Order effects in the translation process

Christopher D. Mellinger & Thomas A. Hanson
University of North Carolina at Charlotte
Minnesota State University Moorhead

This study investigates how the order in which various translation memory match-types occur in a target language version of a text may influence translator behavior and cognition. Empirical research designs often attempt to mitigate for possible confounds from order effects, yet explicit recognition of the time-series nature of data collection can yield a better understanding of the influence that translation technologies have on the translation task. Data are drawn from a previous study that investigated technical, temporal, and cognitive effort in a translation memory environment, and here we analyze the time-series data for potential order effects. Findings show that order effects are present in some instances during the sequential progression through the target text, particularly with respect to technical and cognitive effort. Results are also suggestive of a potential first impression effect.

Keywords: order effects, time series, translation memory, editing behavior, perception of difficulty

1. Introduction

Translation technologies regularly figure into the daily work environment of professional translators. Survey data often indicate the prevalence of the adoption of translation memories to support professional translation work (e.g., Christensen & Schjoldager 2015) and, to an extent, language industry stakeholders drive the adoption of these tools in order to reuse previously-translated content (Dunne 2012, 2014). In addition, research has shown that the introduction of translation memories into the translation process alters the task paradigm (Dragsted 2008; Pym 2011; Mellinger & Shreve 2016). These alterations to the translation task and the ways in which translation technologies shape their work are often anecdotally described by translators, and scholarship on the topic is now revealing some of the constraints of which translators may themselves already be aware (e.g., Ruokonen & Koskinen 2017).

doi: 10.1075/tcb.00001.mel
Translation, Cognition & Behavior 1:1 (2018), pp. 1–20. issn 2542-5277 | e-issn 2542-5285

Translation process research, in particular, has investigated the ways in which translation memories change the translation task. For instance, cognitive effort—defined by Krings (2001, 179) in the context of post-editing research as "the type and extent of those cognitive processes that must be activated in order to remedy a given deficiency"—has been shown to differ depending on the types of segments with which the translator is presented (O'Brien 2007; Guerberof 2009; Mellinger 2014). Likewise, the necessity to segment the text and present the translation as a series of sentences or phrases has been shown to affect cognitive segmentation across levels of expertise and difficulty (Dragsted 2005; Dragsted & Hansen 2008). Other research suggests that the shift in the task environment when using translation memories results in a tendency to over-edit translations stored in the translation memory (Mellinger & Shreve 2016).

One aspect of the translation process that has yet to receive considerable attention is the order in which tasks are performed during the translation process. In many cases, the goal of translation process research is cross-sectional analysis and comparison; that is to say, comparisons are made between groups of individuals as a way to differentiate translation behavior under various conditions or as the result of specific participant variables. However, these data are necessarily collected over a period of time, and the order in which the tasks are presented might influence a translator's perception of the target text or the TM quality, which could lead to different behavior when editing or translating. For instance, translators may be asked to work with translation memories, requiring editing or revision of stored translations or translation of previously-untranslated source text content. Depending on where in the text these segments are presented, translators may behave differently as a result of more or less editing behavior. In a similar vein, translators may vary in their use of external resources during the translation task. While previous research has shown a variety of usage strategies with respect to using resources during the translation task (e.g., Enríquez 2014), it is unclear whether their use at specific stages in the process impacts future behavior as the translation task unfolds. These possible order effects suggest the value of additional inquiry into the temporal dimension related to the phenomena or constructs being observed.

An explicit perspective recognizing time series data may prove useful when researching translation technologies. As noted above in the case of translation memories, translators often alternate between editing stored translations and translating segments for which there is not a match. The order in which these segments are presented is driven by several factors, including the source text, the level at which the text has been segmented, the availability of previously-stored translations, and the algorithm used to determine how close of a match is available. As such, there is considerable variation with respect to the order in which these decision-making tasks occur when translators use translation memories. The present study aims to examine the impact that this type of variation may have during an experimental task and to then extrapolate on these findings to comment on general non-experimental behavior.

2. Time series in translation

Previous research has described the temporal progression during the translation process as it applies to reading strategies and gaze progression. These studies are often focused on specific aspects of the reading task and compare several groups; however, these scholars recognize that the task unfolds over a defined time period and that the exhibited behavior may change as the result of external or textual factors. For instance, Shreve et al. (1993: 29) note that translators potentially change reading strategies depending on their intention, be it to understand the text or to translate. In their discussion, the authors note a decrease in reading time per clause as the task progresses—i.e., translators took less time reading per clause when reading in anticipation of translating rather than during the translation process—possibly pointing to a common reading comprehension process across the various groups. Jakobsen & Jensen (2008) draw on eye tracking data and, as part of their results, comment on changes in visual attention between the source and target text occurring about every four seconds in one of their participant groups. In doing so, the authors explicitly recognize these data as having been derived as a time series process. However, they do not examine any subsequent changes in translation behavior.

Translation process research has also relied on time series pause data that occurs during the translation task. For instance, Alves & Vale (2009) describe how prototypical translation units can be observed in relation to the notion of cognitive rhythm and cognitive durability.[1] In their study, Alves & Vale recognize how data are generated as part of a time series and that time intervals may be revealing with respect to cognitive processing as the translation task unfolds. Other research involving eye-tracking data (e.g., Jensen 2011; Jakobsen & Jensen 2008; Hvelplund 2017) recognizes the temporal progression of the translation task, allowing researchers to investigate the distribution of attention or coordination of cognitive resources. Time series data also figure into transla-

1. Cognitive rhythm, described by Schilperoord (1996), is related to this temporal aspect of text production, insofar as it describes an attempt to identify pause patterns as the text production process unfolds. This idea has been applied in translation studies by several scholars (e.g., Jakobsen 2002; Alves 2005; Whyatt, Stachowiak, and Kajzer-Wietrzny 2016).

tion competence and pedagogical research. For example, Angelone (2012, 2013) draws on time series data as a potential way to improve translator education. Screen recording, as a means of visually recording and representing the translation process as it unfolds, allows for qualitative analysis of a variety of temporal aspects related to the translation task, such as problem-solving and documentation strategies.

Research on translation technologies has also addressed several temporal aspects of the translation task. In his discussion of how technology changes translation, Pym (2011) describes the impact that translation technologies have on the linearity of a text, particularly in the case of translation memories and machine translation. As a result of the textual segmentation imposed by these tools, "[t]he translating mind is thereby invited to work on one segment after the other, checking for terminological and phraseological consistency but not so easily checking, within this environment, for syntagmatic cohesion" (Pym 2011, 3). Whereas translation without the use of translation technologies may proceed in a more or less linear fashion, with the translator moving from start to finish within the written text, the translation process involving translation technologies is much less linear. Pym argues that the inclusion of technologies that allow translators to consult external resources or to have potential translations presented for evaluation is a disruption in the text's linearity and describes ways in which the writing and translation are altered by the introduction of electronic tools into the process.

Killman (2015) also recognizes the potential change in the translation process brought about by introducing translation memory matches and machine-translated segments into the translation task. He describes the challenges inherent to using these tools with respect to context, which may lead to translators accepting suboptimal translation proposals. These comments echo the disruptive nature described by Pym (2011) and also show the potential for a non-linear progression during translation.

Despite the recognition of the temporal progression of the translation task, quantitative research that focuses on the order in which tasks are presented is limited. While the previously-mentioned studies rely on quantitative measures (i.e., reading times or number of pauses), the interpretation of these results with respect to the temporal progression of the task is largely based on qualitative observation or interpretation. Researchers in other disciplines have examined the influence of presentation order on decisions and responses (e.g., Bansback et al. 2014; Bausenhart, Dyjas, & Ulrich 2015). Meanwhile, Michels & Helson (1954) warns that the appealingly simple solution of employing blocking in the experimental design by alternating presentation order does not eliminate order effects. The statistical analysis employed here does not combine identical tasks into separate categories, as would be typical in an ANOVA, for example. Instead, we align the observations based on pairs, essentially analyzing the data in an event time setting, a technique that is perhaps most associated with financial event studies (Brown & Warner 1985). Simply by reconceptualizing the data analysis to consider the prior segment, new insights can be gleaned regarding translator behavior.

Specifically, we draw on data from a previously-conducted study (Mellinger 2014) to answer the following research questions related to translators working with translation memory (TM) matches:

1. Do translator behavior and perception of difficulty change when working with a fuzzy match if the segment is directly preceded by an exact match or no-match that requires translation from scratch?
2. Do translator behavior and perception of difficulty change when working with an exact match based on the type of match that occurs directly before the match?
3. Do translator behavior and perception of difficulty change based on the translator's exhibited editing behavior in the previous segment, regardless of TM match type?

The measures used in the previous study will be summarized below; however, the main goal of the study here is to determine whether there is an effect based on the order in which these segments were presented.

3. Methods

Prior to discussing the dataset used in this study, we first note several methodological challenges in examining order effects. In many experimental designs, researchers aim to eliminate potential order effects or spillover effects to mitigate for potential confounds (Saldanha & O'Brien 2013, 117). Randomization and/or blocking of tasks and participants are often considered best practices since they allow researchers to isolate the phenomenon under investigation and reduce bias (Gile 2016). As such, it can be difficult to analyze existing datasets for order effects since, under ideal circumstances, any influence of order on the study itself has been mitigated. However, as described by Michels & Helson (1954), complete elimination of the order effect in actual practice is impossible. These lingering effects are explicitly investigated in the present study by analyzing the data as aligned pairs based on match type rather than in isolation.

Translation process research, which generally involves relatively few participants, can compound the analysis challenge due to high variability among participants. In this situation the within-group variability can be substantial enough to obscure cross-sectional differences due to low statistical power. The present analysis addresses this challenge in two ways. First, by analyzing pairs of match types that differ across versions of the translation task, the experiment yields multiple replications for each participant. In effect, this reorganization of the raw data increases the number of observations. Second, by employing a mixed effects linear model as one of the analytical tools, the within-subject subject variability can be accounted for using statistical methods, which increases the power of the hypothesis tests. The small sample and inherent variability might still obscure effects that could be found in larger-scale replications of this work, but significant findings in this setting should be encouraging for future research on the topic.

While many researchers randomly allocate participants into treatment and control groups and alternate the order in which stimuli are presented, texts by their very nature lend themselves to an order. The dataset on which we draw employed a Latin square design, which is a common blocking tool to eliminate significant order effects. However, the segmentation and match type triplets described below still provide the opportunity to analyze possible carryover effects. Unless conducting research with more traditional psycholinguistic research protocols of single sentence presentation, researchers present a text in its entirety for the sake of ecological validity. Consequently, texts are presented with multiple sentences or segments in order, allowing the translator to work with the entire paragraph or excerpt.

3.1 Participants

In the original study, nine Spanish-English professional translators with four to seven years of experience were enrolled. Participant recruitment and selection was done online using a demographic questionnaire. All of the participants were native English speakers and regularly translated from Spanish into English.[2] Participants were volunteers and were not compensated for their participation in the study.

3.2 Procedures

The 400-word source text was drawn from an Argentinian newspaper and was segmented at the sentence level into 21 segments.[3] Mellinger (2014) created three

2. FTE figures for participants are not available; however, all participants reported that they derive a minimum of 30% of their overall income each year from translation.

3. Participants worked with an excerpt from a longer, 626-word article.

different target text versions, such that each segment was presented in one of three conditions: (1) an exact match that did not require any revision; (2) a fuzzy match that required revision; and (3) no-match that required translation from scratch.[4] A Latin square design ensured that every segment was translated in each of the three conditions by a different group. In all of the target text versions, the first three segments contained only one of each match type. For example, the first segment of Version A of the target text was an exact match, followed by no match, and then followed by a fuzzy match. Versions B and C had these three match types in a different order, such that none of the versions ever had the same match type being examined in the same segment. The pattern was continued throughout the text so that each match type occurred once in every three segments. This blocking technique varied the type of match for each segment, so that the segments themselves should not have a meaningful effect on the resulting quantitative measures across groups.[5] However, the repeating triplets of match types, as well as the replication of each version by three participants, still allows an examination of any effect due to the presentation order.

The experimental task was completed online using TransCenter (Denkowski & Lavie 2012). Participants were randomly assigned into one of three groups to translate a general source text from Spanish into English. After each segment, participants were asked to use a 5-point Likert-type scale to rate their perception of difficulty to translate or edit each segment, with a score of 1 representing an incomprehensible segment requiring retranslation and a score of 5 representing a perfect translation that did not require any change.[6] The source text was presented in order with the entire text available to participants at all times. The visual presentation of the text using TransCenter mimics a typical translation memory work environment, in which the segmented source text was presented in a column on the left-hand side of the screen with space for the target text in the right-hand column. Participants were not required to use the proposed translation and were allowed to retranslate if they chose to do so. No time limit was imposed during the translation.

4. For details regarding how the instrument was created and to access the three versions of the stimulus, see Mellinger (2014).

5. In statistical parlance, this design does not require segment type to be included as a random effect or co-variate in regression or ANOVA, respectively.

6. The prompt asked participants about the overall difficulty. This phrasing may be akin to asking about effort since the distinction between effort and difficulty may not be readily apparent to participants. However, difficulty is used throughout this manuscript to be consistent with the task instructions.

3.3 Measures

The software used to conduct the study, TransCenter, is a keystroke logging software that allows researchers to record all the keystrokes and mouseclicks made by the participants. The software also associates every action of the participant with a timestamp, which allows researchers to examine temporal information related to translator behavior. In this study, time measures related to the total time on task were collected as well as the total amount of time spent in each segment. The latter measure accounts for the time from the moment that the translator leaves the previous segment to the time he or she enters the next segment. This measure accounts for any initial time spent reading the segment, the time to produce or revise the segment (including pauses and keystrokes), as well as any time reviewing the translation prior to moving onto the next segment.

To borrow Krings' (2001) typology of effort, three types of effort can be observed: technical, temporal, and cognitive effort. Technical effort refers to the physical manipulation or action of an individual while temporal effort is related to the time spent on task. Cognitive effort, in contrast, is not directly observable, nor is it an amalgamation of technical and temporal effort; rather, it is a construct that must be indirectly observed or measured. In the case of keystroke logging, researchers have access to the technical and temporal effort exerted by translators. This data collection method has been common in translation process research (e.g., Jakobsen 1999, 2011), and the challenges and benefits inherent to using web-based software to conduct process-oriented research have been described by Mellinger (2015).

In addition to keystrokes, mouseclicks, and temporal data related to total time on task and time per segment, pause data were collected. In the initial study, pauses were operationalized as being a minimum of one-second in length, and this definition is in line with other studies that have used keystroke logging (e.g., Jakobsen 1998; O'Brien 2006, 2007; Lacruz, Shreve, & Angelone 2012).[7] This threshold also allowed for a measure of cognitive effort to be calculated, namely the average pause ratio (APR), proposed by Lacruz, Shreve, & Angelone (2012). Rather than using total pause time as a proxy for cognitive effort, this measure takes into account the variable lengths of pauses and potential clusters of pauses

7. This minimum threshold for pause duration is in line with the initial study; however, we recognize that other studies account for variation across participants by using a flexible pause threshold (e.g., Dragsted 2004). The present study also accounts for variation, but does so not in a flexible lower boundary threshold, but rather by means of the average pause ratio. This measure incorporates the average pause duration in each segment to be calculated, which will vary naturally as the result of participant behavior. For an extended discussion on pause duration in translation process research, see Kumpulainen (2015).

that may be more indicative of effort. The APR is the ratio of the average time per pause to the average time per word in the segment. Finally, a numeric rating was collected for each segment as to the perceived difficulty.

3.4 Analysis

Statistical analysis was conducted using the R statistical program environment (R Core Team 2016); the estimation of mixed effects models employed the contributed package *lme4* (Bates et al. 2015). For correlational analysis, we follow the recommendation in Mellinger & Hanson (2017) to calculate Kendall's τ given its demonstrated superiority to other nonparametric tests. Likewise, we employ Hedges' *g* for effect sizes. In some cases, effect sizes were calculated using the Excel spreadsheets and calculators provided online by Mellinger & Hanson (2017).

4. Results

The first two research questions address whether editing behavior and perception of difficulty differ due to the match type of the previous segment. Specifically, the first question asked whether the dependent variables would differ for fuzzy match segments, based on whether the previous segment was a new translation or an exact match. Results of this analysis appear in Table 1 in the form of descriptive statistics and associated univariate tests.

The statistics in Table 1 demonstrate that following exact matches participants used on average fewer keystrokes to edit fuzzy match segments. Following a new translation segment, translators used more than two-and-a-half times more keystrokes in editing than following an exact match segment, and Hedges' *g* for this difference was large, which implies the difference is large enough to be meaningful in addition to its statistical significance (Cohen 1988). Cognitive effort, as measured by the APR, also declined following exact matches. The differences between number of clicks, total pause time, total segment time, and difficulty rating were not statistically significant, though the effect sizes suggest a meaningful difference for the two time variables, despite their lack of statistical significance. Furthermore, in all cases the averages differed in a direction to imply less effort for the fuzzy matches following exact match segments. In particular, fuzzy match segments that followed exact matches received a higher quality rating on average.

The second research question explored whether behavior and perception of difficulty would differ for exact matches, based on whether the previous segment was a new translation or a fuzzy match segment. Table 2 presents the results of this analysis. None of the measured variables differ at a statistically significant

Table 1. Statistics for fuzzy match segments

	After New	After Exact	t	df	p	g
Keystrokes	207.19	82.42	2.83	48.03	.007	0.775
	(139.53)	(174.3)				
Clicks	3.70	3.24	−0.36	47.62	.719	0.100
	(4.1)	(5.0)				
APR	1.26	13.21	−2.49	29.04	.019	0.590
	(0.58)	(26.3)				
Total Pause Time	160.89	96.87	1.27	26.71	.214	0.405
	(213.18)	(104.0)				
Total Segment Time	157.58	87.51	1.46	28.56	.154	0.459
	(199.1)	(110.01)				
Rating	4.00	4.23	−0.77	31.07	.446	0.237
	(1.22)	(0.77)				

Descriptive statistics for each of the six variables are mean with standard deviation in parentheses. Univariate tests are independent *t*-tests with exact *p*-values and Hedges' *g* provided as an effect size.

level, though in every case with the exception of number of clicks, the averages point toward more effort following fuzzy matches than new translations. While these non-significant results should not be over-interpreted, they do suggest that exact matches are more likely to be scrutinized and edited following a fuzzy match segment. These results suggest value in further research dedicated to this phenomenon.

The third research question—i.e., do translator behavior and perception of difficulty changes based on the translator's exhibited editing behavior in the previous segment regardless of TM match type—disregards the previous segment's match type in favor of the translator's behavior. Ignoring all segments that required new translations, we examined the subset consisting of all fuzzy-followed-by-exact and exact-followed-by-fuzzy pairs and divided them into two samples, depending on whether any editing keystrokes were recorded in the first segment. A mixed linear model was then employed to examine whether the dependent variables differed in the second segment. Participant and segment number were both included as random effects, and the prior match type was included as a fixed effect. A chi-squared test compared this reduced model to a full model that included previous editing behavior.

The difference between the two models was statistically significant at the 10% level for only APR ($p = .052$) and total segment time ($p = .073$). In the former case, if the prior segment was edited, the following segment had a smaller APR, indicat-

Table 2. Statistics for exact matches

	After Fuzzy	After New	t	df	p	g
Keys	263.73	207.19	1.31	47.42	.196	0.361
	(167.45)	(139.53)				
Clicks	2.9	3.24	−0.31	36.62	.755	0.093
	(3.25)	(4.10)				
APR	1.14	1.26	−0.73	41.72	.468	0.209
	(0.55)	(0.58)				
Total Pause	185.57	160.89	0.44	36.95	.662	0.130
	(171.24)	(213.18)				
Total Segment	170.66	157.58	0.26	34.52	.799	0.077
	(145.99)	(199.11)				
Rating	3.93	4.00	−0.19	41.98	.847	0.056
	(1.17)	(0.12)				

Descriptive statistics for each of the six variables are means with standard deviations in parentheses. Univariate tests are independent t-tests with p-values and Hedges' g provided as an effect size.

ing more cognitive effort. In the latter case, total segment time increased following an edited segment. The model was not statistically significant for keystrokes ($p=.596$), clicks ($p=.514$), total pause time ($p=.293$), nor rating ($p=.190$).

In none of the results were the participants' ratings affected by prior segment type or editing behavior regarding the prior segment at a statistically significant level. These results raise the question of what factors do influence quality ratings. Full analysis of this topic would require a dedicated future study, but preliminary correlation analysis presented in Table 3 reveals a strong association between rating and measures of technical and cognitive effort but a weak association between rating and pause time or total segment time. That is to say, increased keystrokes, clicks, or cognitive effort are all associated with lower quality ratings assigned to the suggested matches. The association also holds true in a regression analysis that includes all of the measurements as independent variables and a random effect due to participant ($F[12,176]=7.488$, $p<.001$, $Adj.\ R^2=.293$). Full results are omitted because they are tangential to the primary research questions of this project, but keystrokes ($p=.050$), clicks ($p=.091$), and APR ($p=.043$) are all statistically significant at the 10% level.

Another topic that is a strong candidate for further research is the question of a first impression effect. That is to say, do translators differ in their editing behavior or perception of translation quality based on the quality of the first segment (or the first few segments) presented? The experiment in this study was not

Table 3. Correlations with quality rating

	Kendall's τ	p
Keystrokes	−0.221	.007
Clicks	−0.236	.005
APR	0.228	.004
Total pause time	−0.131	.436
Total segment time	−0.174	.256

designed to test this question directly. However, we can see some preliminary evidence that a first impression effect might exist. For the three participants whose first segment was a fuzzy match, the overall rating for the 21 segments was lowest ($F[2, 115.4] = 5.07$, $p = .008$, $\omega^2 = .041$) and the number of keystrokes was highest by a wide margin ($F[2, 117.7] = 4.01$, $p = .021$, $\omega^2 = .031$). Descriptive statistics for these two variables for the three groups appear in Table 4. The APR was the lowest for the group that began with a fuzzy match segment, though not at a statistically significant level ($p = .931$), while number of clicks ($p = .181$), total pause time ($p = .792$), and total segment time ($p = .844$) also did not differ to a statistically significant extent.

5. Discussion

The results for the first two research questions suggest that measurable order effects do result from the presentation of the various translation memory match types. The results of the first research question (see Table 1) demonstrate that translators exerted more technical effort to edit segments after having translated a segment from scratch in comparison to segments following an exact match. This distinction can be attributed to two possible influences: overediting behavior following a new translation or under-editing behavior following an exact match. In the first case of possible overediting, the translators' behavior exhibited in fuzzy match segments supports previous evidence related to overediting and cognitive rhythm. Likewise, there was an increase in the amount of cognitive effort exerted by participants in fuzzy matches after translating a segment from scratch. This behavior is indicative of what Mellinger & Shreve (2016, 144) describe as "a mismatch between the previous translation stored in the TM and the participant's internal conception of what this match should be." The analysis presented in this study corroborates these findings and also allows for the idea of cognitive dissonance to be extended potentially to the emerging target text as a whole. As the task

Table 4. First impression effect

	Rating	Keys
First segment fuzzy	3.78	201.65
	(0.97)	(270.91)
First segment exact	4.14	99.3
	(1.11)	(130.33)
First segment new	4.24	97.63
	(0.61)	(148.58)

progresses, translators appear to be influenced by previous segments and, therefore, are inclined to introduce changes based on the type of segment preceding the one in which they are currently working. This influence manifests as over-editing behavior and the results from research questions one and two are suggestive of a relationship between the preceding match type and technical effort exerted in fuzzy matches.

The second case of potential under-editing behavior is also related to the observed difference in technical and cognitive effort in fuzzy match segments; this phenomenon may be related to cognitive rhythm. The results presented in Table 1 show that translators tend to introduce fewer keystrokes and exert less cognitive effort when working with fuzzy matches if the previous segment is an exact match. Taken in conjunction with the non-statistically-significant difference in the amount of temporal effort, these results may indicate more fluid target text production and editing. The rationale for this under-editing behavior cannot be determined from the present analysis since the original data and stimuli were not designed to investigate this type of causal relationship. Nevertheless, the behavior may be the result of reliance or trust on the presented translation memory matches or could be a tendency to exert a similar amount of technical effort from previous segments. Future research explicit to this behavior would need to be conducted to provide a better understanding of this phenomenon. Moreover, future studies could examine if and to what extent professional experience or translator education and training influence translator behavior throughout the translation task.

Analysis related to the second research question examined the analogous order effects for exact match segments, depending on whether the prior segment was a new translation or a fuzzy match segment. While we would caution against the overinterpretation of non-significant results, future studies may confirm the slight tendency to overedit fuzzy matches that follow exact matches (see Table 2). One possible explanation of this effect is an order effect in which the translator remembers the amount of effort expended in one segment (as he or she conceptualizes effort), forms an opinion about the quality of the proposed TM match, and alters behavior accordingly in subsequent segments. As in the previous research question, this directional result would corroborate previous findings of Mellinger & Shreve (2016); however, the present research design cannot allow for a conclusive determination as to whether the observed behavior is the result of an order effect because the study was not designed explicitly for such analysis. The preliminary quantitative analysis suggests the value of such future work.

5.1 Perception of difficulty

An order effect was not found with respect to the perceived difficulty rating provided by translators. The correlation analysis and fixed-effect models indicate a strong relationship observed between ratings and technical effort indicators (i.e., keystrokes and mouseclicks) as well as cognitive effort measures (i.e., APR). Therefore, we might conclude that participants present strong biases regarding what constitutes effort by favoring technical indicators and cognitive effort over time spent on the translation task. Additional research related to the cognitive ergonomics of translation (cf. Ehrensberger-Dow 2017) would help elucidate whether this bias is the result of the stimuli, the data collection methodology, or the translation technology itself. Findings related to these biases would be of particular importance to tool designers of translation technologies; however, we limit our conclusions here based on the analysis to the establishment of a relationship between technical and cognitive effort and the provided ratings.

5.2 First impression effect

The results in Table 4 present preliminary findings of a possible first impression effect. To the best of our knowledge, no previous studies have examined the possible influence that the first sentence of a translation task might have on subsequent translation and editing behavior. However, the power of first impressions has been widely studied in areas ranging from web design (Lindgaard et al. 2006) to personality impression formation (Anderson & Hubert 1963) and, perhaps most closely aligned with translation, the rating of written language performance (e.g., Weigle 1994; Eckes 2008). The statistical analyses presented here are suggestive of such an effect. When the first segment was a fuzzy match, the participants used the most keystrokes and provided the lowest average rating across all segments of the source text. When the first segment required a new translation, the opposite effect occurred, with the highest average rating and fewest keystrokes logged. Due to the small sample size, these results are far from definitive, but statistically significant results from a limited sample are an argument for replicative work exploring this phenomenon directly. By recognizing the possibility of first impressions having an impact on subsequently collected data, researchers may be able to examine self-reported evidence related to poor TM quality affecting the way in which they approach the translation task (cf. LeBlanc 2013; Ruokonen & Koskinen 2017).

5.3 Future research

In presenting our results, we have suggested the value of future work to examine order effects directly, and we argue that the general topic of order effects is an understudied and likely fruitful area of research. More research is needed to explicitly investigate order effects in a wide variety of areas; translation technologies is one area prime for this type of explicit work, but research involving a variety of product- and process-oriented variables could prove useful. We propose at least three topics that merit examination. The first two areas are the over- and under-editing behavior that manifest even in a Latin square design that was intended to minimize such effects. The third area is the possibility of a first impression effect, in which an initial stimulus—be it a written text or oral rendition—might influence the behavior of a translator or interpreter throughout the remainder of the task. Given the well-known influence of first impressions in a range of settings, it is plausible that a measurable effect could exist.

Our results suggest a weak relationship between segment order and perception of difficulty. Therefore, further exploration of order effects related to this variable may not be fruitful. The correlational analysis, however, argues that quality ratings are related more to technical and cognitive effort than to time spent completing the translation task. Future studies could explore the implications of this relationship in translator pedagogy, evaluation, and assessment, the relationship between translators and editors, and even compensation schemes.

Finally, we note that researchers should consider the possibility of order effects when designing future studies. Appropriate experimental design through blocking and randomization are necessary to mitigate order effects, yet the resulting datasets should still be examined for possible carryover effects at a textual/stimulus level. Not only should order effects be studied directly, but their influence on any study involving a segmented text should also be considered as a possible confounding variable.

6. Conclusion

Order effects have previously been considered an undesirable byproduct of experimental design in translation process research, and researchers generally strive to eliminate their influence using various techniques. However, even in studies employing appropriate blocking and randomization, the findings presented here demonstrate that there may yet be statistically significant and meaningful order effects that ought to be considered. In the case of the current study, the order resulting from the use of a translation memory and the text itself influenced translator behavior. Evidence exists that the match type of a segment can lead to translators over- and under-editing subsequent segments. This result is evident not only in segments that directly follow a specific translation memory match type, but also may be present as a first impression effect. While the latter is not conclusive, more research is needed to investigate the extent to which the initial segments in a translation task influence the remainder of the translation. In both cases, experimental and quasi-experimental research could be supplemented by retrospective verbalizations or questionnaires that probe the rationale for specific translation behavior. These research methods may allow a more comprehensive view to determine whether translators adopt a specific approach to translation tool use after having made an initial determination on the quality or ease of use of specific translation aids. The purpose of the current study is to draw attention to some of the potential issues arising from the inherent time series nature of the translation task.

References

Alves, Fabio. 2005. "Ritmo cognitivo, meta-função e experiencia: parámetros de análise processual no desempenho de tradutores novatos e experientes." In *Competência em tradução: cognição e discurso*. Edited by F. Alves, C. Magalhães, and A. Pagano, 109–169. Belo Horizonte: Editorial UFMG.

Alves, Fabio, and Daniel Vale. 2009. "Probing the Unit of Translation in Time: Aspects of the Design and Development of a Web Application for Storing, Annotating, and Querying Translation Process Data." *Across Languages and Cultures* 10 (2): 251–273. doi: 10.1556/Acr.10.2009.2.5.

Anderson, Norman H., and Stephen Hubert. 1963. "Effects of Concomitant Verbal Recall on Order Effects in Personality Impression Formation." *Journal of Verbal Learning and Verbal Behavior* 2 (5–6): 379–391. doi: 10.1016/S0022-5371(63)80039-0

Angelone, Erik. 2012. "The Place of Screen Recording in Process-oriented Translator Training." *Rivista Internazionale di Tecnica della Traduzione (RITT)* 14: 41–55.

Angelone, Erik. 2013. "Watching and Learning from 'Virtual Professionals': Utilizing Screen Recording in Process-oriented Translator Training." In *New Prospects and Perspectives for Educating Language Mediators*. Edited by D. Kiraly, S. Hansen-Schirra, and K. Maksymski, 139–155. Tübingen: Günther Narr.

Bansback, Nick, Linda C. Li, Larry Lynd, and Stirling Bryan. 2014. "Exploiting Order Effects to Improve the Quality of Decisions." *Patient Education and Counseling* 96 (2): 197–203. doi: 10.1016/j.pec.2014.05.021.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. doi: 10.18637/jss.v067.i01.

Bausenhart, Karin M., Oliver Dyjas, and Rolf Ulrich. 2015. "Effects of Stimulus Order on Discrimination Sensitivity for Short and Long Durations." *Attention, Perception, & Psychophysics* 77 (4): 1033–1043. doi: 10.3758/s13414-015-0875-8.

Brown, Stephen J. and Jerold B. Warner. 1985. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics* 14 (1): 3–31. doi: 10.1016/0304-405X(85)90042-X

Christensen, Tina Paulsen, and Anne Schjoldager. 2015. "Computer-aided Translation Tools: The Uptake and Use by Danish Translation Service Providers." *The Journal of Specialised Translation* 25: 89–105.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Denkowski, Michael, and Alan Lavie. 2012. "TransCenter: Web-based Translation Research Suite." *AMTA 2012 Workshop on Post-Editing Technology and Practice Demo Session*.

Dragsted, Barbara. 2004. *Segmentation in Translation and Translation Memory Systems: An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM-System into the Translation Process*. Copenhagen: Samfundslitteratur.

Dragsted, Barbara. 2005. "Segmentation in Translation: Differences Across Levels of Expertise and Difficulty." *Target* 17 (1): 49–70. doi: 10.1075/target.17.1.04dra

Dragsted, Barbara. 2008. "Computer-aided Translation as a Distributed Cognitive Task." In *Cognition Distributed: How Cognitive Technology Extends our Minds*. Edited by I. E. Dror and S. Harnad, 237–256. Amsterdam: John Benjamins. doi: 10.1075/bct.16.16dra

Dragsted, Barbara, and Inge Gorm Hansen. 2008. "Comprehension and Production in Translation: A Pilot Study on Segmentation and the Coordination of Reading and Writing Processes." In *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Edited by S. Göpferich, A. L. Jakobsen, and I. Mees, 9–30. Copenhagen: Samfundslitteratur.

Dunne, Keiran J. 2012. "The Industrialization of Translation: Causes, Consequences, and Challenges." *Translation Spaces* 1: 143–168. doi: 10.1075/ts.1.07dun

Dunne, Keiran J. 2014. "Localization and the (R)Evolution of Translation." In *A Companion to Translation Studies*. Edited by S. Bermann and C. Porter, 147–162. Oxford, UK: Wiley Blackwell. doi: 10.1002/9781118613504.ch11

Eckes, Thomas. 2008. "Rater Types in Writing Performance Assessments: A Classification Approach to Rater Variability." *Language Testing* 25 (2): 155–185. doi: 10.1177/0265532207086780

Ehrensberger-Dow, Maureen. 2017. "An Ergonomic Perspective of Translation." In *The Handbook of Translation and Cognition*. Edited by J. W. Schwieter and A. Ferreira, 332–349. Malden, MA: Wiley Blackwell.

Enríquez Raído, Vanessa. 2014. *Translation and Web Searching*. New York: Routledge.
Gile, Daniel. 2016. "Experimental Research." In *Researching Translation and Interpreting*. Edited by C. V. Angelelli and B. J. Baer, 220–228. New York: Routledge.
Guerberof Arenas, Ana. 2009. "Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation." *Localisation Focus* 7 (1): 11–21.
Hvelplund, Kristian. 2017. "Eye Tracking in Translation Process Research." In *The Handbook of Translation and Cognition*. Edited by J. W. Schwieter and A. Ferreira, 248–264. Malden, MA: Wiley Blackwell.
Jakobsen, Arnt Lykke. 1999. "Logging Target Text Production with Translong." In *Probing the Process in Translation: Methods and Results*. Edited by G. e Hansen, 9–20. Copenhagen: Samfundslitteratur.
Jakobsen, Arnt Lykke. 2002. "Translation Drafting by Professional Translators and by Translation Students." In *Empirical Translation Studies: Process and Product*. Edited by Gyde Hansen, 191–204. Copenhagen: Samfundslitteratur.
Jakobsen, Arnt Lykke. 2011. "Tracking Translators' Keystrokes and Eye Movements with Translog." In *Methods and Strategies of Process Research: Integrative Approaches to Translation Studies*. Edited by C. Alvstad, A. Hild, and E. Tiselius, 37–55. Amsterdam: John Benjamins. doi: 10.1075/btl.94.06jak
Jakobsen, Arnt Lykke, and Kristian T. H. Jensen. 2008. "Eye Movement Behavior Across Four Different Types of Reading Task." *Copenhagen Studies in Language* 36: 103–124.
Jensen, Kristian T. H. 2011. "Distribution of Attention Between Source Text and Target Text during Translation." In *Cognitive Explorations of Translation*. Edited by S. O'Brien, 215–237. London: Continuum.
Killman, Jeffrey. 2015. "Context as Achilles' Heel of Translation Technologies: Major Implications for End-users." *Translation and Interpreting Studies* 10 (2): 203–222. doi: 10.1075/tis.10.2.03kil
Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Edited by G. S. Koby. Kent, OH: Kent State University Press.
Kumpulainen, Minna. 2015. "On the Operationalization of 'Pauses' in Translation Process Research." *Translation & Interpreting* 7 (1): 47–58.
Lacruz, Isabel, Gregory M. Shreve, and Erik Angelone. 2012. "Average Pause Ratio as an Indicator of Cognitive Effort in Post-editing: A Case Study." *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice*. San Diego, CA.
LeBlanc, Matthieu. 2013. "Translators on Translation Memory (TM): Results of an Ethnographic Study in Three Translation Services and Agencies." *Translation & Interpreting* 5 (2): 1–13.
Lindgaard, Gitte, Gary Fernandes, Cathy Dudek, and Judith M. Brown. 2006. "Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression!" *Behaviour & Information Technology* 25 (2): 115–126. doi: 10.1080/01449290500330448
Mellinger, Christopher D. 2014. *Computer-Assisted Translation: An Empirical Investigation of Cognitive Effort*. Ph.D. dissertation, Kent State University, OH. Available at http://bit.ly/1ybBY7W
Mellinger, Christopher D. 2015. "On the Applicability of Internet-Mediated Research Methods to Investigate Translators' Cognitive Behavior." *Translation & Interpreting* 7 (1): 59–71.
Mellinger, Christopher D., and Gregory M. Shreve. 2016. "Match Evaluation and Over-editing in a Translation Memory Environment." In *Reembedding Translation Process Research*. Edited by R. Muñoz, 131–148. Amsterdam: John Benjamins.

Mellinger, Christopher D., and Thomas A. Hanson. 2017. *Quantitative Research Methods in Translation and Interpreting Studies*. New York: Routledge.
Michels, Walter C., and Harry Helson. 1954. "A Quantitative Theory of Time-order Effects." *The American Journal of Psychology* 67 (2): 327–334. doi: 10.2307/1418635
O'Brien, Sharon. 2006. "Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output." *Across Language and Cultures* 7 (1): 1–21. doi: 10.1556/Acr.7.2006.1.1
O'Brien, Sharon. 2007. "Eye-tracking and Translation Memory Matches." *Perspectives: Studies in Translatology* 14 (3): 185–205.
Pym, Anthony. 2011. "What Technology Does to Translating." *Translation & Interpreting* 3 (1): 1–9.
R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.
Ruokonen, Minna, and Kaisa Koskinen. 2017. "Dancing with Technology: Translators' Narratives on the Dance of Human and Machinic Agency in Translation Work." *The Translator* 1–14. doi: 10.1080/13556509.2017.1301846
Saldanha, Gabriela, and Sharon O'Brien. 2013. *Research Methodologies in Translation Studies*. New York: Routledge.
Schilperoord, Joost. 1996. *It's about Time: Temporal Aspects of Cognitive Processes in Text Production*. Utrecht: USI &C.
Shreve, Gregory M., Christina Schäffner, Joseph H. Danks, and Jennifer Griffin. 1993. "Is There a Special Kind of 'Reading' for Translation? An Empirical Investigation of Reading in the Translation Process." *Target* 5 (1): 21–41. doi: 10.1075/target.5.1.03shr
Weigle, Sara Cushing. 1994. "Effects of Training on Raters of ESL Compositions." *Language Testing* 11 (2): 171–195. doi: 10.1177/026553229401100206
Whyatt, Bogusława, Katarzyna Stachowiak, and Marta Kajzer-Wietrzny. 2016. "Similar and Different: Cognitive Rhythm and Effort in Translation and Paraphasing." *Poznań Studies in Contemporary Linguistics* 52 (2): 175–208. doi: 10.1515/psicl-2016-0007

Authors' addresses

Christopher D. Mellinger
Department of Languages and Culture Studies
College of Liberal Arts and Sciences, UNCC
9201 University City Blvd
Charlotte, NC 28223-0001
USA

cmelling@uncc.edu

Thomas A. Hanson
Paseka School of Business, MSUM
1104 7th Ave South
Moorhead, MN 56563
USA

thomas.hanson@mnstate.edu

Biographical notes

Christopher D. Mellinger is assistant professor of Spanish interpreting and translation studies at the University of North Carolina at Charlotte. He is also the managing editor of *Translation and Interpreting Studies* and co-author of *Quantitative Research Methods in Translation and Interpreting Studies* (Routledge 2017).

Thomas A. Hanson is assistant professor of finance at Minnesota State University Moorhead. He is co-author of *Quantitative Research Methods in Translation and Interpreting Studies* (Routledge 2017).