Identification and Analysis of Post-Editing Patterns for MT

Declan Groves
Traslán
Co. Wicklow,
Ireland
dgroves@traslan.ie

Dag Schmidtke
Microsoft Corporation
Dublin,
Ireland
dags@microsoft.com

Abstract

For this work we have carried out a number of analysis experiments comparing raw MT output produced by Microsoft's Treelet MT engine (Quirk et al., 2005) with its human post-edited counterpart, for English–German and English–French. Through these experiments we identify a number of interesting post-editing patterns, both textual (string-based) and constituent-based. In this paper we discuss our analysis methodologies, present some of our results and provide information on how this type of analysis can be of benefit to translation systems and post-editors, with a view to improving initial MT output and consequently post-editor productivity. In addition, we also discuss the MT and post-editing workflow at Microsoft and results from MT post-editing pilots for a number of different language pairs.

1 Introduction

Post-editing can be considered the correction and perfection of content already automatically translated (in contrast to the task of 'revision', which although similar in some aspects, deals with the error correction of human-produced draft translations). It is the task of the post-editor to edit, modify and/or correct pre-translated text that has been processed by a machine translation system from a source language into (a) target language(s) (Allen, 2003). Such edits can involve correcting errors involving punctuation, inappropriate glosses, misconstructions of meaning, misspellings, mistakes in numerals, incorrect attachment and incorrect ambiguity resolutions amongst other things (McElhaney & Vasconcellos, 1988).

Post-editing is essential in ensuring high-quality translation output wherever MT is used to provide initial draft translations. Once we have invested time and money into developing and implementing an MT system many of the costs of translation fall on any pre- and post-processing necessary to produce high-quality output translations (cf. Figure 1).[1] Therefore, if we wish to maximize the value of MT, and decrease translation costs, we need to look at these areas at the front- and back-end of the translation workflow.

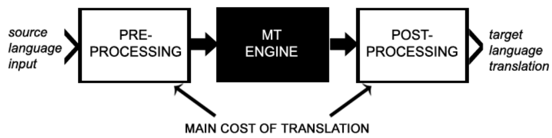

Fig. 1: The sources of translation costs when using MT

An MT system will output the same error whenever it is given the same input, assuming all other conditions are also the same. If the MT output is used other than for gisting, it is therefore left up to the post-editor to correct these reoccurring errors on a regular basis, which can prove to be a time-consuming task and significantly reduce the productivity of the post-editor, thus increasing the cost of translation. Several organisations, including PAHO, Caterpillar, General Motors and EC Translation Services, have carried out internal studies with regard to the post-editing process with the aim of reducing the human effort, and therefore the cost involved. However, information and results about the methodology used in these studies have very rarely been made available publicly. For this research we carry out a number of experiments focusing on comparing raw MT output with its post-edited counterpart so as to provide potentially useful information which can then, in turn, be ultimately exploited to improve the translation process, thus reducing post-editing effort, time and cost.

The remainder of this paper is organised as follows: in Section 2 we give an overview of the MT and post-editing workflow at Microsoft and the methodology used to measure post-editor productivity, together with some pilot experiment results on post-editor productivity gains.

[1] Of course, other cost factors are involved in the use of MT including system customisation and post-editor training.

In Section 3 we describe the data sets used in this research and Section 4 presents string-based post-editing analysis experiments and results. Section 5 outlines additional structural- (constituent-) based comparative experiments. We conclude by summarising our findings before giving indications towards possible future research in this area.

2 MT and Post-Editing Workflow at Microsoft

MT post-editing is used in the localisation process for several Microsoft products. It is typically used in conjunction with translation memory and recycling tools, so that only text which cannot be recycled is passed to MT. The Microsoft Treelet system (Quirk et al., 2005) is used, trained on a technical domain (Microsoft's own localised products).

To measure and evaluate post-editing results, Microsoft has developed a methodology for productivity tracking, intended for larger translation projects with substantial volumes. In our experience raw-MT evaluation metrics such as BLEU (Papineni et al., 2002), and human evaluations and ratings of smaller samples, are not in themselves sufficient or representative indicators of how useful MT will be on specific larger post-editing projects.

Our approach is based on translators logging time taken to translate documents, with low levels of recycling, with and without use of MT. This establishes both a human translation productivity baseline for a specific translator, and the post-editing productivity for the same translator. The measurements are taken for 3 representative translators, as selected by the translation vendor; one of average productivity, one new to the project (lower productivity), and one expert translator, of higher productivity, which helps to give a representative average and account for productivity variations between translators. For each translator, productivity is measured with and without MT and results are averaged to calculate an overall productivity gain, from measured average human translation productivity to measured average post-editing productivity, for a specific project and product.

In addition to productivity tracking, feedback is also gathered on specific issues encountered in post-editing. Typical issues reported by translators include grammar and terminology errors, and incorrect handling of markup and formatting. For some of these issues special handling has been developed to ensure terminological consistency (Itagaki et al., 2007; Itagaki & Aikawa, 2008). For Microsoft Office post-editing special handling has also been added to the MT processing workflow for Help documentation to accurately translate User Interface terminology, which is marked up with <ui>< /ui> tags in text.

Microsoft have found that MT quality improves over time and we can observe related productivity increases as translators become more familiar with MT and post-editing. These improvements have resulted in productivity gains increasing from 5-10%, to 10-20%, for selected languages.

We also have an increased understanding of the factors that influence MT quality and translator productivity. There can be substantial variations in post-editing productivity, for the same language, between different projects and products, different handoffs during the same project, and between different translators. This indicates that MT language quality itself is only one of several important factors influencing productivity. How closely the text to be MT'd correlates with training data is an obvious quality factor. Certain types of text work better for MT than others, and a formal writing style helps, although this may be more a matter of what text types are most common in training data rather than inherent issues with MT itself.

With respect to source text quality, using controlled language improves both MT quality and post-editing productivity (Aikawa et al., 2007) and authoring support tools in the form of lightweight style guide checkers are in use in several groups in Microsoft, to help ensure consistent, more readable 'Global English'. These tools have been demonstrated internally to have a positive impact on MT quality. Additional areas of the authoring and translation life-cycle which influence post-editing productivity include the MT processing stage, and the post-editing stage itself. Microsoft has developed specific post-editing guidelines and training sessions are frequently held at the start of translation projects to ensure translators are familiar with how to make the best use of MT. Translator attitudes to MT are an important factor. Close engagement and dialogue, and incremental improvements, are important factors in establishing confidence and a positive attitude to post-editing (Groves, 2008).

Aggregated post-editing results for a number of selected projects are shown below in Tables 1 and 2. Table 1 shows selected results for Office Online end-user Help documentation.

Language	Productivity Gain
French	14.5%
Brazilian	20.0%
Swedish	8.0%
Danish	28.6%
Czech	6.1%
Dutch	14.7%

Table 1: MT post-editing pilot results, Office Online Continuous Publishing

Considerable variation between similar languages such as Swedish and Danish for this specific project is more likely caused by factors at the translation stage rather than significant core MT system quality differences between the languages. This illustrates why productivity measurement is necessary to asses suitability of MT for a specific project and why standard MT evaluation metrics are not always reliable.

Table 2 shows selected results for Office technical documentation for MSDN and TechNet, with some variations in productivity gain from the end-user documentation.

Language	**Productivity Gain**
French	6.6%
Chinese (Simplified)	5.9%
German	16.0%

Table 2: MT post-editing pilot results, Office MSDN and Technet

While the post-editing projects and the tracking methodology described in this section have provided a lot of useful information for Microsoft and significantly improved our understanding of what it takes to make post-editing successful, there are still some issues in feedback gathering. Translators provide a lot of useful feedback on specific issues, and also indicate which areas cause difficulty. It can be difficult though to quantify how frequent a problem is, and hence where the priority should be placed in improving the system. This is the main motivation for the current study; to complement post-editing tracking and translator feedback with quantifiable data analysis on what actually gets changed in post-editing.

3 Data Sets Used

For this research, we made use of data for English–German and English–French taken from the larger 1.3M word data sets used for the pilot experiments from Table 2. Each data set consists of source language (English) sentences together with raw-MT output produced by the Microsoft Treelet MT system in the target language and the same MT output post-edited to meet human quality standards.[2] Details of the data sets, in terms of the number of sentences and words are outlined in Table 3, for the English source ('SOURCE'), raw MT output ('MT') and post-edited MT text ('PEMT'). Additional information is also given regarding minimum, maximum and average sentence lengths in terms of words.

Source language sentence length can be an important factor to consider when employing MT and can be useful in predicting the likelihood of the MT engine producing a good quality target language translation, with the risk of potential translation ambiguity, and therefore translation error, increasing as sentence length increases. The frequency distributions across source language sentences,

[2]Note that we refer to the data in terms of sentences throughout the paper, when in fact the input, MT output and post-edited output may consist of multiple sentences or single individual terms.

		GERMAN	FRENCH
# Sentences		9,454	5,400
# Words	SOURCE	138,578	67,924
	MT	124,831	82,497
	PEMT	135,728	84,364
Min. Sent. Length		1	1
Max. Sent. Length		74	81
Avg. Sent. Length	SOURCE	14.66	12.58
	MT	13.20	15.28
	PEMT	14.36	15.62

Table 3: Details for English–German and English–French data sets, with sentence lengths given in words.

based on source sentence length, for both language pairs are given in Figure 2. From the graph in Figure 2, we can see that the distribution of sentence lengths follows similar patterns for English–German and English–French, with approx 72.5% of source sentences from the English–German data set and 76% of the English–French source sentences set having lengths of less than 20 words.

Fig. 2: Frequency Graph of source sentence lengths for English–German and English–French data sets

4 String-Based Edit-Distance Analysis

Taking the data described in Section 2, we carried out some initial analysis using 'edit distance', or Levenshtein distance (Levenshtein, 1965). We made use of standard dynamic programming techniques to calculate the edit distance between the MT and post-edited MT strings, building a distance matrix, D, where the distance between the i^{th} word in the MT string, S, and the j^{th} word in the post-edited string, T, is calculated according to the formula given in Equation 1.

$$D_{i,j} = \min \begin{cases} D_{i-1,j} + 1 \\ D_{i,j-1} + 1 \\ D_{i-1,j-1} + (\text{if } S_i = T_j \text{ then } 1, \text{else } 0) \end{cases} \quad (1)$$

Using this formula we calculated the edit distance across all MT and post-edited MT tuples. The resulting average word-based edit distance for French was 5.60, whereas German received an average edit distance score of 8.81, indicating that translating into German seems to present a more difficult post-editing task than French.

Figure 3 displays graphs plotting the normalised edit distance against source sentence length for English–German and English–French, with edit distance scores normalised according to the formula in Equation (2), where $|S|$ and $|T|$ are the lengths of the MT output sentence S and post-edited output T.

$$\frac{distance(S,T)}{\max|S|,|T|} \tag{2}$$

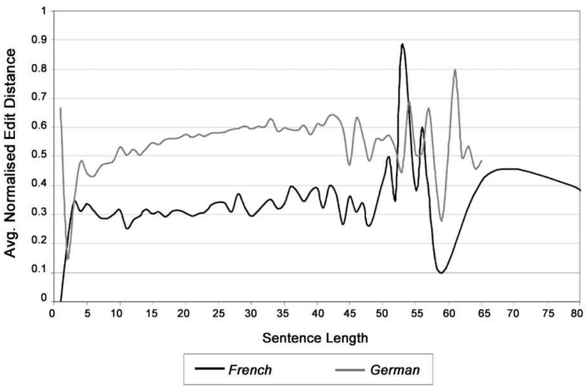

Fig. 3: Graph displaying average normalised edit distance per source sentence length for English–German and English–French data sets

The graph in Figure 3 highlights that post-editing effort for German appears to be significantly greater (and less consistent) than that for French, in terms of edit distance. For English–German we get an average normalised edit distance score across the MT of 0.58, compared with 0.34 for French. The normalised edit distance remains relatively constant across sentence lengths, but we do observe an increase in the edit distance as we move towards longer sentences, with values becoming less stable as we reach sentences greater than 40 words in length, more so for German than for French. Interestingly, we observe a peak for very short sentences (<5 words) which we have found in previous pilot experiments can be problematic for translation, often caused by lack of context to better inform the MT system.

4.1 Extraction of Post-Editing Patterns

In addition to calculating edit distance scores, we carried out some evaluations on the types of edits that were being performed by the post-editors by tracing the minimum edit distance path through the resulting distance matrix.

When filling the distance matrix we hold onto back-pointers which allow us to trace a path back through the matrix. A non-diagonal move indicates that either a deletion or insertion edit has been performed. If we move diagonally within the matrix without any reduction in the current edit distance score (i.e. where we have $D_{i,j} = D_{i-1,j-1}$), no edits have been performed. However, if when moving diagonally there is a reduction in the distance this indicates that a substitution has been made.

4.2 Post-Editing Patterns Identified

Using this method we extracted editing operations across the full data sets. Table 4 gives the top 10 most common edits according for English–German and English–French.

	GERMAN	FRENCH
1.	, ⟶NULL	NULL ⟶de
2.	NULL ⟶die	de ⟶NULL
3.	Sie ⟶NULL	les ⟶NULL
4.	der ⟶die	le ⟶NULL
5.	die ⟶NULL	NULL ⟶les
6.	NULL ⟶in	la ⟶NULL
7.	die ⟶der	des ⟶les
8.	der ⟶NULL	NULL ⟶le
9.	NULL ⟶,	á ⟶NULL
10.	zu ⟶NULL	l' ⟶NULL

Table 4: The top 10 most frequent edits (based on the minimum edit distance path) for German and French

From the table we can see that for both German and French, edits involving the deletion and insertion of function words appear to be by far the most common type of editing action carried out (in Table 4 NULL represents the empty token), with 70% of edits for French and 42% for German involve determiners alone. Edits in punctuation also appear to be amongst the most frequent, with the removal and insertion of commas being particularly prevalent.

Terminology is frequently mentioned as a very important issue for MT. However, internally Microsoft have found that for a small sample evaluation for one particular product, inconsistencies in terminology have only accounted for approx. 20% of errors. In line with this observation, for the data used in this particular study, lexical patterns involving NOUN ⇔NOUN edits account for 15% of the total patterns for German and just over 11% for French.

4.2.1 English–German

In German, the top 3 most common edits occurred with very similar frequency. The removal of commas and the insertion of the determiner *die* were the most common edits. German nouns require articles, but common computer string usage has led to the tendency to drop leading articles in English resulting in not entirely ungrammatical translations, but less formal and more direct translations into German which often results in the absence of required articles, such as *die* as in the examples in (1).

(1) a. Development for Windows SharePoint Services 3.0 technology ⇔
Entwicklung für Windows SharePoint Services 3.0 Technologie ⟶
Entwicklung für <u>die</u> Windows SharePoint Services3.0-Technologie

b. As shown in the following graph, data connections can have a significant impact on performance ⇔
Wie in dem folgenden Diagramm dargestellt, können Datenverbindungen eine erhebliche Auswirkungen auf Leistung haben ⟶
Wie in dem folgenden Diagramm dargestellt ist, können Datenverbindungen eine erhebliche Auswirkung auf <u>die</u> Leistung haben

The deletion of the pronoun *Sie* also proves to be particularly interesting. Further analysis reveals a prevalence for the deletion in the translation of imperative sentences, where the post-editor has dropped *Sie* as subject in the final translation, as in the examples given in (2). On further inspection, 60% of the MT-produced translations of imperatives contain *Sie* where as 21% of the equivalent post-edited sentences do not, giving it a relatively high overall deletion rate.

(2) a. Print the labels ⇔
Drucken <u>Sie</u> die Etiketten ⟹Drucken der Etiketten

b. Change the shaded background of fields ⇔
Ändern <u>Sie</u> den schattierten Hintergrund der Felder ⟹Ändern des schattierten Hintergrunds von Feldern

4.2.2 English–French

As with German, all of the most frequently occurring edits involve function words, with the top 2 candidates involving the deletion and insertion of the function word *de*, which is the most common word in the French language and can be used as a preposition, determiner and in the formation of compounds, often occurring in French where no equivalent direct translation is needed in English. 43.2% of MT-produced French sentences undergo either an insertion, deletion or substitution involving the word *de* (accounting for 40.33%, 21.39% and 75.83% of these edits, respectively) during post-editing. Edits involving *de* are approx. 58% more frequent than the second most common candidate word for editing, *les* thus further indicating the prevalence of *de*-related edits. Examples taken from the English–French corpus illustrating the insertion, deletion and substitution of *de* are given in (3)a, (3)b and (3)c, respectively.

(3) a. Web Part Discovery Service ⇔
Service de Découverte Partie Web
⟹Service de Découverte <u>de</u> les Composants Webpart

b. Specifies options for a drop-down list ⇔
Spécifie <u>de</u> les options pour une liste déroulante
⟹Spécifie les options pour une liste déroulante

c. Specifies the display size for the field ⇔
Spécifie la taille d'affichage <u>pour</u> le champ.
⟹Spécifie la taille d'affichage <u>de</u> le champ.

(3)a illustrates how *de* is often inserted when translating noun compounds into French. This type of translation accounts for the large majority of edits involving *de*. This example also highlights a particular gap in terminology for French where the translation of *webpart* should be *les composants webpart*, rather than the more literal *partie web*. The remainder of changes listed in Table 4 for French and German all occur with similar frequency.

5 Structural Based Comparison

To attempt to gather more informative editing patterns than those described in Section 4.2, we carried out some structural-based analysis, making use of the parse structure of the MT and post-edited MT (PEMT) data. We first extracted word and phrase alignments between the MT and PEMT strings using standard statistical phrase extraction methods (Och & Ney, 2003). We then identified corresponding constituents (or sub-trees) within the MT and PEMT parse trees which spanned the extracted phrases. This gave us a set of constituent-based alignments which we then collected over the entire data set. From the resulting tree-based alignments, we determined those alignments which represented post-edits. Those extracted constituents with identical tree structures (isomorphic) represent lexical, morphological or orthographic changes, whereas non-isomorphic alignments identify more significant changes in structure. In the following sections we discuss the types of patterns extracted for both the German and French data sets.

5.1 English–German

The top 10 most common constituent-based editing patterns for the German data set are given in Table 5.

Changes in part of speech for commas rank highly in the list, and often indicate significant structural differences between the MT and post-edited data, for example, changes in attachment of adjuncts and prepositional phrases. Changes in the determiners *die*, *der* and *den* account for 6 of the top 10 most common edits. Due to the complex linguistic nature of German case and gender, it is not surprising that such edits rank so highly, and correlates closely with those presented previously in Table 4.

	MT	PE-MT
1	(CHAR (,))	(CONJ (,))
2	(CONJ (,))	(CHAR (,))
3	(DETP (ADJ (die)))	(DETP (ADJ (der)))
4	(DETP (ADJ (der)))	(DETP (ADJ (die)))
5	(DETP (ADJ (eine)))	(DETP (ADJ (einer)))
6	(NP (PRON (die)))	(DETP (ADJ (die)))
7	(DETP (ADJ (die)))	(DETP (ADJ (den)))
8	(ADJ (der))	(PRON (der))
9	(PRON (der))	(ADJ (der))
10	(NOUN (sites))	(NOUN (websites))

Table 5: German MT & PEMT constituent-based editing patterns

Generally, this type of edit is also very much dependent on the surrounding context which indicates which form of the article is to be realised, as can be seen in the examples in Figures 4 and 5.

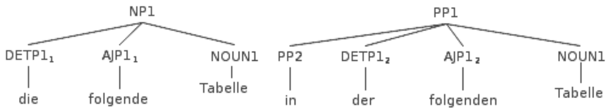

Fig. 4: Substitution of *die* with *der* in German; deriving a dative.

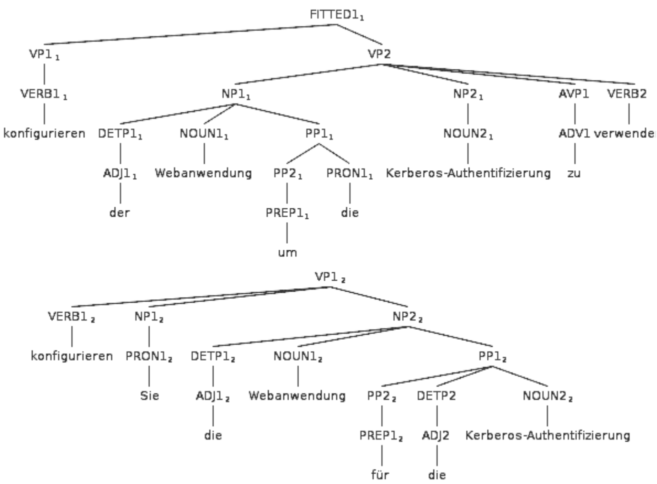

Fig. 5: Substitution of *der* with *die* in a German imperative

The pattern in Table 5 (10) indicates a preference of terminology, where *websites*, rather than *sites*, is the preferred German translation of the English word *sites* by the post-editors in the majority of cases.

Table 6 generalises over the extracted constituent-based patterns, ignoring the lexical items, and presents those patterns which represent changes in structure, rather than just lexical-based changes. This allows us to see more general trends over the significant edits made by post-editors. From the patterns in Table 6, we can see that all of the most common patterns describe transformations involving nouns. Patterns 3, 4, 5 and 7 involve the compounding of nouns (e.g. Figure 6). Pattern 1, 6 and 9 deal with the insertion of determiners (e.g. Figure 7) or adjectives into noun phrases, whereas pattern 10 represents the insertion of a preposition into a noun phrase.

	MT	PE-MT
1	(NP (NOUN))	(NP (DETP) (NOUN))
2	(NP (NOUN))	(PP (PP (PREP)) (NOUN))
3	(FITTED (NP (NOUN)) (NP (NOUN)) (CHAR))	(NP (NOUN) (CHAR))
4	(FITTED (NP (AVP (NOUN) (CHAR)) (NOUN)) (NP (NOUN) (NAPPOS (NOUN))))	(NP (AVP (NOUN) (CHAR)) (NOUN) (NAPPOS (NOUN)))
5	(NP (NOUN) (NAPPOS (NOUN)))	(NP (NOUN))
6	(AVP (NOUN))	(NP (NOUN) (POSS (DETP) (NOUN)))
7	(NP (NOUN) (NAPPOS (NOUN)))	(NP (NOUN))
8	(PP (PP (PREP)) (NOUN))	(PP (PP (PREP)) (DETP) (NOUN))
9	(NP (NOUN))	(POSS (DETP) (NOUN))
10	(NP (NOUN) (NAPPOS (NOUN)))	(NP (NOUN) (PP (PP (PREP)) (NOUN)))

Table 6: German MT & PEMT structural editing patterns: where there are changes in parse-tree structure. FITTED indicates a robust parse for an ungrammatical sentence.

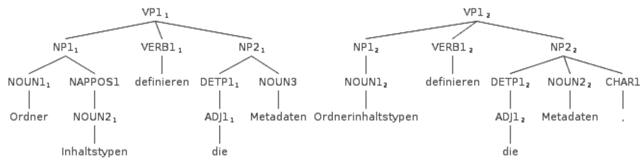

Fig. 6: Compounding of nouns in German.

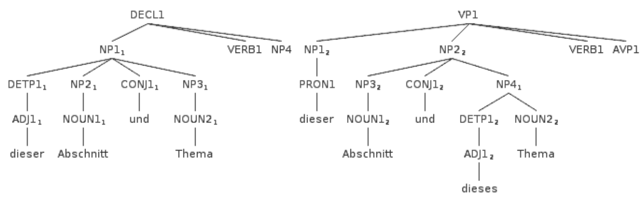

Fig. 7: Insertion of determiner in German NP.

5.2 English–French

Table 7 lists the most common constituent-based editing patterns for the French data set, where we can see that, as with German, changes in the part of speech tag for commas are amongst the most commonly observed edits, together with changes in the use of function words.

	MT	PE-MT
1	(DETP (ADJ (des)))	(DETP (ADJ (les)))
2	(CHAR (,))	(CONJ (,))
3	(DETP (ADJ (les)))	(DETP (ADJ (des)))
4	(CONJ (,))	(CHAR (,))
5	(NOUN (document))	(NOUN (documents))
6	(PP (PREP (dans)))	(PP (PREP (de)))
7	(NOUN (workflow))	(NOUN (flux))
8	(PP (PREP (de)))	(AJP (ADJ (de)))
9	(ADJ (l'))	(CHAR (,))
10	(VERB (consultez))	(VERB (voir))

Table 7: French MT & PEMT constituent-based editing patterns

The most common pattern is the replacement of the plural determiner *des* with *les*, and vice-versa as seen in patterns 1 and 3. The word *document* in French is also frequently replaced by its plural equivalent *documents* in the translation of noun phrases (which also often includes the insertion of the determiner *de*) as in (4)a. We can see that terminology comes into play in pattern 7, where although *workflow* is indicated as being replaced by *flux* in Table 5, in actual fact on further analysis it is edited to create the corresponding French noun compound *flux de travail* (cf. (4)b). Pattern 10 suggests a stylistic preference for the use of *voir* over *consultez* as the translation of the English word *see*, as in the example in (4)c.

(4) a. Document Parser Interface ⇔
l'Interface de l'Analyseur Document ⟹
l'Interface de l'Analyseur <u>de documents</u>.

b. Business processes are represented by workflows⇔
Processus d' entreprise sont representes par les workflows⟹
Les processus metiers sont representes par les flux de travail

c. For more information, see WPSC services ⇔
Pour plus d'informations, <u>consultez</u> WPSC services ⟹
Pour plus d'informations, <u>voir</u> WPSC services

	MT	PE-MT
1	(NP (NOUN) (NP (NOUN)))	(NP (NP (NOUN)) (CONJ) (NP (NOUN)))
2	(NP (DETP) (NOUN))	(NP (NOUN))
3	(PP (PP (PREP)) (NOUN))	(PP (PP (PREP)) (DETP) (NOUN))
4	(NP (NOUN) (NP (NOUN)) (CHAR))	(NP (NOUN) (CHAR) (NAPPOS (NOUN)))
5	(NP (NOUN) (PP (PP (PREP)) (NOUN)))	(NP (NP (NOUN)) (CONJ) (NP (NOUN)))
6	(NP (NOUN))	(NP (DETP) (NOUN))
7	(NP (NOUN) (NP (NOUN)))	(NP (NOUN))
8	(NP (NOUN) (PP (PP (PREP)) (NOUN)))	NP (NOUN) (PP (PP (PREP)) ((DETP) (NOUN)))
9	(NP (DETP) (NOUN) (PP (PP (PREP)) (NOUN)))	(NP (NOUN) (PP (PP (PREP)) (NOUN)))
10	(NP (NOUN) (PP (PP (PREP)) (NOUN)))	(NP (NOUN) (PP (PP (PREP)) (DETP) (NOUN)))

Table 8: French MT & PEMT structural editing patterns: where there are changes in parse-tree structure

Looking at the generalised structural patterns in Table 8, we can see that the insertion of conjunctions into noun phrases is the most common operation (pattern 4 and pattern 5 also reflects this, where conjunctions may be alternatively tagged as CHAR rather than CONJ, which is generally due to the fact that the segment is recognised as a system command by the parser). Looking at the data, we observe that these instances consist mostly of commands where the post-editor inserts a comma into a noun phrase as in the examples given in (5).

(5) a. doc_version Paramètre. ⟹doc_version, paramètre.
b. mèthode dialogview. ⟹dialogview , mèthode .

The insertion of determiners (Figures 8, 9 and 10) is the most common editing operation overall (occurring in patterns 3, 6, 8 and 10), which correlates with the edit-distance patterns extracted previously (cf. Table 4).

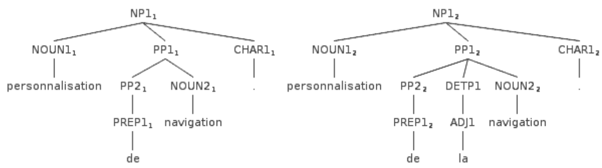

Fig. 8: Insertion of the determiner *la* in French

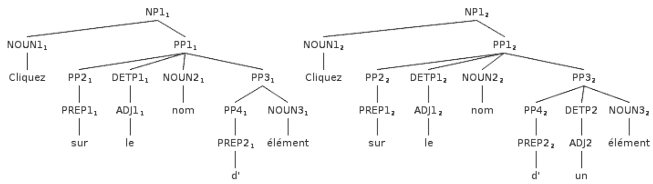

Fig. 9: Insertion of the determiner *un* in French

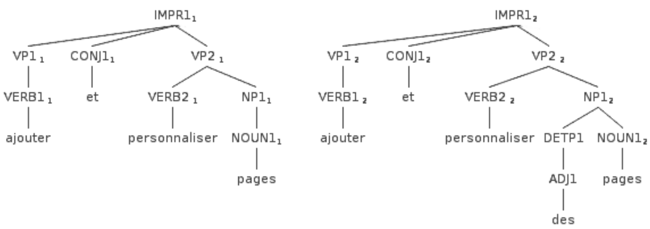

Fig. 10: Insertion of the determiner *des* in French

The deletion of function words is present in patterns 2 and 9. Examining the data reveals that this deletion involves determiners exclusively and occurs occasionally in the translation of single nouns, which can exist as standalone segments in menus and lists, as in the examples given in (6).

(6) a. Parameters ⇔Les Paramètres ⟹Paramètres
b. Indexers⇔Les Indexeurs⟹Indexeurs
c. Adding ⇔L'ajout ⟹Ajout
d. Document Libraries. ⇔Les Bibliothèques de Documents. ⟹Bibliothèques de Documents.

The last remaining pattern in Table 8, pattern 7 describes the compounding of nouns, as illustrated in Figure 11, which can suggest potential improvements to the MT system's termbanks.

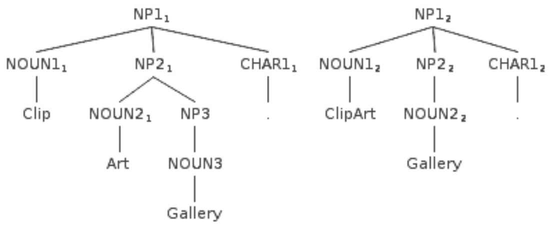

Fig. 11: Compounding of nouns and resulting changes in structure in French

6 Summary and Future Work

In this paper we have given an introduction to how MT together with post-editing is implemented at Microsoft, and have given a description of how the productivity of post-editors can be evaluated. We have also introduced both string-based and constituent-based methods of identifying post-editing changes made when perfecting draft MT-produced translations. Both methods provide us with some interesting insights into the common types of edits carried out by post-editors.

We plan to carry out further investigation of the patterns generated. One area of interest is correlating raw-MT output with training data to see how it may influence some of the patterns identified. A related aspect is assessing the impact of the source sentence and how the patterns can help guide selection of controlled language rules. Another area of assessment is the evaluation of the validity and necessity of the post-edit changes. We would also like to investigate statistical post-editing techniques which can help resolve consistent MT errors and reduce post-editing effort (Simard et al., 2007; Aikawa & Ruopp, 2009). We feel the pattern analysis techniques described in this paper can help us to evaluate the benefits of such a post-processing step.

References

Aikawa, T., & Ruopp, A. (2009). Chained System: A Linear Combination of Different Types of Statistical Machine Translation Systems. In *Proceedings of MT Summit XII.* Ottawa, Canada (to appear).

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. (2007). Impact of Controlled Language on Translation Quality and Post-editing in a Statistical Machine Translation Environment. In *Machine Translation Summit XI* (pp. 10–14). Copenhagen, Denmark.

Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation: A Translators Guide.* Benjamins Translation Library, 35. Amsterdam: John Benjamins.

Groves, D. (2008). Bringing Humans into the Loop: Localization with Machine Translation at Trasln. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas* (pp. 11–22). Waikiki, Hawaii.

Itagaki, M., & Aikawa, T. (2008). Post-MT Term Swapper: Supplementing a Statistical Machine Translation System with a User Dictionary. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)* (pp. 1584–1588). Marrakech, Morocco.

Itagaki, M., Aikawa, T., & He, X. (2007). Automatic Validation of Terminology Consistency with Statistical Method. In *Machine Translation Summit XI* (p. 269-274). Copenhagen, Denmark.

Levenshtein, V. I. (1965). Binary Codes Capable of Correcting Spurious Insertions and Deletions of Ones. *Problems of Information Transition*(1), 8–17.

McElhaney, T., & Vasconcellos, M. (1988). The Translator and the Postediting Experience. In *Technology as translation strategy* (Vol. 2, pp. 140–148). State University of New York at Binghamton.

Och, F., & Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, *29*, 19–51.

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.

Quirk, C., Menenzes, A., & Cherry, C. (2005). Dependency Treelet Translation: Syntactically Informed Phrasal SMT. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (pp. 271–279). Ann Arbor, MI..

Simard, M., Goutte, C., & Isabelle, P. (2007). Statistical Phrase-based Post-editing. In *Proceedings of NAACL.* Rochester, New York.