

Vitrine-forum sur le traitement informatique des langues

Montréal, Canada 14-15 juin 2003

Table ronde

Les solutions susceptibles de mener à une intensification des travaux de recherche universitaires dans le domaine du TIL et à un accroissement de leur qualité

TIL, Langues nationales & partenariats francophones

Mokhtar BEN HENDA

Mokhtar.benhenda@isd.rnu.tn

CEM-GRESIC, Université de Bordeaux 3-FRANCE
ISD, Université La Manouba-TUNISIE

1. Introduction

- L'histoire biblique de la tour de Babel : la diversité linguistique
- Langues dominantes, langues minoritaires,
- Langue et culture : la polémique du contenant / contenu
 - ◆ Une seule langue pour plusieurs cultures
 - ◆ Plusieurs langues pour une seule culture

Nouvelles données des NTICs → TIL

- Nouveaux défis :
 - ◆ Normalisation (contenants, contenus = Interopérabilité)
 - ◆ I18n
 - ◆ L10n



Problématiques de recherche

2. Problématique

- Comment gérer l'aspect linguistique intensif du contenu informationnel sur le réseau (codage des langues) ?
- Comment ont été abordés les deux principes de l'Internationalisation (I18n) et de la Localisation (L10n) dans les systèmes et les ressources d'information multilingues (langues complexes) ?
- Quels sont les apports de l'I18n pour les langues « complexes » (i.e.arabe) au sein des systèmes d'information multilingues ?
- Quelle forme de partenariat peut exister entre le français comme langue pivot et l'arabe comme sa première langue partenaire du Sud ?

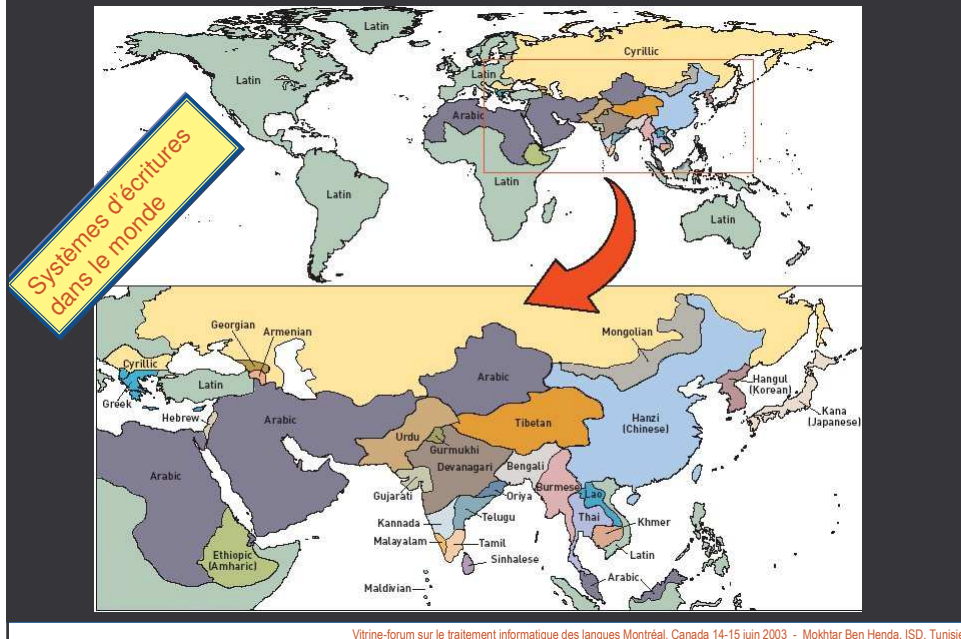
Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

3. Objectifs

- Renforcer le volet culturel dans le traitement automatisé des langues « minoritaires »,
- Révoquer les solutions clé en main de l'I18n et de la L10n proposées par les multinationales,
- Faciliter davantage la transparence linguistique dans les communautés caractérisées par un « Multilinguisme Lourd »
 - ◆ Bidi (RTL / LTR)
 - ◆ Rendus graphiques
- Renforcer le partenariat français-arabe dans le cadre de R&D (AIESI = Cyberprof)

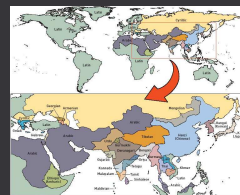
Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

4. Contexte linguistique mondial



4. Contexte linguistique mondial

- ❑ Porteurs de valeurs linguistiques et culturelles diverses
- ❑ État polémique accentué par les écritures complexes (minoritaires)



Language	Number of Countries
Arabic	18
Farsi	1 (Iran)
Urdu	2 (India, Pakistan)
Hebrew	1 (Israel)
Yiddish	Israel, North America, South America, Russia, Europe

Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

5. Caractéristiques propres des langues dites « complexes » : cas de la langue arabe

- **Codage des caractères**
 - ◆ ASMO (+ d'une douzaine de pages de codes)
 - ◆ ISO/UNICODE
- **Terminologie**
 - ◆ Pauvreté lexicale en Sc. & Tech. (emprunts, translittérations)
 - ◆ Variantes terminologiques (synonymie, polysémie)
 - ◆ Absence des sigles et acronymes
- **Langues informelles (modernes ou médianes)**
 - ◆ Phonologies (Intonations)
 - ◆ Terminologie (interférences langagières, translittérations)
 - ◆ Langues d'information et de communication (Publicité, Presse)
- **I18n, L10n (diversité culturelle : calendriers, chiffres, Bidi)**

Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

6. Partenariats linguistiques

- **Le Français, langue partenaire**
 - ◆ Peut-on développer l'usage du français et celui de la langue du pays d'accueil ?
 - ◆ OUI : Développer des rapports directs avec le maximum de langues nationales pour éviter au maximum l'interférence de langues pivots
- **Formes de partenariat (Projets)**
 - ◆ Développement de la TAO (Industrie de la langue)
 - ◆ Développement d'applications multilingues
 - ◆ Projets internationaux de recherche en TIL
- **Enjeux du partenariat numérique multilingue**
 - ◆ Réduction de la fracture numérique dans les PVD
 - ◆ Renforcement des ressources humaines locales
 - ◆ Lutte anti-uniformisation linguistique et culturelle de la mondialisation
 - ◆ Préserver les identités culturelles et linguistiques locales

Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

7. Démarches, recommandations

- **Collaboration internationale entre équipes de recherche et d'application** (complémentarité d'approches pour réduire le coût du développement des outils technolinguistiques)
- **Développement & adaptation des logiciels de TAL aux caractéristiques linguistiques et culturelles des partenaires du Sud** (phonétique, syntaxe, sémantique, terminologie, didactique etc.)
- **Renforcer les initiatives de solutions localisées contre l'hégémonie des multinationales**

Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie

8. Conclusion

En tenant compte du fait que ...

- Quand on produit un logiciel quelconque, 2/3 des utilisateurs potentiels sont en dehors des limites du monde anglo-saxon,
- Les utilisateurs peuvent refuser l'usage d'un programme qui ne répond pas à leurs exigences linguistiques et/ou culturelles les plus subtiles,
- Un logiciel est inutile pour un utilisateur qui n'y trouve pas les ressources dont il a besoin dans la forme qu'il attend.

Il est donc nécessaire de ...

- Localiser sur une base linguistique et culturelle fournie par des personnes ressources qui vivent au sein du milieu cible plutôt que par des experts qui n'ont que la connaissance théorique de la langue et du milieu

Vitrine-forum sur le traitement informatique des langues Montréal, Canada 14-15 juin 2003 - Mokhtar Ben Henda, ISD, Tunisie