A Minimal Cognitive Model for Translating and Post-editing

Moritz Schaeffer moritzschaeffer@gmail.com
Gutenberg University, Mainz, Germany

Michael Carl m.gummiball@gmail.com
Renmin University of China, and Copenhagen Business School, Denmark

Abstract

This study investigates the coordination of reading (input) and writing (output) activities in from-scratch translation and post-editing. We segment logged eye movements and keylogging data into minimal units of reading and writing activity and model the process of post-editing and from-scratch translation as a Markov model. We show that the time translators and post-editors spend on source or target text reading predicts with a high degree of accuracy how likely it is that they engage in successive typing. We further show that the typing probability is also conditioned by the degree to which source and target text share semantic and syntactic properties. The minimal cognitive Markov model describes very basic factors which play a role in the processes occurring between input (reading) and output (writing) during translation.

1 Introduction

We build a cognitive model of the translation process (from-scratch translation and post-editing) which aims at predicting where translation problems occur. We ground the model in translation activity data that consists of keystrokes and gaze data that was captured during translation sessions. We decompose the translation process into minimal cycles of iterative reading and writing. We assume that the typing activities represent the solution to a translation problem that emerged during the preceding reading event. We show that the complexity (i.e. non-literality) of the produced translation as well as the duration and distribution of gaze activities on the source and target texts has an effect on the probability of a successive typing event.

Schaeffer et al. (2016); Hvelplund (2016); Carl et al. (2016); Läubli and Germann (2016) describe methods to decompose the stream of eye movements and keystrokes into sequences of minimal activity units. In this paper we relate the duration of activity units with properties of the translation product — the degree of translation literality — to predict the probability when post-editors and translators will type after reading either the source (henceforth ST) or the target text (henceforth TT).

Carl et al. (2016) show that a measure of *translation literality* has a great predictive power for behavioral observations in the translation process. According to this definition, a translation is literal if:

1. Word order is identical in the ST and TT

2. ST and TT items correspond one-to-one

3. Each ST word has only one possible translated form in a given context

A translations which completely fulfills all three criteria is an *absolutely literal translations*. A *literal* translation consists of the same number of ST and TT tokens where each TT token corresponds to exactly one ST token, and tokens in both texts are ordered in the same way. A change in word order or a situation in which one ST word is aligned to more than one TT word or vice versa weakens literality criteria 1 and 2 and makes a translation less literal. Criteria 1 and 2 thus measure the *syntactic similarity* of an ST and its translation. The third criterion describes the *semantic similarity* in both languages. If a word (or phrase) is consistently translated in the same way by different translators, we assume that the ST word and its translation also have large overlapping semantic properties. The more a source word (or phrase) can be rendered into different translations, the weaker is also the semantic overlap between the two languages (with respect to this word or phrase). In this paper we show that the degree of translation literality has an effect on the reading activities prior to translation typing.

In section 2 we introduce an operationalization of the literality metric as described above. We introduce a metric "HCross" which measures the entropy of word-order choices that are observed in alternative translations, and which is strongly predictive for reading time duration during the translation process. Section 3 presents the material of our empirical study. In section 4 we introduce translation units and translation states, as well as the topology of a minimal cognitive model for from-scratch translation and post-editing. We review similar work which used transition networks of activity units to model novice and expert translators. We review a proposal that defines different translation styles and map these onto sequences of translation states of our minimal cognitive model. In section 5 we analyze our data and develop a minimal model of translation and post-editing.

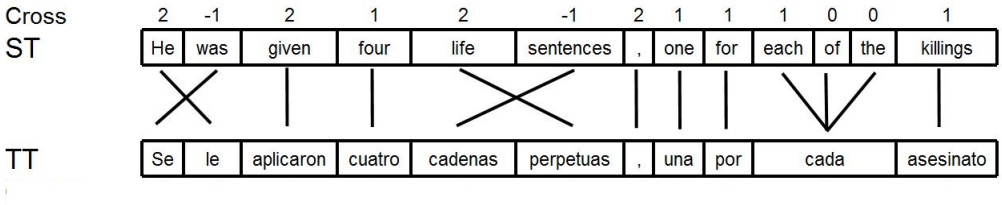


Figure 1: An English-Spanish alignment with Cross values

2 Operationalising Translation Literality

2.1 Word-order Distortion (Cross)

From a given translation and its word alignment relations we compute Cross values (see Figure 1). For any two successive source words s_{k-1} and s_k, we follow the alignment links to their translations ($s_{k-1} \rightarrow t_{k-1}$ and $s_k \rightarrow t_k$}) and compute the distance between the position of words t_k and t_{k-1} in the translation (i.e. $\text{position}(t_k) - \text{position}(t_{k-1})$) as the value for $\text{Cross}(s_k)$. We thus obtain a vector of relative alignment distortions for word positions in the ST and the TT, indicating the word order similarity of the two sentences. In the case of an (absolutely) literal translation, we say that each successive word aligns with the next one in the target language, which provides the Cross vectors with values 1.

For instance, the word [He] in Figure 1 occurs at position 1 on the English source side while its Spanish translation [le] occurs one word ahead at position 2 in the translation. [He] thus has a $\text{Cross}(s_1)$ value of 2 in that sentence. In order to generate the translation [aplicaron] for the English [given] we need to jump from the previous alignment [was-Se] two words to the

right, which produces a $\mathrm{Cross}(s_3)$ value of 2. In this way, Cross values are generated for each word position in the text, for the source and the target sides. If a word is aligned to more than one word (e.g. $s_k \rightarrow \{t_{k_1} \dots t_{k_n}\}$), $\mathrm{Cross}(s_k)$ is the signed value of the maximum absolute difference between the two translations, i.e. $\max(abs(\{t_{k_1} - t_{k-1}), \dots, abs(t_{k_n} - t_{k-1})\})$. In this way, t_{10} which has the alignment [cada → "each of the"] has an alignment distortion value $\mathrm{Cross}(t_{10}) = 3$.

2.2 Word Translation Entropy (HTra)

Carl et al. (2016) introduce word translation entropy as a measure to quantify observed translation choices. Entropy, H, represents the average amount of non-redundant information provided by each new item. It is computed based on the sum of the probability of the items and their information. The information of a probability p is defined as $I(p) = -\log_2(p)$ The entropy H is the expectation of that information as defined in equation (1):

$$H = \sum_{i=1}^{n} p_i I(p_i) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{1}$$

We adopt this notion to assess the entropy of word translation choices for a given ST word s_k into its n possible translations $t_{i\dots n}$ as shown in equation (2)

$$\mathrm{HTra}(s_k) = -\sum_{i=1}^{n} p(t_i|s_j) \times \log_2(p(t_i|s_j)) \tag{2}$$

The word translation entropy $\mathrm{HTra}(s_k)$ in equation 2 is computed for each source word s_k and in every segment. The translations $t_{i\dots n}$ are taken only from the aligned alternative translations of this segment. That is, the word translation probabilities $p(t_i|s_k)$, as computed according to equation (3), represent the ratio of the number of observed translations $s_k \rightarrow t_i$ separately for each source segment in which s_k occurs. Thus, while in language modeling, the entropy indicates how many possible continuations for a sentence exist at any time, we deploy the metric to assess how many different translations an ST word has in a given context.

$$p(t_i|s_k) = \frac{\mathrm{count}(s_k \rightarrow t_i)}{\mathrm{count}(s_k)} \tag{3}$$

We take it, that HTra reflects the semantic similarity between a source word and its translation(s): low HTra values indicate a high amount of agreement between translator choices, and thus a high degree of semantic similarity according to literality criterion 3 above.

2.3 Word-order Entropy (HCross)

The choices that a translator has to re-order translations of a source word s_k in the target language is captured by the metric HCross, as given in equation 4.

$$\mathrm{HCross}(s_k) = -\sum_{i=1}^{n} p(\mathrm{Cross}(s_k)) \times \log_2(p(\mathrm{Cross}(s_k)) \tag{4}$$

The probability for each relative translation word-order distortions $p(\mathrm{Cross}(s_k))$ for a source word s_k is computed as the ratio of the number of the distortions $\mathrm{Cross}(s_k)$ for alternative translations $s_k \rightarrow t_{1\dots n}$ divided by the total number of observed alternative translations $\mathrm{count}(s_k)$, similar to equation 3.

HTra and HCross values correlate to a high degree (*r*=.79, $p < .001$). That is, semantic and syntactic variation seem to correlate highly in translation. More variation in syntactic (i.e.

word-order) rendering of the translation seem to come along with more variation in lexical choices, and vice versa: Low cross-lingual semantic similarity (i.e. high HTra values) are correlated with high syntactic variation and complexity (i.e. high HCross values).

3 Experimental material

As a basis for our investigation in this paper we use the *multiLing* subset of the TPR-DB (Carl et al., 2016). The *multiLing* set consists of six short English source texts (together 849 words, 40 ST segments) and a large number of alternative translations into Danish (da), Spanish (es), German (de), Hindi (hi), Chinese (zh) and Japanese (jp) each by several translators. It contains currently more than 1500 text production sessions, for from-scratch translation (T), post-editing (P), monolingual editing (E), translation dictation (D) and text copying (C). However, in this study we only make use of from-scratch translation and post-editing, which amounts to approximately half the data, 124 hours productin time. For each text production session, keystroke and gaze data were collected and stored. A real-time gaze-to-word mapping tool (Carl, 2012) was used to map the gaze samples on the words, so that it is known which word was gazed at, at any time during the translation sessions. The tool also computes which keystroke contributes to the production (or modification) of which word. The STs and TTs were manually aligned using the YAWAT tool (Germann, 2008). Aligners were adviced to align each segments as compositional and complete as possible. The aligned data were further post-processed into a set of summary tables, which integrate and describe the data of the translation process and the translation product by means of currently more than 300 features (Carl et al., 2016).

	SText:#Seg		1:6	2:7	3:5	4:5	5:10	6:7	STtok	STseg	
	ST Token		160	154	146	110	139	139	848	40	
Task	Study	TL	Alt	Alt	Alt	Alt	Alt	Alt	TTtok	TTseg	Dur
P	BML12	es	10	12	10	12	8	12	10216	431	5.22
	ENJA15	ja	13	12	14	12	13	12	14447	519	16.81
	MS12	zh	3	5	3	3	3	2	2561	129	3.18
	NJ12	hi	7	12	8	10	12	11	9365	409	18.2
	SG12	de	8	7	7	8	7	8	6470	305	8.78
T	BML12	es	11	10	8	10	12	8	9938	411	10.34
	ENJA15	ja	12	13	12	13	13	13	14134	525	22.46
	KTHJ08	da	24	23	22	0	0	0	10667	523	7.7
	MS12	zh		3	3	3	3	3	1916	89	4.12
	NJ12	hi	7	7	5	7	6	6	5783	266	14.84
	SG12	de	6	8	8	8	7	8	6777	305	12.46
	Total		**101**	**112**	**100**	**86**	**84**	**83**	**92274**	**3912**	**124**

Table 1: Subset of the TPR-DB *multiLing* corpus with the post-editing (P) and from-scratch translation (T) data. The table shows for each of the six English source texts the number of segments and the number of words, as well as the total number of ST segments (STseg:40) and words (STtok:848). It also shows for each language the number of alternative translations (Alt) the total number of target text tokens (TTtok), segments (TTseg) and duration (Dur) per target language and for each of the translation modes.

Table 1 shows some figures of *multiLing* Corpus. The length in words for each of the six STs is given in the first row in Table 1 (ST1-6). For each of the six STs, the table indicates the number of participants (#Part), and for each of the six STs the number of alternative translations (Alt) and their total number in tokens (TokT). The total number of target words (TtokT) and

target sentences (Ttsg) is also provided, together with the total production duration in hours (Dur). The data is freely available. For more information on this dataset, please consult the CRITT website.[1]

4 Translation states

We extend the work of Schaeffer et al. (2016), who introduce Activity Units as a means to segment the stream of translation (and post-editing) activity into distinct units. Similar to Carl et al. (2016) they make a distinction between 6 different basic types of activities [2]:

- type 1: ST reading
- type 2: TT reading
- type 4: translation typing (no gaze data recorded)
- type 5: ST reading and typing (touch typing)
- type 6: TT reading and typing (translation monitoring)
- type 8: no gaze or typing activity recorded for more than 2.5 seconds

In this study we simplify the 6 types of activity units into four translation states. We collapse activity units 4, 5 and 6 into writing activities (W), irrespectively of whether reading activities are also recorded at the same time. This leaves us with the following four translation states:

	Post-editing						From-scratch translation					
	# OBS	%Dur	S_2	T_2	W_2	P_2	# OBS	%Dur	S_2	T_2	W_2	P_2
S_1	15695	26	0.00	**0.81**	0.16	0.02	17756	29	0.03	**0.52**	0.42	0.03
T_1	19275	**40**	**0.56**	0.01	0.41	0.03	17417	19	0.42	0.00	**0.54**	0.03
W_1	13092	27	0.35	**0.44**	0.14	0.07	26187	**44**	**0.36**	0.28	0.30	0.05
P_1	1723	8	0.19	0.28	**0.53**	0.00	2303	8	0.18	0.21	**0.60**	0.00
Total	49785	38.76 hours					63663	42.44 hours				

Table 2: Distribution of translation states in number of total observations (#OBS) and duration (%Dur), as well as a transition matrix for post-editing and from-scratch translation. The data represents translation states during the drafting phase of the data from Table 1

- S: ST reading (with no concurrent writing activity)
- T: TT reading (with no concurrent writing activity)
- W: Writing (with or without concurrent gaze activity on the source or target window)
- P: Pausing (no activity recorded for more than 2.5 seconds)

Each of the translation states (i.e activity units) can be described by a number of features (excluding P which has only a duration), including the number of keystrokes (deletions and insertions), the word(s) produced by the keystrokes, the number and duration of fixations, the fixation scanpath (i.e. sequence of fixations) within a state, including the number of different words fixated, their average distance etc. (cf. Schaeffer et al. (2016)).

[1] sites.google.com/site/centretranslationinnovation

[2] The Activity Unit of type 7, as suggested in Carl et al. (2016), which entails concurrent type 1, 2 and 4 behaviour is not assumed here. Instead the activities were split into the six types above.

4.1 State transitions in translation and post-editing

The data in Table 2 shows the distribution of translation states from the *mulitLing* data which were introduced in Table 1. The total dataset was segmented into 49,785 and 63,663 activity units for the post-editing and translation experiments respectively.

The data represented in Table 2 only accounts for the activities during the drafting phase. This amounts to 38.76 hours post-editing and 42.44 hours translating. The column #OBS shows the number of observations per translation state, while the %Dur column gives their percentage of the total production duration. In the post-editing mode, most activities (19,275 units) were observed in the TT reading (T_1) mode, as well with respect to the number of units and with respect to their duration. In the translation mode, the translators were 44% of the total time involved in writing activities.

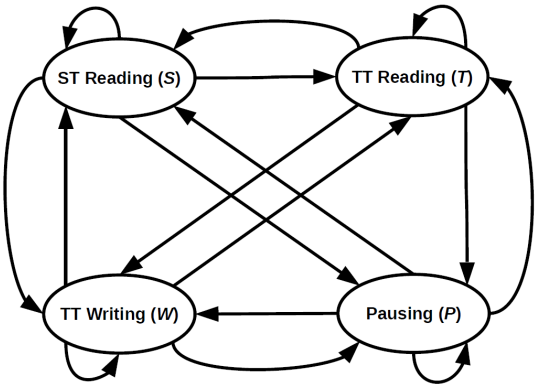


Figure 2: A fully connected translation process transition network with four states.

The columns S_2, T_2, W_2 and P_2 provide the likelihood of the next state to which post-editors or translators will switch.[3] For instance, if a post-editor is involved in an ST reading state (S_1), there is a high chance of 81% that next he or she will switch to TT reading (T_2). Once in the T_1 state, the highest probability (56%) is to switch back to ST reading (S_2). This is different in the translation mode, where the translator will most likely turn to writing (W_2) after a T_1 activity. Table 2 provides thus a transition table which can be represented in the form of a completely connected transition network as shown in Figure 2. Each state in the network in Figure 2 is connected to each other state in the network, and the transition from one state to the successor states are weighted by probabilities, such that the sum of all outgoing archs sums to 1.0. Two possible instantiations of the transition network are shown in Table 2, which produce slightly different behavior for post-editing and for from-scratch translation.

4.2 Novice and expert translators

Hvelplund (2016) reports that novice and expert translators exhibit different behavior with respect to the length and the sequencing of translation activities. His study is restricted to the English to Danish data collection which is gathered in the KTHJ08 study in Table 2. According to Hvelplund, experienced translators shift more often from ST reading (S_1) directly to writing (W_2) than student translators; in 65.5% and 52.2% of the cases respectively. Student

[3]There are also transitions in the diagonal e.g. $W_1 \rightarrow W_2$ which result from the fact that we have collapsed activities of type 4,5 and 6 into one state. We will ignore them here, since we are not concerned with these transitions in this paper.

translators show more occurrences of TT reading than professionals, which suggests that students aim more often at confirming meaning hypotheses, rather than allocating the cognitive resources directly to writing once a meaning hypothesis has been established. Hvelplund also finds a higher variability in the unit duration of professional translators as compared to student translators. Hvelplund sees this as an indicator for greater ability to adapt to the situation by the professional group.

While Hvelplund investigates the impact of the level of translation expertise on the activity transition probabilities of $S_1 \rightarrow W_2$, we will show below in section 5 that the inner structure of the preceding units (i.e. S_1) themselves seem to determine to some extent the transition to the next state.

4.3 Post-editing styles

Based on a taxonomy that overlaps to some extent with our six activity units, Mesa-Lao (2013) suggests six post-editing steps and develops a minimal model of post-editing with spells out four translation styles. His first two post-editing styles are:

- style$_1$: The post-editor first reads the TT segment, detects an MT error, reads the ST segment, and fixes the MT error.
- style$_2$: The post-editor first reads the ST segment, then the TT segment, detects an MT error, and fixes it.

Translation style$_3$ in Mesa-Lao's taxonomy is a variations of style$_2$ (omit ST reading) and in style$_4$ the post-editor reviews first a previous segment before fixing the MT error. Post-editing style$_1$ seems to be the most preferred among his participants. However, in order to simulate Mesa-Lao's translation styles based on the available data that we have (keystrokes and fixations) and the the four translation states, we cannot know when a translator actually detects an MT error. Skipping the step "detect an MT error" leaves us thus with two post-editing patterns that we can map on sequences of the translation states: style$_1$: $T \rightarrow S \rightarrow W$ and style$_2$: $S \rightarrow T \rightarrow W$. In the following section we reduce these two patterns even further and examine the minimum translation cycles $T \rightarrow W$ and $S \rightarrow W$, which represent the question: what happens before typing?

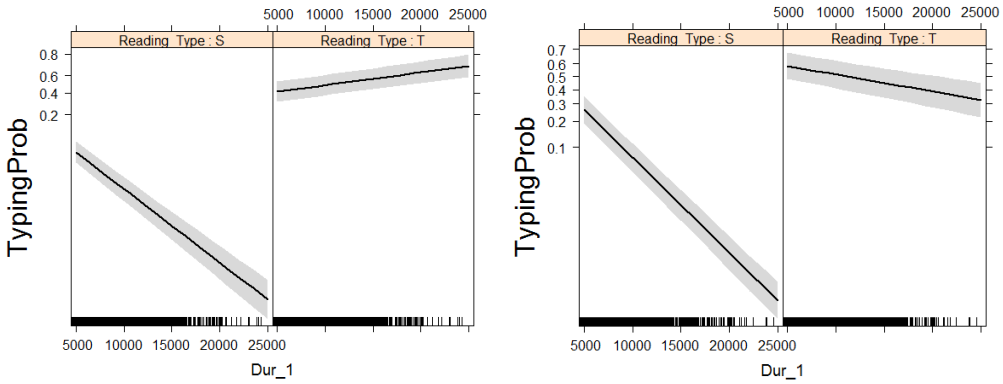
5 Determinants of writing probability

In this section we analyze where and for how long the gaze was observed prior to writing. We will also test to what extent the HCross value (i.e possibility for syntactic choice) of the typed text has an effect on S_1 and T_1 reading duration, prior to typing W_2. The analysis tells us something about the processes which take place between the input (S_1 and T_1 reading), the output (W_2 writing activity) in the cognitive system.

For all the analyses in the present study, R (R Development Core Team, 2014) and the lme4 (Bates et al., 2014) and languageR (Baayen, 2013) packages were used to perform generalized linear mixed-effects models. To test for significance, the R package lmerTest (Kuznetsova et al., 2014) was used. Two separate models (one for post-editing and one for from-scratch translation) with reading duration and HCross as predictors and their interaction with reading type were tested. Both models had participant and target language as random factors.

5.1 The effect of reading duration on writing probability

Increased S_1 reading duration during post-editing (Figure 3a, left) and from-scratch translation (Figure 3b, left) decreases the probability of successive writing (TypingProb). Increased S_1 reading duration thus increases the chances that post-editors and translators engage in successive T_2 reading, instead of writing (W_2). The reason might be due to ST comprehension



(a) **Post-editing** (b) **Translation**

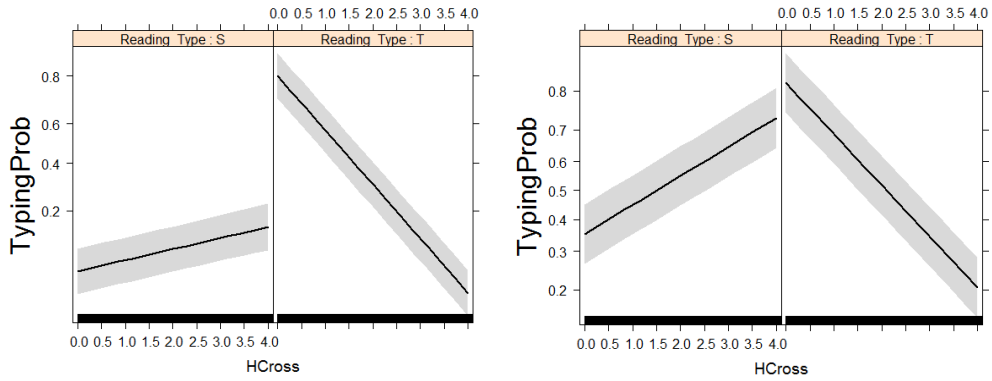
Figure 3: The effect of S_1 and T_1 reading duration (**Dur_1**) on the probability (**TypingProb**) that participants engage in successive writing activity W_2. The gray shadow represents the standard error.

or translation difficulties, which require longer S_1 reading reading times, for both post-editing and from-scratch translation. The more information is processed during ST reading (long S_1 reading), the stronger is the need to first cross check the emerging translation hypothesis with the existing TT, before typing in the translation solution - possibly due to working memory limitations. We thus see more likely a transition $S_1 \rightarrow T_2$ for longer S_1 reading times. That is, a translation hypothesis gathered during S_1 reading needs to be integrated with the existing TT before writing a solution. If the ST information intake is long (long S_1 reading), memory on the status of the TT might first need to be refreshed through (re-newed) TT reading in order for the new solution to be properly integrated. Accordingly, the 16% and 42% of $S_1 \rightarrow W_2$ transitions in post-editing and from-scratch translation respectively (see Table 2) take mainly place if S_1 reading durations are short ($<$ 5000ms, see section 5.3).

Longer T_1 reading activities increase the probability of a successive writing (W_2) for post-editing (Figure 3a, right) but decrease the probability of successive writing for from-scratch translation (Figure 3b, right). This difference in TT reading patterns might be due to the fundamental difference between post-editing and from-scratch translation. In post-editing a TT already exists and some modifications can be made without consultation of the ST. W_2 activities after longer T_1 reading times during post-editing might relate to the correction of (relatively minor) fluency errors which can be corrected without consultation of the TT.

In from-scratch translation, information from the ST needs to be retrieved and integrated with the existing translation in order to continue producing the emerging TT. The longer from-scratch translators read the TT, the more likely they will need to retrieve new information from the ST in order to continue translation production

There were highly significant main effects for reading type, for post-editing (β=4.02, SE=0.06, t=62.54, $p < .001$) and for from-scratch translation (β=2.11, SE=0.04, t=54.65, $p < .001$), such that writing (W_2) was more likely after TT reading (T_1). There were also highly significant main effects for reading duration for post-editing (β=-0.91, SE=0.045, t=-20.38, $p < .001$) and for from-scratch translation (β=-0.68, SE=0.02, t=-33.25, $p < .001$), such that longer reading activities (**Dur_1**) made writing less likely. The interaction between reading type (S_1/T_1) and reading duration (**Dur_1**) was highly significant for post-editing (β=1.07, SE=0.05, t=22.65, $p < .001$) and for from-scratch translation (β=0.56, SE=0.03, t=21.41, $p < .001$).



(a) **Post-editing** (b) **Translation**

Figure 4: The effect of the HCross on the probability that participants type immediately after the reading activity (TypingProb), depending on whether the source (S) or the TT (T) is read prior to the writing event.

5.2 The effect of HCross on writing probability

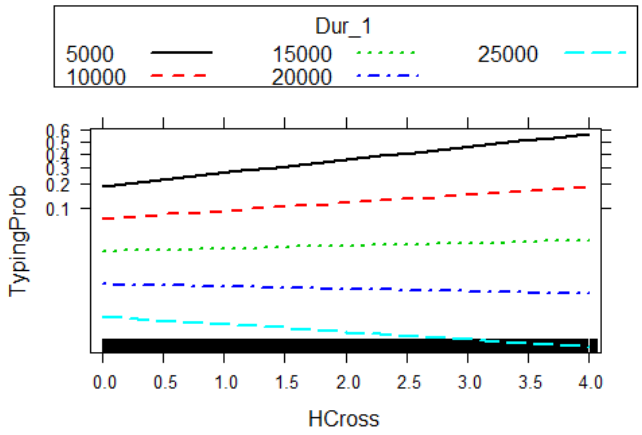
As discussed in section 2, HCross represents the possibility for the translation of a word or phrase to occur in different syntactic positions in the target text segment. HCross is highly correlated with cross-lingual semantic similarity - HTra and HCross correlate to a high degree (r=.79, $p < .001$). The more likely it is that different word orders are realized (high syntactic complexity), the more likely it is that different lexical items are used (high semantic complexity).

For both post-editing (Figure 4a) and from-scratch translation (Figure 4b), HCross had a positive effect on the probability that writing follows ST reading. Thus, higher HCross values increase the probability of a $S_1 \rightarrow W_2$ transition. This effect was more pronounced for from-scratch translation then on post-editing. However, for both post-editing and from-scratch translation, HCross had a negative effect on the probability that writing follows TT reading. Again, this effect was more pronounced for from-scratch translation.

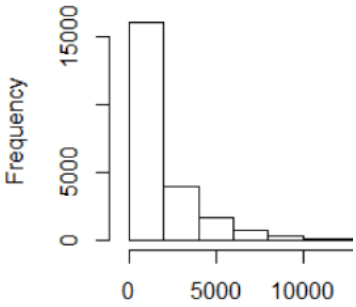
An explanation of this observation might be that items with higher HCross values can be seen as particularly challenging to translate and that solutions for difficult translations emerge during ST reading. The more complex the translation is, i.e. semantically and syntactically less similar, (the less literal), the more likely both post-editors and translators are to refer back to the ST and the less likely they are to type a translation solution immediately after reading the TT.

That is, the 41% and 54% of $T_1 \rightarrow W_2$ transitions in post-editing and from-scratch translation respectively (see Table 2) take preferably place if HCross values are low (the translation is easy). The solutions of more complex translation problems are preferably typed in after S_1 reading.

There were highly significant main effects for HCross, for post-editing (β=0.23, SE=0.02, t=9.73, $p < .001$) and for translation (β=0.41, SE=0.017, t=24.15, $p < .001$), such that writing (W_2) was more likely for higher HCross values. The interaction between reading type (S_1 / T_1) and HCross was highly significant for post-editing (β=-1.34, SE=0.03, t=-44.28, $p < .001$) and for translation (β=-1.12, SE=0.02, t=-49.84, $p < .001$).



(a) Interaction between HCross * Dur_1



(b) Distribution of S_1 duration (Dur_1)

Figure 5: Interaction between the duration (Dur_1) of an S_1 event and the complexity (HCross) of successive translation production W_2 on the probability of a $S_1 \rightarrow W_2$ transition. The typing probability increases with short S_1 reading times and high HCross values. Typing probability decreases with long S_1 reading times and high translation complexity (HCross).

5.3 Interaction of S_1 reading duration and HCross value on W_2 probability

As discussed in the previous sections, the $S_1 \rightarrow W_2$ typing probability during translation depends (among other factors) on the:

- expertise of the translator (section 4.2)
- S_1 reading duration (section 5.1)
- HCross value of the W_2 event (section 5.2)

Figure 5a shows the interaction effect between S_1 reading duration (Dur_1) and the complexity of the translation (HCross) that follows the reading event. In line with the findings discussed in Figures 3a and 3b (left) it shows that short ST reading activities ($< 5000ms$) are followed with high probability by typing events. As shown in Figure 4a and 4b (left) the typing probability is even more likely if the produced translation solution is more complex. This suggests that complex translations are preferably produced immediately after a short ST consultation, presumably to relieve working memory by flushing out probably intermediate and incomplete translation solutions that are later to be revised and thus to avoid building up and keeping more complex structures in mind. In contrast, less complex translation problems may still be integrated with more information gathered during successive TT reading before a typing event occurs.

This trend is reversed for longer ST reading duration, where the typing probability decreases if the translation problem becomes more complex. It suggests that long S_1 reading duration in combination with complex translation problems requires additional T_2 reading, and presumably additional ST-TT integration cycles.

In combination, these observations suggest that difficult translation problems are cross-checked and resolved after reading the ST, while simple translation problems may be rectified after TT reading.

6 General Discussion

According to Dillinger (2014, xi), a key ability for post-editors (and translators) is their ability to compare sentences (and texts) across languages, in terms of both literal meaning and the culturally determined patterns of inference and connotation that different phrasings will entail. Patterns of keystrokes and gaze behavior make it possible to trace the origin of problems translators face to establish equivalence across languages. We have shown that bigrams of translation states, i.e., reading the ST or the TT and writing, constitute minimal and coherent problem identification and solution cycles. The degree of complexity (i.e. syntactic choice) clearly predicts subsequent activities, both during translation and post-editing. Remarkable in this regard is the fact that the effect of word-order choices in the target language (HCross) is similar in both tasks, suggesting that post-editors engage in processes which are not unlike those during from-scratch translation, when the raw MT output is faulty. Both post-editors and translators refer back to the source text when the produced TT is semantically and/or syntactically complex or non-literal. We hope that these minimal and coherent problem identification and solution cycles will constitute the building blocks for a more fully fledged model of both post-editing and from-scratch translation.

References

Baayen, R. H. (2013). languageR: Data sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics". Technical report.

Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (2014). {lme4}: Linear mixed-effects models using Eigen and S4.

Carl, M. (2012). *Translog-II: a Program for Recording User Activity Data for Empirical Translation Process Research*. Paper presented at The Eighth International Conference on Language Resources and Evaluation.

Carl, M., Schaeffer, M. J., and Bangalore, S. (2016). The CRITT Translation Process Research Database. In Carl, M., Bangalore, S., and Schaeffer, M., editors, *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*, New Frontiers in Translation Studies, pages 13–54. Springer International Publishing, Cham, Heidelberg, New York, Dordrecht, London:.

Dillinger, M. (2014). Introduction. In O'Brien, S., Winther Balling, L., Carl, M., Simard, M., and Specia, L., editors, *Post-editing of Machine Translation: Processes and Applications*, pages IX–XV. Cambridge Scholars Publishing, Newcastle upon Tyne.

Germann, U. (2008). Yawat: Yet Another Word Alignment Tool. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*, HLT-Demonstrations '08, pages 20–23, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hvelplund, K. T. (2016). *Cognitive efficiency in translation*, pages 149–170. John Benjamins.

Kuznetsova, A., Christensen, R. H. B., and Brockhoff, P. B. (2014). lmertest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package). R package version 2.0-6.

Läubli, S. and Germann, U. (2016). Statistical Modelling and Automatic Tagging of Human Translation Processes. In Carl, M., Bangalore, S., and Schaeffer, M. J., editors, *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB2*, pages 155–181.

Mesa-Lao, B. (2013). *Eye-tracking Post-editing Behaviour in an Interactive Translation Prediction Environment*, volume 6, page 541. Lund University.

R Development Core Team (2014). R: A language and environment for statistical computing.

Schaeffer, M. J., Carl, M., and Lacruz, I. (2016). Measuring Cognitive Translation Effort with Activity Units. *Baltic Journal of Modern Computing*, 4(2):331–345.