

Evaluation of thesis

Thesis title: Lost and added in translation: A corpus-based study on Chinese numeral classifiers in translation between Chinese and English

Thesis author: Saizhu Hu

Thesis university: Uniwersytet im. Adama Mickiewicza w Poznaniu

Thesis reviewer: Marc Allasonnière-Tang, PhD

General comments:

This dissertation provides a corpus-based analysis of Chinese numeral classifiers in the context of Chinese-English translation. The thesis investigates how Chinese numeral classifiers and other elements of the noun phrase contribute to its semantics and how they correspond to equivalent elements in English translation based on their semantic and discourse functions. This investigation is based on three corpora. Two of these corpora are extracted from the BCC corpus. They consist of 6700 pairs of Chinese-English numeral noun phrases without adjectives and 523 pairs of Chinese-English numeral noun phrases with adjectives respectively. The third corpus consists of 645 pairs of noun phrases derived from five chapters of the novel *The Three-Body Problem*. The results are relevant for linguistic analyses within Chinese and within a translation context. Numeral classifiers contribute additional meanings related to quality and quantity to noun phrases, beside denoting semantic features of noun referents. Moreover, classifiers can also be referential to different degrees depending on how frequent they are in collocations with nouns. In terms of translation from Chinese to English, the results show that English uses measure words, the plural form, articles and demonstratives, nouns and pronouns to translate meanings conveyed by classifiers. The study also demonstrates that the semantic functions of Chinese numeral classifiers are rarely reflected in English translation. However, discourse functions of Chinese numeral classifiers are more likely to be reflected in English translation. As an example, English articles and demonstratives can be used to express definiteness and referentiality, which are related to the function of reference management, while nouns and pronouns can be also used to represent the functions related to reference management as well as reference identification and re-presentation of referents.

In terms of theoretical background, the thesis provides a very detailed overview of nominal classification systems, discussing the two main types of systems, which are grammatical gender and classifiers. This helps the reader to understand the existing theoretical framework of classifiers, which are the focus of this thesis. However, background on translation studies could have been further developed since it is also what the thesis is looking at. This question is also reflected in the research aims, which are clearly listed and answered in the thesis. However, the motivation for choosing these questions could be further linked more directly with translation. For example, how are the research questions related to issues encountered by translators? As for methodology, the choice of corpora is well described. Examples are also given from the gathered corpora. Nevertheless, additional details could help the reader to better understand how the data was cleaned. For example, additional details about how classifiers and measure words were distinguished in the data would be helpful. See my detailed comments in the rest of the text. Nevertheless, the results of the analyses are clearly presented with tables and examples, which make it easy for the readers to follow through. The formal layout and language used are also extremely clear and precise. The

reading of the thesis was really smooth and enjoyable. Congrats on the author for having these parts very well represented. Finally, the overall original contribution of the dissertation can be illustrated, as an example, by its quantitative data on the frequency of classifiers in a translation context, which is rare in the literature.

Specific comments:

In the following text, I provide comments on the few main points I would like to discuss during the defense. These points are more curiosity questions and do not impact the general appreciation of the thesis.

My first main comment relates to the balance between the linguistic analyses and the context of translation. The analyses provided in this thesis are interesting from a linguistic point of view. However, while the title of the thesis mentions ‘translation’, I found that the literature review was mostly discussing linguistic questions, while more details and explanation could have been provided about translation. For example on p2 "However, relatively little research has been done on the comparison of numeral classifiers with their corresponding forms in English translation," it would be nice to know more about why it is important to look at this and if it is important, why so few people did it so far. Was it a technical limitation? In the same spirit, quite a few research questions seem to be relevant for linguistics but less for translation, e.g., p3, the first question investigates the use of classifiers with and without adjectives in Chinese. How is this question relevant to translation? Likewise, on p4 "comparing Chinese numeral classifiers with their corresponding forms in non-classifier languages based on their functionality can contribute to the studies on the acquisition and translation of numeral classifiers in applied linguistics.": It would be interesting to have more details or examples about how concretely this is something relevant for translation or applied linguistics. As an example, what are the difficulties faced by translators with regard to classifiers? This main question is also related to the ratio of linguistic description of nominal classification systems in the thesis. A lot of space is given to the discussion of gender (p11 – p23), types of classifiers, and the diachrony of classifiers, while very little space is dedicated to the literature review of translation studies. For example, extended paragraphs about translation arrive at p135 of the thesis, which is more than half of the entire manuscript i.e., 135/211 pages.

My second main comment is related to data gathering and data cleaning. First, while examples from the corpora are provided, a few more details about some methodological choices would be useful. For example, on p91 "In every 4,000 pairs of sentences, 50 of them were chosen as samples for the study.": how was this decision made? e.g., why 50 sentences? and how were these 50 sentences selected? Likewise, on p92 "Numeral noun phrases in the sample sentences were manually annotated in terms of five categories in Chinese and their equivalents in English translation": I might have missed it, but how did you define and choose these categories? As an example, whether there are adjectives or not in Chinese is also a theoretical question (http://crlao.ehess.fr/docannexe/file/1552/jet_adjectives_proofs_2010_04pau.pdf). What is the author’s position on this? Second, some additional statistical analyses would have been beneficial to further support the quantitative trends mentioned in the thesis. For example, p102 "Compared with numeral classifiers used without adjectives, far fewer numeral classifiers are used in the context of adjectives in Corpus 2": this is quite expected considering that you have a difference of size of 6000 vs 500 between the two corpora. As another example, p104 “which shows that the presence of adjectives affects the choice of numeral classifiers.”: this is not really statistically proven. In general, in Section 5, the author could have used some statistical analyses (such as simple regression or mixed models) to identify statistically significant interactions. However, this is an

optional comment. Third, I have some questions about the definition of classifiers. While the author provides a very detailed literature review about the definition of classifiers, some more details about how the author concretely differentiated classifiers and measure words would be good. For example, p144: does the author consider *kuai* as a numeral classifier, i.e., specific entity classifiers denoting shape? Or a measure word? In the text, *kuai* is glossed CLF:LUMP.LIKE in (80b) 'a piece of cake'. If the author considers that it is a measure word, I suggest to change the gloss. To be sure of how the author labelled classifiers and measure words, I would need some examples of if and how the author applied the tests mentioned on p59 section 3.2.3 when annotating the data? if yes, I would expect that *kuai* is a measure word.

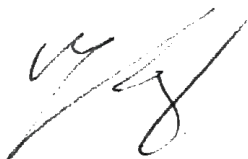
My third main comment relates to the choice of novel for the third corpus. On p5 "Qualitative studies were made based on a parallel corpus of about 645 pairs of noun phrases in more than 411 pairs of sentences derived from five chapters of The Three-Body Problem": It would be interesting to know more about how and why this novel was chosen? Some answers are given on p98 but these motivations could be questioned. For example, the author suggests that a fiction novel is a better choice. However, why this novel and not another fiction novel? With regard to the justification about the Hugo award, other novels probably received this prize or other prizes as well, why not those novels? Third, it is probably not the only novel that can reflect the use of classifiers. Finally, it is a bit expected that both the author and the translator are good writers in Chinese for all novels being translated in this way, so this is not really a motivation specific to this novel either. Of course, I am not asking the author to re-make a study. I think that the current results are already interesting. Nevertheless, I thought it would be good to point out this potential question.

My final comment is a curiosity question about the methodology. You investigated the translation from Chinese to English, but did you think about consider translation from English to Chinese? Since the choice of classifiers when translating from English could also be relevant. What would be your opinion on this matter?

Conclusion

To sum up, the dissertation (despite the critical comments) receives a positive assessment from the reviewer and the PhD candidate can proceed to the final stages towards earning a PhD title.

Sincerely,
Marc Allasonnière-Tang



CNRS researcher
Museum of Natural History/
French National Centre for Scientific Research/
University Paris City
marc.allasonniere-tang@mnhn.fr