon moi, moins stratégiques pour l'Europe que le problème des langues. Les technologies du langage sont probablement le sujet qui s'adapte le mieux à l'idée d'une action européenne coordonnée, vu que l'effort pour couvrir les différentes technologies et les différentes langues européennes est trop onéreux pour être porté seulement par la Commission, et qu'il peut facilement être partagé avec les États membres de l'UE, avec leurs langues et leurs programmes. Alors que les États membres devraient en premier lieu s'occuper des spécificités de leur langue, telles que les ressources linguistiques (corpora de texte et parole, dictionnaires), les adaptations spécifiques pour cette même langue, la Commission européenne pourrait d'abord affronter les aspects généraux indépendamment d'une langue particulière : la coordination générale, le développement de la technologie de base, l'évaluation, les standards.

L'Europe doit encore mettre au point une infrastructure pour évaluer les technologies du langage comparable à celle existant aux États-Unis telle que NIST, et les agences de financement devraient prendre en considération le fait d'évaluer les prestations des technologies de base, quand elles sélectionnent les projets, afin d'éviter de financer le développement d'applications se basant sur une technologie non adéquate.

### ***Ceci compléterait l'Association européenne pour les ressources linguistiques, l'ELRA, qui vient de fêter son dixième anniversaire.***

Oui, en effet. La situation concernant les aspects des ressources linguistiques et l'évaluation des technologies du langage s'est de beaucoup améliorée grâce à l'existence d'entités permanentes telles que NIST et LDC aux États-Unis, ou ELRA en Europe, mais le défi pour l'Europe requiert bien davantage. Personnellement j'appuie la création d'une Agence pour les technologies du langage au niveau de la Commission européenne, une structure permanente qui coordonne les efforts de la Commission Européenne et des États membres vers les technologies du langage nécessaires pour une Europe multilingue, et qui place l'Europe à la frontière des TLH dans le monde, tirant un avantage de son défi linguistique.

## 4. Où nous en sommes aujourd'hui

### 4.1. Le travail de traduction aujourd'hui

La traduction et la localisation sont des fournitures de services même si elles sont supportées par la technologie.

Le travail de traduction est très diversifié. Il va de la traduction du diplôme d'un étudiant-hôte à la gestion de l'énorme site web de la société Fortune 500, de couples de langues similaires tels que le néerlandais et l'allemand à des langues totalement différentes tels que l'italien et le chinois, et il peut recouvrir des langages standard, de la prose ou des langages hautement spécialisés.

La traduction se déroule, d'une part, au sein des instituts ou des sociétés qui en ont besoin et, d'autre part, elle est déléguée à l'extérieur à des fournisseurs de services. Les 20 premiers fournisseurs de services ne couvrent que 16,3% du marché<sup>31</sup>, il résulte par conséquent extrêmement fragmenté. En comptant les sociétés ayant au moins cinq employés, il existe au monde 5.000 offres de services de traduction. De nombreux traducteurs exercent la profession libérale, et même les plus grands fournisseurs de services de traduction et les plus grands usagers des services de traduction utilisent des professionnels pour leurs travaux de traduction. Si nous prenons

comme exemple le DGT (Directorat Général de la Commission Européenne pour la Traduction), la proportion du travail exécuté par les professionnels est augmentée de 11.8% en 1992 à 23.0% en 2004.

### 4.2. Les technologies utilisées dans la traduction professionnelle

**Les principales technologies actuellement employées par les services de traduction sont la traduction automatique et les ressources terminologiques.**

La technique sophistiquée du travail du traducteur dépend beaucoup de la société pour laquelle il travaille et de la nature du matériel qu'il doit traduire. La situation est très différente si les traductions diffèrent et les textes sont courts, ou si l'on veut localiser de grands paquets software dans de nouvelles langues, à partir d'une version existant déjà. Dans ce cas, il est important de prendre en considération la traduction précédente. C'est pour cette raison que la traduction automatique est largement utilisée dans certains secteurs de l'industrie, et il existe des méthodes de contrôle automatique qui garantissent que les paroles soient traduites de façon cohérente. L'emploi d'une terminologie cohérente et de haute quali-

<sup>31</sup> Source : Common Sense Advisory [CSA]

té peut être supporté par des ressources terminologiques.

La traduction automatique n'est pas encore employée à grande échelle, mais elle est utilisée dans certaines applications. La traduction automatique entre en jeu lorsque l'on a besoin de traductions sommaires à exécuter très rapidement. Dans l'Union Européenne, par exemple, les traductions sommaires sont utilisées pour faciliter les communications internes. Mais quand le matériel est présenté à l'extérieur, la traduction est entièrement exécutée à la main ou le résultat de la traduction automatique est revu par un traducteur.

**Les technologies telles que la traduction automatique et les ressources terminologiques non seulement réduisent les coûts mais elles améliorent aussi la qualité et la cohérence de la traduction.**

Il est important de comprendre que la traduction automatique et les ressources terminologiques augmentent non seulement l'efficacité d'un traducteur mais aussi la qualité et la cohérence du travail. Nous avons vu que la traduction automatique est utilisée pour fournir une traduction sommaire à utiliser comme point de départ pour la traduction humaine, et ceci afin d'en améliorer l'efficacité. En tout cas, l'avantage principal de la traduction automatique réside dans la traduction rapide de textes qui doivent être compris pour des raisons de travail, et que les coûts de production, ou plus souvent les temps de traduction, empêchent l'utilisation d'autres approches. La quantité de ces textes *en ligne de la dernière minute* qui doivent être traduits dans les plus brefs délais et en quantités plutôt limitées (par exemple le *flash d'informations*) croît rapidement.

## Interview de Kevin Bolen, Chief Marketing Officer, Lionbridge

Kevin Bolen est responsable du marketing global chez Lionbridge et il définit les directions du marché, les solutions de management, le placement compétitif, et la promotion du label. Mr. Bolen s'est uni à Lionbridge en 2005 suite à l'achat de Bowne Global Solutions dont il avait été VP du Marketing dès 2002. Outre directeur du marketing global, Mr. Bolen a été responsable des ventes pour la côte est. Avant d'entrer chez Bowne, il a géré le marketing chez LexiQuest Inc., une société de développement software spécialisée dans la technologie du langage naturel. Avant de faire partie de LexiQuest, Mr. Bolen a passé six ans chez IBM Global Services où il a rempli plusieurs rôles: du conseil, à la croissance des entreprises, au marketing. Lors de sa dernière fonction, il était Senior Marketing Manager responsable du développement des solutions mondiales et de la planification stratégique dans les secteurs de la vente au détail, du packaging, du transport et de la production.



**Kevin Bolen**  
Chief Marketing Officer  
Lionbridge  
Waltham, MA, USA

Mr. Bolen a un Bachelor en Business Administration en International Business de la Pace University et un Master en Business Administration en Marketing and Management de la Stern School of Business de la New York University.

### *Pourriez-vous décrire en quelques mots le marché de la localisation?*

Le marché de la localisation est extrêmement fragmenté avec des milliers de petits concurrents travaillant au niveau régional, vertical ou fonctionnel. Certains offrent toute la gamme de support à la localisation, alors que d'autres offrent le simple service de traduction, souvent en tant que sous-contractants de fournisseurs plus grands. Quelques-unes parmi les plus grandes sociétés offrent aussi des services complémentaires, tels que authoring, création design, et interprétations ou software de support pour gérer le processus de localisation et les patrimoines.

### *Quels sont les principaux facteurs entraînant le marché, même pour ce qui concerne sa croissance?*

Les sociétés continuent à s'étendre globalement pour atteindre de nouveaux consommateurs ou pour réduire leurs frais opérationnels en déplaçant le travail dans des régions à bas prix. Entrer sur ces nouveaux marchés requiert la localisation du produit, le marketing, le packaging, la documentation de support selon les lois locales, le matériel d'entraînement pour les ventes et pour les clients, et les communications aux employés. Les marchés de consommation telles que la Chine ou l'Inde poussent la localisation pour une vaste gamme de produits et de services dans des langues qui ne sont pas traditionnellement servies par des entreprises consoli-

dées, ils alimentent donc de nouvelles possibilités pour ces fournisseurs de services. D'autre part, les cycles de produits accélèrent et ceci signifie que plus les spécificités des produits augmentent, plus d'informations seront distribuées plus fréquemment. Ceci génère de nouvelles possibilités de localisation, mais requiert que les fournisseurs de services revoient leurs modèles de production comme par exemple le délai total de livraison du matériel traduit et le transfert des données, qui sont des facteurs cruciaux pour la concurrence.

### **Quelles sont les technologies les plus importantes dans la traduction des langues ?**

La *Traduction Automatique* (TA) demeure la technologie leader en termes d'efficacité productive. De toute façon, les modèles traditionnels de *licensing* et les applications au niveau du desktop sont en train de devenir obsolètes à l'époque des modèles *software on-demand, web-enabled*. À travers le web, les TA permettent à plusieurs traducteurs de travailler simultanément sur un projet unique, en exploitant le travail effectué par chacun à chaque instant afin d'améliorer la prestation globale.

Les dictionnaires terminologiques représentent eux aussi un élément clé, du moment qu'ils facilitent une plus grande cohésion au sein d'un grand nombre de contenus et de traducteurs. L'habileté à inclure le contrôle de la langue dans le processus de production garantit un travail soigné et la réduction du temps de production, un point crucial dans un monde *on-demand*.

### **Quelle est votre expérience avec la traduction automatique?**

Avec l'achat de Bowne Global Solutions, Lionbridge possède l'un des moteurs de traduction automatique basé sur les règles les plus utilisées par l'industrie, "Barcelona". Bien que les systèmes statistiques et basés sur les exemples se soient révélés prometteurs dans la recherche, l'ensemble des langues disponibles et des domaines demeure trop limité pour en faire un usage pratique commercial sur le marché actuel. Lionbridge a utilisé avec succès son moteur Barcelona sur une variété de projets en réduisant les temps et les coûts associés à l'élément traduction de tout un projet de localisation. Ces projets doivent être examinés attentivement du moment que le système requiert de personnaliser les dictionnaires et les règles avant d'obtenir une qualité suffisante pour permettre que le travail soit revu par un post-éditeur plutôt que par un traducteur. Il faut un volume suffisant de paroles pour assurer un ROI positif sur le coût initial de personnalisation.

### **La localisation de haute qualité et la fourniture de traduction automatique à bas prix visent plusieurs marchés. Entend-elle aussi couvrir les "technologies de rupture"?**

Lionbridge offre déjà un site web gratuit pour la traduction basée sur la traduction automatique et il a habilité ses propres clients à utiliser TA pour traduire des contenus qui, à cause des coûts associés à TH, n'auraient pas été traduits. À l'avenir, au fur et à mesure que TA deviendra une alternative plus accessible, les clients auront une variété de niveaux de solution parmi lesquels choisir, vu que la quali-

té, de toute façon, est très subjective. Nous sommes convaincus que les clients sont en train de chercher une solution plus intégrée qui fasse pression sur les MT web-based, TA, et TH/post-editing pour distribuer la TH de qualité plus rapidement et à des coûts inférieurs par rapport au modèle unique actuel de MT/TH. Ceci permettrait de s'étendre sur plusieurs marchés, en de plus brefs délais et avec un meilleur retour d'investissement.

### 4.3. La recherche sur les technologies à la base de la traduction parole-parole

Ces dernières décennies, le langage, que ce soit dans sa forme parlée ou écrite, a fait l'objet d'une recherche scientifique de la part de milliers de chercheurs dans le monde entier. Voyons les différentes technologies concernées :

**L'identification automatique de la voix**, c'est-à-dire que la création d'une séquence écrite de paroles à partir d'un flux de langage parlé s'est révélée être un problème scientifique très difficile, sûrement plus compliqué que jouer aux échecs au niveau des performances humaines. La communauté des chercheurs a d'abord remporté un succès en affrontant des problèmes plus limités, puis de plus en plus difficiles. Au début, la technologie abordait un vocabulaire restreint, des systèmes dépendant du locuteur en milieu tranquille et des paroles isolées, mais elle est désormais en mesure de gérer un vocabulaire plus étendu, et de reconnaître le parlé en continu et ceci indépendamment du locuteur. Alors que le public imagine que cette technologie en est encore à un bas niveau, en réalité la tâche de dicter un texte dans un contexte professionnel est souvent supportée par des identificateurs automatiques du langage parlé, même s'il existe encore des erreurs d'identification. La communauté des chercheurs s'est concentrée sur le défi suivant, à savoir le langage parlé provenant de personnes non coopé-

rantes; c'est-à-dire que le locuteur ne parle pas dans l'intention d'être reconnu par un système automatique. Les recherches-types concernent l'identification de conversations entre personnes parlant d'une façon naturelle et spontanée. Désormais, l'identification du langage parlé est souvent une des composantes d'une interface multimodale.

**Traduction texte-parole (TTP)**, appelé également synthèse texte-parole, génère le langage parlé à partir du texte écrit. Alors que les systèmes de TTP sont intelligibles depuis plusieurs années, la dernière amélioration significative, qui exploite la concaténation d'unités à longueur variable, a conduit à des systèmes produisant une voix synthétique plus naturelle et fluide qu'il y a quelques années. Les systèmes de synthèse jouent un rôle important dans la communication parlée homme-machine. Alors que des phrases pré-enregistrées – ou leurs concaténations – peuvent être essentiellement utilisées pour un contenu de type statique, le TTP devient une nécessité en cas de contextes variables. Les efforts actuels dans le développement de la synthèse texte-parole se focalisent sur des aspects telles que la génération rapide de voix différentes et la possibilité d'enrichir la voix synthétique d'émotions et d'intonations contrastantes<sup>32</sup>.

La **traduction automatique** traduit un texte écrit d'une langue source à une langue cible. Au cours de ces dernières décennies de nombreux efforts ont été investis dans ce difficile

<sup>32</sup> Comme dans les projets européens CHIL (Computers in the Human Interaction Loop, <http://chil.server.de/>) et AMI (Augmented Multi-party Interaction, <http://www.amiproject.org/>).

problème. De même que pour l'identification du langage parlé, où vers 1990 le paradigme méthodologique a changé et les méthodes basées sur les règles ont été remplacées par une approche statistique, il existe pour la traduction automatique deux approches fondamentalement différentes. Historiquement, la première approche se base essentiellement sur des méthodes linguistiques et sur la codification explicite de la connaissance humaine, alors que l'approche récente est plus guidée par les données et exploite des variantes de méthodes statistiques qui se sont révélées efficaces dans le domaine de l'identification du langage parlé. Pour chacune de ces approches, les systèmes de dernière génération produisent des traductions de texte – dans des domaines non circonscrits – qui demeurent de basse qualité si on les compare aux traductions humaines et qui peuvent être utilisées seulement pour des traductions sommaires fournissant le sens premier, mêlé à des erreurs.

Alors que les méthodes linguistiques tendent à bien traduire certaines phrases mais trébuchent misérablement sur d'autres, les méthodes statistiques distribuent les erreurs plus uniformément, en traduisant toute chose sans échec total mais avec peu de cas de traduction exemplaire. Ceci dit, l'on remarquera que le progrès dans ce secteur est rapide, et que des compétitions internationales et des mesures objectives sur des ensembles de données de test montrent des progrès notoires, en

particulier dans les systèmes qui utilisent des approches se basant sur des données.<sup>33</sup>

- La parole de la langue source est convertie en texte par un module d'identification automatique (ASR).
- Ensuite, l'on passe ce texte à un composant de traduction automatique (TLP/SLT) qui est spécifique car le texte contient des erreurs d'identification et exploite les caractéristiques de la parole par rapport au texte – phrases qui ne suivent pas la grammaire, hésitations, faux départs.
- Son output, un texte dans la langue cible, constitue l'input du module de synthèse (TTP/TTS) qui génère la parole dans la langue cible. Afin de préserver les caractéristiques de la voix originale, d'ultérieures informations sont transférées au module de synthèse (TTP/TTS).

Le **traduction du langage parlé** (TLP) traduit une langue de la forme parlée à la forme écrite dans une autre langue. Comme vous pouvez l'imaginer, les systèmes TLP mêlent essentiellement un identifieur vocal et un traducteur automatique, mais il y a plus : le texte contient des erreurs d'identification qui doivent être prises en considération. Une autre difficulté vient du fait que les contextes où travaillent les chercheurs prennent en considération le langage parlé qui n'est pas directement destiné à

<sup>33</sup> Cf. [Ney]

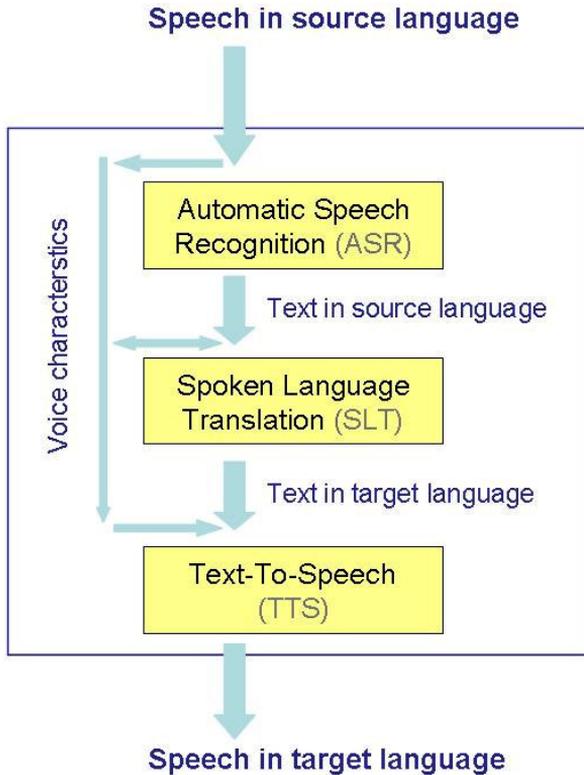


Fig. 5: Diagramme fonctionnel d'un système de traduction de parole à parole (TPP).

être reconnu par un système, c'est-à-dire le langage parlé que l'on entend à la télévision, durant un sommet ou lors d'une conversation de personne à personne.

La **traduction parole-parole** (TPP) mêle toutes les méthodes susmentionnées. Ici, le langage parlé de la langue source est traduit en langage parlé dans la langue cible, de façon à ce que les personnes puissent parler entre elles en différentes langues, en utilisant l'ordinateur comme interprète. Pour atteindre ce but, un système de TLP (traduction du langage parlé) est mis en cascade avec un système TTP (traduction texte-parole), et l'on adopte quelques mesures en surplus. En pre-

mier lieu, l'identité du locuteur devrait être préservée dans le langage parlé de la langue cible, avec les caractéristiques prosodiques qui apparaissent dans la langue source. Naturellement, ces caractéristiques ne sont visibles dans aucune forme écrite intermédiaire. Ce problème, de toute évidence très difficile, fait l'objet du projet européen TC-STAR, qui travaille sur la traduction de langage parlé à langage parlé lors des débats du parlement européen. Bien que la tâche soit ambitieuse, il est intéressant de voir que la qualité des traductions est déjà assez bonne pour comprendre le sens des discours parlementaires dans une langue qui n'est pas familière.

#### 4.4. Le projet TC-STAR

La traduction parole-parole est un ouvrage de recherche ambitieux. Pour simplifier cette tâche et la rendre plus faisable, les premiers projets de recherche qui ont affronté le sujet dans les années 90 (déjà avec la parole spontanée) ont travaillé sur des domaines limités comme la planification des rendez-vous<sup>34</sup>.

#### **TC-STAR est le premier projet conjoint de recherche affrontant la traduction de langage parlé à langage parlé dans un domaine illimité.**

Il est difficile de décrire la différence en termes de complexité nécessaire pour passer d'un domaine limité à un domaine illimité, mais s'il vous est déjà arrivé d'utiliser un guide de poche pour commander un déjeuner au Portugal, puis d'essayer d'écouter un débat parlementaire relatif à un quelconque argument, vous pourriez vous faire une idée. TC-STAR<sup>35</sup> (*Technologies et corpus pour la traduction parole-parole*) est le premier projet conjoint de recherche affrontant la traduction parole-parole dans un domaine illimité. À part la traduction d'émissions télévisées du chinois à l'anglais, qui a été choisie en référence à des indicateurs de mesure internationaux, TC-STAR travaille sur les discours du parlement européen en anglais et en espagnol, première langue

et non. Les thèmes comprennent tous les arguments dont s'occupe le Parlement européen.

Cet effort se focalise sur la recherche avancée dans toutes les technologies clés pour la traduction parole-parole (TPP) : l'identification du langage parlé, la traduction automatique et la synthèse texte-parole. Les objectifs de ce projet sont très ambitieux : atteindre un tournant dans la recherche sur la traduction parole-parole pour réduire de façon significative la différence de prestation entre les personnes et les machines. La première moitié des trois années de ce projet a donné des résultats très satisfaisants, mais il reste beaucoup de chemin à faire.

Outre le travail scientifique, TC-STAR a créé l'infrastructure nécessaire pour accélérer la vitesse du progrès dans ce domaine. Il a recueilli les données nécessaires à la méthodologie statistique poursuivie par le projet, et il a mis au point une infrastructure d'évaluation compétitive. Cette approche garantit que les progrès de recherche soient reconnus comme tels, et que les méthodes développées par les différents partenaires puissent être comparées et validées de façon appropriée. En combinant de façon opportune la coopération et une compétition salutaire, cette approche devrait maximiser le progrès scientifique. Parallèlement, le consortium se mesure au sein de la

<sup>34</sup> Le projet allemand Verbmobil, 1993-2000, avec un financement global de 53M€ a entrepris un gros effort dans cette direction. L'initiative internationale C-STAR, un grand consortium consacré à la traduction parole-parole avec vingt partenaires, a débuté en 1991 et elle est encore en cours.

<sup>35</sup> TC-STAR (<http://www.tc-star.org>) est un projet intégré (PI), c'est-à-dire un grand projet au sein du VI programme-cadre.

communauté scientifique extérieure, et les partenaires du projet ont atteint les meilleures places dans les récentes évaluations internationales IWSLT<sup>36</sup>.

<sup>36</sup> International Workshop on Spoken Language Translation. IWSLT 2005 à Pittsburgh, USA, <http://www.is.cs.cmu.edu/iwslt2005/>. IWSLT 2004 à Kyoto, Japon, <http://www.slt.atr.jp/YIWSLT2004/>

## 5. Le pouvoir d'une technologie catalysante

### 5.1. Les besoins insatiables de l'être humain

La traduction automatique a tout juste commencé à être mise en pratique, la traduction du langage parlé au langage parlé est encore une recherche de frontière comme beaucoup de technologies la concernant. D'autre part, il existe une énorme demande latente qui doit être satisfaite, et nous savons qu'un jour la technologie pourra la satisfaire. Pouvons-nous, même vaguement, prévoir l'avenir?

Pour ne pas nous perdre dans les détails d'analyses de marché qui devraient continuellement être retouchées, essayons d'identifier les forces marquantes qui déterminent l'offre et la demande d'un futur marché. Certaines tendances sont largement acceptées aussi bien par de nombreux spécialistes que par des non spécialistes, comme par exemple la mobilité, le networking et le calcul distribué. Si nous comprenons les variables clés, technologiques et commerciales, les chaînes de valeur (ou, plus précisément les "réseaux" de valeur) ainsi que les dynamiques du marché, nous pouvons presque être sûrs que notre vision de l'avenir n'est pas aussi vague qu'on pourrait le croire. Il existe un stratagème qui rend la prévision encore plus digne de foi : *la meilleure façon de prédire l'avenir, c'est de l'inventer.*<sup>37</sup>

Si l'on observe les vraies histoires du succès des nouvelles technologies installées sur le marché, l'on trouvera deux ingrédients qui reviennent régulièrement : une nouvelle technologie et une vieille exigence humaine. La téléphonie mobile est un exemple typique, qui satisfait l'insatiable besoin de l'homme de communiquer avec d'autres personnes et aussi son désir d'avoir un impact malgré la distance. Dès l'enfance, l'on utilisait la communication presque en temps réel, mais pas isochrone, avec de brefs messages : autrefois à l'école l'on écrivait de petites notes sur du papier – maintenant les enfants utilisent les SMS. Il serait peut-être plus exact de comparer les SMS d'aujourd'hui aux télégrammes du passé? Il existe des ressemblances : dans les deux cas, un court texte est expédié à des destinataires par un canal électronique. Toutefois, vu l'énorme différence de prix et les domaines d'emploi très différents, il n'est pas pensable de considérer le SMS comme une forme moderne de télégramme. Toute tentative de prévoir l'ampleur du trafic des SMS d'aujourd'hui en se basant sur le nombre de télégrammes expédiés il y a vingt ans échouerait fatalement. De même qu'il serait trompeur et tout à fait approximatif d'estimer l'actuel emploi de Google en se basant sur les chiffres du "desktop paper research" ou en

<sup>37</sup> Citation attribuée à Alan Kay, pionnier du PC et spécialiste de l'interface usager.

cherchant dans les bases de données d'il y a vingt ans. En tout cas, il était clair même autrefois que l'on aurait adopté toute technologie allant au devant des besoins fondamentaux et insatiables de l'homme : les personnes ont besoin de communiquer avec d'autres personnes (téléphone, mobile, e-mail, SMS, chat), les personnes ont besoin de dominer les espaces (automobile, avion, et même la communication), les personnes ont soif d'informations et de distractions (internet, télévision).

**Les besoins insatiables de l'homme:  
la communication avec les personnes,  
la domination du temps et de l'espace,  
la soif d'informations et de distractions.**

Le marché actuel de la traduction et de la localisation ne nous dit pas grand-chose sur ce qui arriverait si la traduction était disponible en temps réel et à un coût raisonnable. Dans de nombreux cas, payer un euro pour chaque ligne de texte traduit n'a pas de raison d'être par rapport à la valeur ajoutée. Faire baisser d'un ou deux crans consentirait une énorme croissance de l'utilisation de la traduction, même si la qualité n'était pas parfaite. L'être humain a une grande envie de communiquer et d'obtenir des informations. Il est vrai que la traduction est déjà possible avec des traducteurs professionnels, mais tout comme il y a vingt ans nous n'aurions pas pu imaginer l'incroyable succès des SMS en nous basant sur l'utilisation des télégrammes, de même il nous est difficile d'imaginer aujourd'hui comment sera la vie après l'élimination universelle des barrières linguistiques.

Les marchés changent constamment.

La valeur abandonne certaines chaînes pour entrer dans d'autres. Les vieilles demandes croisent de nouvelles possibilités, et naissent de nouveaux produits et de nouveaux marchés.

Qu'arriverait-il si la traduction en temps réel était disponible à un coût raisonnable? Quels seraient les effets de cette traduction de texte et de parole, de documents, de sites web, de vidéo, de transmission de contenus? Qu'arriverait-il s'il n'y avait plus de barrières dues aux langues?

Nous n'aurions plus besoin de baisser nos estimations sur la situation économique de la traduction comme aujourd'hui. La quantité actuelle de traductions n'est que la pointe de l'iceberg, par rapport à la demande latente de traduction. Laissez-moi vous montrer ce que *moi* je voudrais faire.

Je voudrais regarder Al Jazeera – sous-titrée dans ma langue – pour avoir une idée de ce que pense le monde arabe. En parlant anglais avec mes pairs italiens ou espagnols, nous pourrions quelquefois apporter une petite aide pour élargir la bande de notre canal de communication. Il m'est difficile de lire un document en français, spécialement quand je ne connais pas la terminologie – une traduction approximative en allemand ou en anglais rendrait la lecture fondamentalement plus rapide. Durant des vacances au Portugal, je voudrais comprendre les menus et être en mesure d'échanger quelques paroles avec les habitants.

**Imaginez un monde avec la traduction  
en temps réel mise à disposition  
de façon économique et immédiate.**

Ne regarderiez-vous pas un site web en chinois ? Ne cuisineriez-vous pas cette merveilleuse recette de vos vacances en Espagne et qui, malheureusement, est écrite en espagnol ? Quelles sont vos requêtes ?

Nous Européens, nous pourrions communiquer plus facilement les uns avec les autres et nous pourrions nous sentir plus proches.

L'emploi des technologies du langage fournira un avantage économique significatif. Le contact avec les clients finaux est en train de s'automatiser, là où s'est possible, y compris le traitement des FAQ list multilingues – voilà les secteurs où parviennent les TLH. La plupart des biens non tangibles des sociétés résident, en plusieurs langues, dans les cerveaux et dans les files de leurs employés. Pouvoir accéder facilement à l'information, indépendamment de la langue, est une véritable nécessi-

té pour toute société opérant de façon globale. Voilà qui serait un marché intéressant, et les sociétés telles que IBM et SAP, Yahoo et Google se préparent.

La traduction en temps réel de toute source de données – *newswire*, contenus web dynamiques, vidéo; la recherche de documents en différentes langues; l'accès simplifié aux marchés étrangers, en particulier pour les professionnels et les PME, sont des opportunités illimitées pour l'Europe.

L'avènement de la traduction automatique sera soutenu par trois grands facteurs dynamiques du marché :

- les bas prix
- les opérations en temps réel
- les propriétés de l'élaboration automatique

Chacun de ces trois facteurs dynamiques du marché pourrait être, à lui seul, suffisamment fort pour soutenir un *business*.

## Interview de Dimitris Sabatakakis, CEO, SYSTRAN

SYSTRAN est leader de marché pour les fournitures de produits et solutions software pour la traduction dans les différentes modalités: "desktop", "enterprise" et "Internet". Les solutions qui sont offertes facilitent la communication dans 40 directions de traduction (20 couples linguistiques) et 20 contextes. Avec plus de trois décennies d'expérience dans la recherche et le développement, le software de SYSTRAN est utilisé par des sociétés leaders globales, des portails et des agences publiques. L'emploi des produits et des solutions SYSTRAN améliore la communication et augmente la productivité de l'utilisateur et fait épargner du temps aux marchés B2E, B2B et B2C, étant donné qu'ils distribuent des solutions en temps réel pour la recherche, la gestion du contenu, le support on line au client, les communications au sein des sociétés et l'e-commerce.



**Dimitris Sabatakakis**  
CEO de SYSTRAN  
Paris, France

Dimitris Sabatakakis est né en 1962 à Athènes, Grèce. Diplômé à l'Université de Strasbourg en Sciences économiques, il a débuté dans la finance, pour passer ensuite à l'industrie. Il s'est uni aux investisseurs, il a relevé et géré la récupération de la société Gachot, qui a été vendue au KEYSTONE/TYCO Group en 1995. Dimitris Sabatakakis est CEO de SYSTRAN depuis février 1997.

### *Quel rapport y-a-t-il entre le marché de la traduction automatique et celui de la traduction traditionnelle?*

Le marché traditionnel de la traduction est stable et représente un processus humain tout comme écrire. Le marché de la TA (traduction automatique) est très différent. L'offre comprend des services et des applications de traduction en temps réel et basés sur le web. Il est donc pratique d'adopter la TA pour les publications multilingues dans des situations où l'on peut combiner la demande de traduction de grands volumes de contenus et le contrôle de l'utilisateur sur le texte original. 99% du marché actuel des applications de traduction automatique concerne la recherche d'informations importantes dans les milliers de pages des portails tels que Yahoo! et Google qui sont traduits quotidiennement.

TA ou les software de traduction du langage sont des processus automatiques permettant à l'utilisateur de:

- comprendre dans sa propre langue, en temps réel et sans frais, le contenu exprimé en langues étrangères;
- publier le contenu dans différentes langues en temps réel, en «contrôlant» naturellement le texte d'origine.

***Les personnes comprennent-elles les capacités des systèmes automatiques? Ont-elles tendance à les surévaluer ou à les sous-évaluer?***

La valeur de la TA utilisée pour comprendre le sens général d'un texte est évidente. La preuve en est les milliers de pages qui sont traduites chaque jour.

La valeur de l'utilisation de la TA pour la publication en plusieurs langues est, elle aussi, évidente, mais l'investissement significatif requis pour "contrôler ou pour structurer le texte d'origine est souvent sous-évalué ou négligé. Pour l'évaluer correctement, tout le flux de travail de publication doit prendre en considération les problèmes de la production multilingue dès le début. En réalité, les organisations produisent un contenu (tels que fiches techniques de produits, matériel de marketing, informations de support technique, bases de connaissance etc.) dans une langue source qui est généralement l'anglais et elles ne considèrent la localisation du contenu que comme un pas en plus, de cette façon à cause de restrictions budgétaires, le montant de texte localisé est très inférieur au texte original.

Enfin, le maintien de versions de contenu localisé est un défi du moment qu'il est coûteux, lent (il s'agit d'un processus humain) et qu'il ne résoud pas complètement les problèmes d'inconsistance de la terminologie.

***Voyez-vous la TA plus comme un instrument de rationalisation pour la traduction humaine ou plus comme une nouvelle modalité de traduction?***

TA est à la fois un instrument pour la traduction humaine et une traduction d'un type nouveau. Les traducteurs humains devraient adopter la TA du moment qu'elle accroît la productivité de l'utilisateur et réduit les temps de traduction. Cette solution n'a pas encore pris pied, car l'ergonomie et l'interface utilisateur n'ont pas été explicitement créées pour les traducteurs humains. SYSTRAN a fait de gros efforts pour fournir ces instruments de façon à ce que la TA puisse être utilisée au sein des services de traduction traditionnels afin d'en augmenter la productivité.

## 6. Conclusions

Des instruments rapides, fiables et économiques pour communiquer, transporter et élaborer des données représentent l'épine dorsale d'une société de l'information moderne. Ce qui est vrai pour une normale technologie de l'information et de la communication l'est aussi pour les systèmes traduisant d'une langue à l'autre, que ce soit pour la communication ou pour l'accès à l'information. Pour l'Union Européenne, avec ses vingt langues officielles et encore plus nombreuses langues parlées, la disponibilité de traductions rapides, fiables et économiques est une nécessité, et la technologie de traduction devrait être considérée comme importante d'un point de vue stratégique.

D'autres régions du monde ont besoin de services automatiques de traduction, mais dans des contextes différents. Ils ne résoudre pas nos problèmes; nous ferions mieux de les résoudre nous-mêmes. Et nous avons les moyens de le faire.

En Europe, à cause des besoins de traduction automatique et de traduction du langage parlé, il existe des conditions de marché favorables pour les sociétés ayant l'intention d'offrir une technologie et des services de traduction. Certains grands groupes euro-

péens de recherche peuvent jouer le rôle de fournisseurs de technologie de traduction. En plus du marché européen, la demande de traduction automatique sera également élevée au Japon, en Chine et en Corée ainsi qu'en Inde.

L'industrie des technologies du langage croît rapidement, mais le plein potentiel de la traduction automatique ira au-delà d'une amélioration de l'efficacité de la traduction humaine actuelle. Elle s'ouvrira aux segments de marché qui, pour le moment, existent à peine, et qui ont des marges de profit tellement réduites qu'ils n'auront jamais aucun attrait pour les services de traduction humaine. Tout comme il aurait été impossible d'extrapoler l'utilisation des SMS de l'utilisation des télégrammes, il est difficile de prévoir le volume de traductions pour l'avenir. Il sera très grand, et il aidera à développer cette partie de notre potentiel économique qui est encore limité par les barrières linguistiques.

Toutes ces innovations demandent du temps, mais elles sont en marche et elles auront un impact énorme. Tâchons, entre-temps, de dessiner l'avenir d'une façon avantageuse pour nous. Lui donner une forme ne dépend que de nous.

## 7. Bibliographie et autres renseignements complémentaires

### 7.1. Bibliographie

- [Chr] C. Christensen : The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail. Harvard Business School Press, Boston, Mass., 1977.
- [CRa] C. Christensen, M. Raynor : The Innovator's Solution. Creating and Sustaining Successful Growth. Harvard Business School Press, Boston, Mass., 2003.
- [Cry] D. Crystal : English as a Global Language. Cambridge University Press, 2eme edition, 2003 (1ere edition 1997).
- [CSA] R. Beninatto, D. DePalma : Ranking of Top 20 Translation Companies. Common Sense Advisory, Inc., Juin 2005. Peut être téléchargé sur Common Sense Advisory website <http://www.common senseadvisory.com> à [http://www.common senseadvisory.com/members/res.cgi.php/050701\\_QT\\_top\\_20.php](http://www.common senseadvisory.com/members/res.cgi.php/050701_QT_top_20.php)
- [DGT] The Directorate-General for Translation of the European Commission (DGT). [http://europa.eu.int/comm/dgs/translation/index\\_en.htm](http://europa.eu.int/comm/dgs/translation/index_en.htm)
- [EB1] Europeans and languages. Eurobarometer 54 Special. INRA Report, 60 pages, février 2001. Ce rapport peut être téléchargé sur [http://europa.eu.int/comm/education/policies/lang/languages/barolang\\_en.pdf](http://europa.eu.int/comm/education/policies/lang/languages/barolang_en.pdf) et est mentionné dans [http://europa.eu.int/comm/education/policies/lang/languages/index\\_en.html](http://europa.eu.int/comm/education/policies/lang/languages/index_en.html)
- Notez que les tableaux se réfèrent à la vieille Europe des 15.
- [EB5] Europeans and languages. Eurobarometer 63.4, septembre 2005. L'enquête a été menée dans l'Union Européenne des 25 États membres et, de surcroît, dans les pays entrant (Bulgarie et Roumanie), dans les pays candidats (Croatie et Turquie) et au sein de la communauté turco-chypriote. Ce rapport peut être téléchargé sur [http://europa.eu.int/comm/public\\_opinion/archives/ebs/ebs\\_237.en.pdf](http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_237.en.pdf) et est mentionné dans <http://europa.eu.int/languages/en/document/80/20>
- [ELP] Languages of Europe. Sur le site web officiel de la Commission européenne concernant les langues parlées dans l'UE. [http://europa.eu.int/comm/education/policies/lang/languages/index\\_en.html](http://europa.eu.int/comm/education/policies/lang/languages/index_en.html)
- [Gra] D. Graddol : The Future of English? A guide to forecasting the popularity of the English language in the 21st century (commissionné par The British Council). The English Company (UK), 64 p., 2000 (1ere edition 1997), ISBN 0-86355-356-7.

- [Gri] B. Grimes : Ethnologue Language Database. <http://www.sil.org/ethnologue/>
- L’Ethnologue est un catalogue de plus de 6,700 langues parlées dans 228 pays.
- [Gor] R.G. Gordon (ed.) : Ethnologue : Languages of the World. 15ième édition, Dallas, Tex., 2005, SIL International. Version on-line : <http://www.ethnologue.com/>
- [Ney] H. Ney : One Decade of Statistical Machine Translation : 1996–2005. Dans : Proceedings of the MT Summit X, pp. i-12 – i-17, Phuket, Thaïlande, septembre 2005. Ce document peut être téléchargé sur <http://www-i6.informatik.nwth-aachen.de/web/Publications/index.html>
- [TMC] Translating for a multilingual community. Directorate-General for Translation of the European Commission (DGT), avril 2005, 19 pages. La brochure pdf peut être téléchargée sur [http://europa.eu.int/comm/dgs/translation/bookshelf/brochure\\_en.pdf](http://europa.eu.int/comm/dgs/translation/bookshelf/brochure_en.pdf)
- [TTW] Translation tools and workflow. Directorate-General for Translation of the European Commission (DGT), avril 2005, 25 pages. La brochure pdf peut être téléchargée sur [http://europa.eu.int/comm/dgs/translation/bookshelf/tools\\_and\\_workflow\\_en.pdf](http://europa.eu.int/comm/dgs/translation/bookshelf/tools_and_workflow_en.pdf)

## 7.2. Autres lectures

en français :

portail Europa consacré aux langues <http://europa.eu.int/languages/fr/home>

en anglais :

Common Sense Advisory <http://www.commonsenseadvisory.com/>

LISA – Localization Industry Standards Organization <http://www.lisa.org/>

Multilingual Computing, Inc. <http://www.multilingual.com/>

EAMT – European Association for Machine Translation <http://www.eamt.org/>

ELRA – European Language Resources Association. <http://www.elra.info/>

ELSNET – European Network in Language and Speech. <http://www.elsnet.org/>

GALA – The Globalization and Localization Association. <http://www.gala-global.org/>

## 7.3. Tableaux

Tableau 1 : Conditions commerciales pour le commerce international au sein de l’UE pour les biens et l’information. Pour les services de l’information, la barrière de la langue représente un obstacle de taille à franchir.

Tableau 2 : Les principales langues du monde en millions de locuteurs natifs se-

lon deux différentes sources, (A) The English Company's engco model [Gra] et (B) chiffres comparatifs tirés de Ethnologue ([Gri ] cf [Gra]).

Tableau 3 : "Influence globale" des 12 principales langues selon le modèle engco (cf tableau 2). Le classement 100 représente la position de l'anglais en 1995 [Gra].

Tableau 4 : Disciplines dans lesquelles les universitaires allemands affirment que l'anglais est leur langue de travail [Gra].

Tableau 5 : Les 20 langues officielles de l'Union Européenne et leurs abréviations [ELP]. L'Irlandais sera la 21e langue officielle de l'UE à partir du premier janvier 2007.

Tableau 6 : Capacité de parler des langues étrangères dans l'UE : pourcentage de citoyens européens parlant leurs langues respectives dans l'UE (comme langue maternelle /comme deuxième langue ou langue étrangère / une autre langue). L'enquête a été menée dans l'Union Européenne des 25 États membres et, de surcroît, dans les pays entrant (Bulgarie et Roumanie), dans les pays candidats (Croatie et Turquie) et au sein de la communauté turco-chypriote - Source : [EB5].

Tableau 7 : Placement des 20 premiers fournisseurs de services linguistiques - revenu 2004. (Ces chiffres prennent en considération les deux importantes acquisitions de 2005, l'acquisition de Bowne Global Solutions par Lionbridge et l'acquisition de TRADOS par SDL.)

Tableau 8 : Exemples d'innovations de rupture, et technologies correspondantes réalisées. - Source : [Chr], [CRa].

## 7.4. Figures

Fig. 1 : Proportion des livres publiés chaque année dans le monde dans chaque langue. L'anglais est la langue la plus utilisée pour la publication des livres : plus de 60 pays publient leurs titres en anglais [Gra].

Fig. 2 : Les langues du monde. Chaque point représente la principale localisation des langues vivantes citées dans Ethnologue [Gor], un catalogue de plus de 6.700 langues parlées dans 228 pays.

Fig. 3 : Le changement technologique déterminé par les innovations de soutien et de rupture. Source : [CRa].

Fig. 4 : Budget du DoD (Département de la Défense) des Etats-Unis pour les technologies du langage concernant la traduction. Budget prévisionnel 2005 pour la re-

cherche, le développement, le test et l'évaluation (RDT&E), défense globale, chez DARPA (Defense Advanced Research Projects Agency). – Source : DARPA.

Fig. 5 : Diagramme fonctionnel d'un système de traduction parole-parole (TPP).

## 7.5. Liste des acronymes

|         |  |
|---------|--|
| AFNLP   | Asian Federation of Natural Language Processing  |
| ASR     | Automatic Speech Recognition (RAP - Reconnaissance automatique de la parole)   |
| BLEU    | Mesure statistique de qualité pour les traductions en correspondance avec le jugement humain (plus nombreuses sont les correspondances, mieux c'est) |
| CCE     | Commission de la Communauté Européenne   |
| DARPA   | Defense Advanced Research Projects Agency  |
| DGT     | Direction Générale de la Traduction  |
| DTMF    | Dual Tone Multiple Frequency   |
| ELRA    | European Language Resources Association (Association Européenne pour les Ressources Linguistiques)   |
| GALE    | Global Autonomous Language Exploitation  |
| IVR     | Interactive Voice Response   |
| LISA    | Localization Industry and Standards Association  |
| MT      | Mémoire de Traduction  |
| PI      | ici : Projet Intégré<br>(dans d'autres contextes : Protocole Internet; Propriété Intellectuelle)   |
| PME     | Petites et Moyennes Entreprises  |
| RI      | Recherche d'Informations   |
| ROI     | Return On Investment (Retour sur Investissement)   |
| SMS     | Short Message Service  |
| TA      | Traduction Automatique   |
| TC-STAR | Technology and Corpora for Speech-to Speech Translation  |
| TH      | Traduction Humaine   |
| TI      | Technologie de l'Information   |
| TLH     | Technologie du Langage Humain  |
| TLN     | Traitement du Langage Naturel  |
| TLP     | Traduction Langage Parlé   |
| TPP     | Traduction Parole-Parole   |
| TTP     | Traduction Texte-Parole  |
| UE      | Union Européenne   |



Finito di stampare  
nel mese di dicembre 2006  
dalla Tipolitografia TEMI - Trento

