Bringing Humans into the Loop: Localization with Machine Translation at Traslán

Declan Groves
Traslán
Co. Wicklow,
Ireland
dgroves@traslan.ie
http://www.traslan.ie

Abstract

Traslán makes full use of MT during our translation workflow, where the raw output from our Machine Translation (MT) system is passed onto human translators who perform post-editing (if necessary) to arrive at the final translation.

Within Traslán we have found that using MT has enabled us to increase the speed, accuracy and consistency of translation - elements which allow us to process larger amounts of translation; with quicker turnaround times, which in turn has resulted in overall savings of approx. 20% so far.

One of the main challenges in using MT within a commercial setting is getting human translators to adopt and make full use of the technology. Within Traslán we overcome this obstacle by working closely and intensively with our translators, getting them involved directly in the development process. Doing so enables translators in turn to train new users of the system and to communicate effectively to other translators the benefits of integrating MT into the translation pipeline.

1 Introduction

As the demand for commercial translation increases the demand for automation increases with MT, together with post-editing, presenting itself as a practical solution. User acceptance of MT is vital in implementing this type of technology in a commercial setting. Although the quality of the MT output plays an important role in this acceptance, it is not the sole decisive factor; user perception is key (Hutchins, 2001). However, historically, translators are not known for their open acceptance and use of MT technology. Common arguments from translators against the use of MT include a dislike for correcting repetitive errors, a fear of losing language proficiency by working with poor quality output, and a dislike of having one's freedom of expression limited (O'Brien, 2006). Often, MT is wrongly perceived as a replacement for translators, when in fact it needs to be viewed as one of several tools in the users' work environment to help their productivity and efficiency. It is important that translators have a positive view of MT and that they are not in any way made to feel redundant (Valderrábanos et al., 2003). After all, there will always be a need for human translators; MT will only ever serve to complement human translation, not to replace it (DePalma, 2006).

As there is relatively little research available discussing the successful implementation of MT within translation companies, in this paper we aim to shed some light on our own experiences with MT and how we manage to keep our human translators in the loop when translating as part of a localization task. In MT research, systems are often developed without any consideration for who the final user may be, however, in a commercial setting the user is paramount. We are not seeking translation perfection, but instead we need to base our evaluation of system performance on the potential savings it can provide in practical situations (Hutchins, 2001). Rather than striving for FAHQT (Full Automatic High-Quality Translation), our company presents a working example of FAUT (Fully Automatic Useful

Translation) (van deer Meer, 2006). At Traslán, the use of MT to produce draft translations which are subsequently post-edited by our translators (where necessary) has resulted in significant savings, producing high-quality translations at a fraction of the cost and time needed for traditional human translation.

Similar to PAHO (Vasconcellos, 1985), we have the advantage of developing our own MT system in-house, tailoring it to fit within our translation workflow. The added benefit of developing our own bespoke system, and of being a small company, is that communication between our translators and development team is easy to facilitate and we are able to absorb integration costs and deal with any system difficulties efficiently.

Consulting directly with our team of translators is also invaluable to the adoption of the technology. If we are to maximize the benefits of using MT, translators need to have an understanding of the technology, of its capabilities and its limitations. If translators are made fully aware of these aspects of MT they are less likely to become frustrated when using the technology and their increased understanding, in turn, facilitates the provision of more beneficial and valuable feedback directly to the MT developers.

The remainder of this paper is organised as follows: we provide a description of Traslán, the type of translation work we do and the benefits we've gained from using MT in Section 2. In Section 3 we give a brief overview of our MT system before describing how we bring our human translators into the loop in Section 4. Despite the high quality of our MT output as indicated by standard automatic evaluation metrics, there are still some areas where the MT output requires a large amount of post-editing effort and time to perfect. Often these types of errors are not reflected in the automatic metrics and yet can be resolved relatively simply and automatically. To this end, Section 5 describes some evaluations we have carried out on our system performance, the type of errors identified by our translators together with resulting system improvements. Finally we present our conclusions and outline our plans for further research and development.

2 Company Background

Traslán was formed in 2004 as a technology-driven translation service provider. In contrast to many translation companies, Traslán was founded with the aim of making full use of MT. Accordingly, we have developed and adapted our own MT technology which we use within our translation workflow to provide translation services to both the public and private sectors, using post-editing where necessary. In addition to full end-to-end translation services, we also offer various interpretation and consultancy services, catering primarily for the English–Irish language market but also facilitating additional language pairs.

We are a small wholly Irish-owned company currently employing a small number of full-time human translators. As we have a limited number of in-house translators we occasionally outsource translation and post-editing work to freelance translators. In addition to our translators we have a small team dedicated solely to the development of our MT technology and an additional marketing officer.

2.1 Irish & The Official Languages Act

Irish, or Gaeilge (GA), is an Indo-European Celtic language and is the national and first official language of the Republic of Ireland. It was awarded official status with the European Union in 2005, ensuring it has equal status to other European languages. It is also officially recognised as a minority language in Northern Ireland. The Irish language has a number of characteristics which present interesting challenges in the development and application of MT (cf. Figure 1).[1]

Today, according to the 2006 Census,[2] 41.9% of the population in the Republic of Ireland (approx. 1.7 million people) can speak Irish competently, aided in part by being an obligatory subject in schools within the Republic. In addition, 10.4% of Northern Ireland have some knowledge of Irish. Irish still survives as a community language in certain rural areas

[1] Note that the different **spoken** dialects of Irish do not, in themselves, present any major difficulties to the process, as the texts to be translated are generally expected to be translated into a form of standardised, official Irish.

[2] http://www.cso.ie/census/

Syntax

Perhaps the most striking feature of Irish syntax is its basic VSO order:

The man	hit	the ball
Bhuail	an fear	an liathróid
[hit]	[the man]	[the ball]

However, in subordinate phrases SVO applies:

The man	who	hit	the ball
An fear	a	bhuail	an liathróid
[The man]	[who]	[hit]	[the ball]

Morphology

The most significant features of Irish morphology are initial mutations (lenition and eclipsis) and final changes. Word-initial consonants such as [c,g,t,d,p,b] etc in nouns and verbs can be lenited or eclipsed:

cara	⇔	[a friend]	caithim	⇔	[I throw]
mo chara	⇔	[my friend]	ní chaithim	⇔	[I do not throw]
a cara	⇔	[her friend]	an gcaithim	⇔	[do I throw?]
a chara	⇔	[his friend]			
ár gcara	⇔	[our friend]			

A similar change may occur with word-initial vowels:

úll	⇔	[an apple]	ithim	⇔	[I eat]
m'úll	⇔	[my apple]	an n-ithim	⇔	[do I eat?]
a h-úll	⇔	[her apple]			
a úll	⇔	[his apple]			
a n-úll	⇔	[their apple]			

Word-final changes in nouns and adjectives are case-driven. With some exceptions, there are just two cases in Irish, nominative and genitive, singular and plural. Examples of word-endings are as follow:

an fear	⇔	[the man]	an fear mór	⇔	[the big man]
hata an fhir	⇔	[the man's hat]	hata na bhfear mór	⇔	[the big men's hats]
na fir	⇔	[the men]	na fir mhóra	⇔	[the big men]
hataí na bhfear	⇔	[the men's hats]	hata an fhir mhóir	⇔	[the big man's hat]

Definite Article

One other area of interest for us in the MT testing was the absence of a definite article in Irish:

fear	⇔	[a man]	an fear	⇔	[the man]
fir	⇔	[men]	na fir	⇔	[the men]

Figure 1: Characteristics of the Irish Language

in Ireland, referred to as the Gaeltacht. These areas have a population of roughly 90,000 people collectively. In addition, there is a steadily-increasing number of urban-based Irish speakers, closely connected to the development and expansion of Irish-medium primary and second-level schools. There is an associated development of third level courses taught entirely through Irish in a number of universities.

One of the original motivational factors behind the foundation of Traslán was the introduction of the Official Languages Act by the Irish Government in 2003, which provided for a greater availability and higher standard of public services through Irish. As a result, Government Departments and public bodies in Ireland are now required to publish any key documentation simultaneously bilingually – both in English and Irish – and to make services available to the public in both languages.

Similar to the wider translation industry, within Ireland demand greatly exceeds supply in terms of translation needs and the number of professional human translators available. In order to meet the demands of the Irish market, a number of degree-level and graduate-level course have been established. In addition, the body responsible for promotion of the Irish language throughout the whole island of Ireland, Foras na Gaeilge[3], has established an official accreditation scheme for professional Irish translators. At present approximately 150 such accreditations have been awarded, with all of Traslán's full-time translators holding this accreditation. While it is not a statutory requirement, an increasing number of official bodies require all work to be performed by such appropriately accredited translators.

The lack of Irish translators is also evident at a European level. When the Irish government sought recognition of Irish as a full official EU language (upgrading its status from that in which only a certain number of key documents were available in Irish) the government reassured the EU that sufficient translators, interpreters and jurist-linguists would be available to serve the EU requirements. However, to date only approximately 5 Irish language experts have been employed so far, with a further 14 on a panel at varying stages of the recruitment process.

So, the demand for high-quality Irish translation services continues to grow. In order to help meet this increasing demand Traslán was established with the aim of making full use of MT technology during our translation workflow. As an industry, if we are ever to address the demand for translation services the use of MT is inevitable and necessary.

2.2 Domain Type & The Localization Task

Following the introduction of the Official Languages Act, some examples of our customers include Government Departments, Local Authorities, Education Services and Committees, Universities and County Enterprise Boards. As a result, a large majority of the texts that we receive for translation consist of official Government documentation, which although complex, lend themselves readily to MT as they contain lots of repetition and are reasonably domain-specific. The legal nature of many of the texts, however, also emphasises the importance of having professional translators on hand to ensure the accurate post-editing of the MT output.

In addition to these types of texts, we also deal with localization projects. One such localization project we are currently involved in is a large-scale research project at the Centre for Next Generation Localization (CNGL), focusing on carrying out research into standards, interoperability and automation of language and digital content management technology.[4]

As Hutchins (2001) points out, it is now widely accepted that MT works best within domain-specific and controlled environments. Thus, MT lends itself readily to use during commercial localization projects where the data to be localized comes from a specific domain, such as software and user-based documentation; two domains in which we have large multinational customers.

This type of large-volume documentation is highly repetitive and translations need to be produced rapidly. In order to work within these time constraints while controlling costs, MT provides the ideal solution. The consistent nature of MT helps to ensure the consistency in translation output, in particular in terms of product-specific terminology

[3] http://www.gaeilge.ie

[4] http://www.cngl.ie

and glossaries. Due to its repetitive nature and rich domain-specific terminology this type of documentation is not particularly interesting for translators, and in any case MT is capable of taking care of a large proportion of the effort involved, especially those elements which prove to be particularly tedious for the human translator. It is this localization work which we focus on in this paper.

2.3 The Value of MT

Our MT system has been in deployment since 2006, and already we have seen overall savings of approximately 20%. In 2008, we have used the system to help in the translation of over 2M words of text, from a variety of sources. On one localization task, we successfully used our technology to process over 68K sentences (approx. 400K words) of text from English to Irish.

Using MT technology has allowed us to increase the volume of our translation work, something that is extremely valuable for a small company like Traslán. It enables us to take on extra translation work and bid for new jobs at lower and more competitive rates, yet still ensuring the high quality of our translations. This gives us a competitive edge over similar-sized companies allowing us to compete at the same level of larger translation service providers, and providing us with savings across the board.

Our own experiences and findings are in line with other research which has shown evidence that MT and post-editing can be faster and cheaper than manual translation. Allen (2004) gives a prime example of how MT plus post-editing can result in much faster translation speeds, showing that translation speeds of nearly 3 times that of the average translation speed can be achieved using MT and that an individual can achieve production rates that are 25%–30% of the time expected using traditional translation methods (without the use of MT). PAHO, similarly, have proved that MT can be successful, reporting that MT plus post-editing can produce standard quality output two to three times faster than without the use of MT technology (Vasconcellos, 1985). The PaTrans system has resulted in savings of approx. 50% (Ørsnes et al., 1996) and research from Senez (1998) cements the claim that MT together with post-editing can deliver faster throughput rates than traditional human translation, with reports of increases in translation throughput of over 260%.

3 MT at Traslán: System Overview

As mentioned previously, at Traslán we have devoted a large amount of resources for the development of our own MT system. Currently the system is only used in-house but we are in the process of widening access to the system with the development of a translation web service.

Our MT technology employs a hybrid approach, based on state-of-the-art corpus-based methods drawing from the works of Groves & Way (2005, 2006a,b), Armstrong et al. (2006) and Groves (2007). As a primarily data-driven system, during training, our MT system automatically generates its translation resources from existing bilingual aligned EN–GA corpora, producing both aligned sub-sentential segments together with statistically-weighted dictionaries. In addition, the system has the ability to make use of existing translation resources, depending on user preferences, such as bilingual glossaries, terminological databases and translation memories making it customizable to customer requirements.

During translation, the input sentence is passed through a number of pre-processing modules which deal with tokenisation, punctuation processing and formatting issues. The processed input sentence[5] is then passed onto our system's decoder which makes use of the generated translation resources, together with existing resources, to translate the source input into our initial target output.

During the decoding process, the MT system makes use of a number of different segmentation techniques prioritizing those segmentations that allow us to retrieve equivalent segments in our translation resources that maximize coverage. In favouring those segments which give us the greatest context we minimize the risk of agreement errors due to contextual conflicts, such as boundary friction (Nirenburg et al., 1993; Way, 2003), in our MT output. At the most basic level, the decoder is able to back-off to word-level translation where we do not have any suitable segments contained within our translation

[5]Note that the input to the MT system may consist of multiple sentences.

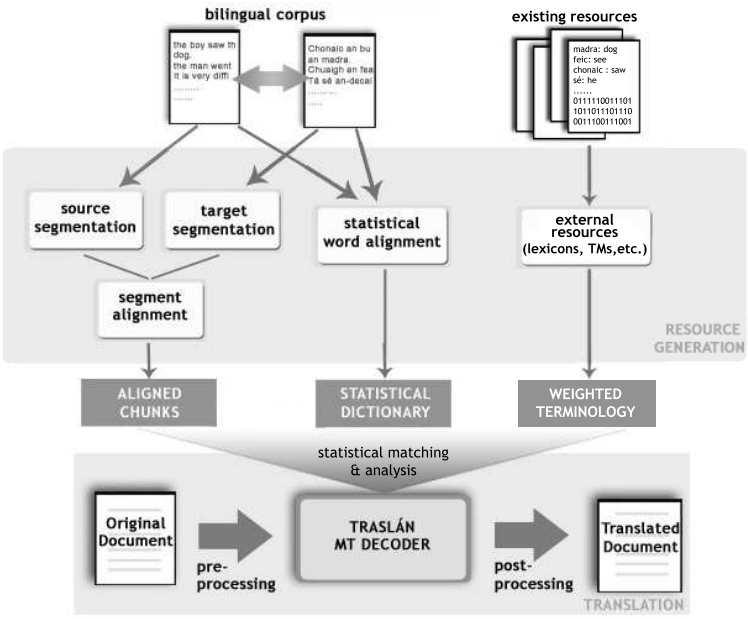

Figure 2: Overview of Traslán's MT System

resources.

The resulting MT output passes through a final post-processing module before being relayed onto our human translators to perform any post-processing necessary to produce a completely accurate translation of the source.

For smaller, individual jobs, the translators can access the MT system directly, but for the larger localization tasks the source documentation is batch processed to increase the efficiency of the translation workflow, with the output generated in a word processing format familiar to, and preferred by, our in-house translators. An overview of our system's architecture can be seen in Figure 2.

4 Bringing Humans into the Loop

Ultimately our aim in using the MT system is, of course, to save money. Once our translators use our MT system in any way, we save money. Therefore the success of an MT system is heavily dependent on human factors, such as the attitude of translators or managers to MT technology in general (O'Brien, 2006). The successful implementation of MT in the future depends not only on technological advances, but also on the training translators receive.

4.1 The Importance of User Feedback

Following this, an important and essential element of the translation workflow at Traslán is the translator feedback loop. During this stage, we gain invaluable insight from the actual users of the MT system, the translators. By collecting this feedback we can implement improvements to our technology, prioritising those issues that the translators deem to be the most important for improving the quality of the MT output and ultimately to their productivity.

Feedback from users of MT is of course beneficial during the development and evaluation of a system, but there are additional secondary benefits, as observed by Flournoy & Callison-Burch (2001). Increased feedback helps to educate the translators about the various strengths and weaknesses of the technology and enables them to make better use of the system to generate accurate translations. By observing improvements in the translations produced thanks to their feedback, their confidence in the MT output increases along with their overall acceptance of MT, something which is invaluable if we are to ensure that translators make full use of the MT technology. Educating the translators about MT and the technology involved helps the translators and post-editors understand what is behind much of the sys-

tems' behaviour and helps to explain why certain errors occur consistently. It also increases the translators' appreciation of the technology together with its limitations (O'Brien, 2006). Involving the translators also has implications for system deployment, as research has suggested that the earlier a translator is involved with the implementation of MT, the faster a usable system can be developed (Ryan, 1988).

4.2 Our Feedback Workflow

During our feedback stage of development we provide our translators with a development set of English sentences together with the output from our MT system. We encourage the translators to carry out their own internal discussions concerning the quality of the MT output before we have a direct consultation with the translators concerning any issues they have with system performance.

In general, we attempt to get translators to focus on consistent errors that the system may be producing. In doing so it allows the developers in turn to focus on resolving those issues that can be rectified and overcome by the MT system. It is equally important to educate our translators to only correct actual translation errors. If the MT system produces a perfectly valid translation, but the translator feels that it should have an alternative translation, we do not want the translator to perform any changes to the MT output; they must try to remain as objective as possible rather than give into their subjective opinions or preferred translation style. Carrying out such unnecessary post-editing effort reduces the benefits and potential cost savings of the MT system. We must remember that translation will never be perfect; for any given sentence there is often multiple possible translations, any of which can be deemed correct depending on who is doing the evaluation. We do not impose any post-editing guidelines on our translators, and instead, through our consultancy directly with the translators, these guidelines manifest themselves organically.

After collecting this feedback from the translators, further technology development is carried out. Subsequently, a further (but much more brief) evaluation is carried out by the translators to ensure that the improvements have been successful and are reflected in the translation output. An overview of this feedback loop as part of the system development is illustrated in Figure 3. At each re-development stage one or two particular issues are focused upon and then re-evaluated.

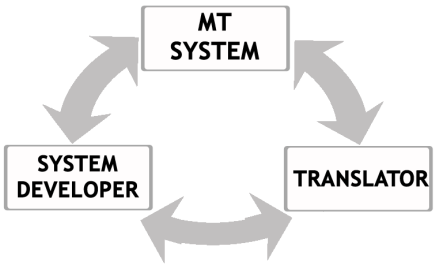

Figure 3: Feedback Workflow

This feedback process also provides our translators with training regarding the MT engine. Once familiar with the system, understanding its capabilities and accepting its limitations, our existing translators can pass on their training and knowledge to new translators. We have found that this type of peer-to-peer training is far more effective than our developers carrying out training themselves.

5 Experiments & Evaluation

In order to evaluate the performance of Traslán's MT system and to illustrate the benefits of translator feedback to system development, in this section we describe some investigative experiments on an EN–GA localization task, involving the localization of software (user-interface documents etc.).

For these experiments we made use of two of our in-house translators to provide us with feedback on the MT system performance. In addition we compared the performance of our MT system against that of a baseline state-of-the-art SMT system.

5.1 Data Resources

In order to evaluate our system performance on the localization task, from our aligned corpus we randomly generated a test set consisting of 1,000 EN–GA sentence pairs. We created an additional development set consisting of 855 randomly extracted sentence pairs. In training our MT system we made use of the remaining EN–GA aligned sentence pairs, together with an existing terminology dictionary and glossary. Details of these resources, together with our test set, are described in Table 1.

Data Resource	#Entries	#EN Words	#GA Words
Test Set	1,000	6,398	7,992
Dev Set	850	5,967	6,786
Term Dict.	2,047	3,439	5,607
Glossary	33,915	285,838	359,915
Corpus	135,192	602,179	687,919

Table 1: Data Resources used in the localization task. The number of entries specifies the number of bilingual entries, whether that be aligned sentences, aligned phrases etc.

5.2 System Performance

For our experiments we trained our system using our aligned corpus, terminology database and glossary and performed translation for EN–GA using our 1,000 sentence test set, evaluating our system performance in terms of BLEU score (Papineni et al., 2002), Precision, Recall and F-Score (Turian et al., 2003). For our comparative baseline experiments we used the same training and test sets, this time training up a standard state-of-the-art SMT system built using the MOSES decoder (Koehn et al., 2007) for which we performed Minimum-Error-Rate (MER) training on our development set using the standard set of feature functions (Och, 2003).

The results of these experiments are given in Table 2.

	BLEU	PREC	RECALL	FSCORE
Baseline	45.26	0.7578	0.7409	0.7493
Traslán	50.33	0.7558	0.7479	0.7518

Table 2: System Performance for EN-GA

From the results of our initial experiments as given in Table 2 we can see that Traslán's MT system outperforms the baseline phrase-based SMT system across almost all metrics, apart from Precision where the SMT system performs slightly better. The difference in performance is most evident when looking at BLEU scores, where we outperform MOSES by 5.07 absolute (11.2 % relative) BLEU score. The difference in performance is even more significant when we take into account that we do not perform any MER optimization on our in-house system. These results indicate that our MT engine is more than capable of producing high-quality draft translations for this particular localization task.

5.3 Translator Feedback

Following our baseline experiments, we passed the MT output onto our translators and subsequently collected their feedback and observations concerning any errors or issues with these initial automatically produced translations. We were surprised to discover that the errors most often identified by the translators referred not to translation or linguistic errors, but more to the family of formatting errors. The main errors which concerned translators were translation and realisation of punctuation, issues with spacing and capitalisation and case issues (cf. Table 3).

On reconsideration, this is probably not all that surprising. These types of errors are not particularly challenging or difficult to rectify, but prove to be tedious and time-consuming for the translator when working as post-editor and potentially take up a large proportion of the post-editors' time which could be put to more productive use. They are typical of the type of issues more often than not ignored by MT researchers, who often forget about the ultimate users of their final products, due to the fact that the overall impact of these types of errors on automatic evaluation metrics are minimal. For translators, serious and glaring translation errors are often the easiest to spot when post-editing, rather than, for example, the misplacement of a comma. Therefore it is even more important to attempt to deal these type of inaccuracies automatically.

Additionally, in the translation task for software localization, these variety of errors are extremely important, as they often are used as markers or refer to meta-information contained within the software text.

As a result of the feedback received from the translators, we improved the systems' overall handling of punctuation and restoration of spacing by developing a number of pre- and post-processing scripts. In addition to these scripts, we developed further modules for the identification and processing of do-not-translate items and for truecaseing the draft MT output. In what follows we give brief descriptions of the implementation of these two modules and their resulting affect on translation quality.

ERROR TYPE	EXAMPLE
Punctuation & Spacing	$</A> \Leftrightarrow <\ /\ A\ >$ Font{Tahoma,8pt} ⇔Clófhoirne { Tahoma , 8 pt } &Time: ⇔Am : 20pt;;;Segoe UI⇔20pt ; ; ; Segoe UI
Capitalization	<A ID=HelpMe> ⇔<a id=helpme> Expecting'.SchemeName' ⟹Ag súil le '. schemename' Image Files ∗.BMP|∗.bmp, ∗.GIF|∗.gif ⇔comhaid íohma ∗.bmp |∗.bmp , ∗.gif |∗.gif Login Options ⇔roghanna logála isteach
Specialised tokens	The '%s' file, is either missing or corrupt ⇔tá and comhad % s ar iarradh nó tuaillthe. Error Report:|0 from '%Application%' ⇔tuairisc ar earráid: | 0 as % feidhmchlár% '

Table 3: Examples of errors identified by translators

5.3.1 Truecasing & Capitalization Issues

One of the main errors which irritated our translators is the matter of truecasing. Truecasing is the process of restoring case information to badly-cased or noncased text (Lita et al., 2003). For MT, in particular for statistical data collection, it is easier and more beneficial to have the majority of your training text lowercased. In this way *The* and *the* can both contribute to statistics for the word type *the*.[6] In general, automatic evaluations are carried out disregarding case information, therefore rather than being an important enhancement for improving evaluation results, truecaseing is important for the practical application of the system in real-world tasks.

In order to address the problem of truecasing we implemented a simple solution, similar to that of Lita et al. (2003), treating truecasing as a type of reranking step. For a given top-ranking output translation we made use of a truecased language model to determine the most likely casing for the output sentence. Lita et al. (2003) have shown that making use of a language model-based approach outperforms that of a unigram model approach as it does not take into account case variations in the surrounding context. For our experiments we made use of a trigram language model created from the truecased version of the target language corpus.

[6] Although it must be realised that lowercasing is not always ideal, illustrated by the well known existence of differences in semantics based on orthography e.g. *Polish* and *polish*

During the truecasing step we build a lattice containing the case variations for each token in the sentence, considering the possibility that a token may be realised in all lowercase, first letter uppercase, all letters uppercase, and mixed case. We then maximize the score of the lattice, picking the casing sequence which gives us the highest language model probability. The truecaseing decoding is essentially a Viterbi search that computes the highest probability sequence, as indicated in Equation 1 (taken from Lita et al. (2003)):

$$q_\tau^* = argmax_{q_{i1}q_{i2}...q_{it}} P(q_{i1}q_{i2}...q_{it}|O_1O_2...O_t, \lambda) \quad (1)$$

where $P(q_{i1}q_{i2}...q_{it}|O_1O_2...O_t, \lambda)$ is the probability of a given sequence conditioned on the observation sequence and the model parameters.

This approach to truecasing, although reasonably simple, proved to be extremely efficient and managed to handle the majority of case restoration.

5.3.2 Variablization of Do-Not-Translate Items

Despite the truecaseing step, the system still failed somewhat in restoring case to those tokens which contained mixed case. These items mostly include filenames and URLs, elements that are not particularly common in more general-domain text but which occur frequently in the software localization domain. Not only do these items cause problems

when we take case restoration into account, when processed by the MT system they are often mistranslated. In actual fact, elements such as URLs and filenames should ideally be treated as do-not-translate (DNT) items, as they are generally left untranslated.

To deal with these elements we implemented a variable substitution pre-processing module within our MT system. When a sentence is input to the system, it is scanned for any DNT units (including items such as HTML tags, formatting information, filenames, URLs, placeholders etc.) by making use of various glossaries together with regular expression patterns. These DNT units are then replaced by dummy variable tokens (2). The resulting variablized sentence is translated as normal (3) and during a post-processing phase the original DNT units are re-inserted (4).

(2) < A HREF="hcp: //local/?id=d36105DcAd">
What is a Digital ID?< /A>
⟶X What is a Digital ID? Y

(3) X What is a Digital ID? Y
⟹X Cad is aitheantas digiteach ann? Y

(4) X Cad is aitheantas digiteach ann ? Y
⟶< A HREF="hcp: //local/?id=d36105DcAd">
Cad is aitheantas digiteach ann?< /A>

5.3.3 Resulting System Improvements

Taking the feedback from the translators, and incorporating the improvements outlined above, we performed additional EN–GA experiments, using the same training and test data as before. The results for these experiments along side our initial baseline experiments are shown in Table 4, where:

- TraslánII: The original system, but including improvements to the treatment of punctuation, spacing etc.
- TraslánII + VAR: Making use of the new variablization module in addition to the improvements to the treatment of punctuation.

From the results in Table 4, we can see that incorporating the developments resulting from feedback we received from our translators, we observed improvements in translation quality across all of the automatic evaluation metrics, for both the enhancements within our punctuation pre-processing module and the addition of our variablization technique. Improvements in the treatment of punctuation and spacing showed the most increase, with BLEU score displaying a 5.3% relative improvement (2.65 absolute) and F-Score increasing from 0.7493 to 0.7518. We observed further increases when applying our variablization techniques, with slight improvements across all metrics. The TraslánII + VAR configuration outperformed the baseline MOSES-based system in terms of BLEU score by 10 points absolute with a 3% relative difference in F-Scores between the two systems. More importantly, however, our translators observed that the use of variablization techniques were effective in preserving case information and helped to reduce the amount of post-editing effort required for DNT items.

	BLEU	PREC	RECALL	FSCORE
Baseline	45.26	0.7578	0.7409	0.7493
Traslán	50.33	0.7558	0.7479	0.7518
TraslánII	52.98	0.7783	0.7641	0.7711
TraslánII +VAR	53.30	0.7796	0.7652	0.7724

Table 4: System Performance after Translator Feedback for EN–GA

6 Conclusions

MT together with post-editing presents itself as a practical solution to address the ever-increasing demand for translation services. At Traslán, developing our own translation technology has enabled us to generate average savings of 20% and has allowed us to cope with greater volumes of translation and offer quicker turnaround times, without sacrificing translation quality.

As MT researchers and developers it is essential that we keep the end user in mind. If translators are ever to adopt and make full use of MT, we need to involve them in the development process. In doing so we can raise their knowledge and understanding of MT, which in turn increases their confidence in the technology, encouraging them to accept it and exploit it as an invaluable aid in the translation process. At Traslán we keep our translators in the loop during the development of our in-house engine; their

feedback and evaluation feeding directly into system improvements. In turn, their acceptance of MT, together with their knowledge its capabilities and limitations, has grown resulting in increased efficiency and productivity. From the results shown in Section 5, we can clearly see the direct benefits of including this feedback in the development cycle, benefits which also further encourage translator interaction.

7 Future Work

Although we work primarily within the EN–GA translation domain, we also provide translation services for additional language pairs, in particular Eastern European languages. Since the accession of additional Eastern European states into the European Union, in May 2004, the number of people speaking these languages in Ireland has steadily increased. Taking Polish as an example, today, between 70,000 – 150,000 native Polish speakers reportedly reside in Ireland, signifying over a 90% increase on the figures for 2002, with the Polish population making up the largest minority within Ireland today. There has also been a significant increase in other Eastern European residents, the largest increases being within the Latvian and Lithuanian communities, with a large proportion of citizens from these countries arriving in Ireland in 2005 or later.

As part of our plan to branch out into these new languages, we have carried out some preliminary experiments into the use of MT for Polish, employing data from the JRC-Aquis corpus[7] (Steinberger et al., 2006) as part of a pilot study. Our initial experiments have been extremely promising encouraging further research and development in this area.

In terms of system development, we plan to further improve the integration of our MT engine into our translation workflow by developing a feature-rich GUI-based translation environment. Currently our MT engine is only employed in-house, but we are in the process of making our MT system available as a web service, facilitating easy integration and use by external clients.

[7] http://langtech.jrc.it/JRC-Acquis.html

Acknowledgements

The author would like to thank Andy Way and Doncha Ó'Cróinín for their contributions to this paper and Mike Dillinger for his invaluable suggestions and keynote invitation. The pilot English–Polish experiment was kindly funded by Enterprise Ireland.[8]

[8] http://www.enterprise-ireland.com/

References

Allen, J. (2004). Case Study: Implementing MT for the Translation of Pre-sales Marketing and Post-sales Software Deployment Documentation at Mycom International. In *Machine Translation: From Real Users to Research.* Springer Berlin / Heidelberg.

Armstrong, S., Flanagan, M., Graham, Y., Groves, D., Mellebeek, B., Morrissey, S., et al. (2006). MaTrEx: Machine Translation Using Examples. In *TC-Star Open-Lab on Speech Translation.* Trento, Italy. (http://www.computing.dcu.ie/∼away/PUBS/2006/Matrex.pdf)

DePalma, D. A. (2006). Evaluating MT prior to deployment. In *7th Conference of the Association for Machine Translation in the Americas, "Visions for the Future of Machine Translation".* Cambridge, Massachusetts. (presentation on user track)

Flournoy, R., & Callison-Burch, C. (2001). Secondary Benefits of Feedback and User Interaction in Machine Translation Tools. *Workshop paper for MT2010: Towards a Roadmap for MT of the MT Summit VIII*.

Groves, D. (2007). *Hybrid Data-Driven Models of Machine Translation*. PhD Thesis, Dublin City University, Dublin, Ireland.

Groves, D., & Way, A. (2005). Hybrid Example-Based SMT: the Best of Both Worlds? In *Proceedings of the ACL 2005 Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond* (pp. 183–190). Ann Arbor, MI..

Groves, D., & Way, A. (2006a). Hybrid Data-Driven Models of MT. In *Machine Translation, Special Issue on EBMT* (Vol. 19, pp. 301–323).

Groves, D., & Way, A. (2006b). Hybridity in MT: Experiments on the Europarl Corpus. In *Proceedings of the 11th Conference of the European Association for Machine Translation* (pp. 115–124). Oslo, Norway.

Hutchins, W. (2001). Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, *13*(1-2), 5–20.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., et al. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions* (pp. 177–180). Prague, Cz. Rep..

Lita, L., Ittycheriah, A., Roukos, S., & Kambhatla, N. (2003). tRuEcasIng. In *41st Annual Meeting of the Association for Computational Linguistics (ACL-03)* (pp. 152–159). Sapporo, Japan.

Nirenburg, S., Domashnev, C., & Grannes, D. J. (1993). Two Approaches to Matching in Example-Based Machine Translation. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation TMI '93: MT in the Next Generation* (pp. 47–57). Kyoto, Japan.

O'Brien, S. (2006). *Machine Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis*. PhD Thesis, Dublin City University, Dublin, Ireland.

Och, F. (2003). Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of 41st Annual Meeting of the Association for Computational Linguistics (ACL-03)* (pp. 160–167). Sapporo, Japan.

Ørsnes, B., Music, B., & Maegaard, B. (1996). PaTrans: A Patent Translation System. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96) – Volume 2* (pp. 1115–1118). Morristown, NJ.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)* (pp. 311–318). Philadelphia, PA.

Ryan, J. P. (1988). The Role of the Translator in Making an MT System Work: Perspective of a Developer. In M. Vasconcellos (Ed.), *Technology as Translation Strategy, American Translators Association Scholarly Monograph Series, Vol. II* (pp. 127–132). State University of New York at Binghamton (SUNY).

Senez, D. (1998). Post-Editing Service for Machine Translation Users at the European Commission. In *Translating and the Computer 20.* London: Aslib. (no page numbers)

Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D., et al. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (pp. 2142–2147). Genoa, Italy.

Turian, J., Shen, L., & Melamed, I. D. (2003). Evaluation of Machine Translation and its Evaluation. In *Machine Translation Summit IX* (pp. 386–393). New Orleans, LA.

Valderrábanos, A. S., Esteban, J., & Iraola, L. (2003). TransType2 - A New Paradigm for Translation Automation. In *Machine Translation Summit IX* (pp. 23–27). New Orleans, LA.

van deer Meer, J. (2006). The Emergence of FAUT: Fully Automatic Useful Translation. In *Keynote at the 11th Conference of the European Association for Machine Translation.* Oslo, Norway.

Vasconcellos, M. (1985). Management of the Machine Translation Environment: Interaction of Functions at the Pan American Health Organization. In V. Lawson (Ed.), *Tools for the Trade, Translating and the Computer 5* (pp. 115–129). London: Aslib.

Way, A. (2003). Translating With Examples: the LFG-DOT Models of Translation. In M. Carl & A. Way (Eds.), *Recent Advances in Example-Based Machine Translation* (p. 443-472). Dordrecht, The Netherlands: Kluwer Academic Publishers.