



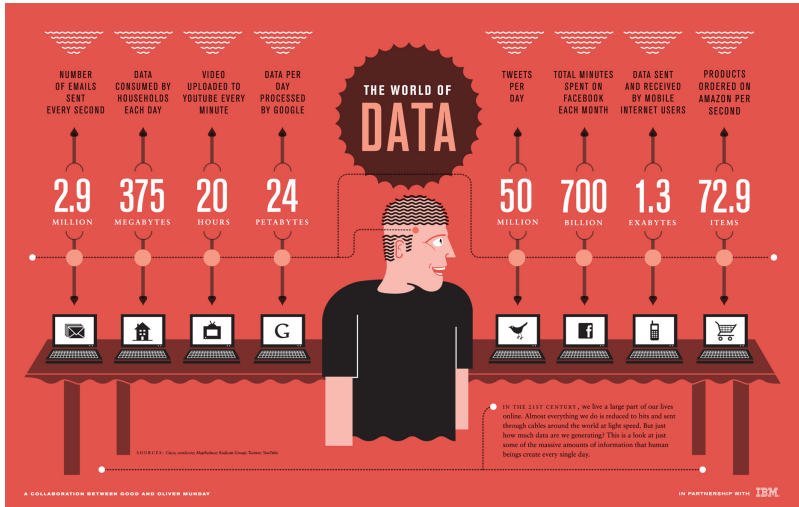
Traditional and Emerging Use-Cases for Machine Translation

Andy Way

Lingo24

`andy.way@lingo24.com`

Too much data to translate . . .



Not enough translators ...



Not enough translators ...



... so not enough stuff gets translated!

MT is the (only) answer!



MT is the (only) answer!



Not everyone agrees . . .



MT is being used every day . . .



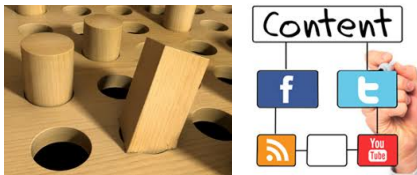
- Provides a billion translations a day for 200 million users,
- 92% of the usage is from people outside the US,
- Amount of text translated daily is more than what's in a million books,
- Surpasses what professional translators handle in a full year.

Lingo24's customised engines



- Improve productivity,
- Translate content previously not feasible due to time or cost constraints,
- Reduce time to market, and
- Reduce translation costs.

Use-cases involving MT



Need to take into account:

- *Fitness for purpose of translations*: amount of human involvement required – or warranted – for a particular task will depend on the purpose, value and shelf-life of the content.
- *Perishability of content*: in all cases, the degree of post-editing or human input should be clearly correlated with the content lifespan.

Traditional Uses of MT



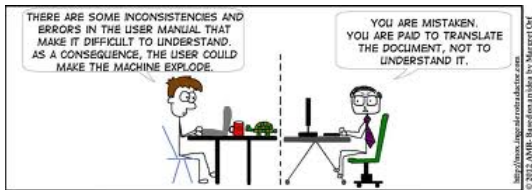
- Raw MT,
- MT with light post-editing,
- MT with full post-editing.

Light post-editing



- Review and post-edit of MT by a native linguist to make the output an understandable reflection of the source text content, but ignoring stylistic niceties.
- Essentially this is a fast service to ensure that the translation is correct and fit for purpose.

Full post-editing



- Review and post-edit of MT by a native linguist to produce a text that is not only understandable, but also presented in a stylistically appropriate way.
- This offering should produce output comparable to human translation quality.

Raw MT across vertical sectors



- Internal communication ('assimilation'): translation of emails, online chat, international communication across hubs, FAQs, repetitive product descriptions (e.g. listings),
- Website translation: rapid translation of critical updates (and gisting),
- Bids/Tenders: translation for gisting purposes.

PEMT-light across vertical sectors



- Online Help,
- Knowledge Forums,
- Support Documentation.

PEMT-full across vertical sectors



- External communication ('dissemination'),
- Sensitive documentation translation, especially where Security and Health & Safety are important,
- Client-facing documentation translation.

Use-Cases for Raw MT

Industrial Sector	Raw MT Use-Cases
Technology	UGC Multilingual Search
Manufacturing	—
Finance/Legal	Real-time translation of stock reports (gisting) Forensic investigation (gisting)
Marketing	Sentiment analysis Responses to surveys & questionnaires
E-Commerce	Basic product descriptions (cf. eBay)

PEMT-light Use-Cases

Industrial Sector	PEMT-light Use-Cases
Technology	Manuals (with few security or health & safety risks) Online help and product support
Manufacturing	Manuals (with few security or health & safety risks) Online help and product support
Marketing	Market research: for information, or basic understanding only

PEMT-light Use-Cases

Industrial Sector	PEMT-light Use-Cases
Technology	Manuals (with few security or health & safety risks) Online help and product support
Manufacturing	Manuals (with few security or health & safety risks) Online help and product support
Marketing	Market research: for information, or basic understanding only

- Finance, Legal, E-Commerce: more likely to require *full* post-edit

PEMT-full Use-Cases

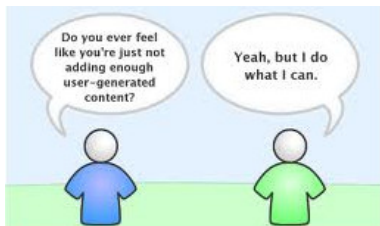
Industrial Sector	PEMT-full Use-Cases
Technology	Manuals (security/health & safety to be considered)
Manufacturing	Manuals (security/health & safety to be considered)
Finance/Legal	Contracts/Patents Reports where certification needed
Marketing	—
E-Commerce	Detailed product information (with Marketing spin)

Which level of PEMT do clients choose?



- 'Light' post-editing: one-off, large-scale jobs,
- 'Full' post-editing: more typical requirement where a long-term partnership exists between client and LSP.

Emerging Use-Case: User-Generated Content



- Online chat, tweets, blogs, hotel or product reviews, social media posts etc.,
- *Extremely* perishable content,
- No budget for human translation of such data, but tremendous added value if available multilingually on company websites.

Translating UGC is hard!

Much of the source-language data is of 'poor' or at least 'non-standard' quality. This comes in many forms:

- ① content produced by non-native speakers,
- ② content produced by native speakers containing non-deliberate typos,
- ③ content produced by native speakers which deliberately departs from spelling norms to bring about some linguistic effect.

Other emerging use-cases



- Verification of potential user demand on multilingual website versions,
- Translation of course syllabi documentation and other educational information (cf. *Bologna FP7* project).

The changing role of the translator



Given the challenges they face in their day-to-day work, most human translators today would acknowledge the critical role of technology in their workflow. However, it is fair to say that some of this technology is more highly regarded than others. For example, most translators are happy to use Translation Memory tools (Heyn, 1998), while Machine Translation has met with much less widespread acceptance to date. (Bota et al., 2013:313)

Translators love TM . . .

- . . . at least they do *now*! Is that justified?
- As a productivity tool it's limited, certainly compared to the potential gain with MT,
- Assume that TM facilitates the translation of maybe 10—20% of a new document, leaving the rest to be translated by the translators,
- (Leave aside for now that translators get paid very little (or nothing at all!) for exact matches, ICE matches, and high fuzzy matches coming from *the(ir)* TMs . . .)
- That's not a high quality threshold by any measures!

Translators love TM . . .

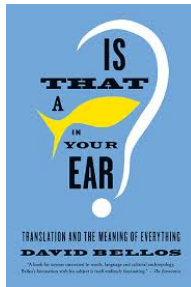
- . . . at least they do *now*! Is that justified?
- As a productivity tool it's limited, certainly compared to the potential gain with MT,
- Assume that TM facilitates the translation of maybe 10—20% of a new document, leaving the rest to be translated by the translators,
- (Leave aside for now that translators get paid very little (or nothing at all!) for exact matches, ICE matches, and high fuzzy matches coming from *the(ir)* TMs . . .)
- That's not a high quality threshold by any measures!

Why, then, do translators love TM?



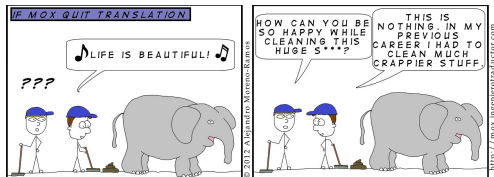
- They trust it,
- They have confidence in it,
- They've put investment into it,
- It's predictable: if you ask for all matches above a 75% fuzzy match, that's what you'll get.

(Some) translators hate MT,
but also (some) translators!



Translation commentators lead the field in throwing most of its work in the direction of the garbage dump . . . it seems implausible that anyone would ever make such a statement about any other human skill or trade. (Bellos, 2011: 329)

Why are translators so nasty to us MT developers?



To me, this was a moment of enlightenment in the book, although probably not one intended (and certainly not mentioned) by Bellos: at last, all translators (or at any rate, those less enlightened than Bellos) have something else to pick on, namely MT! They are so inured to this level of talking about translation, that they naturally use it against us. (Way, 2012: 260—261)

Not all translators hate MT!

- ① “stop seeing technology as our enemy and rather start embracing it”,
- ② “becom[e] co-creators of the tools we will be using in the next decades”,
- ③ “understand that future death by inaction is possible”,
- ④ “make sure we remain relevant”,
- ⑤ “cannot continue hiding in the sand and think that just because we do not want it, it will not happen”,
- ⑥ “face the scary challenge posed by progress and run to catch up for the decades we have been complacent while the rest of the industry became digital, mobile and instant”. (*Claudia Brauer*)

Not all translators hate MT!

- ① “acute need for translators to educate themselves much more thoroughly about what MT is all about and where and how it can be applied”,
- ② “if we reach out, I am certain we will receive a warm welcome”,
- ③ “greatest impediment to translator involvement is a mixture of ignorance (‘MT is just Google Translate’) and fear (‘The machine will eventually replace me, so I will stick my head in the sand’)”,
- ④ “If we can change our attitude to ‘This is interesting, so how can I make this work for me?’ – in short, embrace change – then we are in a much better mindset to take advantage of MT”,
- ⑤ “look at MT as an opportunity to generate more income without working harder or longer in a market that is continually being squeezed on price”. (*Charlotte Brasler*)

Not all translators hate MT!

- ① “cannot stop the deluge of content that is being generated and the scarcity of translators to translate it all”,
- ② “many companies are looking into MT to save money and time and still get their material translated”,
- ③ “Humans are slow and expensive but creative; machines are fast and cheap but stupid. In between stands the quality factor”,
- ④ “That is where we as translators come in: we can affect this quality that the ‘stupid’ machines cannot figure out themselves”,
- ⑤ “use the machines to help us increase our speed and post-edit the MT output to a human translation quality level, we will all reap the financial benefits”,
- ⑥ “clients will get their content translated faster and cheaper, and we will make more money by producing more high-quality words faster. That is a win-win situation”.
(Charlotte Brasler)

Not all translators hate MT!

- ① “form an opinion about MT based on its actual merits and shortcomings, rather than on some pre-formed ideas . . . the greatest impediment to translator involvement right now”,
- ② “then we can determine whether becoming more involved in MT has any benefit to us individually”,
- ③ “already an unmet need for MT post-editors, and that need will continue to increase”,
- ④ “a lot of good opportunities for some of us, but I also think that many of us will not encounter post-editing MT for a long time to come”.
(*Jost Zetzsche*)

But understanding SMT is *hard!*



Mary Hearne & Andy Way. 2011. Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass* 5:205—226

An important thing to remember ...



I always find it helpful to remind myself that my job is not to provide machine translation but to provide translation. The application of appropriate workflows and technologies to that end is simply the smart way to go about it. Smart, forward-thinking translators and translation companies will thrive in this changing industry. And because MT is not an appropriate solution for every use-case, there will also continue to be opportunities for traditional translators. (Jay Marciano)

Observations

- MT is being used in a number of ways by many users on a daily basis.
- The point of questioning whether MT is useful or not is moot.
- Many *translators* find MT to be useful on a daily basis, but that's all it is – a tool in their armoury – and all it ever will be ...
- No threat to translators' jobs from MT, despite ongoing scaremongering from (some) translators.
- More influential translators willing to stick their heads above the parapet and sing the praises of MT.

New dynamic levels of quality

- Days of 'one size fits all' when it comes to quality are gone.
- New emerging use-cases, especially due to the huge increase in UGC.
- Increasing clamour for tools which facilitate customised levels of quality.



andy.way@lingo24.com