

Measuring the Cognitive Effort of Literal Translation Processes

Moritz Schaeffer
Dalgas Have 15
Copenhagen Business School
Denmark
ms.ibc@cbs.dk

Michael Carl
Dalgas Have 15
Copenhagen Business School
Denmark
mc.isv@cbs.dk

Abstract

It has been claimed that human translators rely on some sort of literal translation equivalences to produce translations and to check their validity. More effort would be required if translations are less literal. However, to our knowledge, there is no established metric to measure and quantify this claim. This paper attempts to bridge this gap by introducing a metric for measuring literality of translations and assesses the effort that is observed when translators produce translations which deviate from the introduced literality definition.

1 Introduction

In his seminal paper, Ivir (1981: 58) hypothesises that:

“The translator begins his search for translation equivalence from formal correspondence, and it is only when the identical-meaning formal correspondent is either not available or not able to ensure equivalence that he resorts to formal correspondents with not-quite-identical meanings or to structural and semantic shifts which destroy formal correspondence altogether. But even in the latter case he makes use of formal correspondence as a check on meaning - to know what he is doing, so to speak.”

Related to this notion of “formal correspondence” is the *law of interference* which accounts for the observation that “in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text” (Toury, 1995: 275).

However, context or cross-linguistic differences may make it necessary to abandon formal correspondence: it is often necessary to depart from a one-to-one correspondence between source and target text items, levels or ranks, which is confirmed by the statement “without it [formal correspondence], there would be nothing to shift from” (Malmkjær 2011a: 61).

Tirkkonen-Condit (2005) reformulates Ivir’s formal correspondence translation hypothesis into a *monitor model*: “It looks as if literal translation is a default rendering procedure, which goes on until it is interrupted by a monitor that alerts about a problem in the outcome.” Tirkkonen-Condit (2005:408)

Thus, the *formal correspondence hypothesis*, the *literal translation default rendering procedure*, the *law of interference* and the *monitor model* are all related concepts which seem to assume that one-to-one literal translation correspondences are easier to produce than translations that formally deviate from the source text, as the latter would require more effort, and hence will take longer for a translator to produce.

While it has been difficult to describe in what exactly consist literal translation (Malmkjær 2011b), we define (ideal) literal translation in this paper by the following criteria:

- a) Word order is identical in the source and target languages
- b) Source and target text items correspond one-to-one

- c) Each source word has only one possible translated form in the given context

Although this definition of literality ignores a wide range of phenomena and kinds of equivalence, it allows for quantification and comparison across multiple languages. Any (voluntary or structural) deviation from these criteria would imply a relaxation from a literal translation and thus lead to greater effort, as measured by e.g. longer production times and more gaze activities.

In this paper we assess this hypothesis by

notion of *translation choices*, derived from a corpus of alternative translations to account for criterion (c) above. In section 3, we correlate the predictions of the *literal translation default rendering procedure* with observed translators' behavior. Section 4 discusses the results.

2 Operationalizing literal translation

In this section, we first present a quantification of translation choices (literality criterion c) and then describe the computation of alignment cross values which account for literality criterion (b) and (c).

Killer		nurse		receives		four		live		sentences	
11	asesino	7	el_enfermero	15	recibe	28	cuatro	12	perpetuas	13	cadenas
6	el_asesino	5	enfermero_asesino	3	es_condenado			12	cadenas	11	perpetuas
3	el_enfermero	4	enfermero	3	condenado_a					2	asesino
2	enfermero_asesino	4	asesino	2	recibe_a						
		3	un_enfermero								
		2	enfermera								

Figure 1: Translation choices and numbers of occurrences as retrieved from 31 En -> ES translations in the TPR-DB

analyzing the gazing behavior of translators. As a basis for our investigation we use the TPR-DB (Carl, 2012), which currently contains more than 940 text production sessions (translation, post-editing, editing and copying) in more than 10 different languages¹. For each translation and post-editing session keystroke and gaze data was collected and stored, and translations were manually aligned. The TPR-DB is therefore ideally suited for answering aspects of the cognitive processes during translation which are shared across individuals and language combinations.

In section 2 we operationalize *literal translation* from a process point of view. We describe a transducer to measure the similarity of word order in the source and target language strings, to account for criteria (a) and (b). We introduce the

2.1 Translation Choices

A source word can often be translated in many different ways. In order to quantify such translation choices, Choice Network Analysis has been suggested (Campbell, 2000) as a method to infer cognitive processes from the different choices made by different translators: the more choices and the more complex choices a translator has to consider, the more effortful the translation of this particular item is. Campbell (2000) argues that translations by different translators of the same source text can be used to draw inferences about the cognitive processes during translation.

In line with these considerations, to estimate the translation effort for lexical selection, we count the number of different translation realizations for each word. We use the TPR-DB (Carl, 2012, Carl et al. 2014) which contains (among others) a large number of different translations for the same source text. For instance, Figure 1 shows the number of Spanish translation choices

¹ The figures relate to TPR-DBv1.4 which can be downloaded from:
http://bridge.cbs.dk/platform/?q=CRITT_TPR-db

produced by 31 different translators for the same English source sentence. Figure 1 only shows translations which occur at least twice. Figure 2 shows one of the realized translations.

There is a considerable variance in the number of translation variants for different words. In 11 out of 31 translations “Killer” was aligned with “asesino”, in 6 cases with “el asesino” etc. while for 28 out of 31 cases “four” was translated as “cuatro”. Thus, according to the above hypothesis, the translation production of “Killer” would be more effortful than it would be to translate “live” than the translation of “four”.

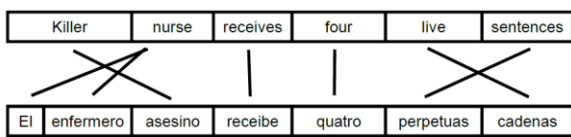


Figure 2: Oracle translation with word alignments

2.2 Alignment crossings

In order to quantify translation locality criterion (a) and (b), we adopt a local metric to quantify the similarity of the source and target language word order, relative to the previous alignment position. The metric is implemented as a transducer which produces translations word by word, writing the correct target language word order into an output buffer, while a reading device successively scans the source text to find the reference word(s) for the next word in the translation.

Given a reference source text (ST), an output oracle translation (TT), and the ST-TT alignments (as in Figure 2), the *CrossT* values indicate the distance between ST reference expressions of successive TT words, in terms of progressions and regressions.

For instance, assume the English source sentence “Killer nurse receives four live sentences” was translated into Spanish with the alignment relations as shown in Figure 2. In order to produce the first Spanish TT word “El”, two English words (“Killer” and “nurse”) have to be consumed in the reference text, which results in a

Cross value of 2. Since the second source word (“nurse”) emits two adjacent TT words, no further ST word has to be consumed to produce “enfermero”, which results in the value $Cross=0$. To produce the third Spanish word, “asesino”, one ST word to the left of “nurse” has to be processed, leading to the *Cross* value -1. The next Spanish word “receibe” is the translation of two words to the right of the current ST cursor position; “cuatro” one ST word ahead etc. with their respective *Cross* values of 2 and 1. Figure 3 illustrates this process. The inclined reader may continue this example and reconstruct how the *CrossT* values {2,0,-1,2,1,2,-1} are incrementally generated. Thus, *Cross* values indicate the minimum length of the progressions and regressions on the reference text required to generate the output string.

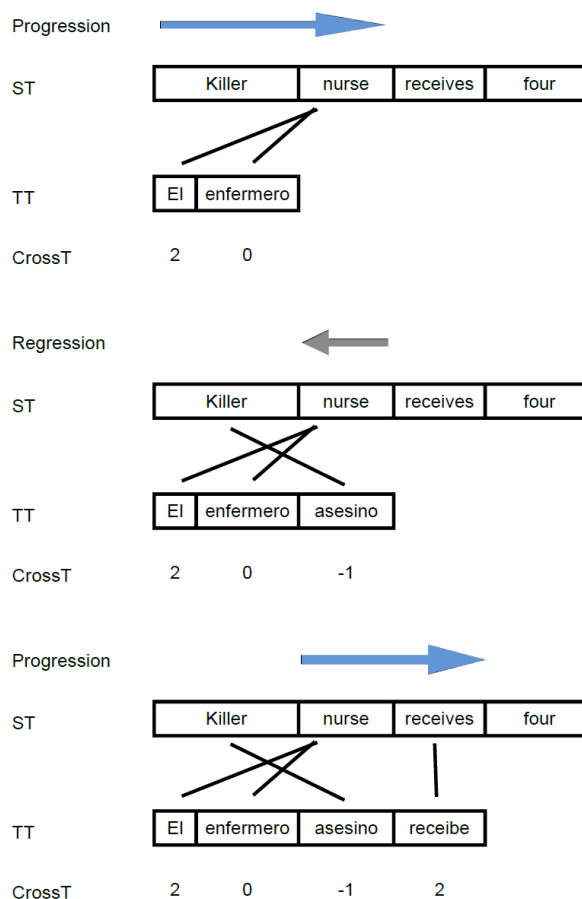


Figure 3: Computation of alignment crossings (*CrossT*) as length of progressions and regressions in the reference ST.

Cross values can also be computed from the source text. For the *CrossS* values we would then

assume the ST text to be the output text and the TT text to be the reference.

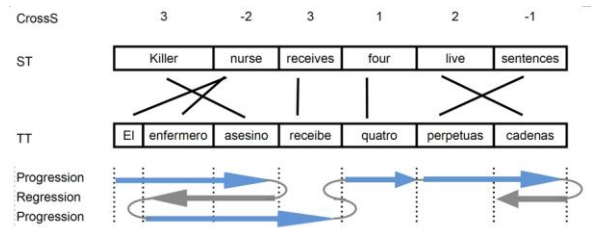


Figure 4: ST alignment crossings (*CrossS*), as generated when checking the ST against the TT

While *CrossT* values reflect the alignment effort for mapping ST tokens on the TT structure, as is required for translation production, *CrossS* values have a reverse interpretation, as they represent the mapping effort of TT tokens on the ST structure, as is more likely the case during revision. Figure 4 shows the *CrossS* values for the sentence in Figure 2. Note that the sequence of *CrossT* and *CrossS* are not symmetrical: in the given example *CrossS*: {3,-2,3,1,2,-1}. In section 3 we will show that both types of effort occur in translation and in post-editing.

The Cross value is small if source and target languages are (structurally) similar, and consists only of one-to-one token correspondences. The more both languages structurally differ or the less compositional the translations are, the bigger will become the Cross values.

Similarly, we expect to observe a larger number of translation choices as semantic shifts are introduced by the translator or if only “not-quite-identical meanings” are available.

3 Translators behaviour

Different parts of the TPR-DB have been used for the different analysis reported in this section. A set of 313 translations have been investigated to map translation crossings in section 3.1; 86 sessions were used for the post-editing experiment in section 3.2, and 24 translations for translation choices reported in section 3.3.

A simple linear regression was carried, to ascertain the extent to which total reading time

(*GazeS* and *GazeT*) can be predicted by *Cross* values in sections 3.1 and 3.2, and by translation choices in section 3.3. The correlation for Cross values in sections 3.1 and 3.2 was calculated from value 1 to the peak in each distribution in the negative and positive directions. Only Cross values from -8 to 8 are reported because items with higher Cross values are very rare, resulting in vastly unequal numbers of items.

3.1 Alignment Crossing

This section reports an analysis of 313 translation sessions with 17 different source texts into six different languages as contained in the TPR-DB. The target languages were Danish, Spanish, English, Chinese, Hindi and German; the source languages were English and Danish.

Figure 5 depicts gazing time on an ST token with a given *CrossT* value, while Figure 6 depicts gazing time on the TT tokens with a given *CrossS* value. These figures show that higher *CrossT* and *CrossS* values are strongly correlated with *GazeS* and *GazeT* and thus more effortful to process than lower *CrossT* and *CrossS* values.

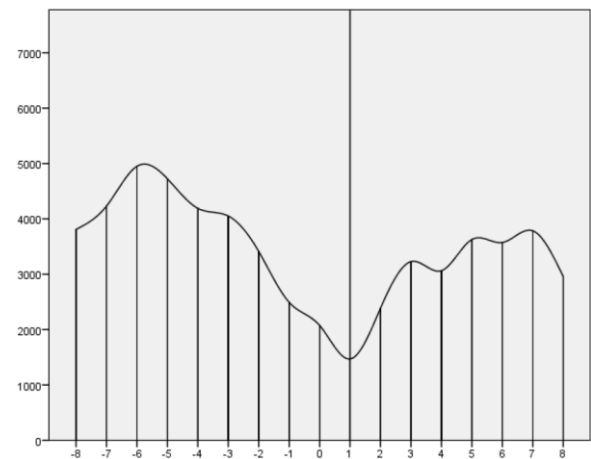


Figure 5: Average gazing time (vertical) on ST token with different *CrossT* values (horizontal)

Correlation between *CrossT* values and Total Reading Time on Source Text

As shown in Figure 5, a strong positive correlation was found between *CrossT* values and total reading time on the source text ($r=.97$ for negative *CrossT* values and $r=.91$ for positive *CrossT* values). The regression model predicted

97% and 82% of the variance for negative and positive values. The model was a good fit for the data ($F=205.7$, $p<.0005$ and $F=22.89$, $p<.005$, respectively). For every single increase in the negative *CrossT* value, the total reading time on the source text increased by 516ms, for positive *CrossT* value, the total reading time on the source text increased by 347ms.

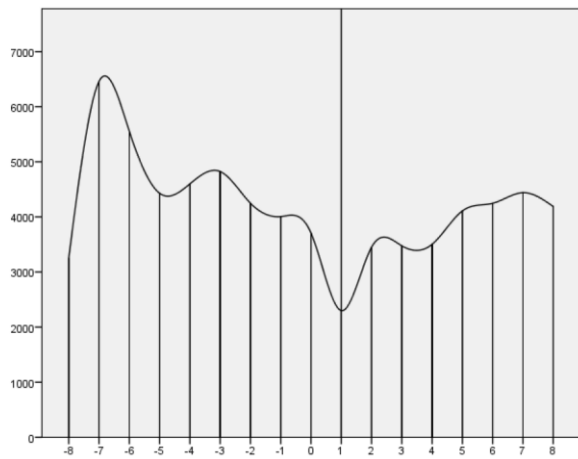


Figure 6: Average gazing time (vertical) on TT tokens for different CrossS values (horizontal)

Correlation between CrossS values and Total Reading Time on Target Text

Also for negative and positive *CrossS* values and total reading time on the TT a strong positive correlation was found ($r=.92$ and $r=.93$, respectively). The regression model predicted 84% and 85% of the variance, and was a good fit for the data ($F=36.97$, $p<.001$, $F=30.69$, $p<.003$). For every single increase in the negative *CrossS* value, the total reading time on the target text increased by 389ms, for positive *CrossS* values the total reading time on the target text increased by 301ms.

3.2 Alignment crossing in post-editing

This section reports an analysis over 86 different post-editing sessions from the TPR-DB of 9 different English source texts which were translated into three different target languages, German, Hindi and Spanish. As in section 3.1 the analysis shows that *CrossT* and *CrossS* values correlate with the total reading time per word (*GazeS* and *GazeT*). Figures 7 and 8 plot gazing

times on ST and TT token with different *CrossT* and *CrossS* values during post-editing.

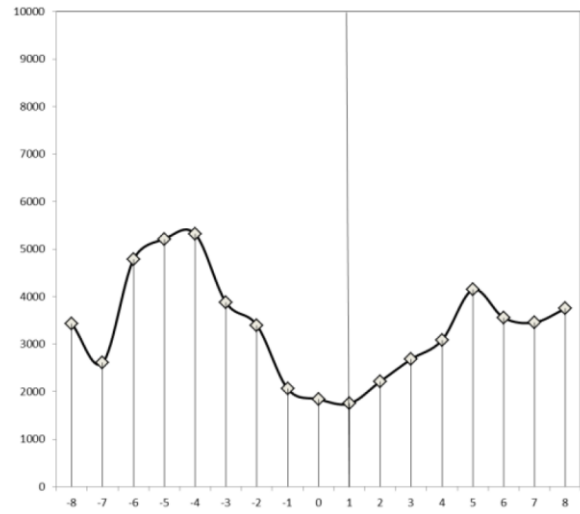


Figure 7: Average gazing time on ST tokens during post-editing for different CrossT values (horizontal)

Correlation between CrossT values and total reading time on source text

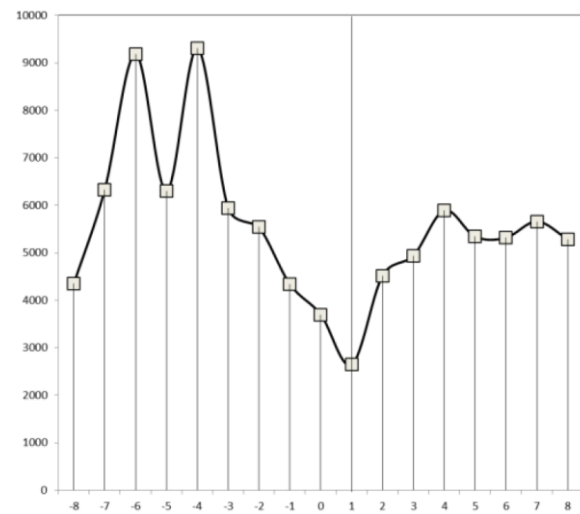


Figure 8: Average gazing time (vertical) on TT tokens during post-editing for different CrossS values (horizontal)

Similarly, a strong positive correlation was found between negative *CrossT* values and total reading time on the source text ($r=.95$ and $r=.98$), and the regression model predicted 88% and 94% of the variance for negative and positive *CrossT* values. The model was a good fit for the data ($F=38.50$, $p<.003$ and $F=67.56$, $p<.004$). For every single increase of the negative *CrossT* value, the total reading time on the target text increased by 723ms, while for positive *CrossT* values reading time increased by 566ms.

Correlation between *CrossS* values and total reading time on target text

A strong positive correlation was found between *CrossS* values and total reading time on the target text ($r=.95$), and the regression model predicted 87% and 89% of the variance for negative and positive *CrossS* values respective. The model was a good fit for the data ($F=35.26$, $p<.004$ and $F=38.50$, $p<.003$). For every single increase in the negative *CrossS* value, the total reading time on the target text increased by 1179ms, while for positive *CrossS* values reading time increased by 1016ms.

3.3 Translation choices

The data used for translation from scratch used for this purpose are 24 translations of 3 different texts from English into Danish and the data for post-editing used for this purpose are 65 post-edited translations of 9 different source texts involving one source language (English) and two target languages (German and Spanish). The number of alternative translations for every source item of the different source texts were counted. Only words which had up to 9 alternative choices were included in the analysis, partly so that a comparison between translation from scratch and post-editing was possible and partly because there are few items with more than 9 alternative translations.

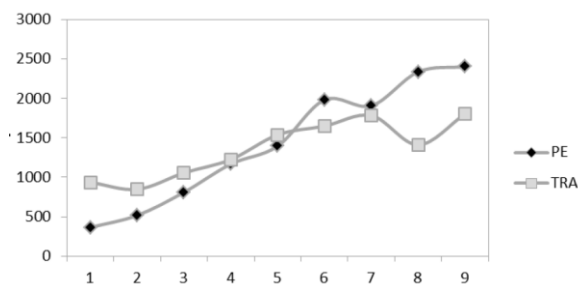


Figure 9: Correlation of alternative translation (horizontal) and average production time (vertical) for translation (TRA) and post-editing (PE).

Correlation between duration and alternatives

As shown in Figure 9, for translation from scratch and for post-editing there was a strong correlation between the time it took participants to produce a target word and the number of

alternatives for every source word ($r=.89$ and $r=.99$, respectively). With few choices post-editors are quicker than translators, but this distance decreases as the number of translation choices increase. The regression model predicted 76% and 97% of the variance and was a good fit for the data ($F=26.14$, $p<.001$) for Translation and ($F=269.50$, $p<.0001$) for post-editing. For every increase in the number of alternatives, the production time increased by 117ms, respectively 278ms for translation and post-editing.

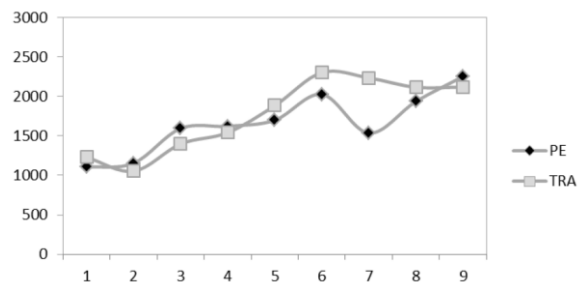


Figure 10: Correlation of alternative translation (horizontal) and average gazing time on TT words (vertical) during translation (TRA) and post-editing (PE).

Correlation between total reading time on the target text and alternatives

Similarly, Figure 10 depicts a strong correlation for translation from scratch and for post-editing between the total reading time on the target text per word and the number of translation choices for every source word ($r=.90$ and $r=.87$ respectively). The regression model predicted 77% and 72% of the variance and the model was a good fit for the data; $F=28.45$, $p<.001$ and $F=21.80$, $p<.002$ for translation and post-editing respectively. For every increase in the number of alternatives, the total reading time on the target text increased by 153ms, and 120ms.

Correlation between total reading time on the target text and alternatives

For translation from scratch, there was a strong correlation between total reading time on the source text per word and the number of alternatives for every source word ($r=.76$), but the regression model only predicted 52% of the variance. The model was a good fit for the data

($F = 9.74$, $p < .017$). For every increase in the number of alternatives, the total reading time on the source text increased by a modest 47ms.

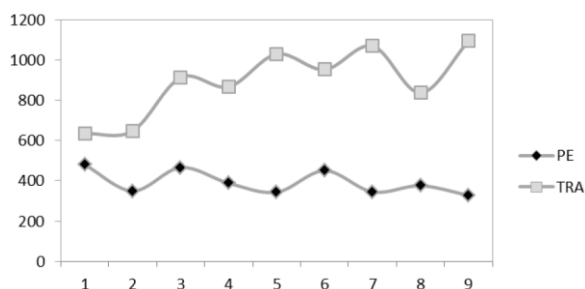


Figure 11 Correlation of alternative translation (horizontal) and average gazing time on ST words (vertical) during translation (TRA) and post-editing (PE).

However, as depicted in Figure 11, for post-editing there was no correlation between total reading time on the source text per word and the number of alternatives for every source word.

4 Discussion

The investigation reported here is not the first of its kind. Dragsted (2012) compared eye movement measures (total reading time and number of fixations) and pauses for words which were translated by 8 participants using the same target word with words for which the eight participants used different words.

She found that the total reading time and the number of fixations on words with many (5-8) alternatives target text items was significantly higher than the number of fixations on words with only one or two different target items. She also found that the pauses prior to critical words were longer for words with many alternatives as compared to words with one or two alternatives.

This seems to confirm the assumption that the more lexical choices a translator has to consider, the more effortful the processing of this item becomes. Campbell (2000: 38) suggests that “the complexity of choices available to the translator to select from” can be taken as a measure of the effort of the related cognitive processes.

Our analysis investigates this suggestion on a larger scale, involving more language pairs and two conditions: translation from scratch and

post-editing. It shows similar results to those of Dragsted (2012), but in addition shows that effect of alternatives on *production time* per word was much stronger for post-editing as compared to translation (171ms for translation vs. 278ms for post-editing). This suggests that highly (translation) ambiguous texts should perhaps not be considered for post-editing. In (Carl and Schaeffer, 2014) we look at this effect in more detail by investigating the word translation entropy in human and machine produced translations and propose a translation ambiguity threshold that might be suitable for post-editing.

The effect of translation choices on total *TT reading time* was comparable for translation and post-editing (153ms for translation vs. 120 for post-editing). For total *ST reading time* there was no effect for post-editing, while every additional translation choice increased the total *ST reading time* by 47ms, however modest compared to the effect on *TT reading time*. This finding suggests that in from scratch translation choices are already processed during *ST reading*, while during post-editing choices are considered mainly when the gaze is on the *TT*.

As a second variable we investigate *ST-TT crossing values*. Higher *Cross* values indicate non-monotonous translation relations such as local distortions of *ST-TT alignment*, discontinuous, idiomatic or multi-word units, all of which require larger sequences of the source and/or target text to be integrated and related, and thus increased effort when maintaining and processing larger numbers of items in working memory. The large increases of total *ST reading time* for tokens with higher *CrossT* values in translation and post-editing suggests that integrating larger *ST chunks* is also more effortful during translation and post-editing.

Similar findings are also reported by Jensen et al. (2010) who investigate gazing time for English-Danish verbal translations when they switch their sentence position (SVO → SOV) vs. they remain in both languages in the same sentence position (SVO → SVO). Our investigation generalizes

these findings to different language pairs and all kinds of relative ST-TT distortion.

Another observation is related to the large increases in total *TT reading time* for higher *CrossS* values, during translation and post-editing. This observation suggests that translators not only read the ST to generate a TT equivalent, but they also check the produced TT whether it corresponds to the ST. As one could expect, this tendency is very pronounced during post-editing, but appears interestingly also during translation from scratch. The observation is in line with a previous assumption of Carl and Dragsted (2012: 141) who find that source text related processes are “triggered by problems associated with text production rather than” during source text reading.

Note that for all analysis, both translation and post-editing, reading time increased much more with negative *Cross* values than this is the case for positive *Cross* values. This coincides with the finding that regressions - which negative *Cross* values reflect - are more effortful to process than progressions, since regressions often mirror misunderstanding and imply the integration of already parsed input text (e.g. Reichle et al 2009).

5 Conclusion and outlook

There has been some discussion in the translation process research (TPR) literature on the “tendency of the translating process to proceed literally to a certain extent” Tirkkonen-Condit (2004: 183), where a deviation from the ideal default translation would result in higher effort. However, to our knowledge the literal default translation hypothesis has never been quantified and empirically assessed in a larger context. In this paper we bridge this gap. We provide a quantifiable definition of *literal translation* as a continuous concept involving alternative translation choices and source-target distortions, apply it to a collection of translation and post-editing sessions from the TPR-DB and assess translation effort by measuring gazing and translation time. We find that gaze activity and

production time is inversely proportional to the literality of the produced translations. Using linear regression we find in particular:

- More translation choices lead to longer reading and processing time
- Longer relative source-target language distortions increase gaze activity.
- Regressions are more effortful than progressions
- Translators and post-editors map not only the source text against the target, but also the target against the source text

These findings suggest a model in which, paradoxically, translators already know the translations which they produce; they merely refer to the ST - and to the TT for cross-checking - to verify the translation hypothesis which they already have in mind.

A number of issues remain open for further research. For instance, the impact of the target language and the (syntactic) similarity of the source and target languages. According to the hypothesis supported here, closely related languages with similar word order and similar conceptual repository will more likely have more literal translations. They will more often consist of monotonous one-to-one translations, approaching an *ideal literal translation* (Schaeffer, 2013). The more syntactic reordering between source and target text take place the more it will become non-literal.

Another set of questions relates to whether and how the methods discussed here can be used to assess the cognitive effort for translating and/or post-editing entire sentences and texts and the impact on post-editing practice.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant

agreement no 287576. We are grateful to all contributors to the TPR database for allowing us the use of their data.

References

- Campbell, Stuart. 2000. "Choice Network Analysis in Translation Research." In *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*, edited by Maeve Olohan, 29–42. Manchester: St Jerome.
- Carl, Michael. 2012. "The CRITT TPR-DB 1.0: A Database for Empirical Human Translation Process Research." In *Proceedings of the AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*, edited by Sharon O'Brien, Michel Simard, and Lucia Specia, 9–18. Stroudsburg, PA: Association for Machine Translation in the Americas (AMTA).
- Carl, Michael, Mercedes García Martínez, Bartolomé Mesa-Lao, Nancy Underwood. 2014. "CFT13: A new resource for research into the post-editing process." *Proceedings of LREC*
- Carl, Michael, and Barbara Dragsted. 2012. "Inside the Monitor Model: Processes of Default and Challenged Translation Production." *Translation: Computation, Corpora, Cognition* 2 (1): 127–145.
- Carl, Michael, and Moritz Schaeffer (2014) "Word Transition Entropy as an Indicator for Expected Machine Translation Quality", *Proceedings of LREC*
- Dragsted, Barbara. 2012. "Indicators of Difficulty in Translation — Correlating Product and Process Data." *Across Languages and Cultures* 13 (1) (June 1): 81–98. doi:10.1556/Acr.13.2012.1.5. <http://www.akademai.com/openurl.asp?genre=article&id=doi:10.1556/Acr.13.2012.1.5>.
- Ivir, Vladimir. 1981. "Formal Correspondence Vs. Translation Equivalence Revisited." *Poetics Today* 2 (4): 51–59.
- Jensen, Kristian T.H., Annette C. Sjørup, and Laura W. Balling. 2010. "Effects of L1 Syntax on L2 Translation." In *Methodology, Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by F. Alves, S. Göpferich, and Mees Inger M., 319–336. Copenhagen: Samfundslitteratur.
- Malmkjær, Kirsten. 2011a. "Linguistic Approaches to Translation." In *Oxford Handbook of Translation Studies*, edited by Kirsten Malmkjær and Kevin Windle, 57–70. Oxford: OUP.
- Malmkjær, Kirsten. 2011b. "Translation Universals." In *The Oxford Handbook of Translation Studies*, edited by Kirsten Malmkjær and Kevin Windle, 83–94. Oxford: Oxford University Press.
- Reichle, Erik D., Tessa Warren, Kerry McConnell. (2009). "Using E-Z Reader to model the effects of higher level language processing on eye movements during reading." *Psychonomic Bulletin & Review*, 16(1), 1–21.
- Schaeffer, Moritz. 2013. *The Ideal Literal Translation Hypothesis: The Role of Shared Representations During Translation*. PhD Thesis. University of Leicester.
- Tirkkonen-Condit, Sonja. 2004. "Unique Items: Over- or Under-Represented in Translated Language?" In *Translation Universals: Do They Exist?* Amsterdam and Philadelphia: John Benjamins.
- Tirkkonen-Condit, Sonja. 2005. "The Monitor Model Revisited: Evidence from Process Research." *Meta: Translators' Journal* 50 (2): 405–414.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Benjamins Translation Library V4. Vol. 75. Amsterdam and Philadelphia: John Benjamins.