

# Errors in Human and Machine Translation

Natalie Kübler (EILA/Paris Diderot), Guillaume Wisniewski (LIMSI/Paris Sud) & François Yvon (LIMSI/CNRS)

# TransRead: Bilingual Reading



le cnam



CONTINT  
2012-2015



Improved visualization devices for bilingual documents

- Augmented electronic books
- Enhanced translation checking environments

# Automatic detection of Human Translation Errors

Vast literature on **MT Errors** and **MT evaluation**

*« more has been written about machine translation evaluation than about machine translation itself » (Yorick Wilks)*

What about **Human Errors** ?

- Do they actually exist ? [Short answer: yes they do !]
- Can we hope to automatically spot them ?

# Assessing Humans Translation Errors

Basic approach: look at what humans do

Study (human) error classifications & their use

Long term : use these also for MT evaluation [getting ready for HQMT]

What is human and what is not ?

Most human translator use machines (word processors, spell checkers, translation memories, pre-translations): take « human » with a grain of salt...

# Outline

---

- ① Human error analysis
- ② Sources of inspirations
- ③ Towards spotting human errors

# Human Translation Quality Assessment

- Pragmatic translation assessment: readership's reception of the translation (Newmark 1988/133) focussing on the perlocutionary effect (Hickey 1998) of the translation in the TL
- Professional and educational
- No universal framework: TQA is context-dependent
- Most efficient way => categorization of errors

# Why are typologies useful?

- Obviously, providing a framework to assess translation
- Time gaining
- Legal matters: avoid « serious » errors
- Helping students understand the evaluation
- Helping students improve their translation skills

# Some models: ATA, ITI

- ATA: quality and certifying device
  - 22 error types (terminology, register, diacritical marks)
  - Weight: 1 to 16, BUT, no guidelines how to evaluate
- ITI: quality
  - 18 error types
  - No weighting



# Some models: SAEJ2450 (2001)

SAE J2450 Translation Quality Metric

Main Category:	(abb.)	Sub-Category: (abbreviation)	Weight: serious/minor
Wrong Term	(WT)	serious (s)	5/2
Syntactic Error	(SE)	minor (m)	4/2
Omission	(OM)		4/2
Word Structure or Agreement Error	(SA)		4/2
Misspelling	(SP)		3/1
Punctuation Error	(PE)		2/1
Miscellaneous Error	(ME)		3/1

# Some models: Blackjack (Eckersley 2002)

- 21 error categories with a weight on each
- Some refer to the source text => looking for the origin of the error
- Hierarchy, style and content, easy to use => successful tool
- => important to know the origin in case the source text is wrong

# Some Models: EU contractors

## Linguistic recommendations:

- the delivered target text is complete (no omissions nor additions are permitted);
- the target text is a faithful, accurate and consistent translation of the source text;
- the terminology and lexis used are consistent throughout the text and with any relevant reference material;
- sufficient attention has been paid to the clarity and register of the target text;
- the target text contains no syntactical, spelling, punctuation, typographical or other grammatical errors;

# The MeLLANGE project (2007)

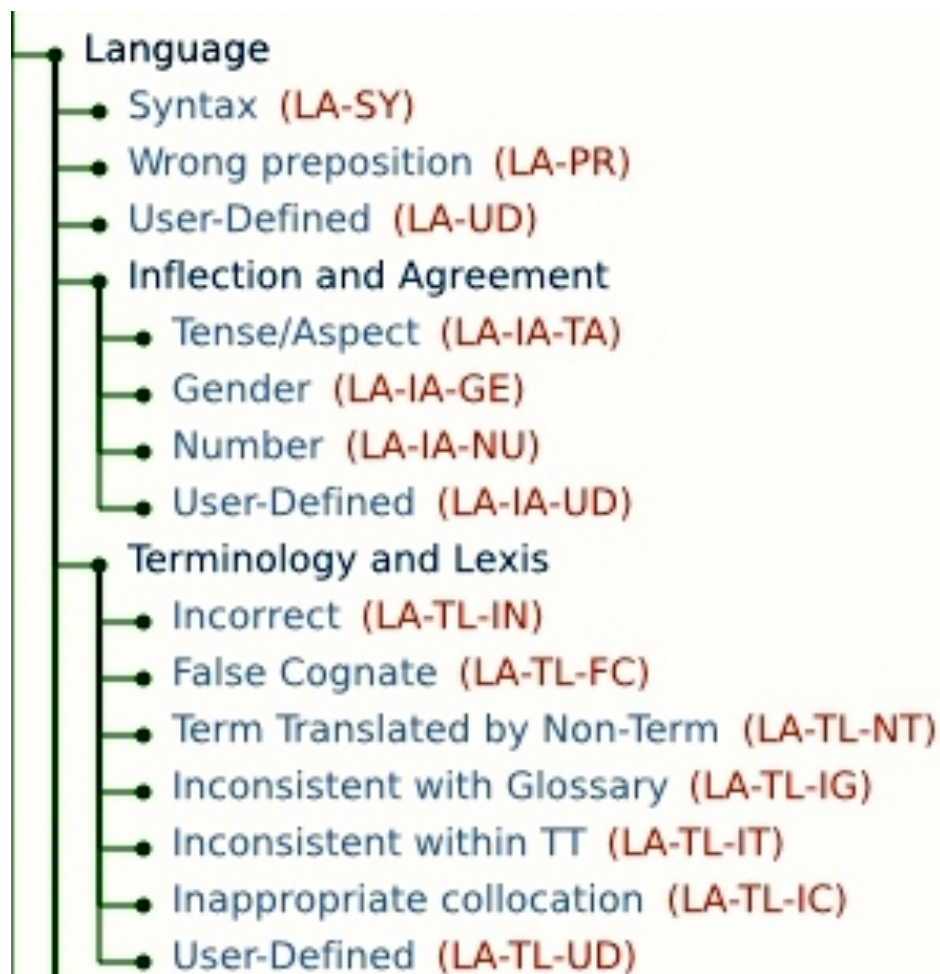
- to adapt “vocational training for translators and other language professionals to meet the new needs that flow from globalisation and increased cross-cultural communication: an increased need for the management of intercultural language resources »
- LTC: Learner translator Corpus
- Manual annotation of errors

# MeLLANGE Typology: Content

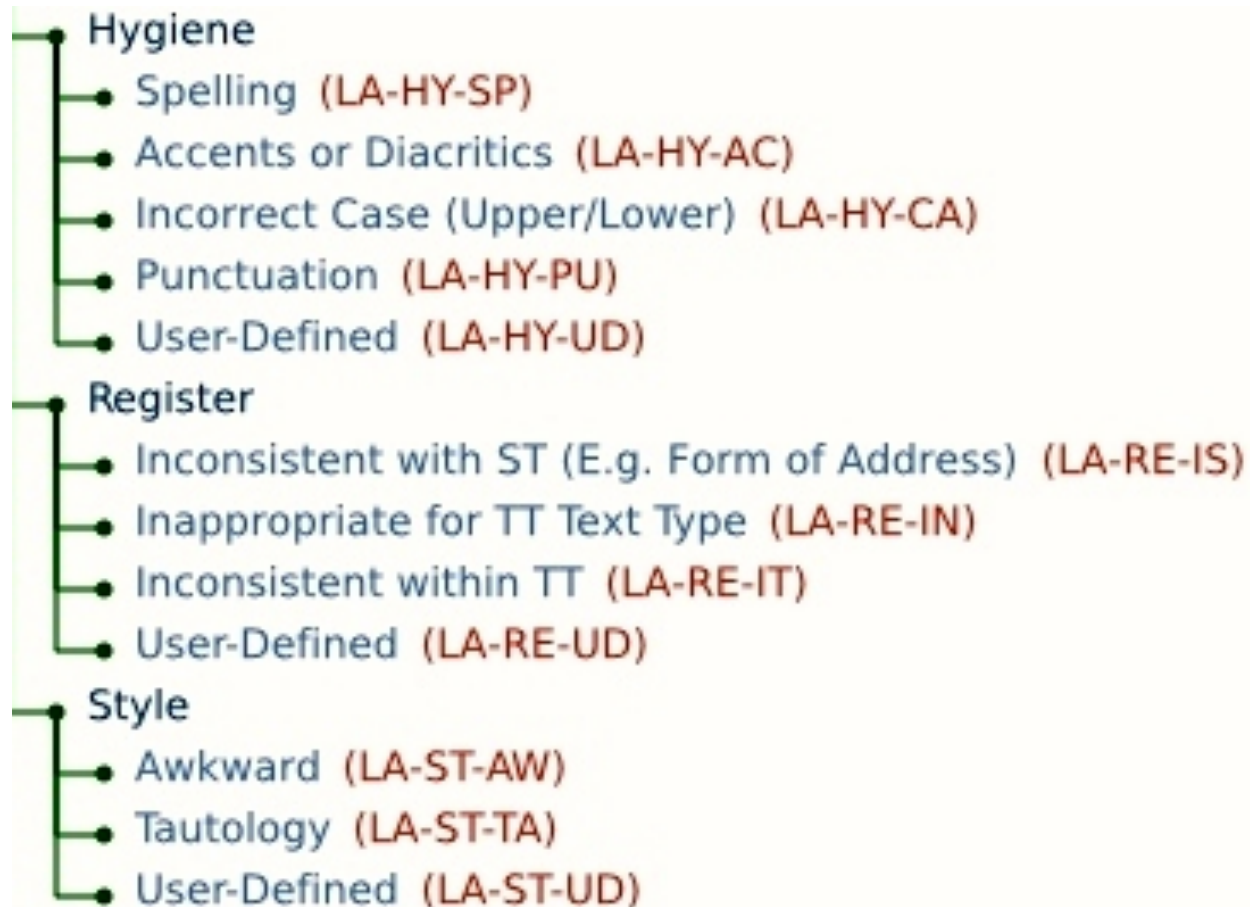




# MeLLANGE Typology: Language



# MeLLANGE Typology: Language (continued)



# Content and language errors

- Errors TR-DI: distortion

En He is entitled to the same social and **tax benefits** as national workers

Fr [Il a droit aux]la-st-aw mêmes bénéfices sociaux et aux mêmes [réductions]tr-di fiscales que les travailleurs nationaux .

Ref Il y bénéficie des mêmes **avantages sociaux et fiscaux** que les travailleurs nationaux.

- ErrorsLA-TL-IN: language error, incorrect term in TT

En The examination may be carried out either by one of the organisation's **medical officers** or by a medical practitioner chosen by the person concerned.

Fr L' examen peut être effectué soit par un [agent médical]la-tl-in de l' organisation , soit par un médecin généraliste choisi par la personne concernée

REF Cet examen peut être effectué, soit auprès de l'un des **médecins-conseil** de l'Institution, soit auprès d'un médecin au choix de l'intéressé(e).



# Legal text

**En** He may be excluded from the management of bodies under public law and from the exercise of an office **under public law**.

**Fr** [Il] **la-hy-pu** peut être exclu de la gestion d' [organismes] **tr-om** et [être démis de ses fonctions] **au nom du droit public] tr-di**.

**REF** il peut être exclu de la participation à la gestion d'organismes de droit public et de l'exercice d'une fonction **de droit public**.

---

**En** Any national of a Member State is entitled to take up and engage in gainful employment on the territory of another Member State in conformity with the relevant regulations applicable to **national** workers.

**Fr** Tout [citoyen] **la-tl-in** d' un État membre a le droit d' exercer et de [conserver] **tr-di** un [emploi rémunéré] **la-tl-ig** sur le territoire d' un autre État membre conformément [aux [règlementations] **la-ia-nu** ] **la-hy-ac** en vigueur s' appliquant aux travailleurs [[**locaux] tr-di ] la-tl-in**

**REF** Tout ressortissant d'un État membre a le droit d'accéder à une activité salariée et de l'exercer sur le territoire d'un autre État membre, conformément à la réglementation nationale pertinente applicable aux travailleurs **nationaux**.

# Technical text

**En** system must make it easy to attach extension work surfaces (horizontal extension) and a structure supporting the superstructure (vertical extension); it shall also allow computer peripherals to be attached on/under the work surfaces or laterally, to the cantilever legs,

**Fr** le système doit rendre simple (TR-SI-TL: too literal) (TR-OM:omission)l'ajout (LA-TL-IN) aux surfaces de travail (LA-TL-NT: term translated as non-term) d'une rallonge, horizontale d'une part, et verticale d'autre part (structure supportant **la structure dans son ensemble** (TR-DI)) ; il doit également permettre le rattachement (LA-TL-IT: incoherence with TL) de périphériques (TR-OM) aux pieds en porte-à-faux (LA-TL-NT), que ce soit par le dessus, le dessous ou même sur le côté,

**Ref** le système permet la fixation aisée de plateaux annexes (extension horizontale) et d'une structure supportant les éléments du 3e niveau (extension verticale). Il permet aussi la fixation de périphériques informatiques sur/sous les plateaux ou latéralement aux piétements,

# Fuzzy category?

**En** Nevertheless a small scale test project **might be justified** given the current bleaching crisis.

**Fr ? Il pourrait cependant être justifié** de mener un projet de test à petite échelle, **au vu de la crise actuelle du blanchissement des coraux.**

**Ref Il serait cependant utile** de mener une évaluation à petite échelle, au vu de la **gravité** de la situation actuelle.

# Another fuzzy one

Omission ?

**En ...no quantitative estimate exists  
of the amount of these gases =>**

**Fr il n'existe aucune estimation  
\_\_\_\_\_ de la quantité de gaz...**

# Omission/ Addition

**En** In this paper we will **(1)** discuss functional requirements for the biorobotic fin and fin sensorimotor system

**Fr** Dans cet article nous déterminerons **tout d'abord** les caractéristiques techniques d'une nageoire biorobotique

# Summary

Human errors:

- ① From obvious to subtle error types
- ② Errors can be precisely localized
- ③ Spotting error does not require a reference translation
- ④ Various levels of seriousness

How much hope can we have ?

# Practice of spell checking

Modern spell checkers are **data-based**:

① “Language Models”:  $P(\text{right}_1 \dots \text{right}_n)$

n-gram model (local context), stochastic grammars (sentence level), topic models (long range, semantic)

② “Error Model”:  $P(\text{wrong} | \text{right})$

frequent letter confusions, phonetic similarity, homophones (trained on genre / register / style)

Fully automatic spell checking is difficult: long range dependencies, balance False Positive vs. False Negative, etc.

# Practice of spell checking

Translation checkers could be data based

① “Language Models” - OK

② “Error Model”: ??

Frequent learner errors (“faux amis”, syntactic analogies, etc) ?

Non literal translations (idioms & terms) ?

Would have to be even more resourceful: domain / register/  
genre / dependent



# Practice of Quality Estimation

- ① QE systems are **data-based** (cf. Lucia's talk)
- ② Annotated corpus:  
automatic translations ( $\mathbf{e}, \mathbf{f}$ ) + quality measures ( $y$ ) [BLEU, hTER, PET time, etc]
- ③ Typical spans : sentences (rare: words, very rare: document)
- ④ Feature extraction:  $(\mathbf{e}, \mathbf{f}) \rightarrow \Phi(\mathbf{e}, \mathbf{f})$
- ⑤ Machine learning:  $\hat{y} = f_{\theta}(\Phi(\mathbf{e}, \mathbf{f}))$

# Towards Human Quality Estimation ?

- ① HQE systems could be **data-based**
- ② Annotated corpus:  
human translations  $(\mathbf{e}, \mathbf{f})$  + human errors  $(y)$
- ③ Candidate spans : word, group, sentences
- ④ Feature extraction:  $(\mathbf{e}, \mathbf{f}) \rightarrow \Phi(\mathbf{e}, \mathbf{f})$
- ⑤ Machine learning:  $\hat{y} = f_{\theta}(\Phi(\mathbf{e}, \mathbf{f}))$

Could this actually work (provided appropriate data) ?

# Comparing translations

- Corpus TRACE of post-edited translations<sup>1</sup>
- 3 types of translations
  - Fully automatic translations (RBMT, SMT)
  - Post-edited versions
  - “Free” translations
- ① How different are these translations?
- ② Can they be automatically separated ?

---

<sup>1</sup>Wisniewski et al, MT Summit 2013.

# Comparing translations

- Task: discriminate two kinds of translations  $\Rightarrow$  binary decisions
- Typical QE features : language models, IBM1 models, POS distribution, length ratio, etc.
- Main Results:
  - Harder to distinguish MT from post-edited MT (score on test : 59%)
  - Than to distinguish MT from free references (66%) and [free references from post-editions \(67%\)](#)

So there may be some hope...

# On going activities

- On going experiments with MeLLange Corpus
- Data collection (post-edited human translations):
  - Cochrane reviews
  - Scientific texts
  - Translation learner corpus

# References

- Castagnoli, Sara, Ciobanu, Dragos, Kübler, Natalie, Kunz, Kerstin, Volanschi, Alexandra. 2011. Designing a Learner Translator Corpus for Training Purposes. In N. Kübler. (ed) *Corpora, Language, Teaching, and Resources : From Theory to Practice*. Bern: Peter Lang. 221-248.
- Direction Générale de la Traduction, Union Européenne : checklist for outgoing translations: [http://ec.europa.eu/translation/french/guidelines/fr\\_guidelines\\_fr.htm](http://ec.europa.eu/translation/french/guidelines/fr_guidelines_fr.htm) (consulté le 18/11/13)
- Eckersley, Helen. 2002. Systems for Evaluating Translation Quality. In *Multilingual Computing and Technology*, 3(3), 39-42
- Hickey, L. 1998. Perlocutionary equivalence: marking, exegesis and recontextualisation, in L. Hickey (ed.) *The Pragmatics of Translation*, Clevedon: Multilingual Matters.
- Kübler, Natalie & Aston, Guy. 2010. Using Corpora in Translation. In M. Mc Carthy & A. O'Keefe (eds) *The Routledge Handbook of Corpus Linguistics*, London: Routledge. 505-515
- Newmark, P. 1988. Pragmatic translation and literalism. *TTR: Traduction, Terminologie, Rédaction* 1(2): 133-145.
- Secară, Alina. 2005. Translation Evaluation – a State of the Art Survey. Proc. of the eCoLoRe/MeLLANGE Workshop, Leeds, UK.