

# Moving Beyond CAT tools - The MT Paradigm Shift from the Translators' Perspective

Tomáš Svoboda



Charles University  
Prague



# MT Paradigm Shift

Comparison of TM vs. MT in four areas:

1. System
  2. Process
  3. Product
  4. User's View
- 



# MT Paradigm Shift

## 1. System



	CAT	MT / PEMT
History	90s	2010
Development	Mature	Developing (Google Time Loop), "Point of Degradation"
Future	Less important	Big demand
Paradigmatic	No	Yes
Proliferation	Professionals	Ubiquitous
Technology	One: TM	Various (EBMT, RBMT, SMT, hybrid...)
Supported language pairs	Democratic	Uneven / Non-democratic
Data retrieval	Off-line / On-line	Off-line / On-line
Cost	0 ≥	0 ≥

	CAT	MT / PEMT
Confidentiality	Little difficulty (File based / Server based)	Complex issue (Server / generally accessible)
Speed	Reasonable (Time-consuming: converting the bilingual file, Import/Export, Re-Organising)	Pre-processing: "Low" ( $\alpha$ -version of MyMemory: 1,000 wpm) "Interactive" mode: There can be longer response times
Usability	Restricted (text type, domain, repetitiveness), narrower	Restricted (text type, domain), much wider
Customisability	Partial (Match rate threshold, Penalties)	Yes/No (depending on the engine)

	CAT	MT / PEMT
<b>Customisability</b>	Partial (Match rate threshold, Penalties)	Yes/No (depending on the engine)
<b>Maintainance</b>	Yes	No
<b>“Intelligence”/Training</b>	Static: Memory-feature	Dynamic/real-time: MT training
<b>Evaluation</b>	No	Yes (Metrics)
<b>Harmonisation</b>	TM Exchange (*.tmx)	PEMT Guidelines
<b>Tertium Comparationis</b>	No (i.e. is direct)	Yes/No (depending on the engine)



# MT Paradigm Shift


1. System
  2. Process
- 

	CAT	MT / PEMT
<b>Where in the process</b>	Help in the decision-making process	Generating text
<b>On-Off</b>	Yes	Yes
<b>Eliminating redundancy?</b>	Yes	No (translates from scratch)
<b>Assigning penalties</b>	Yes	Yes
<b>Quality of original (Authoring)</b>	Non-Key If there are matches: Can be bridged by fuzzy logic	Key Eliminates typos in Original
<b>Productivity</b>	Varying, increase	Varying, increase
<b>(Re)Search Feature</b>	Limited to available material	Includes (re)search (GT)
<b>Concordance search</b>	Yes	Yes





# MT Paradigm Shift

1. System
  2. Process
  3. Product
- 

	CAT	MT / PEMT
Quality	Varying (zero/low match)	Varying (defective language)
Quality Evaluation	Match rate	e.g. BLUE-Score
Context match	Yes	No
Subject matter	Not relevant	Relevant (depending on corpus)
Text type	Very relevant	Not relevant
Terminology	Consistent	Not consistent
Phrases	Consistent	Not consistent, however idiomatic
Idiomatic expressions	-	Yes
Can become counterproductive	Yes	Yes



# MT Paradigm Shift

1. System
  2. Process
  3. Product
  4. User's View
- 

	CAT	MT / PEMT
MMI	Complex	Limited features
User-friendliness	No	Limited
Psychology	Few objections	"Rage against the machine"
Skill degrading potential ("Verlernen"); addiction rate	Low	Very high
Ergonomics	Less harmful	More harmful
Source of Inspiration	Rarely	Yes, at times
Wow factor	Limited (negative)	Yes

# User's View

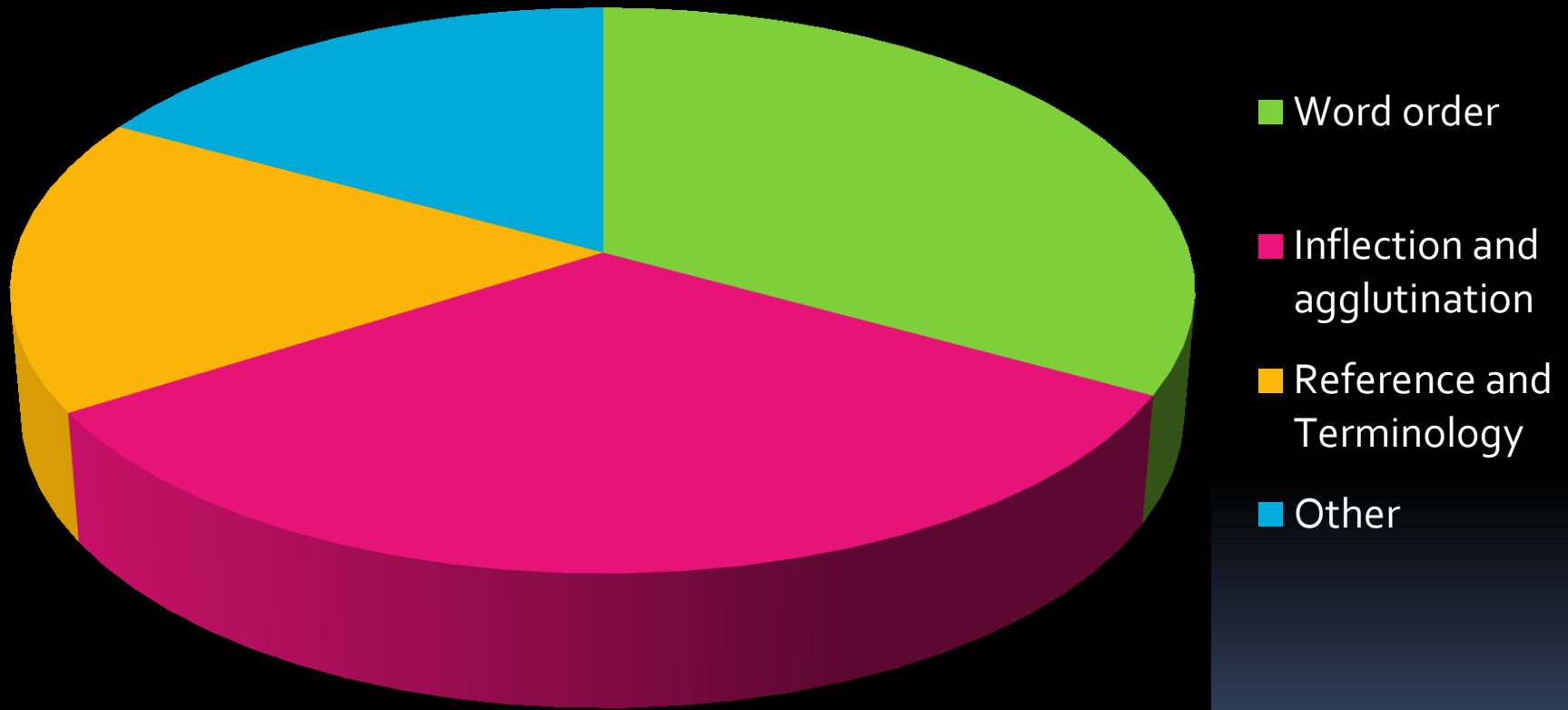
## DGT SMT (mid 2011)

- 60+ linguists
- 9 documents
- EN-X, all EU official languages
- 16,000+ individual judgements
- Usability

# Outcomes

- Romance languages (ES, FR, IT, PT, RO): most optimistic
- Germanic (DA, DE, NL, SV): relatively convinced, except for DE
- Slavic (BG, CS, PL, SK, SL): Less convinced yet, except for BG
- Hellenic (EL), Semitic (MT), Celtic (GA), Baltic (LT, LV): Less convinced yet
- Finno-Ugric (ET, FI, HU): Badly served by the SMT technology

# Most complex issues



# User's View

- Special characters badly treated
- Capital letters badly treated
- Placables badly treated
- Numbers changed
- Punctuation wrong
- Sub-sentences mismatched/misplaced
- Words  
misplaced/missing/added/untranslated



# User's View

- Words (prefixes, suffixes)/vocabulary wrong
- Syntax wrong
- Congruence (subject – verb) wrong
- Grammatical word classes misused
- Compounds wrong
- Unconventional synonyms
- Inconsistent terminology
- Reference documents not respected



# Beyond CAT tools

**Past: “Archaic” client-server solutions**

**Today: Taking the best of both technologies: MT-assisted TM**



**Verge: Match threshold, up to which TM matches are preferred to MT (85%, 75%, 70%)**



# Beyond CAT tools

**Future: TM-assisted MT**

**TMT Prime (CNGL) – a Recommender System for TM and MT Integration**





# Sources

**Champollion, Y.** (2003). Convergence in CAT: Blending MT, TM, OCR & SR to boost productivity. *Proceedings of the International Conference Translating and the Computer 25*, 20-21 November 2003, London. London: Aslib.

**Guerberof Arenas, Ana** (2008). *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation*. Masters Dissertation. Universitat Rovira i Virgili

**Lagoudaki, Elina** (2008). "The value of Machine Translation for the professional translator." 8th AMTA Conference, Hawaii, 262-269



# Sources

**Languages and Translation, No 6, 02/2013, DGT**

**Fiederer, Rebecca and Sharon O'Brien. 2009.** Quality and Machine Translation - A Realistic Objective?. *Journal Of Specialised Translation*, 11,

**Sharon O'Brien. 2011.** Towards Predicting Post-Editing Productivity. *Machine Translation*, 25, 3, pp197-215

**Translation Pricing and PEMT Process Management,**  
Highlights from ELIA Munich ND