

**CNRC-NRC**

*Institut de  
technologie  
de l'information*

# La traduction automatique : de la technologie aux traducteurs

**Michel Mellinger**

**Collaborations et transfert technologique**

**CNRC-Gatineau**



Conseil national  
de recherches Canada

National Research  
Council Canada

**Canada**

## Plan

**Rappel – TABR et TAS**

**Rappel – Fidélité et fluidité**

**TABR + TAS + ?**

**... scénarios de combinaisons, quelques exemples**

**La TA au service des traducteurs professionnels ?**

**... la TABR, la TAS**

**Constats et sujets d'actualité en TAS**

## Plan

**> Rappel – TABR et TAS**

**Rappel – Fidélité et fluidité**

**TABR + TAS + ?**

**... scénarios de combinaisons, quelques exemples**

**La TA au service des traducteurs professionnels ?**

**... la TABR, la TAS**

**Constats et sujets d'actualité en TAS**

## Rappel – TABR et TAS

Les techniques de traduction automatique (TA) suivent l'une de deux approches :

- systèmes à base de règles (TABR)
- systèmes statistiques (TAS)

Question : quelle est la différence fondamentale entre ces deux approches ?

## Plan

### Rappel – TABR et TAS

#### > Rappel – Fidélité et fluidité

TABR + TAS + ?

... scénarios de combinaisons, quelques exemples

La TA au service des traducteurs professionnels ?

... la TABR, la TAS

Constats et sujets d'actualité en TAS

## Rappel – Fidélité et fluidité

**La performance d'un système de TA est jugée selon deux aspects : fidélité et fluidité**

**- fidélité ... contenu = «s'attacher au texte de départ tout en respectant la destination de sa traduction» (*Translation Journal, October 2001*)**

**St Jérôme : «c'est le sens qu'il faut rendre et tout le sens et non les mots» («*Non verbum e verbo sed sensum exprimere de sensu*»)**

**- fluidité ... forme = la traduction du texte source se lit aisément**

## Rappel – Fidélité et fluidité

La TABR produit typiquement une bonne fluidité ; mais elle est difficilement adaptable à un domaine (limitée aux dictionnaires), donc fidélité « générique »

La TAS vise à la fois une bonne fidélité (entraînement pour le domaine ciblé) et une bonne fluidité (utilisation de modèles de langue),

*... mais ceci dépend grandement de la qualité de l'entraînement (corpus utilisés et soin apporté à l'entraînement)*

## Plan

**Rappel – TABR et TAS**

**Rappel – Fidélité et fluidité**

**> TABR + TAS + ?**

**... scénarios de combinaisons, quelques exemples**

**La TA au service des traducteurs professionnels ?**

**... la TABR, la TAS**

**Constats et sujets d'actualité en TAS**



## TABR + TAS + ?

**La combinaison de différents systèmes d'aide à la traduction est un domaine de r-D très actif, qui vise à mettre à profit les caractéristiques complémentaires de ces systèmes**

### **1. TABR améliorée par la TAS :**

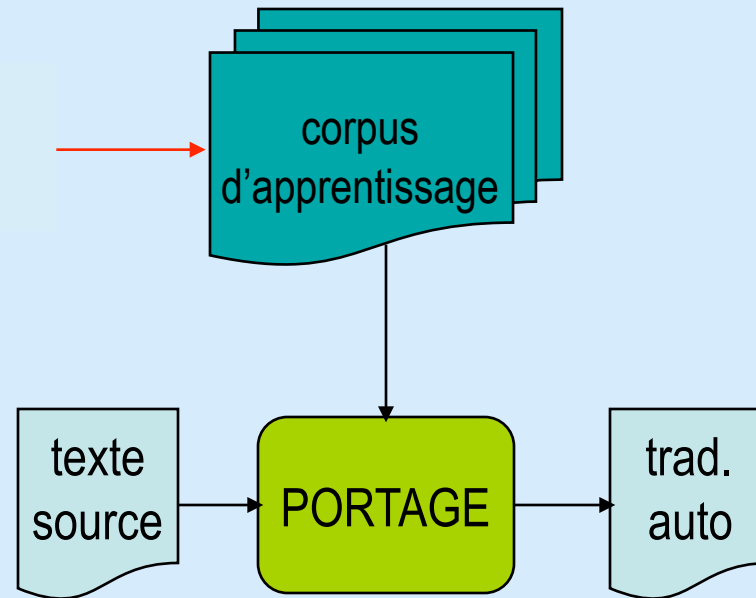
**= systèmes « hybrides »**

**= « post-édition statistique/automatique »**

**Exemple : « SYSTRAN Enterprise Server 7 », dont la sortie a été annoncée en juin 2009 ...**

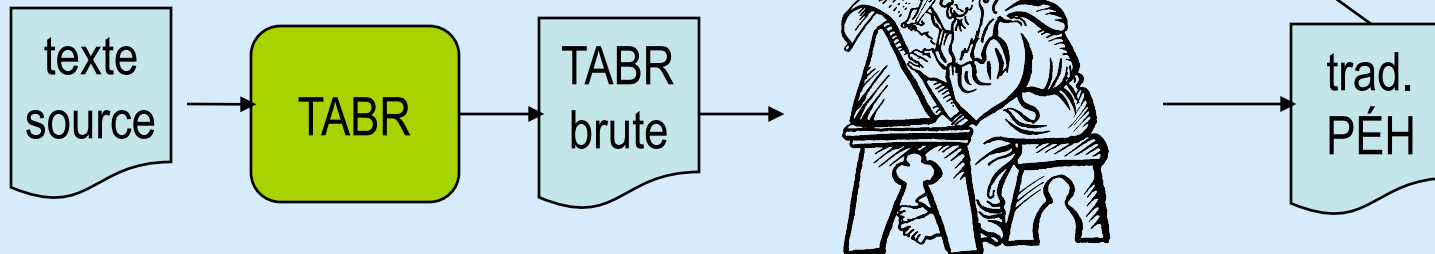
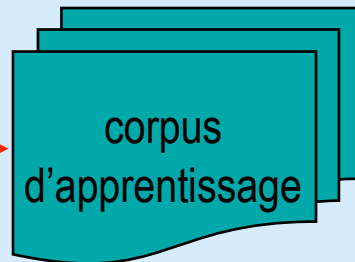
# TAS simple

Paires  
<S,T>



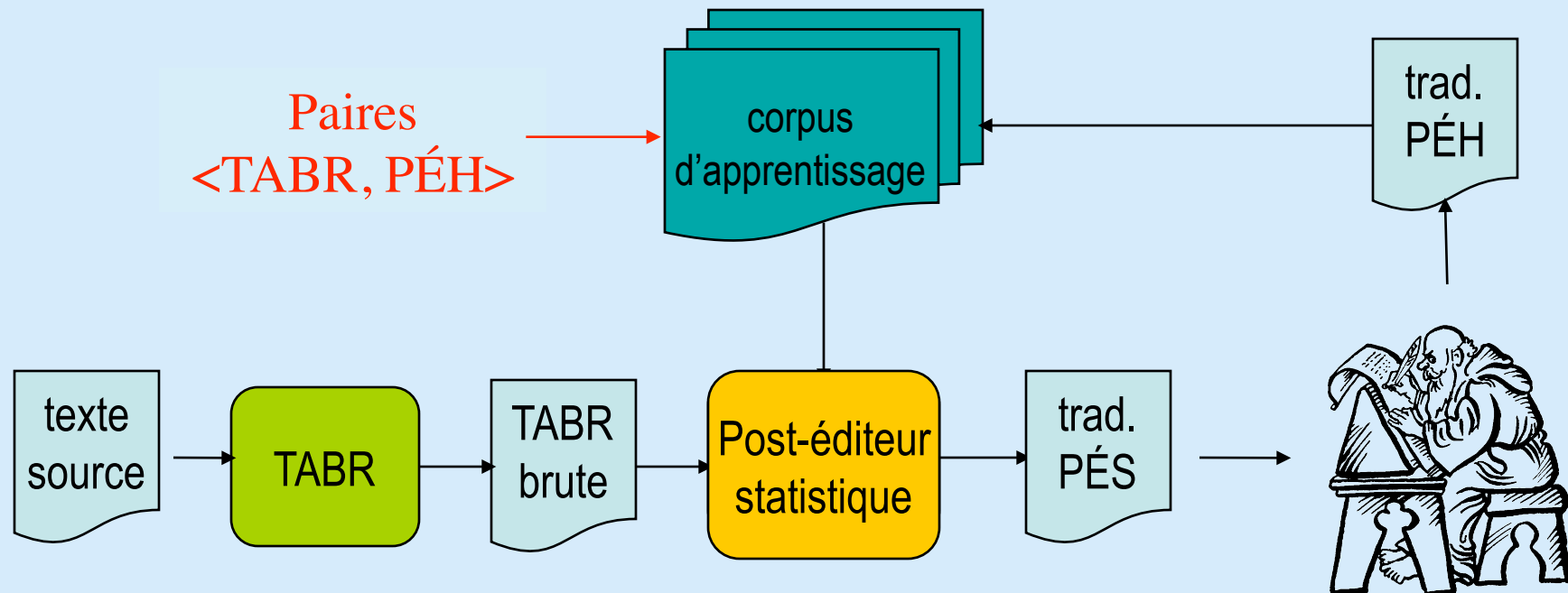
# TAS pour PÉS utilisant la post-édition humaine

Paires  
<TABR, PÉH>



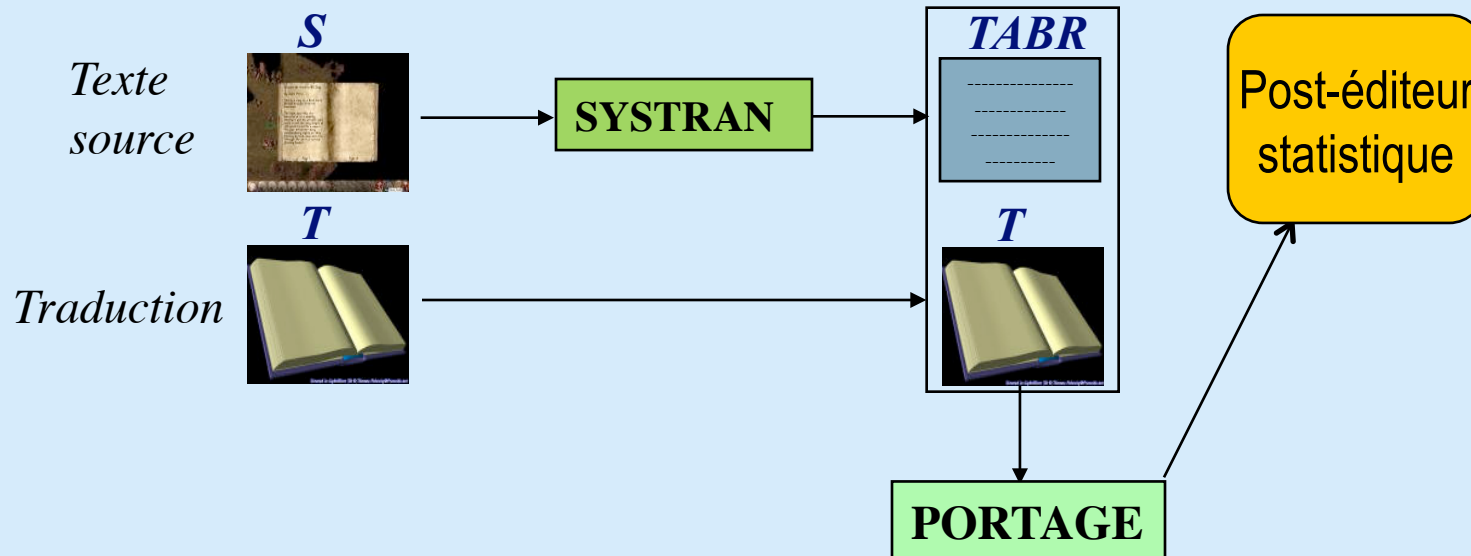
# TAS pour PÉS

## utilisant la post-édition humaine



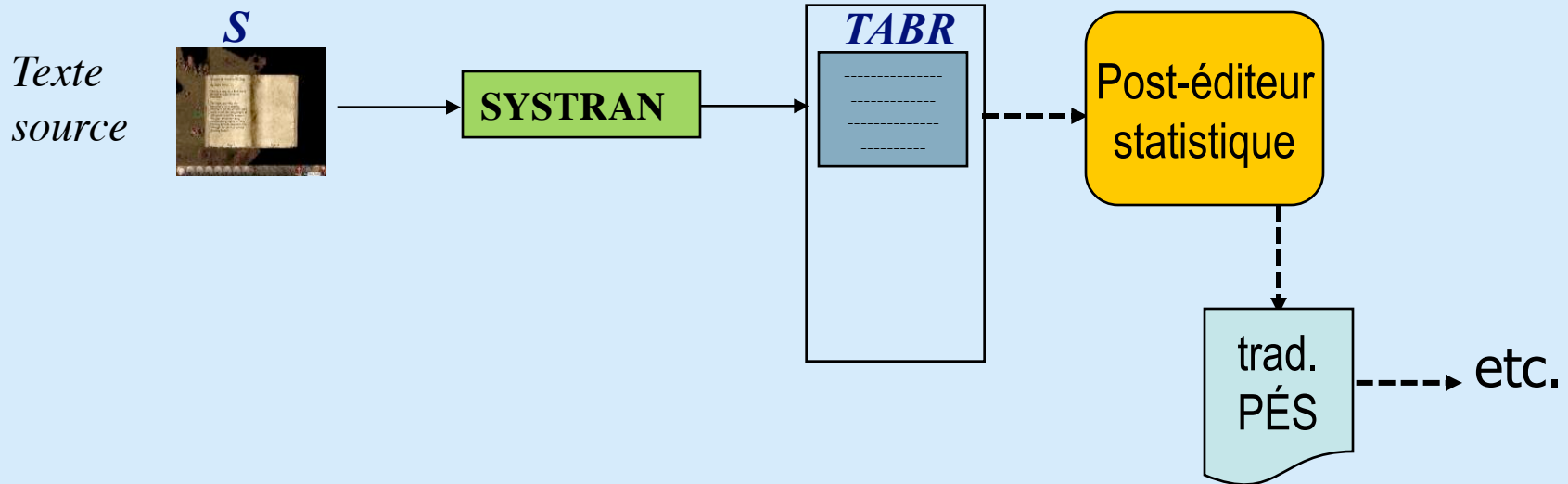
# TAS pour PÉS *sans post-édition humaine*

Il est difficile d'obtenir des corpus de TABR post-éditée par des traducteurs, mais nous pouvons utiliser un corpus habituel : on entraîne alors la TAS à "traduire" une sortie TABR de la source connaissant la traduction faite indépendamment par un traducteur ... et ... ça marche !



# TAS pour PÉS *sans post-édition humaine*

Il est difficile d'obtenir des corpus de TABR post-éditée par des traducteurs, mais nous pouvons utiliser un corpus habituel : on entraîne alors la TAS à "traduire" une sortie TABR de la source connaissant la traduction faite indépendamment par un traducteur ... et ... ça marche !



## TABR + TAS + ? SYSTRAN Enterprise Server 7

Présentation (site Web) de SYSTRAN Enterprise Server 7 :

- « Premier moteur de traduction « hybride » du marché qui combine les qualités de la technologie à base de règles (“rule-based”) et du traitement “statistique”.
- « Nouveaux outils : SYSTRAN *Training Manager* permet aux entreprises d'entraîner le logiciel pour lui apprendre la terminologie spécifique de l'entreprise. SYSTRAN *Corpus Manager* permet de collecter et gérer les contenus utilisés dans l'entraînement statistique. Nouvelle interface de révision des traductions intégrée au Translation Project Manager pour une meilleure productivité en post-édition.»

## TABR + TAS + ? SYSTRAN Enterprise Server 7

- « Comprendre en temps réel [ *compréhension = TABR* ]
- « Produire rapidement des traductions cohérentes. Des outils puissants de révision et d'assurance-qualité aident à contrôler la qualité et l'homogénéité des traductions. [ ? ]
- « Réduire les coûts de traduction. Les traducteurs peuvent augmenter considérablement leur productivité en révisant les traductions SYSTRAN pour produire leurs traductions finales.» [ *grâce à la TAS : post-édition productive* ]
- « Réutiliser et collaborer. Les bases terminologiques et les traductions déjà validées sont stockées sur le serveur.  
» [ *fonction de mémoire de traduction ?* ]



## TABR + TAS + ? SYSTRAN Enterprise Server 7

- « Personnaliser pour une entreprise ou un secteur d'activité. Le module d'entraînement permet d'apprendre au logiciel la terminologie spécifique à chaque client et d'améliorer ainsi considérablement la qualité de la traduction.  
[ *TAS >> TABR réf. fidélité dans un domaine spécifique* ]
- « Accroître la productivité. Workflow efficaces pour la traduction, la création de terminologie et la post-édition.  
[ *aspects intégration système* ]
- « Traiter les volumes. Déploiement sur plusieurs serveurs, ... Intégrer la traduction dans toutes les applications métiers de l'entreprise. » [ *intégration dans l'environnement applicatif du client* ]

# TABR + TAS + ?

1. TABR améliorée par la TAS ✓
2. TABR au secours des mémoires de traduction : la TABR intervient lorsque les correspondances sont  $<$  seuil  $x\%$

**Exemple : étude de cas de SYSTRAN concernant leur projet avec Symantec ...**

## TABR + TAS + ? SYSTRAN chez Symantec

**SYSTRAN : « Dans Idiom WorldServer, le *Fuzzy Score* est défini à 75%. SYSTRAN traduit automatiquement les segments situés en dessous de ce seuil. Les traducteurs révisent les segments colorés dans l'interface Browser Workbench.»**

**Symantec : « Nos indicateurs affichent une augmentation de 100% du volume géré par notre équipe interne de traduction depuis l'introduction des technologies SYSTRAN dans notre workflow de localisation. La traduction automatique est une innovation très performante lorsqu'elle est intégrée dans un flux de travail rationalisé. »**

## TABR + TAS + ?

1. TABR «corrigée par» la TAS ✓
2. TABR au secours des mémoires de traduction ✓  
(TABR lorsque les correspondances sont  $<$  seuil  $x\%$ )
3. TAS au secours des mémoires de traduction : la TAS est intégrée à l'intérieur même d'un environnement de mémoire de traduction  
Exemple : travaux actuels au CNRC (incluant les «estimations de confiance» TAS vs. MT)

## Plan

**Rappel – TABR et TAS**

**Rappel – Fidélité et fluidité**

**TABR + TAS + ?**

**... scénarios de combinaisons, quelques exemples**

**> La TA au service des traducteurs professionnels ?**

**... la TABR, la TAS**

**Constats et sujets d'actualité en TAS**

## La TA au service des traducteurs professionnels ? < la TABR >

La TABR n'a guère réussi à s'imposer aux traducteurs professionnels et elle n'évolue plus ou peu d'elle-même : l'effort requis n'est pas rentable, *mais* la TABR peut néanmoins fournir une traduction suffisamment compréhensible.

Exemple : étude de cas de SYSTRAN concernant leur projet avec le Lincolnshire County Council ...

## La TA au service des traducteurs professionnels ? < la TABR >

... Lincolnshire County Council (réf. la TABR SYSTRAN ) :

« We quickly concluded that whilst manual translations by a specialist translator or translation agency would be more grammatically perfect, with around 10,000 documents and 10 languages to address, the *cost* would be astronomic. To date, the quality of output has been *good*. We were never naive enough to believe that our pages would be as fluent or literal as a manual translation, but *if visitors understand enough of the web content to get a flavour of its meaning, then there can be no argument for not using it.*»

## La TA au service des traducteurs professionnels ? < la TABR >

**PAHO (Pan American Health Organization) a développé un environnement « élargi » autour de son système de TABR et utilise la TABR en production; leur système, appelé PAHOMTS® est un système convivial qui évolue grâce à l'intégration des traducteurs dans le contrôle de qualité :**

**« PAHOMTS® is a fully automatic Machine Translation (MT) program, developed and maintained by PAHO's computational linguists, translators, and programmers.**

**« The translation unit has used PAHOMTS® to process over 88 million words since 1980. Staff and freelance translators postedit the raw output to produce high-quality translations with a 30-50% gain in productivity.»**



## La TA au service des traducteurs professionnels ? < la TABR >

### PAHOMTS® ...

«The software includes a completely trilingual graphical user interface. The extensive context-sensitive online Help in English, Spanish, and Portuguese simplifies the process of browsing and updating the PAHOMTS® dictionaries. Each of the MT dictionaries contains over 150,000 words, phrases, and context-sensitive rules. Both the programs and the dictionaries are constantly being enhanced based on feedback from the users of the systems.»

# La TA au service des traducteurs professionnels ? < la TABR >

## PAHOMTS® ...

### Cycle d'amélioration continue chez PAHO :

- **rétroaction de la part des traducteurs : les traducteurs participants sont une source d'amélioration continue**
- **équipe de linguistes : est-ce que les difficultés observées sont systématiques et pourraient être exprimées en règles plus ou moins générales, ou spécifiques ?**
- **équipe de programmeurs : implantation d'améliorations selon les instructions des linguistes (= nouvelles règles)**
- **équipe de linguistes : validation des améliorations avec les traducteurs participants**

## La TA au service des traducteurs professionnels ? < la TAS >

La TAS a dépassé la TABR en performance et évolue vers des résultats « post-éditables », *mais* :

- (négatif) la mise en œuvre de la TAS n'est pas simple : *la phase d'entraînement* est cruciale et requiert un degré avancé d'expertise :
  - = TI (environnement TI, virtualisation, TI distribuée),
  - = collecte et gestion de grands corpus multilingues (plusieurs dizaines/centaines de millions de mots),
  - = linguistique computationnelle (pré-traitement de corpus),
  - = compréhension de l'univers de la traduction, excellent service technique et excellent service aux traducteurs.

## La TA au service des traducteurs professionnels ? < la TAS >

La TAS a dépassé la TABR en performance et évolue vers des résultats « post-éditable », mais ...

- (positif) les progrès de la TAS sont constants et les traducteurs professionnels commencent à l'adopter

Exemples : Google Translate, essais PORTAGE

- (positif) les techniques de post-édition sont en plein développement

Exemple ... ici, à l'UQO !

- (positif ou négatif ?!) le métier de traducteur va évoluer en conséquence ... Question : qu'en pensez-vous ?

## Plan

Rappel – TABR et TAS

Rappel – Fidélité et fluidité

TABR + TAS + ?

... scénarios de combinaisons, quelques exemples

La TA au service des traducteurs professionnels ?

... la TABR, la TAS

> **Constats et sujets d'actualité en TAS**

## Constats et sujets d'actualité en TAS

(Réf. Roland Kuhn et Pierre Isabelle, MT Summit XII, Ottawa, 2009)

- de plus en plus de recherche par des laboratoires de plus en plus nombreux à travers le monde
- des logiciels libres fournissent aux chercheurs un environnement ouvert de TAS (essentiellement « Moses »)
- Google Translate est devenu l'étalon en traduction disponible publiquement ; il est utilisé assez largement par les étudiants ... et par certains traducteurs
- la qualité de la TAS est maintenant acceptable, et on peut envisager une post-édition productive, pour certains couples de langues : En-Fr, En-Sp

## Constats et sujets d'actualité en TAS

(Réf. Roland Kuhn et Pierre Isabelle, MT Summit XII, Ottawa, 2009)

- la combinaison de systèmes TAS + TABR est efficace (la TABR contribue une base de connaissance linguistique)
- la collecte de grands corpus bilingues et multilingues s'accélère :  
Exemples : Hansard (historique), GC.CA (En-Fr), sites Web EU (23 langues), GigaFrEn (En 575 M mots, Fr 672M mots)
- introduction de concepts syntaxiques dans les systèmes de TAS : entraînement contraint par la décomposition syntaxique des phrases / entraînement de TAS sur des arbres syntaxiques puis recomposition

# Constats et sujets d'actualité en TAS

(Réf. Roland Kuhn et Pierre Isabelle, MT Summit XII, Ottawa, 2009)

- intérêts variés dans le choix des langues et couples de langues :

États-unis : chi, ara, urd, pus, ... **VERS l'anglais**

Europe : langues européennes (diverses directions)

Asie : langues asiatiques

Microsoft et autres multinationales = eng comme source

chi => eng est de plus en plus populaire en recherche

Question : comment définir la « qualité » en traduction ?

Et comment mesurer la « qualité » d'une traduction?





**CNRC-NRC**

*Institut de  
technologie  
de l'information*

# **La traduction automatique : de la technologie aux traducteurs ... ... un peu démystifiée ?**

**La traduction automatique :  
de la technologie aux traducteurs ...  
... un peu démystifiée ?**

**Merci pour votre attention !**

**Questions ? Commentaires ?**

**CMRC-NRC**

*Institut de  
technologie  
de l'information*

La  
Science  
— à l'œuvre pour le —  
Canada



Conseil national  
de recherches Canada

National Research  
Council Canada

Canada