

# Évaluation applicative des terminologies destinées à la traduction spécialisée

Estelle Delpech

Atelier EvalECD

25/01/2011

LINGUA ET MACHINA

AGENCE NATIONALE DE LA RECHERCHE  
ANR

lina



# Plan

1. Contexte, problématiques
2. Protocole d'évaluation
3. Expérimentation et résultats
4. Conclusions



# Plan

1. **Contexte, problématiques**
2. Protocole d'évaluation
3. Expérimentation et résultats
4. Conclusions



# Contexte

- Lingua et Machina
  - commercialise des outils de TAO :
    - mémoires de traduction
    - extraction terminologique bilingue
    - gestion de terminologie
- Metricc
  - exploitation des corpus comparables pour :
    - outils de TAO
    - RI interlingue

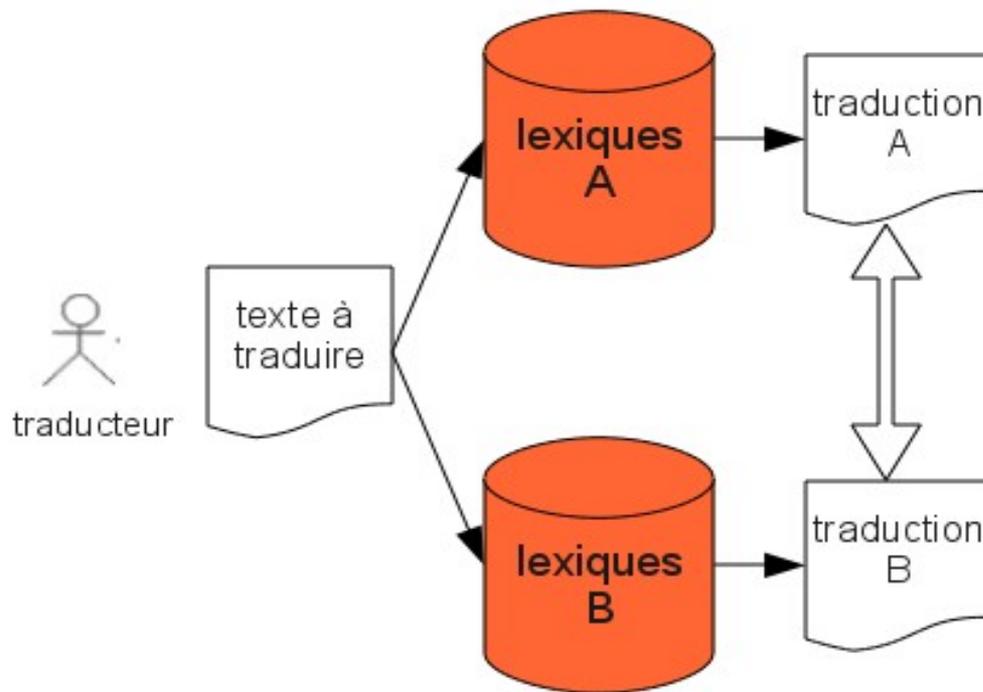


# Problématique

- Comment évaluer ces lexiques bilingues issus de corpus comparables ?
- Souhaits :
  - ne pas se contenter d'une évaluation par rapport à une référence
  - mesurer le véritable apport *en contexte d'utilisation*



# Contextes d'utilisation comparés



# Questions soulevées

- Comment évaluer la qualité d'une traduction ?
- Faut-il évaluer la qualité de la traduction dans son entier ou uniquement la traduction des termes techniques ?
- Comment évaluer une traduction technique sans expert ?



# Comment évaluer la qualité de la traduction ?

- Mesures objectives TA
  - avantage de la reproductibilité faible
- Grilles d'évaluation Traductologie
  - trop complexes, peu documentées
- Évaluation subjective style TA
  - jugement de qualité sur 3 catégories
  - classement



# Évaluer toute la traduction ?

- Qualité d'une traduction = interaction complexe de nombreux paramètres
- Proposition :
  - isoler les éléments sur lesquels les lexiques sont censés aider
  - n'évaluer que la traduction de ces éléments



# Comment pallier l'absence d'expert ?

- Pas d'expert du domaine qui puisse valider les traductions
- Emploi d'une référence
  - les textes à traduire doivent avoir une traduction existante
    - résumés d'articles scientifiques
    - textes de sites web bilingues (vulgarisation)

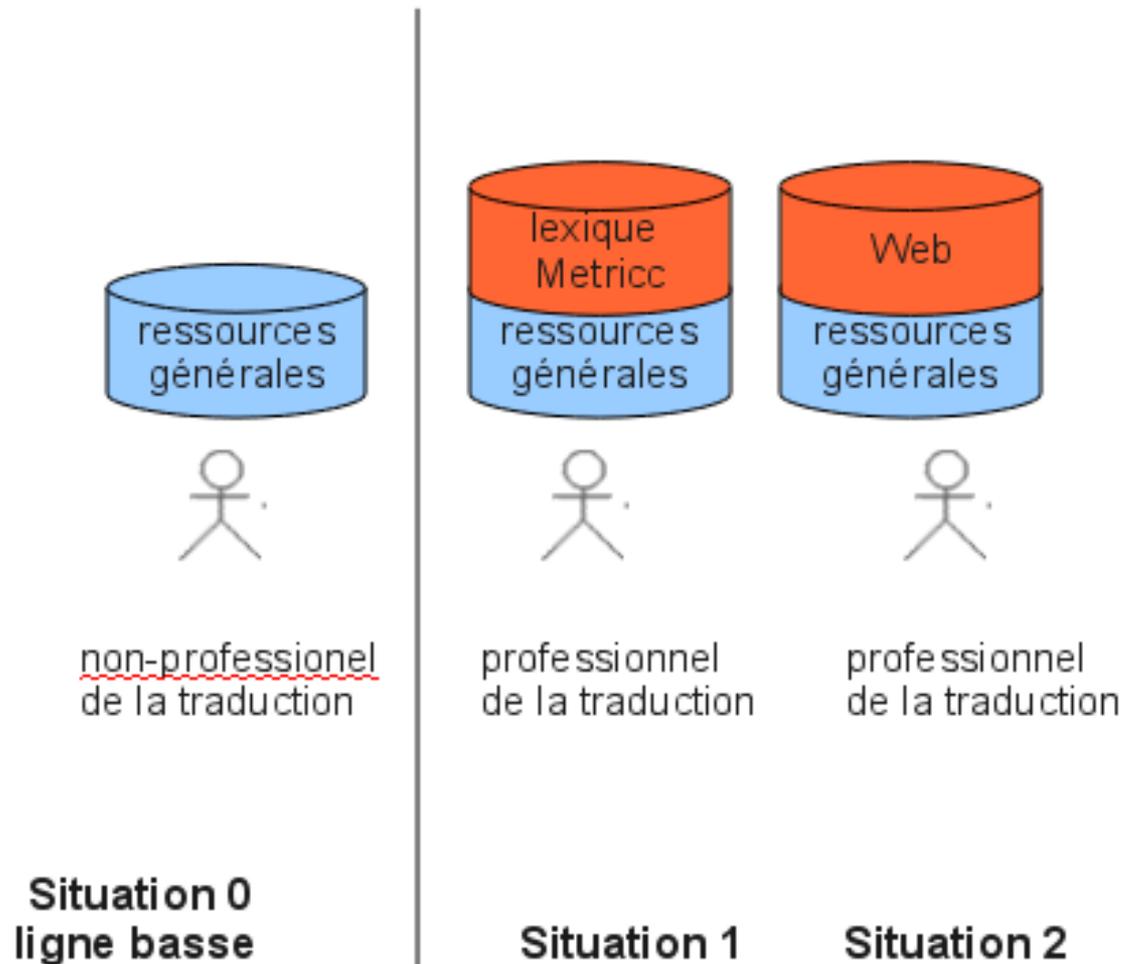


# Plan

1. Contexte, problématiques
2. **Protocole d'évaluation**
3. Expérimentation et résultats
4. Conclusions



# Protocole (1)



# Protocole (2)

	cancer du sein	sciences de l'eau
<i>BASE LINE</i> ress. gén.	non-traducteur	non-traducteur
<i>SITUATION 1</i> ress. gén. + <i>lexique CC</i>	traducteur 1	traducteur 2
<i>SITUATION 2</i> ress. gén. + <i>Web</i>	traducteur 2	traducteur 1

# Protocole (3)

- Le traducteur traduit un texte spécialisé à l'aide d'une ressource donnée
- Une fois le texte traduit :
  - indique les termes ou expressions pour lesquels il a eu besoin d'une ressource
  - indique la traduction retenue
  - indique les ressources utilisées



# Protocole (4)

- Traduction anonymisées, mélangées
- Jugement qualité
  - exact : terme ou expression consacrée
  - acceptable : sens maintenu
- Classement
  - de la meilleure à la moins bonne
  - égalités possibles



# Plan

1. Contexte, problématiques
2. Protocole d'évaluation
3. **Expérimentation et résultats**
4. Conclusions



# Données

	<b>cancer du sein</b>	<b>sciences de l'eau</b>
Corpus acquisition	~400k par langue portail <i>Elsevier</i>	~2 M par langue revues <i>Sciences de l'eau</i> , <i>Water Science and Technology</i>
textes scientifiques	3 résumés d'articles 508 mots <i>portal Elsevier</i>	3 résumés d'articles 499 mots revue <i>Sciences de l'eau</i>
textes vulgarisation	1 page web 613 mots site <i>Société canadienne du cancer</i>	1 page web 425 mots site <i>Lenntech</i>

# Expérimentation de la méthodologie

- Impressions des traducteurs

*En gros, 75% des mots techniques ne figurent pas dans le glossaire Metricc, et sur les 25% restants, 99% ont entre 10 et 20 traductions candidates, mais aucune de validée. Du coup, dans le meilleur des cas on est "à peu près sûr", mais jamais totalement. Et dans le pire des cas (très fréquemment, malheureusement) on y va "à l'instinct"*

- traducteurs pas préparés
- problème d'adéquation du lexique



# Adéquation du lexique aux textes à traduire

- Estimation avec ratio :

$$\frac{\textit{mots lexique} \cup \textit{mots textes à traduire}}{\textit{mots textes à traduire}}$$

	<b>cancer du sein 400k mots</b>	<b>sciences de l'eau 2M mots</b>
Langue source – EN	0,94	0,14
Langue cible – FR	0,67	0,78

- mieux vaut une thématique précise qu'un gros corpus



# Accord inter-annotateur

- Mesure de l'accord

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

P(A) : accord observé

P(E) : accord dû au hasard



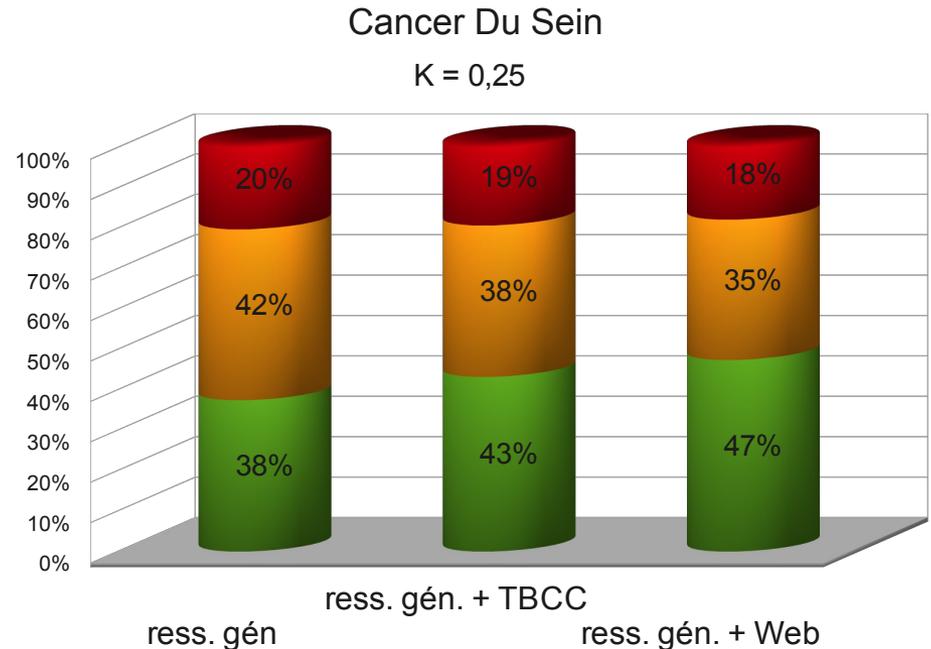
# Accord inter-annotateur

- Mode d'évaluation
  - classement : 0,65
  - jugement : 0,36
- Type de textes
  - vulgarisés : 0,57
  - scientifiques 0,48



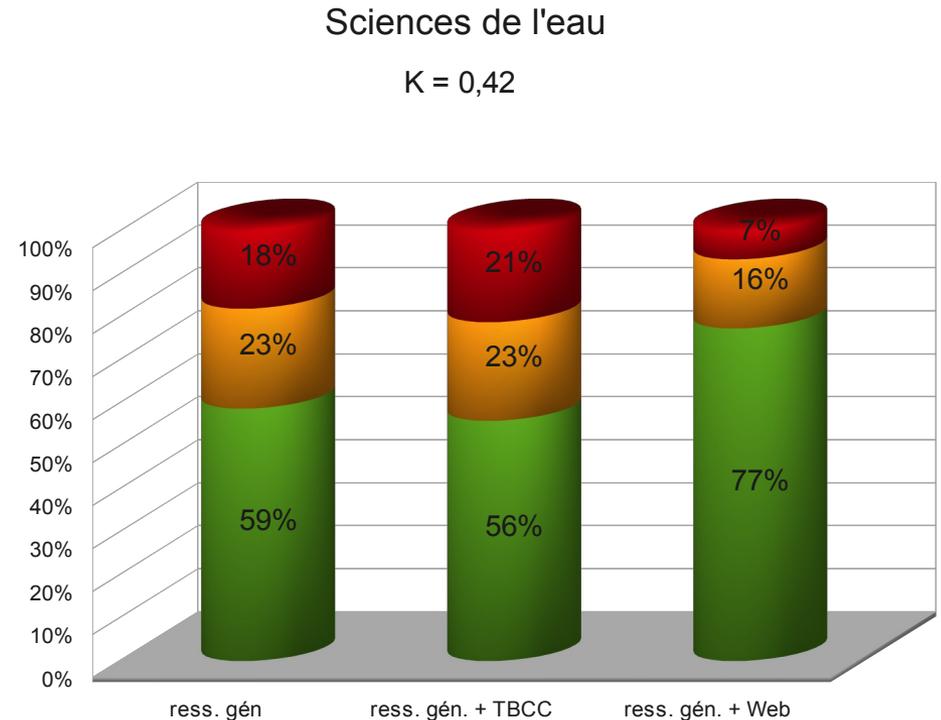
# Jugement de qualité

- Cancer du sein
  - proportion de traductions *fausses* équivalente
  - traductions sont jugées plus *exactes* avec le Web, puis avec Corpus Comp.



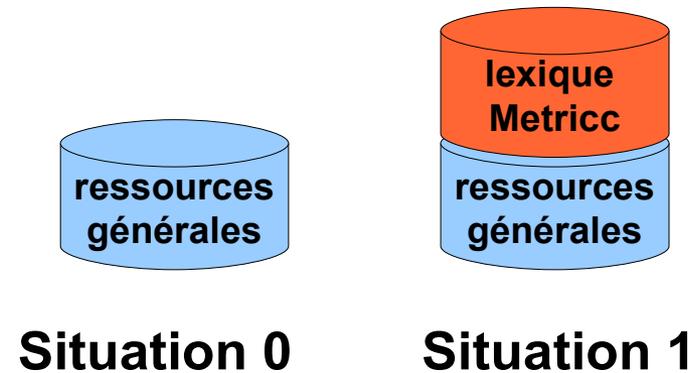
# Jugement de qualité

- Sciences de l'eau
  - traductions nettement mieux jugées pour le Web
  - Corpus Comp. produit des traductions de moins bonne qualité



# Problème dans l'utilisation des ressources

- les traductions de la situation 1 moins bonnes que celles de la situation 0
- Or disposent du même socle de ressources générales



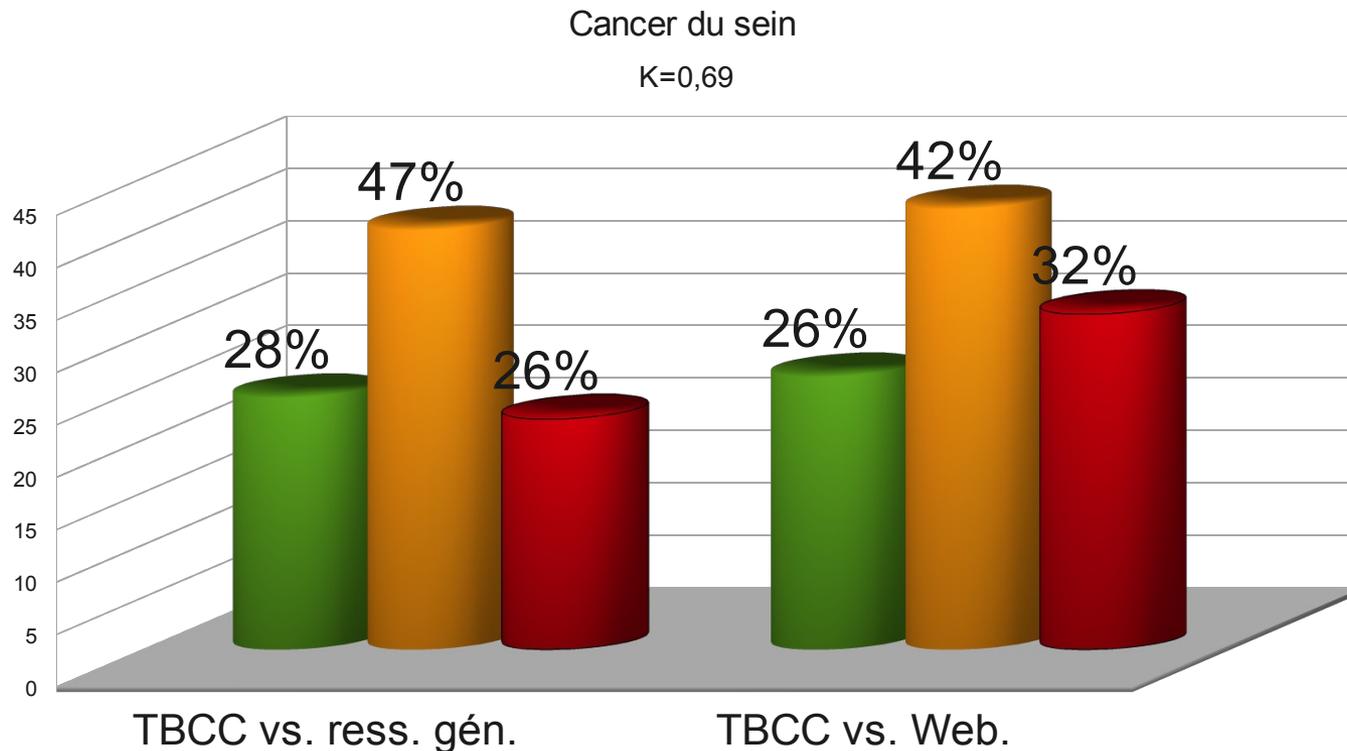
# Problème dans l'utilisation des ressources

- Les traducteurs des situations 1 et 2 ont très peu utilisé les ressources générales

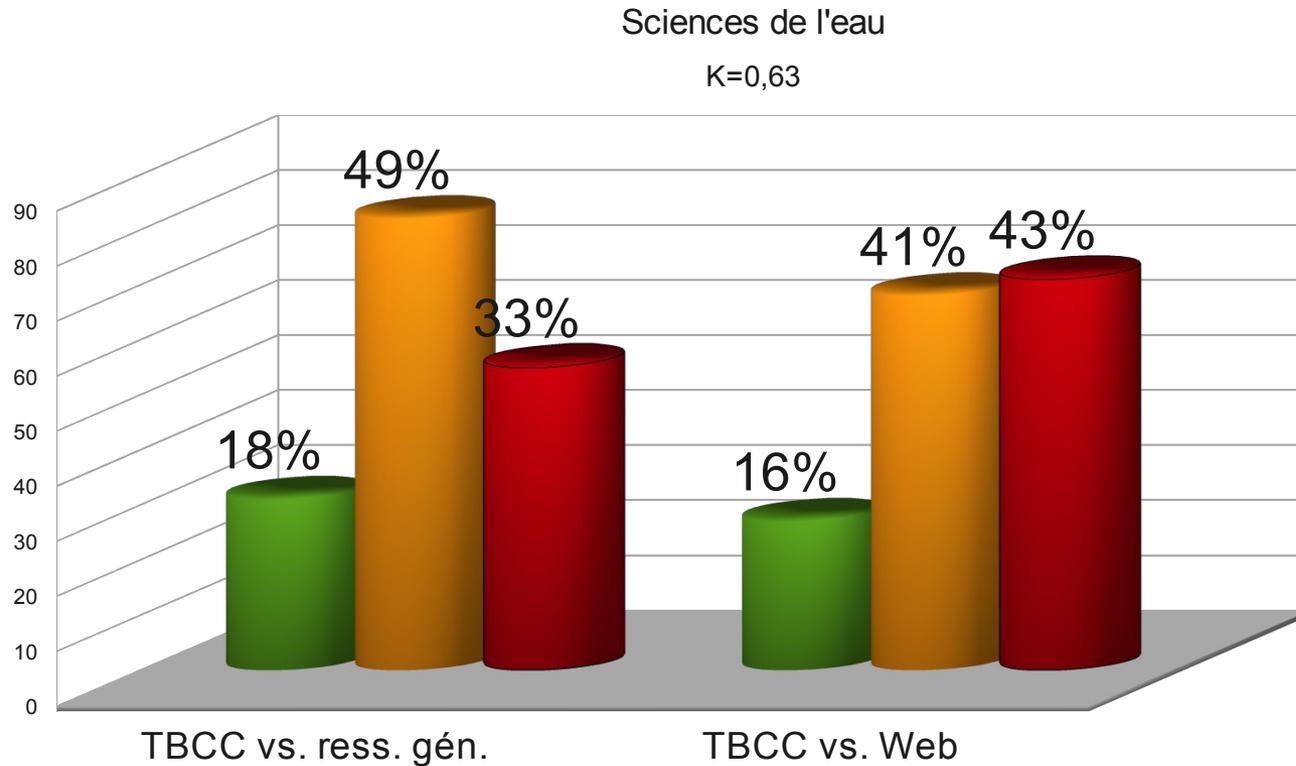
	S0	S1	S2
Ress. Générale	43%	14%	3%
Ress. Spécialisée	-	25%	56%
Intuition	79%	77%	44%



# Classement des traductions



# Classement des traductions



# Plan

1. Contexte, problématiques
2. Protocole d'évaluation
3. Expérimentation et résultats
4. **Conclusions**



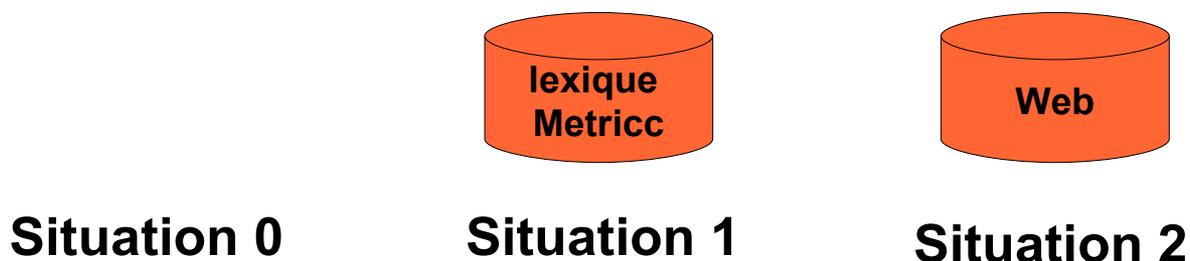
# Conclusion sur la méthodologie

- dépendance entre :
  - apport du lexique bilingue
  - adéquation lexique/textes à traduire
- toute mesure de l'apport doit aussi indiquer l'adéquation entre lexique et textes à traduire
- comment mesurer cette adéquation ? faut-il une adéquation minimale ?



# Conclusion sur la méthodologie

- Problèmes dans l'utilisation des ressources
  - mieux préparer les traducteurs
  - revoir les situations de traductions :



# Conclusion sur la méthodologie

- Faire une traduction « de chauffe » ?
- Idéalement, obtenir plus de données :
  - plus de textes à traduire
  - plus de domaines
  - plus de traducteurs
- Juges distincts des traducteurs !



# Conclusion sur les lexiques

- Utiles si construits à partir d'un corpus à thématique « fine »
- Web plus vaste, contiendra toujours plus de solutions de traductions
- L'idéal :
  - associer le lexique à des routines qui traduisent à la volée et/ou vont chercher les traductions sur le Web



# Remerciements

Ce travail a été financé par :

- ANR, subvention n° ANR-08-CORD-009
- Lingua et Machina, [www.lingua-et-machina.com](http://www.lingua-et-machina.com)

Annotateurs :

- Clémence De Baudus
- Mathieu Delage

