QUANTITATIVE ANALYSIS OF TRANSLATION REVISION: CONTRASTIVE CORPUS RESEARCH ON NATIVE ENGLISH AND CHINESE TRANSLATIONESE

Paul Rayson
UCREL
Lancaster University
Lancaster, UK
paul@comp.lancs.ac.uk

Xiaolan Xu, Jian Xiao, Anthony Wong, Qi Yuan
China Centre for Information Industry Development (CCID)
Beijing, China
yq@trans.ccidnet.com

Abstract: Demand for Chinese-to-English translation has increased over recent years. In contrast, resources for training translators for Chinese-to-English are few although increasing now, relative to English-to-Chinese for example. Corpus-based techniques are now more widely acknowledged as being appropriate for the study of translation. A number of Chinese/English parallel translation corpora have been built and applied to the research of translation practice. While such corpus resources have made a significant impact on these research areas, they suffer from problems due to the skewed nature of translated text, or 'translationese'. Obviously, translators and translation systems trained on these parallel corpora would inevitably inherit these features. Comparable corpora such as news articles, science and technology reports from the same period are more readily available. Studying translation revision carried out by native speakers of English may offer one way in to study Chinese-to-English translationese. However, very few quantitative studies of the products of the translation revision process have been carried out for any language pair. In this paper, we develop a framework using techniques from corpus linguistics, to enable the quantitative study of the translation revision process and describe the initial results we obtained. The research fits within a wider project to train language models in software tools that will assist in searching for non-native features of translated English texts.

Key Words: native English, Chinese translationese, parallel corpora, comparable corpora, multiword expression, automatic language profiling

1. INTRODUCTION

With the increasingly vital role of China's export industry and its steady upward climb in the value chain, science & technology, patents, social and economic development, and even culture and history have become the new content of interests. In the age of globalization, establishment and improvement of communication channels across the globe is always a challenging task. In view of the massive amount of information and data being produced across different channels in China everyday, and the increased and unabated interests in them by the international communities, the need for Chinese-to-English translation has increased significantly in the past few years. This demand has become even more acute in the context of major forthcoming events held in China such as the 2008 Olympics and the 2010 World Expo.

It has been argued that resources in Chinese-to-English translation are much harder to obtain (as compared to the resources in English-to-Chinese), as there are far fewer native English translators with adequate Chinese language skills in the translation industry. A number of Chinese/English parallel translation corpora have been built and applied to the research of translation practice. While such corpus resources are useful, they suffer from problems due to the skewed nature of translated text, or 'translationese'.

'Translationese', translation-specific language, refers to linguistic features that are either specific to translations or occur with a significantly higher or lower frequency in translations than in target-language originals. In the Chinese-to-English translation context, it is sometimes referred as 'Chinglish' colloquially. Human translators and Machine Translation systems trained on these parallel corpora would inevitably inherit these features. It results in mechanical and monotonous translated text at best, and inaccurate and erroneous translation in the worst cases.

As a remedy for this problem, comparable bilingual corpora are receiving increasing attention. These comparable corpora are useful in the study of the translation process by exploring how an idea in one language (Chinese in our case) is conveyed in another language (English) (McEnery, 2004). Comparable Chinese/English corpora such as news articles, science and technology reports from the same period from different news agencies are now more readily available using abundant Web resources, e.g. the English and Chinese Gigaword corpora from the Linguistic Data Consortium (LDC). These are increasingly being used to extract word translations for novel terminology and names (Shao and Ng, 2004).

In this paper, we will investigate translationese resulting from Chinese-English human translation. Our focus here is on using a corpus-based methodology to study Chinese translationese, i.e. using tools and techniques such as keyness (key words and key domains) to study the difference, particularly in the ICT (Information and Communications Technology) domain. Our approach will follow earlier pioneering work on automatic language profiling of learner's data (Granger and Rayson, 1998). This method can be used to discover key words in the corpora which differentiate one corpus from another; for example, to determine significant patterns of over- or under-use. Based on a corpus of manually translated data that has been hand-corrected by native English speakers, we plan to train language models and then apply our tools to automatically search for non-native features of translated English texts.

This paper represents the first portion of this work and reports on the quantitative comparison of manually translated data with its hand-corrected counterpart which has been edited by native speakers of English. Initially, focus will be on key words and a variety of types of Multiword Expressions (MWE) and how they differ from authentic native data to the translated corpus examples.

Section 2 will describe the wide variety of related work. Section 3 will describe the experiment that we have carried out, followed by results in section 4 and finally conclusions in section 5.

2. RELATED WORK

In this section, we establish the theoretical and practical background for the research. We will describe previous corpus-based approaches to the study of translation practice. First, we focus on the later stages of the (human) translation process and in particular the revision of draft translations by the translators themselves or other editors.

Mossop (2007) described 'editing' as making corrections and improvements to translated texts focussing on tailoring the results to a particular audience. He described 'revision' as the same task when applied to draft translations. In our work here, we fold the two definitions together and refer to the whole process as translation revision. Shih (2006) carried out an in depth study of revision from the translators' point of view, although her study was on 'self-revision' i.e. translators correcting their own translation drafts. Shih extended Mossop's revision parameters and divided the aspects that translators check for during revision into three categories. First, target text linguistic problems (e.g. spelling mistakes, grammatical errors, fluency); secondly, other target text problems (e.g. layout, terminology control, logic) and finally, source text meaning transfer (e.g. accuracy, omission, numbers and dates). Issues specific to translation between English and Chinese are not covered by Shih, although this was the scenario examined in the study. In addition, features of native language translator production are not described, although no doubt they impact on all of the categories in the revision checklists presented. Finally, it is worth noting that Shih's method is that of interviewing translators rather than empirical analysis of the product of translation, though the interview method is perfectly valid since "it is exactly these translators' own insiders' views of revision" (Shih, 2006: 300) that her study was aimed at exploring. A similar classification to Shih and Mossop's categories can be found in frameworks for standard error marking in translations, e.g. from the American Translators Association (2002) and others (Secară, 2005).

The equivalent editing process to translation revision, when it is applied to machine translation (MT) output is known as post-editing. Allen (2003) refers to post-editing as the process whereby a human editor modifies and/or corrects pre-translated text that has been produced by an MT system. In certain scenarios, manual editing of automatic MT output can be preferred over full human translation.

Very few research articles have appeared with any quantitative analyses of the editing carried out within either of these two processes (post-editing or revision) in terms of the difference it makes to the final product. Krings (2001) carried out an empirical psycholinguistic analysis of post-editing in order to describe the mental processes involved. In addition, quantitative models have been derived from post-editing output. Simard et al (2007a, 2007b) implemented an automated post-editing system for correcting repeated errors in raw MT output. Their system was based on a statistical phrase-based MT system (PORTAGE) and trained on a parallel corpus of raw MT output (from SYSTRAN) and its human post-edited counterpart. Dugast et al (2007) ran similar experiments using base MT output (again from SYSTRAN) with two different post-editing tools (PORTAGE and Moses). They defined a number of criteria related to lexical changes, grammatical changes and alterations in punctuation, word order and style in order to carry out a linguistic categorisation of post-editing changes. The idea of utilising correction information from post-edited output to improve MT systems is not new, with previous work for example on reversing the MT process from the corrected output to improve the original system (Nishida et al, 1988).

It is worth noting related work from the computational area of evaluation of MT output since this also incorporates analysis of multiword expressions (MWEs). MT evaluation can be done through analysis of post-editing of the output as described above, and a well known metric is the BLEU method. BLEU uses deviation from a reference set of n-grams in order to check the accuracy. N-grams are repeated sequences of words in a text and the method is one way of extracting candidate MWE from corpora. In order to address the problem of legitimate variation in human translations, Babych and Hartley (2004) used a weighted model of n-grams to favour content bearing items (names and terms) over less central information (function words). Their results show higher correlation with intuitive judgements of translation accuracy and fluency than baseline BLEU scores. Multiword expression identification allows grouping of words together for further processing (Sag et al, 2001). There is a very large body of work on this topic from cognitive, linguistic (phraseology), computational and corpus-based approaches. Here, we are interested in MWEs since they might be indicators of regular contexts for translation revision and translationese. In previous work, we have reported on our experiment for automatic extraction of Chinese MWEs using a statistical tool originally developed for English (Piao et al, 2006). Our tool combined with linguistic filters can produce a practically viable tool for extraction of MWEs.

In terms of corpus approaches to human and machine translation, there are a large and ever increasing number and we can distinguish two sub-areas of research. One more computationally focussed and the other more linguistically based. In addition, we can separate corpus approaches which assist in the translation process from those which use corpora to study the translation process. Our work fits clearly into the area of corpus-based approaches to study human translation. Specifically, we will use approaches from corpus linguistics to study the outputs of the translation process. As outlined in Bernardini et al (2003: 2), linguistics based studies have tended to coincide with descriptive translation studies: “Corpus linguistics, as a methodology which focuses on the identification of recurrent patterns of linguistic behaviour in actual performance data, provides the appropriate tool to test hypotheses about norms and regularities in translated texts”. At the computational end of the spectrum, we find computer-aided translation tools such as translator workbenches, terminology management systems and MT systems incorporated into translator practice (Bernardini et al, 2003:3), for example, bilingual alignment tools have been used in the process of computer-assisted revision of translations (Simard, 1993: 68).

Corpora were previously used only indirectly in the human translation process via terminology compilation, the enhancement of monolingual and bilingual dictionaries and in the production of translation memories (Laviosa, 2003: 105). However, more recently, scholars from the fields of

corpus linguistics and translation theory have begun to use corpora of original and translated texts to study the product and process of human translation (Laviosa, 2003: 106). For example, the MeLLANGE project collected a corpus of learner translations with error annotations (Castagnoli et al, 2006). Studies have used parallel corpora, which consist of one or more texts and their translations, and comparable corpora, which contain original texts in two or more languages in the same domain or genre. These studies have used techniques directly borrowed from corpus linguistics such as word frequency lists, concordances and collocations and are explored in the special issue of the Meta journal (Laviosa, 1998) and in particular the innovative research of Munday (1998). Olohan (2004: 62-89) describes these methods in detail in her introductory book on corpora in translation studies. She also discusses the use of the key words technique from corpus linguistics in a case study of translator's lexical choice (Olohan, 2004: 160). Key words are generated by comparing a specialised corpus (translated text) with a more general reference corpus through their respective frequency lists. Common function words occur at the top of the frequency lists but their relative usage may still differ significantly between two texts due to the topic or style or genre differences. Those words which occur at a higher position or with higher frequency in the frequency profile of the first corpus relative to the second corpus (or reference text) are said to be positive (or overused) key words and those which occur relatively less frequently are termed negative (or underused) key words. Statistical metrics can be applied which show how significant or not such key words are. Olohan uncovers the lexical and grammatical features in the extracted key words which show a preference for American spelling conventions, formality, consistency of usage, and allow an exploration of translator's visibility.

Munday (1998) highlights problems of applying comparative corpus techniques across two languages: systematic differences between languages may account for differences in the number of distinct word forms (types) and running words (tokens) as well as type/token ratio. In addition, inflectional variants of the same head-word and multiword units are counted separately where they should be together and multiple meanings of polysemous words are counted together where they should be separated. What is required, we propose, is the use of more advanced corpus techniques which can take account of multiword units, grammatical and semantic differences. It should be noted that in a comparison of the translated text and its revised form the cross-languages issues are clearly not a problem anyway. Techniques from corpus annotation (Garside et al, 1997) may well provide a solution here and we will explore these in this paper. At the word level a process of lemmatisation can group inflectional variants of the same head-word together (Beale, 1987). The most typical form of corpus annotation is part-of-speech (POS) tagging which assigns a grammatical label (or tag) to every word in running text. A variety of tools and techniques for POS tagging exist, ranging from rule-based approaches created manually by linguistics (Karlsson et al, 1995) to probabilistic taggers using language models derived from training corpora that have been marked up or corrected manually, to hybrid approaches combining the two main techniques (Garside and Smith, 1997). Other levels of annotation employ similar combinations of techniques, for example semantic tagging or word-sense disambiguation (Rayson and Stevenson, forthcoming). One tool which incorporates a number of these techniques is Wmatrix (Rayson, 2008) which is a web-based tool allowing automatic corpus annotation and comparison via keyness statistics. With the combination of these two approaches, the tool permits the key words technique to be extended to key grammatical categories and key semantic fields by comparing frequency profiles of annotation tags rather than words.

To summarise the related work presented here, we have identified a lack of previous studies with a quantitative understanding of the changes made by native speakers in a revision step. The empirical methods afforded by corpus linguistics may well provide the solution here, since the techniques allow us to focus on whole texts and in a data-driven manner discover what is key about the differences. This paper will study a large corpus of native speaker revisions in the Chinese-English translation setting and see how corpus linguistics techniques can be used to highlight important trends. Our research described here is descriptive, product-oriented and data-driven.

3. EXPERIMENTAL SETUP

The collection of data for the investigation part of this paper is based on the ICT (Information and Telecommunication Industry) domain, and the source of such data mainly comes from the research work carried out by the Media and Consulting groups from CCID (China Centre for Information Industry Development). The Centre has more than fifteen newspapers and periodicals within its media group and close to 300 ICT research reports published per year by its consulting arm. It produces literature in the ICT domain consisting of around 10-20 million Chinese words annually. As information products, they are translated into English and delivered to global readers. Due to the timeliness requirement of the reports, massive amounts of translation work needs to be completed in a very short time. Computer translation technology (such as machine translation and translation memory) has been deployed extensively to meet the business requirements of the Centre.

The original Chinese text is translated to English by human translators with an ICT professional background (often with Chinese as their mother tongue and a good command of English as their second language), with the aid of translation tools. The translated English text will then be reviewed and edited by native English speakers before publishing.

In process terms, the documents pass through the following outline stages:

1. *Original*: Original Chinese text
2. *Pre-process*: Extraction of a wordlist using Chinese-English MT system called Huan-Yu-Tong (Sun, 2004). This step is useful for Chinese translators in order to ensure that terminology is consistent throughout the text
3. *Translate*: Text translated by Chinese translators (incorporating some 'Chinglish' elements)
4. *Edit*: Resulting text edited by native English speakers.
5. *Train*: Update the MWEs in dictionaries and sentence pairs in translation memory of MT system according to the above text edited by native English speakers.

Our resource consists of a corpus with three components: Chinese, Translated English (Chinglish) and Edited English. For the study described here, we have selected the translated and edited components for automatic analysis and use the Chinese component for reference purposes only. The resulting collection of 102 texts consists of 893,000 words in both translated and edited forms (close to 1,786,000 words altogether). For each text we have two versions produced by steps 3 and 4 above. Our initial corpus was larger by another 8 files (116,000 words), however, we filtered out texts with large differences between the translated and edited pairs (>10 lines different or > 10% difference in lines) since this represents documents with significant amounts of edited introductory or tabular material. The filtering and counting processes were automated with word processor counting functions and cross checked with UNIX tools such as 'wc' (line, word and character count) and 'diff' (compare two files).

We then created two corpora. One of translated material resulting from step 3 above, and one of edited material produced from step 4 above. Each corpus was automatically annotated using the Wmatrix software tool described in section 2 and comparisons were carried out at the key words, key POS and key semantic domain levels. The aim of the experiment was to allow corpus-based techniques to drive the analysis and assist us in selecting distinct translation revision (and therefore translationese) features between the two corpora.

In addition, we used a standard N-gram toolkit called N-gram Statistics Package (NSP) (Pederson, 2008) to extract repeated word sequences of length 2, 3, 4 and 5 from each corpus. Using new software developed at Lancaster University called C-gram (Collapsed-gram) we removed lower order duplicates e.g. those 2 and 3-grams that also appear in 4 and 5-grams. This enabled us to focus on the significant recurrent patterns by applying keyness measures to the n-gram frequencies using Wmatrix.

4. RESULTS

As described in section 3, we have carried out keyness comparisons at the word, POS and semantic levels to derive significant differences between the translated and edited versions of the English texts. In the following, we begin by discussing the key differences found at the word level, and then examine the differences at the POS and semantic levels. Table 1 shows the keyness results at the word level.

Table 1 Keyness Results At The Word Level

Keyword	Translated Corpus Normalised Frequency	Edited Corpus Normalised Frequency	Overused (+) Underused (-)	Log likelihood
Servicservice	471	0	+	547.93
Sever	267	8	+	257.69
Mother	0	126	-	146.85
servicservice_market	122	0	+	141.49
Informationization	167	449	-	112.60
Informatization	515	211	+	110.36
sever_market	88	2	+	86.92
Pieces	143	24	+	79.34
air-condition	0	62	-	72.04
Nt	58	0	+	67.97
sub-contract	98	11	+	67.49
telecommunication_industry	82	236	-	64.90
Conditioners	51	0	+	59.65
front-end	0	45	-	52.65
telecom_industry	228	92	+	50.17
education_industry	54	2	+	48.65
Front-End	42	0	+	48.55
MkW	0	42	-	48.49
air-conditions	0	39	-	45.72
Sourse	38	0	+	44.39
Servicservices	36	0	+	41.62
NT	0	36	-	41.56
Manufactures	104	27	+	39.74
Severs	33	0	+	38.84
Cores	37	1	+	35.49
Modifications	0	30	-	34.64
service_market	116	235	-	34.61
air_conditioner	39	2	+	33.22
Tsingdao	35	1	+	32.84
Diary	32	1	+	30.21
Fig	125	49	+	29.09
middle-sized	29	1	+	26.28
Sets	342	503	-	25.75
Air	137	61	+	25.38
Tsingtao	6	42	-	25.28
Dairy	4	33	-	23.24
Telecom	558	398	+	22.61
Hi-end	19	0	+	22.19
Revisions	57	15	+	21.39
ISVs	0	18	-	20.78

The table shows the words (and in some cases multiword units) which are significantly key when we compare the translated and edited versions of the corpus. This comparison is sorted on the Log

Likelihood value to show which words are most key in one of the two sub-corpora. All forty words in the table are significant. In total there are 58 words that are significant at $p < 0.0001$ (LL over 15.13). The plus and minus signs indicate whether the words are more frequent in the translated corpus (+) or in the edited corpus (-). In fact, the frequencies shown are normalised per million words, so a direct comparison is possible. The Log Likelihood statistic also takes the differing sizes of the two sub-corpora into account. For more details see Rayson and Garside (2000).

Terms indicated with a plus mark are those which have been reduced in frequency by the editing process and terms highlighted with a minus mark are those which have been added during the editing process. Overall, the key words show a number of key pairs where one term is preferred over another. One part of the pair is reduced in frequency whereas the other part is increased in frequency, although the frequency changes do not always exactly match. For example:

- informatization and informatisation replaced by informationization
- air condition(er) replaced by air-condition(er)
- isvs replaced by ISVs
- sever_market replaced by server_market
- servicservice_market replaced by service_market
- nt replaced by NT
- Tsingdao replaced by Tsingtao
- diary replaced by dairy
- telecom_industry replaced by telecommunication industry
- sever(s) replaced by server(s)
- Front-End replaced by front-end
- EBao replaced by ebao

In other cases, less preferred terms are removed completely from the key words list:

- servicservice(s)
- education_industry
- middle_sized
- Sourse
- hi-end

The strange appearance of terms like “servicservice” resulted from tracked changes in Microsoft Word. The word “mother” appears in the edited corpus due to being a preferred term over “main” for “mother board” instead of “main board”. This was later corrected to “motherboard”. On further investigation, we find that the preferred term “informationization” occurred only in one long article from 2002 and was the subject of much discussion during translation at the time. Since then, this non-standard term has been included as one of the top negative examples in the Chinglish term base and is now replaced by “IT application” and similar terms.

Table 2 shows the keyness results at the major word class level. It should be noted that the LL values reported here are not significant. However, it indicates that the largest changes are made to articles and adverbs in the data. These results at the word class level reflect two things. First, reinforcing the observations at the word level that most changes are typographical. Second, that changes to the syntax and grammar of the sentences are very few, indicating that the output from the first translation step (of non-native speakers) is of very high quality.

If we apply the same technique to the full POS tags assigned by the CLAWS tagger, we find that specific tags emerge as being significant e.g. plural common nouns, plural proper nouns and base form of verbs. However, overall word classes masks these differences. Table 3 shows the keyness results at the semantic level. Once again the LL values are not as high as for the word level comparison with only the first two items being significant at the $p < 0.0001$ level. This shows the overall level of similarity between the translated and edited texts. Most differences relate to

specific words that have been edited. The reduction in the number of unmatched (Z99) items shows the reduction in typographical errors resulting from the editing process. Unknown items are those which the semantic analysis system does not recognise from a large coverage modern lexicon.

Table 2 Keyness Results At The POS Level

Key word class	Translated Corpus Normalised Frequency	Edited Corpus Normalised Frequency	Overused (+) Underused (-)	Log likelihood
Article	72430	72968	-	1.65
Adverb	32794	32568	+	0.66
Formula	11620	11753	-	0.63
Verb	117702	117305	+	0.55
Infinitive marker (TO)	9201	9298	-	0.43
Pronoun	14121	14237	-	0.39
Determiner	15330	15211	+	0.38
Existential there	971	999	-	0.33
Before clause marker	940	964	-	0.26
Genitive	1936	1964	-	0.17
Interjection	63	58	+	0.16
Noun	400523	400324	+	0.10
Conjunction	50356	50430	-	0.04
Adjective	98520	98442	+	0.04
Letter	4271	4293	-	0.04
Number	45113	45141	-	0.01
Preposition	122064	121998	+	0.00
Negative	2043	2047	-	0.00

Table 3 Keyness Results At The Semantic Level

Key semantic tag	Translated Corpus Normalised Frequency	Edited Corpus Normalised Frequency	Overused (+) Underused (-)	Log likelihood	Semantic Field
S8+	10229	10896	-	17.66	Helping
F1	2672	3003	-	16.24	Food
B4	172	102	+	14.84	Cleaning
O1.3	228	151	+	13.02	Substances
S4	929	1071	-	8.48	Kin
Z99	45344	44444	+	7.56	Unmatched
A1.1.1	13696	13225	+	6.90	General
L1-	193	149	+	4.82	Dead
O4.6	205	167	+	3.32	Temperature
E6-	332	282	+	3.30	Worry

It is interesting to note that during the editing process, several typical Chinese translationese patterns have surfaced frequently, even from the output of the first translation step (of non-native speakers) that is of relatively high quality. For example, the Chinese character '化' is used extensively to transform nouns to 'action-nouns' in Chinese (e.g. '工业化', '现代化', '信息化', '智能化', '市场化 …. etc.), which may or may not have the corresponding native English terms (former cases like 'Industrialization', 'Modernization' and latter like 'Informatization',

'Marketization' and 'Intelligentization'). The analysis of the keyness results at the various levels can certainly shed more lights on this type of patterns.

We have also computed all recurrent strings of length 2, 3, 4 and 5 in the translated and edited corpora. The full tables are omitted for reasons of space. By comparing them in the same manner as above to show key differences (key clusters or key n-grams in this case), we are able to further explore the edits made during translation revision. For example, 5-grams such as "on Chinese management software market" and "Chinese management software market in" that are overused in the translated corpus are matched by similar phrases "of China's management software market" and "China's management software market in" that are more frequently used in the edited corpus. This indicates consistency checking during the revision process via correcting "Chinese" to "China's" in particular contexts.

5. SUMMARY AND CONCLUSIONS

We have carried out a systematic comparison of texts translated from Chinese to English by Chinese translators with the same texts subsequently edited by native speakers of English. We have created a framework for this comparison by utilising corpus-based techniques such as keywords, corpus annotation and n-gram extraction tools. The results show that techniques from corpus linguistics can be used to assist in a quantitative study of translation revision. Moreover, they show that Chinese-English translationese can be explored using corpus based techniques.

We need to note some possible caveats in our approach. In further work, we need to investigate how accurate the corpus annotation tools are when they are applied to translated texts. Language quality and stylistic differences might affect the results of automatic annotation tools if we use automated MT output for example, but the translated texts that we have used are already of high quality. In addition the tools we have used are robust across a number of domains and when applied to 'non-standard' language such as learner language (Leech and Smith, 2000; van Rooy and Schafer, 2003). Also, we may need to adopt weighted n-gram approaches for comparison between translated and edited corpora in a similar fashion to the improvement of BLEU scores in MT evaluation.

The keyness techniques are able to highlight native speaker's editing in the data, and through concordances of key words, we can trace the differences between the translated and edited texts. Our emerging methodology permits studies of Shih's (2006) first two categories of translation revision, namely linguistic problems and text problems. We believe that this shows the potential of our approach for assisting in the study of translationese and the translation revision process itself.

6. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (grant no. 60520130297). We would also like to thank Yuan Xiaojing and Zhang Hui for their help in data preparation at CCID and Andrew Stone at Lancaster University for his help in the development of the c-gram tool.

7. REFERENCES

[1] Allen, J. (2003). Post-editing. In H. Somers (ed.) *Computers and translation: a translator's guide*. John Benjamins, Amsterdam, pp. 297-317.

[2] American Translators Association (2002), *Framework for Standard Error Marking*, ATA Accreditation Program, http://www.atanet.org/bin/view.fpl/12438.html .

[3] Babych, B. and Hartley, A. (2004). Extending the BLEU MT evaluation method with frequency weightings. In *Proceedings of the 42nd Annual Meeting on Association For Computational Linguistics (Barcelona, Spain, July 21 - 26, 2004),* pp. 621-628.

[4] Beale, A.D. (1987). Towards a Distributional Lexicon. In: R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach.* London: Longman, pp. 149 - 162.

[5] Bernardini, S., Steward, D. and Zanettin, F. (2003). Introduction. In Zanettin, F., Bernardini, S. and Stewart, D. (eds) *Corpora in translator education*. St Jerome, Manchester, UK, pp. 1-13.

[6] Castagnoli, S., Ciobanu, D., Kunz, K., Kübler, N., Volanschi, A. (2006) Designing a Learner Translator Corpus for Training Purposes. In Proceedings of TALC2006, Paris, France.

[7] Dugast, L., Senellart, J. and Koehn, P. (2007) Statistical post-editing on SYSTRAN's rule-based translation system. In *proceedings of the ACL 2007 Second Workshop on Statistical Machine Translation. Prague*, Czech Republic. June 23, 2007. pp. 220–223.
[8] Garside, R., Leech, G., and McEnery, A. (eds.) (1997) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Longman, London.
[9] Garside, R., and Smith, N. (1997) A hybrid grammatical tagger: CLAWS4, in Garside, R., Leech, G., and McEnery, A. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Longman, London, pp. 102-121.
[10] Granger, S., and Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (ed.) *Learner English on Computer*. Longman, London and New York, pp. 119-131.
[11] Karlsson, F, Voutilainen, A, Heikkilä, J and Anttila, A (eds.) (1995) *Constraint Grammar: a language-independent system for parsing unrestricted text*. Berlin: Mouton de Gruyter.
[12] Krings, H. (2001). *Repairing texts: empirical investigations of machine translation post-editing processes*, edited by Geoffrey Koby. Translated from German to English by G. Koby, G. Shreve, K. Mischerikow and S. Litzer. Kent State University Press, Ohio.
[13] Laviosa, S. (1998). The corpus-based approach: a new paradigm in translation studies. *Meta*, Volume 43, number 4, pp. 474-479.
[14] Laviosa, S. (2003). Corpora and the translator. In H. Somers (ed.) *Computers and translation: a translator's guide*. John Benjamins, Amsterdam, pp. 105-117.
[15] Leech, G. and Smith, N. (2000) *Manual to accompany the British National Corpus (Version 2) with improved word-class tagging*. http://ucrel.lancs.ac.uk/bnc2/bnc2postag_manual.htm
[16] McEnery, T. (2004). Parallel and comparable corpora: what are they up to? Presented at *CCID & Lancaster University Joint Symposium on Corpus Linguistics and Machine Translation*, Beijing.
[17] Mossop, B. (2007) *Revising and editing for translators (2nd Edition)*. St. Jerome Publishing, Manchester.
[18] Munday, J. (1998). A computer-assisted approach to the analysis of translation shifts. *Meta*, Volume 43, number 4, pp. 542-556.
[19] Nishida, F., Takamatsu, S., Tani, T., and Doi, T. 1988. Feedback of correcting information in postediting to a machine translation system. In *Proceedings of the 12th Conference on Computational Linguistics - Volume 2 (Budapest, Hungry, August 22 - 27, 1988)*. D. Vargha, Ed. International Conference On Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 476-481.
[20] Olohan, M. (2004). *Introducing corpora in translation studies*. Routledge, London.
[21] Pederson, T. (2008). N-gram Statistics Package. http://www.d.umn.edu/~tpederse/nsp.html
[22] Piao, S.L., Sun, G., Rayson, P. and Yuan, Q. (2006) Automatic extraction of Chinese multiword expressions with a statistical tool. In *proceedings of the Workshop on Multi-word-expressions in a Multilingual Context held in conjunction with the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, Italy, April 3, 2006, pp. 17-24.
[23] Rayson, P. (2008) *Wmatrix: a web-based corpus processing environment*, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/
[24] Rayson, P. and Garside, R. (2000). Comparing corpora using frequency profiling. In *proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*. 1-8 October 2000, Hong Kong, pp. 1 - 6.
[25] Rayson, P. and Stevenson, M. (forthcoming) Sense and semantic tagging. In Lüdeling, A. and Kytö, M. *Corpus Linguistics. An international handbook*, Mouton de Gruyter, Berlin.
[26] van Rooy, B. and Schäfer, L. (2003) An evaluation of three POS taggers for the tagging of the Tswana Learner English Corpus. In *proceedings of Corpus Linguistics 2003*, Lancaster University, pp. 835-844.
[27] Sag, I., Baldwin, T., Bond, F., Copestake, A., Dan, F. (2001). Multiword expressions: a pain in the neck for NLP. *LinGO Working Paper No. 2001-03*, Stanford University, CA.
[28] Secară, A. (2005) Translation Evaluation – a State of the Art Survey. Proceedings of the eCoLoRe/MeLLANGE Workshop, Leeds pp. 39-44.
[29] Shao, L. and Ng, H. T. (2004). Mining new word translations from comparable corpora. In *Proceedings of the 20th international Conference on Computational Linguistics* (Geneva, Switzerland, August 23 - 27, 2004). Association for Computational Linguistics, Morristown, NJ, 618.
[30] Shih, C. Y. (2006) Revision from translators' point of view: An interview study. *Target*, Volume 18, Number 2, 2006 , John Benjamins, pp. 295-312.
[31] Simard, M., Foster, G. F., and Isabelle, P. (1993). Using cognates to align sentences in bilingual corpora. In *Proceedings of the 1993 Conference of the Centre For Advanced Studies on Collaborative Research: Distributed Computing - Volume 2 (Toronto, Ontario, Canada, October 24 - 28, 1993)*. A. Gawman, E. Kidd, and P. Larson, Eds. IBM Centre for Advanced Studies Conference. IBM Press, 1071-1082.
[32] Simard, M., Goutte, C., and Isabelle, P. (2007a). Statistical Phrase-Based Post-Editing. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pp. 508-515, Rochester, USA.
[33] Simard, M., Ueffing, N., Isabelle, P., and Kuhn, R. (2007b) Rule-based Translation With Statistical Phrase-based Post-editing. In *proceedings of the ACL 2007 Second Workshop on Statistical Machine Translation*. Prague, Czech Republic. June 23, 2007. pp. 203–206.
[34] Sun, G. (2004). Design of an Interlingua-Based Chinese-English Machine Translation System. In *proceedings of the 5th China-Korea Joint Symposium on Oriental Language Processing and Pattern Recognition*, Qingdao, China. pp. 129-134.