

*Maite Aragonés Lumeras\* & Andy Way\**

## **On the Complementarity between Human Translators and Machine Translation<sup>1</sup>**

### **Abstract**

Many translators are fearful of the impact of Machine Translation (MT) on their profession, broadly speaking, and on their livelihoods more specifically. We contend that their concern is misplaced, as human translators have a range of skills, many of which are currently – with no signs of any imminent breakthroughs on the horizon – impossible to replicate by automatic means. Nonetheless, in this paper, we will show that MT engines have considerable potential to improve translators' productivity and ensure that the output translations are more consistent. Furthermore, we will investigate what machines are good at, where they break down, and why the human is likely to remain the most critical component in the translation pipeline for many years to come.

### **1. Introduction**

Machine Translation (MT) is being used by many people on a daily basis as a productivity tool, with demonstrable success. There is already a wide variety of use-cases, but more are emerging such as real-time online face-to-face communication where MT is the only solution (Way 2013). Many of these new, emerging use-cases for raw MT and post-edited MT (PEMT) – especially involving user-generated content (UGC) – require different levels of human engagement, and different levels of quality (Penkale/Way 2013). In our view, the degree of human involvement required – or warranted – in a particular translation scenario will depend on the purpose, value and shelf-life of the content. More specifically, we assert that in all cases, the degree of post-editing or human input should be clearly correlated with the content lifespan.

More and more people are finding MT to be useful; at the last Google I/O event in May 2016,<sup>2</sup> Google stated that the average daily translation volume for Google Translate<sup>3</sup> is about 143 Billion words a day across 100 language combinations. However, it is impossible to ignore the fact that the introduction of MT into the translation pipeline has been less than wholly embraced by the entire translator community (Way 2013: 2). Nonetheless, we argue that the translation community has little to fear from MT, and that for those use-cases where human input is required, the translator clearly remains the most important link in the chain.

In this paper, we investigate the role played by translators in translation workflows today by appealing especially to the notions of “syntactic” versus “pragmatic” translation (Kay 2014). We provide a range of examples where, in contrast to human translators, MT cannot – and will not for many years to come – generate translations of the latter type. As we stated in Way/Hearne (2011), the only way for progress to be made in MT is for system developers to engage closely with lin-

---

1 The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the World Intellectual Property Organization

2 <https://events.google.com/io2016/>

3 <https://translate.google.com/>

---

\* *Maite Aragonés Lumeras*  
*World Intellectual Property Organization*  
*34, Chemin des Colombettes*  
*CH-1211 Geneva 20*  
*Switzerland*  
*maite.aragonés@wipo.int*

\* *Andy Way*  
*ADAPT Centre*  
*School of Computing*  
*Dublin City University, Dublin 9*  
*Ireland*  
*andy.way@adaptcentre.ie*

guists and translators.<sup>4</sup> Of course, this cuts both ways, and there are many means available nowadays by which translators can try out MT engines for themselves, and to try and understand the details of today's statistical paradigm (see Hearne/Way 2011).

The remainder of this paper is organised as follows. In Section 2, we summarise the state of play with respect to MT today: what are the dominant system types, what are the typical and emerging areas where MT can be successfully applied, and how translators are usually involved. In Section 3, we focus on the strengths of MT, which often are the very issues which human translators find problematic. In Section 4, we contrast this view by detailing what linguists are good at, which mostly are in direct opposition to the capabilities of today's MT engines. In Section 5, we draw these together as an attempt at converging on a way forward for the MT and translation communities to mutually benefit from the apparent synergies. In Section 6, we conclude, and present avenues for further consideration.

## 2. Machine Translation Today

It is widely accepted that MT is already providing a solution to a number of translation problems, some of which include cases where human intervention is impossible. While the recent research landscape has been dominated by statistical approaches to MT (SMT),<sup>5</sup> some of the leading MT systems available for purchase remain largely rule-based (e.g. Systran).<sup>6</sup> Even here, more and more products are coming to the market which are predominantly statistical, from companies offering customised MT engines based on the open-source Moses<sup>7</sup> toolkit (Koehn et al. 2007), to DIY self-serve platforms (e.g. KantanMT<sup>8</sup> and Microsoft's Translator Hub),<sup>9</sup> to free, online MT systems like Google Translate or Bing.<sup>10</sup>

### 2.1. Statistical Models of Translation

There are two main processes in SMT: training, and decoding (or "search"). Training takes place offline, and involves computing three main models: from parallel data, such as in the (partial) Translation Memory (TM, see Heyn 1998) in Fig. 1, a translation model (or "phrase-table"), and a reordering model; and from monolingual data, a target language model.

As Hearne/Way (2011: 205) show, in the original "noisy channel" version of SMT (Brown et al. 1990, 1993), only the translation model and language model played a role:

The translation model effectively comprises a bilingual dictionary where each possible translation for a given source word or phrase has a probability associated with it. However, the model does not resemble a conventional dictionary where plausible entries only are permitted; many of the entries represent translations that are unlikely but not impossible, and the associated probabilities reflect this. The language model comprises a database of target-language word sequences (usually ranging between 1 and 7 words in length), each of which is also associated with a probability.

4 Note the unfortunate, yet widespread, usage of 'linguists' to refer to 'human translators' in the translation industry.

5 We acknowledge the inroads being made by Neural MT, especially in academic circles, yet omit a comparison with SMT in this paper.

6 <http://www.systransoft.com/>

7 <http://www.statmt.org/moses/>

8 <http://www.kantanmt.com/>

9 <http://hub.microsofttranslator.com/>

10 <http://www.bing.com/translator/>

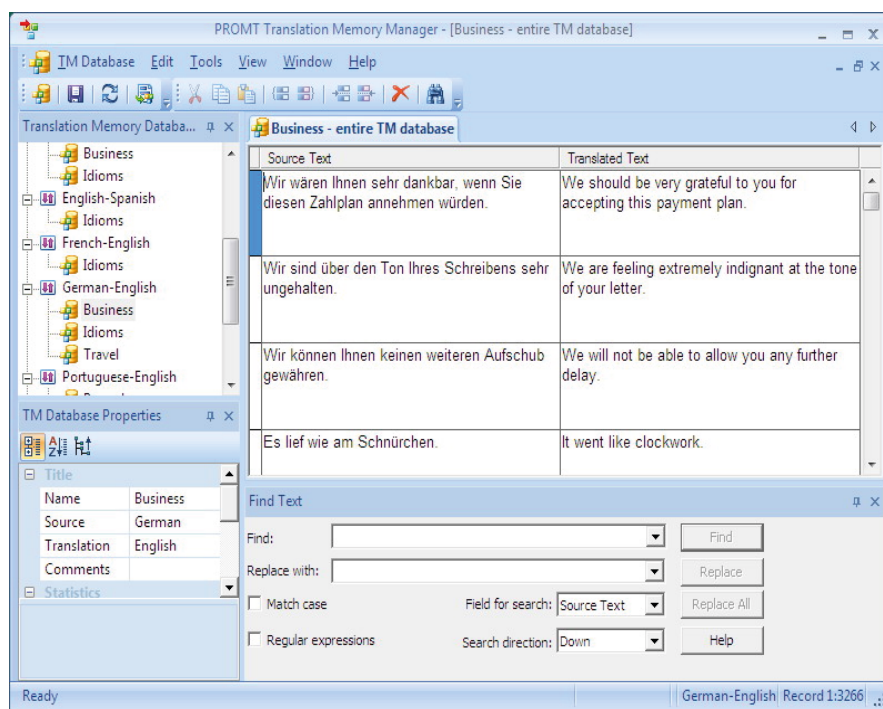


Figure 1. A Translation Memory System

In essence, the translation model suggests those target-language words and phrases which are optimal (based on the inferences computed from the parallel data, and stored in its phrase-table) for the translation of the source sentence, while it is the job of the language model to assemble this bag of target words and phrases into the best (based on the inferences computed from the monolingual data) word order it can find. These two components work hand in hand during the decoding phase so as to output the most likely translation of the source sentence from a *mathematical* point of view.<sup>11</sup> Decoding, then, essentially boils down to searching for the most likely target string given the source string from potentially millions of possible target translations. In this regard, Way/Hearne (2011: 206) make the following useful observation: “the methods used [...] are not intended to be either linguistically or cognitively plausible (just *probabilistically* plausible), and holding onto the notion that they somehow are or should be simply hinders understanding of SMT” (original emphasis).

In recent years, the noisy channel model of SMT has been surpassed by the log-linear model (Och/Ney 2002) of SMT. Here other components – including the reordering model – can be combined with the language and translation models to improve the overall translation quality obtained. Each of these components is assigned a weight in the “parameter estimation” (or “tuning”) phase so that the highest score according to a particular automatic evaluation metric (e.g. BLEU, Papineni et al. (2002)) is obtained on a held-out tuning set.

For both models of SMT, and for each of these components, we refer the interested reader to Hearne/Way (2011), and for the more intrepid to the primary sources cited above.

## 2.2. Machine Translation Usage Today

There is very little need for us to demonstrate that MT is being used widely on a daily basis, but for those who doubt this, Och (2012) notes that just by itself, Google Translate provides a billion translations a day for 200 million users, and the amount of text translated daily is more than

<sup>11</sup> Way/Hearne (2011) go on to show that the traditionally understood terms of ‘adequacy’ and ‘fluency’ can be applied to the translation and language models, respectively.

what's in a million books. As he says, "To put it another way: what all the professional human translators in the world produce in a year, our system translates in roughly a single day". As we noted in Section 1, this translation volume has doubled in the interim.

The business model for Google is well-known, but for translation tool companies to build client-customised engines, there must be an expectation that MT can be a profitable business to be in. TAUS estimate the size of the MT market to be \$250M,<sup>12</sup> which might be somewhat smaller than expected, especially compared to the global value of the translation market for language technology software and services as a whole, which LT-Innovate (2012: 8) estimates as "nearly €20B, and will grow to nearly €30B by 2015, with CAGR of 11.4%". Nonetheless, Jaap van der Meer, one of the co-authors of the TAUS report, notes that "*MT technology is a key enabler and a force multiplier for new services and growth [...] Innovative companies in information technology and other sectors are converging MT technology in new applications and products or they use MT to enhance their existing products.*"

An example of the latter is STAR, a company better known for producing TM products (see Figure 1), who have recently augmented Transit with a Moses-based MT capability.<sup>13</sup> For companies that embrace MT to its fullest extent, they find that it can improve productivity, allow the translation of content which previously was not feasible due to time or cost constraints, reduce the time to market for their products and services, and/or reduce translation costs. Of course, for companies that have SMT system-building expertise, MT can also be used internally as a productivity enhancer, to improve profitability where low margins have been negotiated with clients.

For all such products and services, it has to be openly acknowledged that for SMT, the availability of a large amount of good-quality, representative parallel data is an absolute prerequisite, as stated in the previous section. Of course, such training material comes from human translators, who typically (with or without assistance from TM and/or MT) produce the target-side of the data using traditional means. Without the involvement of human translators, then, we could not even begin to start building our state-of-the-art SMT systems. Note too the vital importance of translators in the SMT system tuning and testing (decoding) phases, where the MT outputs are compared against reference translations provided by human translators.

What has helped SMT development over the past two decades or so, to the point where in research circles it has been completely dominant, and more and more offerings in industry use statistical modelling? Clearly, the existence of open-source tools – such as Moses, the word-alignment software Giza++ (Och/Ney 2003), and language modelling software such as IRSTLM (Federico et al. 2008) – has significantly lowered the barrier to entry into the field of MT. The introduction of automatic evaluation metrics such as BLEU has allowed system developers to measure incremental improvements over time, and while not originally envisaged for that purpose, for side-by-side system comparisons at international bake-offs such as NIST,<sup>14</sup> WMT,<sup>15</sup> and IWSLT.<sup>16</sup>

All these developments have incentivised researchers and industrial teams to build better and better SMT systems, to the point nowadays where there are many successful case studies, with some recent examples including Adobe/ProMT (Flournoy/Duran 2009), Church of Jesus Christ of Latter-day Saints/Microsoft Translator Hub (Richardson 2012), Dell/Safaba/Welocalize (Lavie et al. 2013), DuDu/CapitaTI (Jiang et al. 2012a), and Ford/Systran/SAIC (Plesco/Rychtyckyj 2012). Note the range of MT types – rule-based (ProMT, Systran, LucySoft), statistical (Safaba, CapitaTI) and DIY (Microsoft Translator Hub) – and breadth of industry vertical sectors here: ICT, religion, localisation and automotive. The traditional use-cases for MT regardless of MT engine type or application area have been raw MT and PEMT, whether this be "light" (for assimilation) or

<sup>12</sup> <https://www.taus.net/think-tank/news/press-release/size-machine-translation-market-is-250-million-taus-publishes-new-market-report>

<sup>13</sup> <http://www.star-spain.com/es/blog/transittermstar-nxt-tooltips/transit-nxt-meets-machine-translation-n%C2%BA-1>

<sup>14</sup> <https://www.nist.gov/it/iad/mig/lorehlt16-evaluations>

<sup>15</sup> <http://www.statmt.org/wmt16/translation-task.html>

<sup>16</sup> <https://sites.google.com/site/iwsltevaluation2016/>

“full” (for dissemination). Studies involving post-editing are extremely topical at the current time, and a large body of research exists to demonstrate that PEMT can indeed be faster than translation from scratch (e.g. Flounoy/Duran 2009, Guerberof 2009, Plitt/Masselot 2010). However, results vary across engine types, language pairs, domains, and individuals, so overall conclusions regarding the optimisation of the post-editing pipeline for all situations remain elusive.

While PEMT continues to be an area where MT plays an important role, Way (2013) details a swathe of new, emerging use-cases, typically involving the processing of UGC. Some of these use-cases involve real-time on-demand access to translated content, where no human-in-the-loop is possible. However, other ways in which translators are playing an important role is on joint research on post-editing, using techniques such as Think-aloud Protocols, Keyboard Logging, and Eye-Tracking (Doherty/O’Brien 2013).

Despite the importance of linguists to MT development and system evaluation, and the continuing integration of human translators in research and development in this area, when it comes to post-editing, even though studies show translators to be much faster when post-editing compared to translating from scratch, translators consistently report that they were slower. There are a number of issues pertaining to this, including stress, fear, pride, knowledge, satisfaction, and money (O’Brien 2014).

Some human translators may feel threatened by MT, which is of course one of the main reasons why they are reluctant to use it in their daily work. However, it is clear – we hope, from this paper if nothing else – that on the whole, (i) MT is some way from achieving the same quality as professional human translators do, and (ii) for some, the post-editing task may not appear as intellectually stimulating as translating from scratch. In actual fact, what human translators are really concerned about is the ignorance and disregard towards their profession in a globalised world where multilingual communication is essential for humankind, despite the consensus reached on the fact that MT cannot perform at the level of human translators.<sup>17</sup>

In the remainder of this paper, we attempt to allay these concerns by focussing on the key strengths and weaknesses of the mechanical and human factors in the translation pipeline, and demonstrating that the way forward is a combination of the best of both. Indeed, given the inability of MT engines to do some quite basic things which human translators take in their stride, we assert that translators have, in fact, little to fear from MT, while opportunities exist for forward-thinking translators to benefit considerably from adopting MT and embracing what it has to offer, in the full knowledge of its inherent difficulties.

### **3. What can MT do well, that Human Translators find difficult?**

In this section, we focus on the strengths of statistical models of translation, especially in areas which human translators tend to find most problematic.

#### **3.1. Terminology Consistency**

If no glossary is supplied by a customer for a job involving the building of an SMT system, the engine may still manage to correctly translate a client’s terminology if many examples of how those terms should be translated exist in the parallel corpus. As with regular lexical items, the machine-learning algorithms used in SMT allow patterns to be captured which indicate the likelihood of a target word or phrase being the translation of a particular source word or phrase. These end up in the phrase-table for consultation at decoding time.

However, when no glossary is supplied, the correct translation of a client’s terminology cannot be guaranteed. In contrast, if a glossary is available, we can guarantee adherence for all terms

---

<sup>17</sup> This is compounded by the fact that communication skills in languages other than English appear to be undervalued in today’s globalised society. One clear example can be seen in bookshops, where it is typically the case that translated works are no more expensive than the original English-language versions.



contained therein. Interestingly, translation quality may dip a little (in terms of BLEU score, say) but not to the point that anyone – the client themselves, or a human post-editor – would notice. For clients for whom translation adherence to terminology is of paramount importance, this is a small price to pay.<sup>18</sup>

We will explain why translation quality deteriorates when glossaries are used when adherence to user-provided terminology needs to be assured. Regarding terminology integration, there are two possibilities: (1) in pre-processing, we insert all terms in the user-provided glossaries as partial translations on the target side, and then translate the rest of the sentences in which these terms are contained. This guarantees that the items in the glossary are used, and clearly this is under the control of the user; should they wish to add new terms, or replace old terms with new variants, the look-up procedure simply inserts these in pre-processing on the target side; and (2) the glossary is a translation asset which is available to the MT decoder, as well as phrase-tables, reordering tables and language models (and other parameters), and we let the system decide how best to use all these components in the log-linear model. That is, the glossary itself will be given a score by MERT (Och 2003) in tuning, and the system will use the entries there if overall translation quality in terms of BLEU score (on a per-sentence basis) goes up, and it will ignore it if the BLEU score goes down. Clearly here the user is *not* in control, except that again, prior to translation the glossary can be populated with new terms.

In tests, option (2) has shown itself to lead to improved translation quality. Clearly the terms in the glossary also appear in the training data, and thus in the phrase-tables (and language model). Here much greater context is available than appears in the glossary, so this typically leads to better translations. Of course, where a user insists that the actual terms in the glossary appear in the translation, option (1) should be selected. Translation quality as observed by automatic MT evaluation metrics does deteriorate, but only by a few points. Accordingly, if user control is important, option (1) is a very acceptable solution.

### 3.2. Glossary Creation

Much of our discussion above centred on the availability or not of a glossary as a translation asset. It is well-known (Cabr e 1992/1998) that human translators value such termbanks, which often prove invaluable in helping translators generate accurate target-language texts. International organizations (e.g. UN, UNESCO, EU, WIPO) and the public and private sectors have been involved since the 1960s in creating termbanks to promote multilingual communication in specialised areas. Language services were created to foster exchanges in a globalized world. As Cabr e (1992/1998: 218) states:

The need to ensure communication between languages implies a need to standardize formulae, units and models of communication and to establish reliable equivalents among languages. Our need for information requires the creation of large data banks on various subjects.

Where no such bilingual glossaries exist, they can be created automatically using MT techniques (Haque et al. 2014). Many methods have been developed to identify possible monolingual term candidates. Haque et al. (2014: 46) show how these can be supplemented with SMT word-alignment information to identify target-language equivalents for these source terms. For a range of document types and language pairs, they demonstrate that bilingual multi-word term-pairs can be extracted with high accuracy and in very short periods of time. They observe that “Given the crucial influence of bilingual terminology on quality in translation workflows, we believe that the creation of such assets from scratch in less than 30 minutes may prove to be a significant breakthrough for translators”.

Haque et al. (2014: 49) are convinced that given the reported high levels of performance of their system, verified by human evaluation, “the extracted bilingual multiword termbanks are use-

---

<sup>18</sup> These unpublished experiments were carried out in industrial settings with client-provided TMs.

ful ‘as is’, and with a small amount of post-processing from domain experts would be completely error-free”. Note here that the human translator remains in control: they can edit the glossary, delete terms, or amend terms before actually using it on real translation jobs (with or without MT in the translation pipeline).

It should be noted that such glossaries will contain one target equivalent for each source term. However, the notion of *equivalence* is not always fit-for-purpose, as Gutiérrez (2008: 179-180) suggests:

[...] el hecho de que no se parta del español para introducir términos que luego estarán presentes en estas bases [...] nos obliga a adaptar lo nuestro a lo que hay [...] Sólo un ejemplo: si en España las alopecias se clasifican de acuerdo con la causa, cuando ésta se conoce (medicamentosa, por estrés, etc.), o por las lesiones con las que cursan (areata, marginal, etc.), pero no por su evolución (aguda, subaguda, crónica...), ¿con cuál de todas nuestras alopecias hacer equivaler la alopecia aguda de los textos en inglés, si no tiene correspondiente entre nosotros?<sup>19</sup>

Glossaries created by humans may contain examples of a source term with several equivalents in the target language, e.g. German “Bildschirm” translating as “screen” or “monitor” in English. Such termbanks would be of little benefit to an MT system, as the most likely term would be selected; SMT systems do have the capability of learning under what circumstances such alternatives should be selected, but this would be in the phrase-table rather than in the glossary *per se*.

### 3.3. Spellchecking

At first glance, this may seem like an odd thing to include, as by their very nature, MT systems cannot produce typos; any errors remaining in the MT output must have been there in the training data, or, in the case of PEMT, introduced by the translator in the post-editing stage. It is clear to us that the fact that these exist in TMs is indicative of the fact that some translators (and proofreaders!) do not even rely on basic tools such as spellcheckers; it also points to the fact that clients do not check all of the translations produced for them. Nonetheless, it appears to us that there is a role for MT here to assist the human translator to check spelling and terminology, which is time-consuming, expensive and error-prone.

No matter how large the amount of training data used to seed an SMT engine, it is necessarily finite, so the MT engine is unlikely to ever have complete lexical coverage of the documents to be translated. Such items missing in the training data but contained in the test data are known as out-of-vocabulary items (OOVs), which MT system developers often report as a key indicator of the likelihood of the ability of the system to successfully translate the documents with which it is confronted.

OOVs cannot be processed by an MT system, so typically, such unknown words just get copied into the output.<sup>20</sup> For distantly related languages such as English and Chinese, these are easy to spot, but for closely related languages like English and French, this can be a problem. As an example, consider that the French word “sensible” is not in the training data, but is in the test data (i.e. a new document to be translated). The MT system will just pass “sensible” over to the English side. However, “sensible” is an English word too, of course, but with a completely different meaning; the correct translation ‘sensitive’ is required, here. Such cases will be hard for human translators to spot, in which case errors could quite easily occur in the target text.

Before the translation phase *per se* begins, it is easy to envisage a scenario where all OOVs are reported by the MT system; this is just a case of checking all lexical items in the text to be transla-

19 Our translation: “Since terms entered in databases are in English, we have to adapt Spanish terms to what is already in there [...] Let’s take an example: if in Spain alopecia is classified according to the cause, when it is known (drug-related, stress-related, etc.), or to injuries (areata, marginal, etc.), but not due to its evolution (acute, sub-acute, chronic, ...), which one is equivalent to the English ‘acute baldness’ in English texts, if there is no equivalent in Spanish?”

20 For an input sentence containing an OOV, the situation is very similar to that where a term has been located in pre-processing. The presence of the OOV impacts not just on the translation of that word, but on the whole sentence, especially those elements directly adjacent to the OOV.

ted against items contained in the source side of the SMT phrase-table. Some of these will be real omissions, while others will be real typos. Any such omissions can easily be inserted into a glossary, so that the MT system has full lexical coverage of the new input document, thus leading to improved translation quality, and fewer edits required to be made by the post-editor.

### 3.4. Preprocessing Data

MT developers typically do a lot of pre-processing of the data prior to the translation phase *per se*, including (i) ensuring UTF encoding (as the user's parallel corpus may contain additional non-UTF8 characters that appear as noise in the data); (ii) cleaning parallel data (a number of cleaning techniques are utilised with different heuristics, e.g. sentence length, source—target string length ratio); (iii) removing duplicate sentences; (iv) handling URLs (to prevent these from being broken up into individual terms during tokenisation); (v) handling special characters (e.g. pipes, quotation marks, brackets, ampersands, and XML characters); (vi) tokenisation; and (vii) lowercasing (in order to improve lexical coverage). Accordingly, typically anything from 10–25% of the data supplied by users is deleted in pre-processing.

For UGC, MT practitioners have had to come up with even more data-cleaning operations. Clark/Araki (2011) classified the range of phenomena to be dealt with as shortforms, acronyms, typing errors/misspellings, punctuation omissions/errors, non-dictionary slang, censor avoidance, and emoticons. Jiang et al. (2012a) handle shortforms, punctuation errors and attempts at censor avoidance in much the same way as typographical errors. They utilised a soundex-like algorithm to handle acronyms, and wrote a set of regular expressions to handle the issue of 'wordplay'.

In contrast, human translators do not spend time pre-processing texts in the same way. At most, when there are no pressurised time constraints they first read the whole text to be translated and identify the text genre (Aragónés Lumeras 2008b). However, they *do* understand the rationale behind the communicative event in order to be able to properly construct an appropriate meaning. As productivity is a crucial element in translation, translators ought to account not only for the amount of time spent on actual translation, but also on "pre-translation", if they are to bill clients correctly. Unfortunately, this is not currently the case, so if MT can speed up pre-processing in the ways indicated, this should free up human translators for the more interesting parts of the translation process, as well as help ensure that they are paid for *everything* that they do.

### 3.5. Translation Speed

It goes without saying that MT is much faster than human translation. It is typically accepted that human translators can translate an average of around 2,500 words per day, whereas MT engines can achieve that sort of throughput in around sixty seconds, so MT can be invaluable for translation jobs with very short turnaround times.

Of course, that does not tell the whole story, but the fact remains that more and more use-cases are presenting themselves where no human intervention is possible. As mentioned previously, Way (2013) examines a number of new, emerging use-cases for raw MT and PEMT, in particular involving UGC, and asserts that the days of one level of quality are long gone. Some of these cases include the translation of hotel or product reviews, online chat (Jiang et al. 2012a), and social media posts in the form of tweets, blogs, etc., where huge volumes of material have an extremely short lifespan and so human translation with 'perfect' quality is unwarranted.

Other cases involving UGC noted by Way (2013) as having a slightly longer shelf-life include forum translation (Banerjee et al. 2012), translation of content in online games (Penkale/Way 2012), translation of product listings (Jiang et al. 2012b), and translation of course syllabi documentation and other educational information.<sup>21</sup> For cases where on-demand, real-time translation

---

<sup>21</sup> <http://tramooc.eu>



is required (e.g. multilingual chat), this can only be facilitated by MT, and no human intervention is warranted, or even possible.

### 3.6. Translation Scoring

Bellos (2011) observes that human translators are notoriously bad at coming to an agreement on the acceptability of translations. The very fact that others in the translation pipeline do not just fix clear errors, but also (occasionally) change the translator's initial suggestions ("that's not how I would have said it"), is indicative of this. Bellos (2011: 329) continues that "translation commentators lead the field in throwing most of its work in the direction of the garbage dump". He suggests that in no other profession are workers as nasty to each other as in the translation field. Way (2012: 260-261) interprets this as a possible explanation as to why many human translators are so anti-MT:

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 [system developers].

There are obvious reasons for not finding easily consensus when assessing a translation, including:

1. Language is subjective, so it is extremely arduous, even impossible, to objectively assess a translation. We agree with Hatim/Mason (1994: 4):

Inevitably, translating and discussing translations involve making judgments. But can judgments about translations be made objectively? [...] To replace the impressionism and unsubstantiated opinion which often characterises judgments about the merits and demerits of particular translations, these authors propose methodological and systematic criteria [...] But does this mean that subjectivity can at last be eliminated from translation assessment and that objective evaluations can be made by literary critic and translation teachers alike? [...] Every reading of a text is a unique, unrepeatable act and a text is bound to evoke differing responses in different receivers.

2. Language and drafting is closer to being an art than a science. This also applies to translation, due to the fact that "the translator is under constant pressure from the conflict between form and meaning" Nida (1964: 2). It is, therefore, difficult to find consensus between translators and revisers, and so it is unsurprising to have different evaluations for the same translation depending on the criteria used to assess it.
3. Language, and hence translation, has to deal with ideology and needs to evolve. We write in a specific way according to our beliefs. As an example, human translators will pay a lot of attention to feminising titles in French if they are sensitive to gender issues or if they work in an International Organisation, especially when the client has requested it. A reviser who does not pay attention to gender-sensitive language may prefer, and then correct "Madame l'Ambassadrice" and write "Madame, l'Ambassadeur".

MT can output translations with a confidence score, i.e. how good *it* thinks the translation is based on the knowledge it has, but Moorkens/Way (2016) suggest that this needs to be more robust in order to be trusted reliably. When system developers examine the quality of the current version of their engine, they use a small section (usually no more than 2,000 segment-pairs) of the training material as held-out test data. Here the MT output is compared against the human references in the TM and a score is computed. In runtime scenarios with new input text, no such reference material exists, of course, so quality estimation techniques are used to predict the expected output quality based on prior examination of many previous example-pairs.

This is only now starting to be put to better use in PEMT workflows. We need to be better at presenting only those MT outputs which are ‘good enough’, i.e. good quality output directly by the system, or capable of being post-edited quite quickly and easily. This may mean not showing translators *any* output, but they are used to this sort of behaviour from TM systems, especially when a high fuzzy match threshold is imposed; Moorkens/Way (2016: 44) calculated a match rate of less than 30% even where no fuzzy match threshold was imposed, so match rates when high thresholds are imposed are likely to be much smaller than this figure.

#### 4. What can Human Translators do well, which MT finds difficult?

While the capabilities of MT are often oversold commercially, to the best of our knowledge no MT developers of state-of-the-art engines in academia can be accused of exaggerating their performance. By the same token, some translators take delight in exhibiting where MT falls down, often with spectacular results. While we acknowledge that MT is not (and never will be) perfect, the truth regarding its capabilities lies somewhere in between these two unhelpful extremes.

Nonetheless, some of the things that cause MT to underperform are exactly those areas where humans excel. Accordingly, for the non-expert it can be hard to understand exactly why MT makes the sorts of errors that it does. In this section, we demonstrate the sorts of translations that MT is currently unable to generate, which human translators produce in their stride.

##### 4.1. “Pragmatic” versus “Syntactic” Translation

Kay (2014) gives many examples where MT engines struggle to produce accurate translations. We revisit some of these here. Consider how an MT engine – no matter what type – might translate the French mini-discourse in (i) into English:

- (i) “Est-ce que c’est ta cousine?” “Non, je n’ai pas de cousine.”

While we non-Francophones may indeed have female cousins, or if not know the difference between cousins who are male and those who are female, we don’t have a lexical item to render that distinction, i.e. there is a lexical gap. It would be inappropriate to translate (i) as (ii):

- (ii) “Is that your cousin?” “No, I don’t have a cousin.”

However, this is probably the best an MT engine could do. What human translators would do, in all likelihood, is translate (i) as (iii):

- (iii) “Is that girl your cousin?” “No, she’s not my cousin.”

Kay (2014) calls outputs like (ii) syntactic translation, and outputs like (iii) pragmatic translation. MT engines work as in (ii), while humans freely produce (iii). From a syntactic point of view, (ii) is correct, but of course the translations of both the question and the answer in (ii) fail from a semantic point of view. Kay observes that “More than half the sentences in Europarl require pragmatic translation”; if so, then SMT has a serious problem, as Europarl (Koehn 2005) – the records of the European Parliament – is the basis for most SMT engines that have ever been built for European language pairs.<sup>22</sup>

Of course, MT can only do syntactic translation, as it has no ‘real world knowledge’. Kay defines the Syntactic Model of Translation thus: a long translation is a sequence of short(er) translations, we memorize short translations (lexical items), and these short translations can be reordered.

If MT systems (regardless of their type) find examples like (i) problematic, let us work through how we expect an SMT system to perform on another of Kay’s examples, namely (iv):

<sup>22</sup> We refer the interested reader to Ozdowska/Way (2009), who demonstrate the dramatic effect on translation quality when using inappropriate Europarl data to train SMT engines. Despite their efforts, the bad practice they uncover is still rampant today in MT system development.

(iv) Versichern Sie sich, dass Sie nichts in dem Zug vergessen haben.

In SMT, the first thing to do is word and phrase alignment, as in Figure 2. We can see that while most alignments are one-to-one, some are one-to-many and others are many-to-many. These would all come from entries in the phrase-table, with the patterns having been learnt from many thousands of prior translation examples in the training material. What we have yet to do, of course, is put the suggested target-language words and phrases in the right order. This is the job of the target language model and the reordering model. Assuming again that this works optimally, we generate the target-language word order in Figure 3.

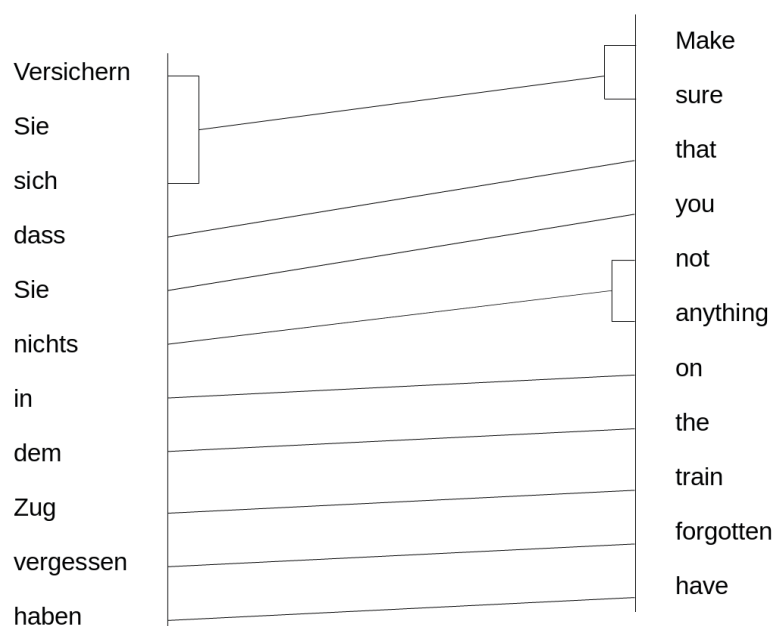


Figure 2. Optimal word- and phrase-alignment for the German sentence *Versichern Sie sich, dass Sie nichts in dem Zug vergessen haben* when translating into English (from Kay 2014)

Now, having generated the correct target-language word order, we simply read off the English translation in (v):

(v) Make sure that you have not forgotten anything on the train.

From a syntactic point of view, there is nothing wrong with this translation; if you were a passenger about to leave the train, and heard this announcement, it might be expected to bring about the correct behaviour, i.e. that you would check to see that you had not left anything behind you. As an MT developer, then, we would be delighted if our engine were able to produce (v) from the input in (iv).

There is only one problem with (v), though it is a significant one, namely that no human translator worth her salt would translate the German sentence in (iv) this way! Again, (v) is a syntactic translation, and unlike (ii), which was a mistranslation, there is no doubt that (v) is acceptable.

What would a human translator do, then, in the face of (iv)? If we revisit Figure 3, things are fine for the first four English words “*Make sure that you*”. The problem then is that probably the best continuation is “*take all your belongings with you*”. It is not clear at all how any MT engine past, present or future would be able to make that inference from the remaining German source words: there is no translation of “Zug” (train), “vergessen” (forgotten), “haben” (have), “nichts” (nothing), let alone the stop-words “Sie”, “in” and “dem”.

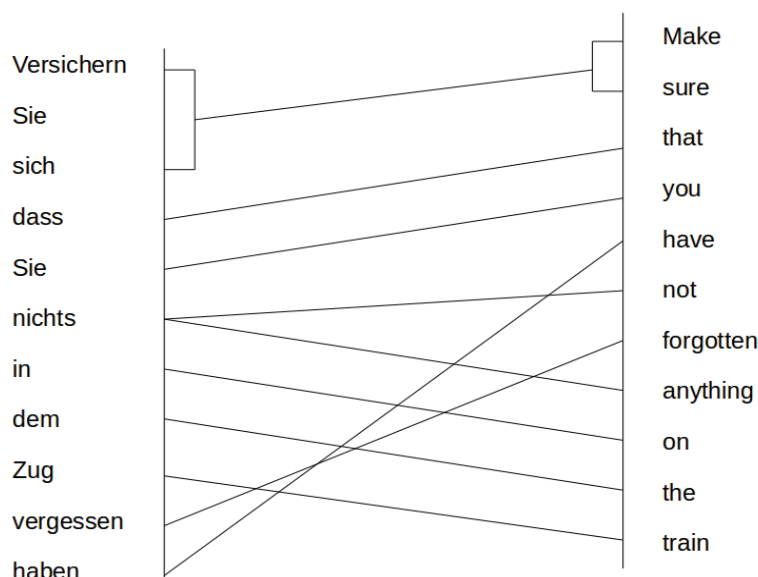


Figure 3. Optimal reordering of the English words and phrases in Fig.2 (from Kay 2014)

Why then do human translators depart from syntactic translation? Common answers would be that sometimes it is to adapt to a translation brief (*skopos*), namely to adapt to clients' wills and needs, while in other cases it is demanded by text type, genre, or communicative conventions. Of course, translators are aware of the relevance of the contextual situation in which a text is produced. Behind the curtain of words and sentences are hidden authors' motivations, intentions and purposes (Aragonés Lumeras 2011: 105). It is common knowledge that every situation requires a specific behaviour, which in French diplomacy is called *étiquette*. Above all, the translator's task is to produce texts that are fit-for-purpose, idiomatic, and which flow smoothly. To do so, translators need to recast sentences using a natural language and make them easily understandable for the expected recipients. This is very difficult for SMT given its restrictions on inferencing solely from static instances in the training data.

#### 4.2. Rephrasing

We begin this section with a clear example of what MT cannot do. In Chinese, “复发率低” (*fufalü di*) (lit. relapse level/rate low) is a common expression in the medical field. MT will encounter plenty of examples of “复发” (*fufang*) and it will be difficult to disentangle the right meaning. For a human translator, it comes naturally to say in English “low risk of relapse”. Here, “率” (*lǜ*) which is usually translated as “level” becomes instead “risk”.

Another example is “复方药” (*fufangyao*), an expression that refers to a composed drug. In English, the translation is “a composed drug containing a mix of natural ingredients, such as herbs and powders used in Traditional Chinese Medicine” (TCM). In the MT literature such examples are known as *periphrastic constructions*. There are many thousands of such examples of unequivalences (see Section 3.2) and deviating conceptualisations determined by language, which pretty much sounded the death knell for (wholly) interlingual approaches to MT.

At this point, it is worth mentioning that computational linguists as well as most translators agree on the fact that the translative act is to find equivalences (Nida 1964). The example above shows that equivalence is not a mere question of wording; one word may have a lot of equivalent

terms with no ‘computational’ mapping between each variant (Aragonés Lumeras 2008a);<sup>23</sup> it is above all a question of how to match different realities.

A clear example of unequivalence is “肾虚” (shenxu), which refers to an imbalance in TCM. To translate this into English, we need to determine the translation brief and the recipients: are we addressing the message to TCM Western doctors, TCM students or patients? Only then can we decide whether to translate it as “kidney deficiency” for TCM doctors, or as “malfunctioning of kidney, an organ related in TCM to sexual disorders” for Western patients and new students.

Furthermore, human translators have developed skills to produce high-quality texts even when the original text is poorly written, ambiguous and nuanced, illogical, contradictory, or inconsistent. This often happens in the globalised context where English has become the *lingua franca* and authors are forced to generate texts in the language despite not having the right skills. It is often the case that the English documents are not the original source language but have been pivoted through a third-party language, whether by a professional translator, an inexperienced translator, an unqualified translator or via MT.

The translated text provided by professional translators conveys the message clearly, captures nuances of the text, adheres to grammatical rules and can adopt specific spellings according to the clients’ instructions (e.g. Latin American Spanish, American English, classical Chinese characters, in-house style conventions etc.). Moreover, any inconsistencies in the translation will already have been detected and resolved (Lafeber 2012: 112-113). Human translators are very good at identifying inconsistencies and contradictions in a text because of their honed reading skills and trained eyes. Professional translators know how to pinpoint errors to their clients in a diplomatic way to avoid confrontations. It is clear to see that, from a linguistic perspective, human translators deliver a final product, and not a by-product as MT does most of the time. In so doing, translators – who are skilled writers in their own right – reconstruct meaning (i.e. abstract away from the source-language form and structure) in drafting a new final text in a different language for a specific new audience. The target text is, therefore, reliable and there is no need to spend additional time and money as the finished product has been delivered.

### 4.3. Domain Expertise and Search Skills

It is well-known that MT systems work best when tested on data that is very similar to the corpora on which they are trained. For example, it would be foolish to translate weather forecasts using an SMT system that had been trained on parliamentary texts. So, for a particular company, it might not make sense to talk about their single English-to-French engine, but rather a whole suite of engines for this language pair depending on the domain (or in the field of Language for Specialised Purposes, what is known as “text genre”), e.g. patents, trials, white papers, personnel texts, product documentation, legal texts, contracts etc. Any experienced MT practitioner can work out quite quickly for a particular client what the best solution is: should we pool all data and build one big system (the more data we have, the more confident the probabilistic decisions made by the system; but the risk is that the more specialised vocabulary and linguistic constructs can be buried by more generic data), or build a range of systems for different sublanguage domains?

However, MT developers have a problem: how much do they rely on the labels assigned to data by humans? For example, a banking client will have all those sorts of documents, but typically they will all be labelled as “banking/financial” by a human ‘expert’, as the only way to *really* know is to look at the data (which does not happen in practice due to the amount of effort that would be involved). This is where machine-learning techniques can help: we can run clustering algorithms on the data and see what optimal subsets of data ensue; from that, we can decide how many engines it makes sense to build. We can also measure the overlap between the manually

23 <http://translationjournal.net/journal/45meaning.htm>



classified data and the automatic clusters. This is why there is currently so much work being conducted in MT in the area of domain adaptation (e.g. Banerjee et al. 2012).

This is one area with a clear distinction between how human translators and MT currently work. An illustrative example of how MT fails due to the lack of domain expertise is with terms like “海绵窦” (haimian dou). “海绵” (haimian) means “sponge” and “窦” (dou) means “sinus”, so MT will probably suggest “sinus sponge” as a ‘good’ translation. Human translators know that there is no such thing and so will then start to search for the correct translation if they are not domain experts. Any online bilingual dictionary will instead quickly suggest “cavernous sinus”.

Of course, if sub-corpora of different genres for a specific domain could be collected in sufficient quantities, then MT systems could be trained to this more fine-grained level. As we pointed out earlier, MT developers rely almost exclusively on parallel data where the target text has been put together by a human translator, so this should be possible. In contrast to MT, for simple texts, human translators do not usually need to be told what domain a particular source text belongs to in order to know how it should be translated, but as we hope is now clear, they *do* need to know what text genre they are dealing with. If a particular source word or expression has two different translations depending on the domain of application, a well-trained translator will use the correct one; it is relatively unproblematic for a human translator if a glossary contains alternative translations for the same source entry because of her domain expertise. Nonetheless, Lagarde/Gile (2009) show how experienced and specialised translators will prefer to use non-terminological sources, such as specialised texts, in order to grasp technolects or sociolects in a contextualised way. Parallel texts that are not translated are a trustworthy source of information for translators with expertise. MT seems to work in a more superficial way putting too much emphasis on terminology, just as trainee translators do when they are just starting out.

The ability to make a quick, productive search is part of the translator’s day-to-day work. Experienced translators are good at searching texts written by native experts (inductive search) and do not spend much time looking for the translated term in an online dictionary (i) because of the huge amount of expressions suggested, and (ii) as sometimes no translation can be found there. Moreover, a document drafted by a native expert will always be more trustworthy than a bilingual database for obvious reasons: it provides contextual information on the searched expression and it shows translators genre conventions (Aragónés Lumeras 2008b: 224-225).

#### 4.4. Resolving Pronominal Reference

Most MT engines built to-date have English on one side of the language pair. It is well-known that MT translating into English is almost always easier than translating from English, from a syntactic perspective, as it is relatively morphologically impoverished, so lexical selection of verbal and nominal forms with the same roots becomes much simpler. In particular, on the whole it has no gender separation, so MT has a real problem in deciding on pronominal referents in order to decide how best to translate such cases.

Kay (2014) uses a couple of nice examples in his presentation which really have their roots in Hutchins/Somers (1992), starting with (vi) (Kay 2014: 95):

- (vi) We gave the monkey the banana as it was
- ripe (sie: die Banane)
  - teatime (es: German pleonastic pronoun)
  - hungry (er: der Affe)

Until very recently, no wide-coverage MT systems had a strategy for coping with such examples, so that if the correct pronominal translation is selected, it was likely to be a matter of chance. Note, of course, that this is just within the bounds of a single sentence! Until recently, all MT systems translate one sentence at a time, with no knowledge of context. Clearly this is not how hu-

mans process language, either monolingually or when performing translation. Currently, therefore, it is *impossible* for MT to consistently process extra-sentential pronominal referents correctly. Let us look at example (vii), adapted from Hutchins/Somers (1992: 96):

- (vii) The soldiers shot at the women. Some of them
- fell (quelques-unes, i.e. the women, les femmes (feminine gender))
  - missed (quelques-uns, i.e. the soldiers, les soldats (masculine gender))

MT developers have only recently begun to think seriously about building systems which go beyond the bounds of the sentence. The work of Hardmeier (2014) is likely to prove influential owing to the development of the open-source Docent decoder,<sup>24</sup> designed specifically to treat whole documents – not just individual sentences – as translation units to allow the building of discourse-level SMT models, including cross-sentential dependencies.

Again, all this is simple for a qualified, experienced human translator, who uses her world knowledge to make the required associations in (vii) to ensure that the correct pronominal referents are made which inform the selection of the appropriate translations. Note also that translators always have context available to them when translating in a TM tool (see Figure 1), so the way we expect MT to work here (at least until recently) is both artificial and doomed to failure from the very start.

As has been shown, even with languages that share similar structure, MT finds it difficult to disambiguate meaning. Chinese is a very good example of how ambiguity is inherent in the language, as there is no plural or conjugation. The extreme simplicity of Chinese grammar makes it more confusing. For languages that are structurally very different, human translators – unlike MT – can guarantee a high-quality translation because they are able to disambiguate or maintain ambiguities where appropriate, omit what is unnecessary, and are able to segment the text in a meaningful way.

Kay (2014) gives a nice example when translating from English to French, as in (viii) and (ix):

(viii) The experiment he described at the end of the lecture was a bit confused. But the result was pretty clear. You can read it in the paper. → FR: *article*

(ix) What happened after the police arrived was a bit confused. But the result was pretty clear. You can read it in the paper. → FR: *journal*

Note that the last two sentences in English are identical, but the translation of *paper* can only be determined from knowledge encoded in the first sentence in both. To return to our previous discussion regarding SMT training data, having two entries in our TM would be fine, as in (x):

- (x) (a) You can read it in the paper ==> Vous pouvez le lire dans l'article.  
 (b) You can read it in the paper ==> Vous pouvez le lire dans le journal.

#### 4.5. Syntactic Agreement

SMT in particular is bad at simple things like Subject-Verb and Determiner-Adjective-Noun agreement (Vanmassenhove et al. 2016), which are requirements one learns quite quickly when learning a foreign language. Accordingly, when post-editors see these sorts of errors, their overall impression of the performance capability of the MT system deteriorates rapidly, and clouds those areas where MT performance is actually quite good, and of benefit to the translator.<sup>25</sup>

Back in the bad old days of MT, people used to publish results of their system's ability to cope with one particular grammatical phenomenon, to the exclusion of all other linguistic challenges,

<sup>24</sup> <https://github.com/chardmeier/docent/wiki>

<sup>25</sup> Rule-based MT is by its very nature somewhat better at this, of course.

so this was the sort of thing we would try to ensure we got right. Now, however, while we can build engines which are by and large robust in the face of *any* input, we can no longer translate the ‘simple’ stuff correctly. Many people are trying to work on incorporating linguistic rules into SMT in order to try to go beyond the current performance ceiling (e.g. de Gispert 2006, Hassan et al. 2008, Marton/Resnik 2008, Haque et al. 2011), but it is by no means as straightforward as one might think, especially from a human translator’s point of view.

It seems to us that this boils down to a problem with the fact that *everything* in SMT is statistical (the glossary entries in Section 3.1 being perhaps one counter-example). Nothing is a fact *per se*; translations of source words and phrases are only so according to a particular probability; no matter how high these might be, they are rarely equal to 1 (see the second Way/Hearne (2011) quote in Section 2.1).

#### 4.6. Semantic Agreement

In addition to the syntactic errors outlined in the previous section, MT can also be quite poor at selecting a word or phrase which is semantically appropriate to the given context. In MT, this is known as word-sense disambiguation (Yarowsky 1995), and quite a lot of work has been done on this topic, especially on its integration with SMT (e.g. Stroppa et al. 2007, Haque et al. 2011).

In the pre-translation phase when using a TM tool, both (a) and (b) in (x) would be suggested as 100% matches, and the human translator would easily be able to select the most appropriate one given the available cotext and context. In MT, this would also be fine, but eventually the system has to select the 1-best translation from the two alternatives in (x), one of which would be more preferred in the SMT log-linear model.<sup>26</sup>

Clearly, when the sense of a source word is wrongly rendered by the MT system in the target language, this causes problems for the post-editor. Since it is almost impossible to know in advance where MT might make a mistake, post-editors have to spend more time comparing the source and target texts. This can make post-editing quite annoying for translators who prefer to focus on linguistic challenges, which may be another reason for their reluctance to work with MT.

We have shown how MT can help human translators create glossaries and text corpora as well as ensure terminology consistency. However, MT cannot function optimally by itself; translators and machines need to work simultaneously, because translators cannot be considered as mere proofreaders, post-editors or revisers. MT system-builders need to keep in mind that translators do not appear solely at the end of the automatic translation process; they have to contribute at appropriate points in-between to ensure quality translations. When a minimum quality requirement is not achieved, it is useless to spend time revising or post-editing the text provided by MT; rather, it would be less time-consuming to translate the particular segment from scratch. Translators need to have control of the machine to ensure quality because MT cannot rely solely on explicit memory. MT cannot, therefore, replace human translators in all use-cases where translation is required; both have to act together and properly coordinate their actions according to their competences.

#### 4.7. Word Order

As mentioned previously, the three main components in the log-linear model of SMT are the translation model, the language model, and the reordering model. While there is a lot of work on the two former models, there is a relative paucity of research papers published on reordering, despite being one of the major sources of errors in SMT today (e.g. Naskar et al. 2011).

---

<sup>26</sup> While some work (e.g. Koehn/Haddow 2009) has addressed the possibility of offering multiple MT outputs to human translators, for the most part the N-best translations produced by an SMT system are almost identical. What has to be done, therefore, is select the N-best ‘interestingly different’ outputs for presentation to the post-editor (Srivastava et al. 2011).

In contrast, human translators are very unlikely to produce output which does not correspond to the correct target-language word order, as best practice stipulates that translators should demonstrate an excellent command of the language(s) into which they translate. Nonetheless, it is widely accepted that SMT systems translate more effectively into languages with relatively impoverished morphology (like English) compared to those which are morphologically more complex (like Basque: Stroppa et al. 2006, Labaka et al. 2007), as in the former case such languages rely on a fairly strict word order in order to convey who is doing what to whom.

In the case of Chinese, word order is paramount for the interpretation of meaning. Chinese is, therefore, a very good example of the difficulties that MT encounters, since word order plays the role of semantic features when gender, plural, verb tenses and prepositions do not exist. The Chinese specificity leads to even more ambiguity compared to other languages. In translation, disambiguation is not advised. MT is not good at deciding whether to disambiguate or not; it relies completely on the evidence it has seen in the training data, exemplified at runtime in its statistical language, translation and reordering models.

Clearly, more work needs to be done in SMT on reordering, and this appears to be happening (see recent examples such as Hannemann/Lavie 2013, Li et al. 2013, Kazemi et al. 2015). This will no doubt result in better translation quality, but what is particularly difficult for post-editors to detect are cases where SMT produces output with ostensibly ‘correct’ target-language word order, but which does not accurately reflect the meaning of the source text.

Language Service Providers are under pressure to reduce their post-editing costs, and consideration is being given to using non-domain experts, and even non-speakers of the source language. Clearly, if they are never exposed to the input string, such post-editors will never be able to detect errors in the target-language output which have been introduced as the result of translation.<sup>27</sup>

Again, this disregard for cross-lingual skills shows that the golden age for human translators has gone. Unsurprisingly, it is more and more difficult to find professional translators with rare language combinations, such as from Asian languages to English or French. Learning Asian languages as L2 requires efforts that stakeholders do not take into account. Nowadays, translators are not better paid when they translate from Asian languages into English due to the globalised market. There is hence no return on investment for human translators who spend 4 years studying translation and/or interpreting at university and at least 5 years in a foreign country studying the foreign (Asian) language.

## 5. The Way Forward: a Synergistic Point of View

It may be unpalatable for some to acknowledge (Penkale/Way 2013), but it is indubitably the case that MT is here, of demonstrable benefit to millions of people on a daily basis, and is not going away. Many enlightened translators have already found a way to make their lives easier by incorporating MT into their translation workflow. By the same token, it has to be recognised by developers of SMT systems that if they do not reach out to linguists and translators, the extent to which the quality of their systems can be improved will be limited.

Accordingly, in this section, we make some recommendations as to how MT developers and human translators can co-operate to their mutual benefit. Way/Hearne (2011) posit that many people find the basic model of SMT hard to understand, and try to describe that model to non-experts in a companion paper (Hearne/Way 2011), as they deem it vital that linguists and translators know how their input is used in SMT. Indeed, Way/Hearne (2011) make the interesting observation that such cooperation was envisaged in the original IBM papers on SMT (Brown et al. 1988a,b, 1990, 1992) – much of which underpins statistical models even today – and regret that greater progress would have been made if such collaboration had taken place.

---

<sup>27</sup> Note that automatic evaluation of MT works exactly like this, with no looking back at the original source text; it goes without saying that no human translator would evaluate her work without recourse to the input text.

Rather than dwell on this, Way and Hearne go on to provide some concrete examples of what different levels of cooperation are required. They contend that only when experts in corpus linguistics, linguistics, computer scientists and translators come together will improvements in today's SMT systems be catalysed. They state with some positivity that the exclusion of linguists can be reversed, but for those SMT developers who do not reach out, Way (2009) similarly concludes that "Those without a linguistic background [...] appear to have two choices: (i) to attempt to include the linguists, so that they may be of help; or (ii) to continue to exclude them, while at the same time trying to make sense out of *their* writings" (original emphasis).

Reducing translation essentially to an activity involving the mastery of two languages has been, in our opinion, a sticking point for MT development and improvement. As we hope to have pointed out, translation requires a pragmatic analysis of the situation, a critical reading competence and a capacity to consider extra-textual factors, such as context, the author's motivations, the recipients' expectations, the conventions pertaining to a ceremony, and much more beyond.

It is essential, therefore, for MT developers to foster cooperation with human translators. Otherwise, the prospects for improving MT over the current performance ceiling are grim. Humans and machines work and 'think' in a different manner. This does not mean that they are irreconcilable, far from it; once they have improved their knowledge of translation, programmers and computational linguists may be able to combine both ways of processing data and information. Implicit and explicit memories are complementary.

The processing of data is closely related to memory. It is no coincidence that MT began to be contemplated after McDougall's (1928) in-depth research on human memory led to a difference between explicit and implicit memory. Acquiring specific skills requires practice (episodic and semantic memory), so how efficient can MT be? MT can process information quickly, has a large storage capacity for information, is limited in performance but has huge processing capacity.

We are convinced, however, that if computational linguists start working closely with translators, consider text genres for creating corpora, and move much more than heretofore towards a semantic as well as a cognitive approach, human translators could benefit considerably from MT.

## 6. Concluding Remarks

In this paper, we have attempted to demonstrate the particular strengths and weaknesses of both human translators and MT systems, especially today's prevalent statistical models. Rather than note this to be problematic, we have shown that it is precisely those areas where humans struggle where SMT systems can help, and in contrast, that where SMT systems go astray, these are cases where human translators are especially efficient. Accordingly, in our view the future is bright, and ripe for collaboration.

We have demonstrated that MT is being used in many ways by many users on a daily basis, so 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 is all it is – a tool in their armoury – and all it ever will be. We very firmly believe that there is no threat to translators' jobs from MT, despite ongoing scaremongering from some translators, especially given the difficulties it faces in Section 4, as well as the fact that MT is not appropriate in all use-cases. Fortunately, more and more influential translators are willing to stick their heads above the parapet and speak about the utility of MT (Way 2013).

In closing, we wonder whether the translative process itself should be rethought in order to be able to adapt MT to other specific needs. For example, in defining MT or translation, we conjecture that the majority would say that it is about rendering ('transferring' is probably the wrong word for many reasons!) the meaning of the source string in the target sentence. Given the time pressures that translators are under nowadays, translators are forced to spend less time drafting versions and searching for help in fixing translation problems; rather, they are selecting from the 'solutions' presented to them by TM and MT, and then adapting the selected solution to the re-



quirements of the target language. Accordingly, “[t]he emphasis has shifted from generation to selection” (Pym 2013: 493).

MT system development can clearly benefit enormously from leveraging such expertise. As we noted in Way/Hearne (2011):

Failure to work together in the recent past has prevented us from making more progress, and the time is ripe for the two communities [MT developers and translators] to come together as a catalyst for further improvements in our translation systems as we go forward together.

We hope that the musings of this human translator and this MT system developer prove to be a useful stepping-stone in this direction.

## Acknowledgements

The second author is supported by the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund. We are very grateful the reviewers for their useful comments which served to improve the paper, and the editors of the journal for their support during its processing.

## References

- Aragónés Lumeras, Maite 2008a: Meaning: The Philosopher’s Stone of the Alchemist Translator? *Translation Journal* 12 [online]. <http://translationjournal.net/journal/45meaning.htm>
- Aragónés Lumeras, Maite 2008b: How to become a patent translator: Tricks and Tips – Notions of Text Genre and Ceremony to the Rescue. In *Meta: Translators’ Journal* 55, 212-236.
- Aragónés Lumeras, Maite 2011: El traductor de patentes: de Sherlock Holmes a James Bond. In *Panace@* 12, 102-107.
- Banerjee, Pratyush/Naskar, Sudip K./Roturier, Johann/Way, Andy/van Genabith, Josef 2012: Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data? In *Proceedings of the 16th Annual Meeting of the European Association for Machine Translation*. Trento, Italy, 169-176.
- Bellos, David 2011: *Is that a fish in your ear: translation and the meaning of everything*. London: Particular Books, Penguin Group.
- Brown, Peter/Cocke, John/Della Pietra, Stephen/Della Pietra, Vincent/Jelinek, Fred/Mercer, Robert/Roosin, Paul 1988a: A statistical approach to language translation. In *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, Hungary, 71-76.
- Brown, Peter/Cocke, John/Della Pietra, Stephen/Della Pietra, Vincent/Jelinek, Fred/Mercer, Robert/Roosin, Paul 1988b: A statistical approach to French/English translation. In *Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*. Pittsburgh, Pennsylvania, 16pp.
- Brown, Peter/Cocke, John/Della Pietra, Stephen/Della Pietra, Vincent/Jelinek, Fred/Lafferty, John/Mercer, Robert/Roosin, Paul 1990: A statistical approach to machine translation. In *Computational Linguistics* 16, 79-85.
- Brown, Peter/Della Pietra, Stephen/Della Pietra, Vincent/Lafferty, John/Mercer, Robert 1992: Analysis, statistical transfer, and synthesis in machine translation. In *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92), Empiricist vs. rationalist methods in MT*. Montreal, Canada, 83-100.
- Brown, Peter/Della Pietra, Stephen/Della Pietra, Vincent/Mercer, Robert 1993: The mathematics of statistical machine translation: parameter estimation. In *Computational Linguistics* 19, 263-311.
- Cabré, M. Teresa 1992/1998: *Terminology. Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins.
- Clark, Eleanor/Araki, Kenji 2011: Text normalization in social media: progress, problems and applications for a pre-processing system of casual English. In *Procedia – Social and Behavioral Sciences* 27, 2-11.
- de Gispert, Adria 2006: *Introducing Linguistic Knowledge into Statistical Machine Translation*. PhD Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain.
- Doherty, Stephen/O’Brien, Sharon 2013: Assessing the usability of raw machine translation output: A user-centered study using eye tracking. In *International Journal of Human-Computer Interaction* 30, 40-51.
- Federico, Marcello/Bertoldi, Nicola/Cettolo, Mauro 2008: *IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models*. In *Interspeech 2008, Proceedings of the 9th Annual Conference of the International Speech Communication Association*. Brisbane, Australia, 1618-1621.

- Flournoy, Raymond/Duran, Christine 2009: Machine translation and document localization at Adobe: from pilot to production. In *MT XII: Proceedings of the Twelfth Machine Translation Summit*. Ottawa, Canada, 425-428.
- Guerberof, Ana 2009: Productivity and quality in MT post-editing. In *MT XII – Workshop: Beyond Translation Memories: New Tools for Translators MT*. Ottawa, Ontario, Canada, 8pp.
- Gutiérrez Rodilla, Bertha M. 2008: La transferencia del conocimiento especializado: el lenguaje biosanitario. In Navarro, C (ed.), *Terminología, traducción y comunicación especializada. Homenaje a Amelia de Irazzábal. Actas del Congreso Internacional*. Edizioni Fiorini, Verona, 177-181.
- Hanneman, Greg/Lavie, Alon 2013: Improving syntax-augmented machine translation by coarsening the label set. In *NAACL-HLT 2013: The 2013 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia, 288-297.
- Haque, Rejwanul/Naskar, Sudip K./van den Bosch, Antal/Way, Andy 2011: Integrating Source-Language Context into Phrase-Based Statistical Machine Translation. In *Machine Translation 25*, 239-285.
- Haque, Rejwanul/Penkale, Sergio/Way, Andy 2014 [online]. Bilingual Termbank Creation via Log-Likelihood Comparison and Phrase-Based Statistical Machine Translation. In *Proceedings of 4th International Workshop on Computational Terminology (Computerm)*, Dublin, Ireland, 42-51.
- Hardmeier, Christian 2014: *Discourse in Statistical Machine Translation*. PhD Thesis, Uppsala University, Uppsala, Sweden.
- Hassan, Hany/Sima'an, Khalil/Way, Andy 2008: Syntactically Lexicalized Phrase-Based SMT. In *IEEE Transactions on Audio, Speech and Language Processing* 16, 1260-1273.
- Hatim, Basil/Mason, Ian 1994: *Discourse and the Translator*. London: Longman.
- Hearne, Mary/Way, Andy 2011: Statistical Machine Translation: A Guide for Linguists and Translators. In *Language and Linguistics Compass* 5, 205-226.
- Heyn, Matthias 1998: Translation Memories – Insights & Prospects. In Bowker, L./Cronin, M./Kenny, D./Pearson, J. (eds.) *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome, 123-136.
- Hutchins, John/Somers, Harold 1992: *An Introduction to Machine Translation*. London: Academic Press.
- Jiang, Jie/Way, Andy/Haque, Rejwanul 2012a: Translating User-Generated Content in the Social Networking Space. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2012)*. San Diego, CA, 9pp.
- Jiang, Jie/Way, Andy/Ng, Nelson/Haque, Rejwanul/Dillinger, Mike/Lu, Jun 2012b: Monolingual Data Optimisation for Bootstrapping SMT Engines. In *Proceedings of MONOMT 2012: AMTA 2012 Workshop on Monolingual Machine Translation*. San Diego, CA, 10pp.
- Kay, Martin 2014: Does a Computational Linguist have to be a Linguist? Invited Talk, *COLING 2014, the 25th International Conference on Computational Linguistics*, Dublin, Ireland [online]. <http://www.coling-2014.org/martin-kay.php>
- Kazemi, Arefeh/Toral, Antonio/Way, Andy/Monadjemi, Amirhassan/Nematbakhsh, Mohammadali 2015: Dependency-based Reordering Model for Constituent Pairs in Hierarchical SMT. In *EAMT-2015: Proceedings of the Eighteenth Annual Conference of the European Association for Machine Translation*. Antalya, Turkey, 43-50.
- Koehn, Philipp 2005: Europarl: A Parallel Corpus for Statistical Machine Translation. In *MT Summit X: Proceedings of the Tenth Machine Translation Summit*. Phuket, Thailand, 79-86.
- Koehn, Philipp/Haddow, Barry 2009: Interactive assistance to human translators using statistical machine translation methods. In *MT Summit XII: Proceedings of the Twelfth Machine Translation Summit*. Ottawa, Ontario, Canada, 73-80.
- Koehn, Philipp/Hoang, Hieu/Birch, Alexandra/Callison-Burch, Chris/Federico, Marcello/ Bertoldi, Nicola/Cowan, Brooke/Shen, Wade/Moran, Christine/Zens, Richard/Dyer, Chris/Bojar, Ondřej/Constantin, Alexandra/Herbst, Evan 2007: Moses: open source toolkit for statistical machine translation. In *ACL 2007: proceedings of demo and poster sessions*, Prague, Czech Republic, 177-180.
- Labaka, Gorka/Sarasola, Kepa/Stroppa, Nicolas/Way, Andy 2007: Comparing Rule-Based and Data-Driven Approaches to Spanish-to-Basque Machine Translation. In *MT Summit XI*, Copenhagen, Denmark, 297-304.
- Lafeber, Anne 2012: Translation Skills and Knowledge – Preliminary Findings of a Survey of Translators and Revisers Working at Inter-governmental Organizations. In *Meta: Translators' Journal* 56, 108-131.
- Lagarde, Laurent/Gile, Daniel (dir.) 2009 : Le traducteur professionnel face aux textes techniques et à la recherche documentaire. Thesis. Université Sorbonne Nouvelle, Paris [online]. <https://tel.archives-ouvertes.fr/tel-00797032>.
- Lavie, Alon/Beregovaya, Olga/Denkowski, Michael/Clarke, David 2013: Analyzing and Predicting MT Utility and Post-Editing Productivity in Enterprise-scale Translation Projects. In *Machine Translation Summit XIV, Main Conference Proceedings*. Nice, France, 305-307.

- Li, Junhui/Resnik, Philip/Daumé III, Hal 2013: Modeling syntactic and semantic structures in hierarchical phrase-based translation. In *NAACL-HLT 2013: The 2013 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia, 540-549.
- LT-Innovate 2012: The Global Language Technology Market: Innovation for the next generation of ICT.
- Marton, Yuval/Resnik, Philip 2008: Soft Syntactic Constraints for Hierarchical Phrased-Based Translation”. In *ACL-08: HLT. 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Proceedings of the conference*. Columbus, Ohio, 1003-1011.
- Moorkens, Joss/Way, Andy 2016: Comparing Translator Acceptability of TM and SMT Outputs. In *Baltic Journal of Modern Computing* 4, 141-151.
- Naskar, Sudip K./Toral, Antonio/Gaspari, Federico/Way, Federico 2011: Framework for Diagnostic Evaluation of MT Based on Linguistic Checkpoints. In *Proceedings of Machine Translation Summit XIII*, Xiamen, China, 529-536.
- Nida, Eugene 1964: *Towards a Science of Translating with Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: E. J. Brill.
- O’Brien, Sharon 2014. Post-Editing. In *MTM 2014: the Ninth Machine Translation Marathon*. Trento, Italy [online]. <http://statmt.org/mtm14/uploads/Main/PostEditingMTM2014.pdf>
- Och, Franz J. 2003: Minimum error rate training in statistical machine translation. In *ACL-2003: 41st Annual Meeting of the Association for Computational Linguistics*. Sapporo, Japan, 160-167.
- Och, Franz J. 2012: Breaking down the language barrier – six years in. Official Google Blog [online]. <http://googleblog.blogspot.ie/2012/04/breaking-down-language-barriersix-years.html>
- Och, Franz J./Ney, Hermann 2002: Discriminative training and maximum entropy models for statistical machine translation. In *40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA., 295-302.
- Och, Franz J./Ney, Hermann 2003: A systematic comparison of various statistical alignment models. In *Computational Linguistics* 29, 19-51.
- Ozdowska, Sylwia/Way, Andy 2009: Optimal Bilingual Data for French-English PB-SMT. In *Proceedings of EAMT-09, the 13th Annual Meeting of the European Association for Machine Translation*. Barcelona, Spain, 96-103.
- Papineni, Kishore/Roukos, Salim/Ward, Todd/Zhu, Wei-J. 2002: BLEU: a method for automatic evaluation of machine translation. In *ACL-2002: 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA., USA, 311-318.
- Penkale, Sergio/Way, Andy 2012: SmartMATE: An Online End-To-End MT Post-editing Framework. In *Proceedings of AMTA 2012 Workshop on Post-editing Technology and Practice*. San Diego, CA, 10pp.
- Penkale, Sergio/Way, Andy 2013: Tailor-made Quality-controlled Translation. In *Proceedings of Translating and the Computer 35*, London, UK.
- Plesco, Craig/Rychtycky, Nestor 2012: Machine Translation as a Global Enterprise Service at Ford. In *AMTA 2012, Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas*. San Diego, California, USA, 9pp.
- Plitt, Mirko/Masselot, François 2010 [online]. A productivity test of statistical machine translation post-editing in a typical localization context. In *Fourth Machine Translation Marathon “Open Source Tools for Machine Translation”*, Dublin, Ireland; *Prague Bulletin of Mathematical Linguistics* 93, 7-16.
- Pym, Anthony 2013: Translation Skill-Sets in a Machine-Translation Age. In *Meta* 58: 487-503.
- Richardson, Stephen 2012: Using the Microsoft Translator Hub at the Church of Jesus Christ of Latter-day Saints. In *AMTA-2012: the Tenth Biennial Conference of the Association for Machine Translation in the Americas, Proceedings*. San Diego, CA, 8pp.
- Srivastava, Ankit/Ma, Yanjun/Way, Andy 2011: Oracle-based Training for Phrase-based Statistical Machine Translation. In *Proceedings of the 15th Annual Meeting of the European Association for Machine Translation*. Leuven, Belgium, 169-176.
- Stroppa, Nicolas/Groves, Declan/Sarasola, Kepa/Way, Andy 2006: Example-based Machine Translation of the Basque Language. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*. Boston, MA., 232-241.
- Stroppa, Nicolas/van den Bosch, Antal/Way, Andy 2007: Exploiting Source Similarity for SMT using Context-Informed Features. In *11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007)*. Skövde, Sweden, 231-240.
- Vanmassenhove, Eva/Du, Jinhua/Way, Andy 2016: Improving Subject-Verb Agreement in SMT. In *Proceedings of the Fifth Workshop on Hybrid Approaches to Translation (HyTra)*. Riga, Latvia, 13pp.

- Way, Andy 2009: A Critique of Statistical Machine Translation. In Daelemans, W./Hoste, V. (eds.), *Journal of translation and interpreting studies: Special Issue on Evaluation of Translation Technology, Linguistica Antverpiensia* (LANS 8/2009). Brussels: Academic and Scientific Publishers, 17-41.
- Way, Andy 2012: Is That a Fish in Your Ear: Translation and the Meaning of Everything – David Bellos, Book Review. In *Machine Translation* 26, 255-269.
- Way, Andy 2013: Traditional and Emerging Use-Cases for Machine Translation. In *Proceedings of Translating and the Computer* 35. London, UK.
- Way, Andy/Hearne, Mary 2011: On the Role of Translations in State-of-the-Art Statistical Machine Translation. In *Language and Linguistics Compass* 5, 227-248.