The development of the TPR-DB
as Grounded Theory Method

Michael Carl & Moritz Schaeffer

Renmin University of China | University of Mainz, Germany

Initial versions of the translation process research database (TPR-DB), were released around 2011 in an attempt to integrate translation process data from several until then individually collected and scattered translation research projects. While the earlier individual studies had a clear focus on *quantitative* assessment of well-defined research questions on cognitive processes in human translation production, the integration of the data into the TPR-DB allowed for broader *qualitative* and exploratory research which has led to new codes, categories and research themes. In a constant effort to develop and refine the emerging concepts and categories and to validate the developing theories, the TPR-DB has been extended with further translation studies in different languages and translation modes. In this respect, it shares many features with Grounded Theory Method. This method was *discovered* in 1967 and used in qualitative research in social science ad many other research areas. We analyze the TPR-DB development as a Grounded Theory Method.[1]

1.   Introduction

The ambition of the TPR-DB has been to make available a corpus of user activity data that allows for open-ended and—as much as possible—unbiased investigation into the human translation process. The objective is to compare differences in the translation product and in the translation process during the production of *alternative translations*, i.e., the translation of the same ST by different translators in different languages and translation modes. Initially, TPR-DB consisted

---

[1]   In the title of their book, Glaser & Strauss (1967) describe GTM as a *Discovery of Grounded Theory*.

of 72 translation sessions (English to Danish) from 12 translation students and 12 translation professionals (Hvelplund 2011), i.e., 24 alternative translations per text. The TPR-DB has grown over the past 5 years to more than than 500 hours of recorded translation time and more than 2000 translation sessions using different source texts (ST), target languages and translation modes, investigating many different research questions, in different research projects. Translation activities were recorded including mouse activities and keystrokes as well as gaze data using various eyetrackers with gaze sampling rates between 60 to 300 Hz.[2] The TPR-DB consists of a data-lake, which contains the raw logging data, and a set of methods (i.e., program scripts) to extract features and summary tables from the raw data. Summary tables and features constitute categories and codes respectively which are analysed to develop models and theories to explain the observed textual and behavioral patterns. An essential part in the TPR-DB are visualizations tools, which allow for a *qualitative* analysis of the primary (raw) and the derived sensor data through translation progression graphs and replay functions.

This paper traces the development of the TPR-DB as Grounded Theory Method (GTM). GTM is a bottom-up approach to unbiased theory induction. One of the fundaments of GTM is to analyze data already during the collection phase, to constantly compare and assess the emerging categories and consolidate them through targeted (i.e., theoretically motivated) codes and categories as well as through additional collections of targeted data sets (i.e., theoretical sampling). Theoretical saturation is reached if no new insights can be gained through further theoretical sampling.

Within the TPR-DB, qualities of the primary sensor data (e.g., the proximity of keystrokes and eye samples in space and time is a quality that can be measured) allow to derive codes and categories that are better suited to describe the qualities of the underlying translation process. That is, the raw logging data consists of single gaze sample points, while a first fixation duration encodes a cluster of gaze samples (i.e., a fixation duration) and their mapping onto a word. Keystrokes are single observations in time, the combination of which produce words and texts. For Glaser (2008) "it is the quantification of subjective responses which is the paradox."[3] However, in the TPR-DB, on the basis of the qualities of the sensor signal we infer codes and categories that are in turn used to develop and quantify theoretical hypotheses. Data visualization (e.g., in the form of translation progres-

---

2. Keystrokes were recorded with Translog-II, and gaze data with various eyetrackers, Tobii, SMI and EyeLink.

3. For Glaser (2008, s.p.) "Qualitative data is credited with providing the meaning and factual interpretation that quantitative data does not …" In the TPR-DB we count and measure the distribution and properties of categories to verify (qualitative) hypothesis.

sion graphs) plays a crucial role in this process by which the qualitative nature of the data can be conceptualized and interpreted and, as we will show below, turned into a narrative.

We follow the maxim *All is data* which is a fundamental property of GTM. For the development of the TPR-DB, this means that descriptive statistics and the outcome of regression analyses and machine learning techniques are also data which may serve as the input to *qualitative* analyses, or to validate abductive hypotheses and theories. GTM has been deployed mostly for interview data. However, Kelle (2007, 204) suggests that GTM is open for "constructing one's own coding paradigm connected to a theoretical tradition one prefers", as long as the emerging theory focuses on how the individual interacts with the phenomenon under study.[4] According to Glaser (2008, s.p.), GTM "inducts abstractions or concepts from whatever data is being used." It is "a general inductive methodology that can be used, with excitement, with quantitative data."[5] However, "without theoretical sampling, constant comparison of data to theoretical categories and theoretical saturation of categories, one should not claim to be using Grounded Theory" (Hood 2007, 164).

In order to show how the development of the TPR-DB can be interpreted as an instance of GTM, in Section 2 we outline some basic GTM concepts. Then we apply these concepts to the TPR-DB. We trace how codes and categories have developed in the TPR-DB and how themes and substantial theories emerge from the data. In Section 3, we trace the development of the "monitor model" in the TPR-DB which has led to a recursive model of the translation process and which, we think, fulfills the criterion of theoretical saturation, where further data records are unlikely to lead to fundamentally new insights. We then describe in Section 4 an example of axial coding which elaborates the relation between the length of ST words, the translation duration and machine translation (MT) quality. We conclude the paper with two sections on "Activity Units" which, we believe, have the potential to unveil many more insightful relations in of the translation process.

---

4. See also Cathy Urquhart: https://www.youtube.com/watch?v=pIFA75IhRGc

5. We consider bottom-up coding and classification a grounded method which generates new qualities, whereas the elaboration of categories and subcategories and the analysis of their distribution has often been instrumental for top-down qualitative and quantitative abductive hypothesis testing and verification.

2.   The grounded theory method

2.1   Basics

Originally developed by Glaser & Strauss in 1967 for sociological research, GT has undergone a number of additions and modifications over the years (Strauss and Corbin 1990, Charmaz 2006). GTM is an approach for theory induction from data, which according to Heist (2012) can be used if the aim is to develop a substantive theory and an explanation for a situation. Starting from an initial set of data, GTM relies on a number of iterative coding and analysis steps that can be summarized as follows:

– Simultaneous collection and analysis of data
– Creation of codes and categories from inspection of the data
– Discovery of the basic processes that created the data
– Inductive construction of abstractions and categories
– Theoretical coding and sampling to refine categories
– The integration of categories and codes into a theoretical framework

The process of coding is separated into various iterating steps: *open coding* is a first access to the data with an open mind and a "vague understanding of the sorts of categories that might be relevant" (Gasson 2004, 82). *Axial coding* (Strauss & Corbin, 1998) is the process of relating codes with the aim to obtain higher order categories (i.e., a grouping imposed on the coded data) and a deeper understanding of the underlying processes. Once initial codes (or *substantive codes,* according to Glaser, 1978) are assigned to the data, axial coding investigates relationships between the coded elements of the data. Gasson (2004, 83) cautions that the insights that can be obtained through axial coding impact the research problem through selecting some categories and not others.

Alternatively, *theoretical codes* may be used to conceptualize how substantive codes relate to each other and may form hypotheses to be integrated into a theory. Theoretical coding categorizes empirical data on the basis of previous theoretical knowledge. However, theoretical codes should only be used if the data itself suggests their use. Glaser (1978, 74) suggests analyzing the data according to the "[…] six C's: causes, contexts, contingencies, consequences, covariances and conditions." Further, *selective coding* strategies may be used to integrate and refine categories until they ultimately become the basis for the grounded theory. This is an iterative process which requires continued "revision or new enactment of past research results" (Walsham 1993, 245).

At some stage, new data collection might be required to examine categories and their relationships in different situations or contexts and to assure the validity

of the categories identified so far. This *theoretical sampling* provides a structure to data collection as well as to the data analysis. Theoretical sampling is "sampling on the basis of concepts that have proven theoretical relevance to the evolving theory" (Strauss & Corbin 1990, 176). In the process of theoretical sampling and re-analysis, Gasson (2004, 84) notes that

> The researcher must continually ask whether the analysis of new data provides similar themes and categories to previous data, or whether other patterns emerge. Constant comparison requires continual research into the meaning of the developing categories by further data collection and analysis.

Finally, *theoretical saturation* is reached when no new categories or relationships emerge and new data confirm findings from previous data. A grounded theory is then "an explanatory scheme comprising a set of concepts related to each other through logical patterns of connectivity" (Birks & Mills 2011, 112–113).

2.2   GTM and the TPR-DB

After an initial data collection in 2012, additional (theoretical) data sampling extended the TPR-DB with further language pairs and translation modes. The idea was that, by keeping STs as well as a maximum of other factors, such as translation interface, translation mode, translation brief, etc. invariant across different translators, the variations observed in the translation product (i.e., the produced text) and in the translation process (i.e., the translation behavior: keystrokes, gaze activity, structure of pauses, etc.) can be analysed in detail. By analyzing within-group and between-group variation between student translators and experienced translators statements could be produced concerning skills and properties of translation expertise (cf. Ericson 2000; Jensen 2009; Diamond & Shreve 2017).

Due to the sheer size of the behavioral gaze and keylogging data (more than 1600 sessions and thousands, if not 10-thousands of lines, per translation session), initial coding such as the annotation of fixations duration, fixation regressions, hesitations, pauses between keystrokes, etc, can only be generated and maintained by means of fully automatized processes. Thus, the TPR-DB provides a set of program scripts to annotate the raw behavioral and textual translation process data, to extract codes from the data and to organize the codes in the form of ten different types of summary tables (Carl, Schaeffer & Bangalore 2016).

Each of the 10 different types of tables can be regarded as categories in themselves. For instance the **st* tables index properties of ST words—and are thus the category for ST words. The **tt* tables categorize target text (TT) words and the **sg* tables contain information concerning translated segments (SG). Each of the tables (categories) contains a number of rows and columns. The rows represent

the units (single observations) of the category—i.e., the ST words, TT words or segments (SG), in the three categories respectively—and the columns encode features describing those units. For example, the labels *TrtS* and *FixS* are codes for the total reading time and the number of fixations on a ST item. In the ST category, *TrtS* and *FixS* refer to the total duration and number of fixations on the word described in the row; in the TT category, the codes refer to the ST word(s) of which the TT word is the translation; and, in the SG category, these codes indicate total reading time and number of fixations on the ST side of the segment. Similar codes exist for TT fixations, production duration, number of insertions and deletions, etc (see Carl, Schaeffer & Bangalore 2016).

Each table row is an independent unit sharing the table's category—a row in the **sg* table describes a translated segment and belongs to the SG category. Subcategories are sets of rows from tables of the same category with some common properties.[6] For instance, all ST words with a fixation duration greater 500 ms is a subcategory which consists of a (possibly empty) set of rows from the **st* tables. The investigation of the category may result in a new research *theme*, e.g., "translation difficulties"—long fixation times (> 500 ms) may be an indicator of this. More complex subcategories are possible, such as sets of sets or sets of lists of rows from one category, like a set of the list of words that form plural noun phrases, or a set of all word bi-grams consisting of a regression starting and landing on a word. Themes may also be organized hierarchically, e.g., a translation difficulty may be a part of a translation problem, and relate to translation expertise. The discovery of "patterned relationships" among categories and themes may then contribute to an evolving grounded theory.

Properties of subcategories may also appear as a code under another category. For instance, assume a subcategory in the **st* tables that groups all words per segment (i.e., an ST category that consists of sets with identical *SegId*). The cardinality (i.e., the size) of the sets that contain the words for each segment may equally be encoded as the length of the segment in the **sg* tables. In this way, new selective codes can be easily introduced under a category which capture properties of subcategories from other tables. This may facilitate further processing.

In the beginning of the TPR-DB project, there were only a dozen codes and a few categories. To date almost 300 codes are extracted from each translation session and listed in 10 different summary tables that jointly describe translation behavior (process) and textual (product) data on a word or segment level as pro-

---

6. In a table with n rows there are, in theory, 2^n different sub-categories, sharing the property that they are in the same category. Only a small subset of the possible categories will in practice be examined.

duction or activity units, etc. The codes can be inspected and analyzed within and across categories, e.g., in a process of axial, theoretical or selective coding.

The distribution of keystroke activities, production times, number of fixations and fixation duration were among the first codes in the TPR-DB and were used to subcategorize the data (e.g., Carl & Kay 2011). Systematic differences in the production time of translation segments shed light on the underlying cognitive translation processes of which the recorded observations are traces (overview of pause analysis in Kumpulainen 2015). During the development of the TPR-DB, further *theoretical coding* was added to the TPR-DB code inventory, including first fixation duration, first pass and regression path duration on words, etc. (Rayner 1998). The finer-grained encoding of fixation behaviour allows, for instance, to distinguish between early (automatized) and later (conscious) translation production processes. The data has also been annotated with linguistic properties of the words, such as word and segment length, word frequency, part-of-speech tags etc. which allows to explain cognitive processes along with different textual properties that are processed.

A number of central, selective codes such as *word-translation entropy, Cross, word-order entropy,* and *pause-to-word ratio* (see below) have emerged and are of great value as they are particularly suited to relate patterns of the translation product and translation processes that cover central themes in TPR:

– Detailed inspection of the data (i.e., visualisations of progression graphs, see below) led to the hypothesis that word order differences in the ST and the TT languages may have an effect on the translation behaviour. The local re-ordering of the word order was calculated based on the ST-TT alignment annotations and regression analyses showed that the amount of local distortion had indeed an effect on gazing and writing behaviour (Schaeffer et al. 2016). Subsequent inspection of the data in regression analyses (e.g., Bangalore et al. 2016) led to a further hypothesis that the number of possible different translation positions in the TT, expressed as the entropy of the *Cross* value *(HCross),* is a suited predictor of translation behaviour (Schaeffer & Carl, forthcoming). In other words, visual inspection of the data led to the formulation of a code (the *Cross* value), which was subsequently generalized as *HCross*. This, in turn, led to the refinement of a model of bilingual memory (Schaeffer & Carl, forthcoming) and its role during translation.

– While the *Cross* feature encodes the syntactic similarities of the ST and TT words and *HCross* the possible word-order choices, *HTra* captures the possible lexical choices of a translation and thus the semantic similarities of ST and TT words (Carl, Schaeffer & Bangalore 2016). These codes are part of the ST and TT categories, but can be generalized to the level of segments (SG cat-

egory). They have proven to be instrumental to categorize translations to be more or less literal.

–   It has long been assumed that longer inter-keystroke times (pauses in the text production) are indicators of cognitive effort (Krings 1986; Jakobsen 2002; Immonen 2006; O'Brien 2006; Carl & Kay 2011). However, how an inter-keystroke pause should be defined (e.g., 300ms—5 sec.) and what exactly it means has been at least contentious. A promising measure is the pause-to-word ratio (*PWR*) which normalizes the number of pauses by the number of words (Lacruz et al. 2012) which has been shown to be a good indicator of cognitive effort (Schaeffer et al. 2016, Vieira 2016).

In the next section we trace in more detail the development of what we think constitutes a *grounded theory* which evolved from the TPR-DB, and which is based on these *selective codes*.

3.   Theoretical saturation: The monitor model

The analysis of variation in translations has revealed that translators produce similar behavioral patterns at similar textual positions. In a "taxonomy of translation styles" Carl, Dragsted & Jakobsen (2011) and Dragsted & Carl (2013) capture translation habits of student and experienced translators on a coarse-grained level in alternative translations. They inspect the visualization of 72 translation sessions in the form of translation progression graphs (see below) and find that translation styles are comparable to styles known from writing research (Hayes & Flowers 1980). Carl & Dragsted (2012, 142) observe that in

> some cases all translators spend increased gaze time on the same ST sequences, for other sequences none seems to invest much reading effort and there are still other sequences at which some translators spend much and others no reading time.

An explanation into the finer structure of the translation process is provided by the translation *monitor model*. This model aims to explain processes of default and challenged translation based on the analysis of the observable traces in the process data. It assumes the existence of 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, 407–408). Based on the observation of instances of sequential or concurrent reading and writing activities during translation production, the monitor model was further extended into a recursive model of the translation process (Schaeffer & Carl 2013). This model explains the rela-

tion between production times and gaze data with a theory of shared cross-lingual representations in the bilingual mind. The recursive model states that translation involves activation of shared lexico-semantic and syntactical representations, i.e., the activation of features of both source and target language items which share one single cognitive representation. We argue that activation of shared representations facilitates automated processing (Schaeffer & Carl 2013, 169)

This implies that, in terms of product and process, more translation variation may be observed at places where automated processing is disturbed. Whereas automated processing enables concurrent reading and writing activities, sequential processing would slow down the translation process, and indicate instances where thus the monitor intervenes. Schaeffer & Carl (2014) show that the number of alternative translations and the cross-lingual alignment distances between the ST words and their translations (the *Cross* feature) has an effect on behavioral measures, such as production times, number of fixations, first fixation duration and total reading times. The further investigation of word translation choices of alternative translations has led to defining *word-translation entropy (HTra),* and the choices necessary to reorder translations in the target language is the basis for *word-order entropy,* or *HCross* (Carl & Schaeffer 2017). These new codes formalize the available syntactic and semantic choices that a translator has in the production of their translations and support a network of categories and themes (literal translation, automated processing, shared representations, monitoring, etc) which add to the age-old discussion on word-for-word and free translations.

We contend that these theories, codes and categories were arrived at by abductive reasoning. According to Reichertz (2007, 221), abductive reasoning is "an attitude towards data and towards one's own knowledge: data are to be taken seriously, and the validity of previously developed knowledge is to be queried." Following Peirce (1929), it relies on a three-stage discovery procedure which consists of abduction, deduction and induction:

> When faced with surprising facts, abduction leads us to look for meaning-creating rules […] The end-point of this search is a (verbal) hypothesis. Once this is found, a multi-stage process of checking begins, […] [which] consists of the derivation of predictions from the hypothesis, which is deduction, and the third step consists of the search for facts that will 'verify' the assumptions, which is induction.   (Reichertz 2007, 222)

In the case of the "taxonomy of translation styles", the abduction was qualitative in the sense that sets of progression graphs were intuitively classified. The development of the "recursive model" made use of quantitative hypothesis verification using the TPR-DB data. Both methods, as Glaser points out (above) are compatible with GTM.

Additional theoretical sampling has confirmed (or at least not contradicted) the hypothesis and theories put forward in the recursive translation models. A comparison of alternative translations from English into very different languages shows that translations into more remote languages (Japanese, Chinese, Hindi) are more difficult than into closer languages (Danish, Spanish, German, in that order), with higher *PWR* values, higher number of translation choices *(HTra)* and more syntactic distortions *(Cross)*. These investigations show that the general underlying difficulties in translation are comparable across different languages and different translation modes, but differ in their severity. To sum up, the line of research presented in this section showed that

1. words (and phrases) that are difficult to translate by a statistical machine translation systems (i.e., the MT system makes likely errors) are also difficult to translate for a human translator (Schaeffer & Carl 2014; Carl & Schaeffer 2017).
2. words (and phrases) which are more difficult to translate into one language are likely also more difficult to translate into any of the another languages, and vice versa (Carl, in print) and
3. the monitor model, extended by a theory of shared cross-lingual representations in the bilingual mind can explain those observations (Schaeffer & Carl 2013).[7]

We feel that this research theme has reached a certain degree of theoretical saturation, so that new samples are unlikely to add substantially new insights, unless the categories and themes are combined with new research methods and new, possibly more fine-grained, research questions.

4.   Axial coding: Translation duration and translation quality

One might wonder whether there is a relation between the length of an ST word and the time needed to produce the translation of the word. Figure 1 shows the effect of the length of an ST word *(LenS)* on translation production duration *(Dur)* in the translation mode and in the postediting mode in the English to Chinese dataset. There is a clear and significant effect, which is stronger in translation from scratch than in postediting. One might ask, why is this the case? Why are longer

---

7. Note that these conclusions are based on experiments with short news and sociology texts, which are relatively easy, and in particular require very little specialized vocabulary. We believe, however, that the conclusions can be extrapolated to translations of similar type.
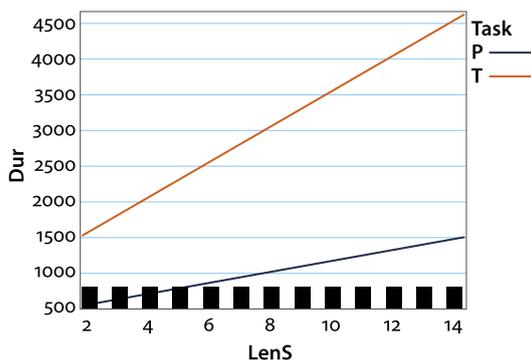
Figure 1.   Effect of ST word length on translation (T) and postediting time (P) duration
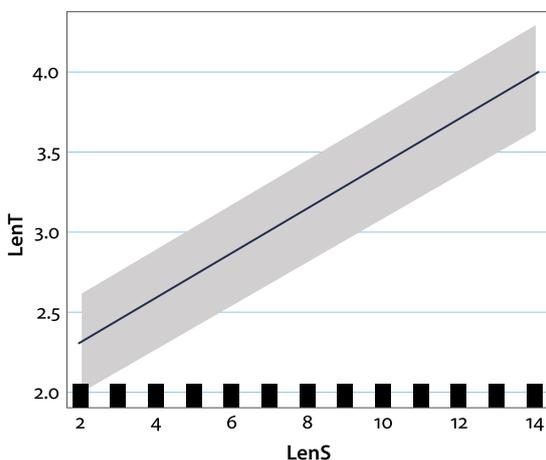

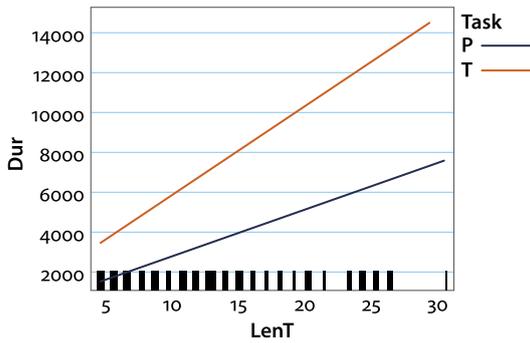**Figure 2.**   Relation between length of English ST word and Chinese translation in characters

ST words more time-consuming to translate than shorter ones, and why is this effect stronger in translation than in postediting? Checking the data reveals that:

1.   longer ST words tend to be translated into longer target words (Figure 2)
2.   postediting is faster than translation from scratch (Figure 3)
3.   longer target words are more time consuming to translate than shorter ones (Figure 4)

While it is not surprising that longer words also need more time to be typed when producing the translation from-scratch (as longer words require more keystrokes) it might be surprising why this effect is not as strong in postediting as compared to from-scratch translation. A hypothesis might be that the MT quality of longer ST words is, on average, better than that of shorter words, so that less
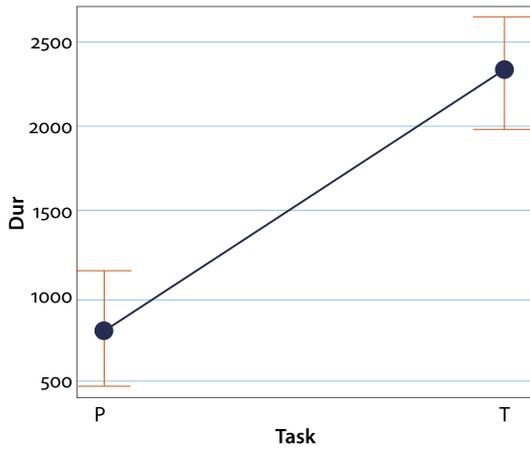
Figure 3. Duration in ms for word translation production in postediting (P) and translation (T)

Figure 4. Length of Chinese TL string and production duration
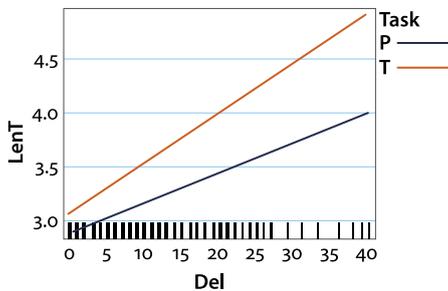
Figure 5. Number of deletions in long and short translations

postediting is needed for longer words. That is, there might be a negative correlation between *LenS* and the average error of the MT output—and indeed transla-
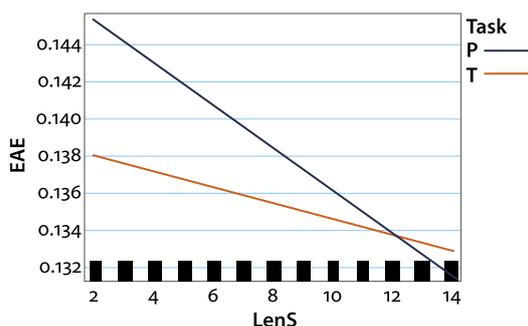
Figure 6. The effect of ST word length (LenS) on MT error (EAE)

tions of longer ST words need relatively less amendments during postediting than shorter ST words. Figure 5 shows that the number of deletions grows more slowly with the length of the target word for postediting as compared to translation from scratch. Figure 6 shows that indeed longer ST words have on average, less MT errors *(EAE)* than shorter ones. This effect is significant ($p < 0.05$) for postediting, but not for translation from scratch.

A closer investigation reveals that these MT errors are mainly fluency errors. Such errors, in turn, are often related to the wrong translation of function words, which are usually shorter than content words. In a further chain of inferences, drawing on the relevance-theory (Blakemore 2002), the distinction between content and function words can be mapped on encoding of procedural and conceptual instructions (Alves et al. 2014), which leads to conclude that procedural encoded content is more time-consuming to postedit than conceptual one (e.g., Alves et al. 2016).

This example highlights potentials of axial coding by putting into relation a number of codes: length of the ST and the TT words; number of deletions during translation production; translation production times; error annotation of MT output; and translation task. The correlations and regression analyses may lead to new categories (e.g., MT quality according to length of ST word) and explains, develops and underpins a number of themes (procedural vs conceptual encoding, good vs bad translations).

The example started out with a question about the relationship between the length of the ST word and translation duration in postediting and translation from scratch. The further exploration of the research was guided by a number of ad-hoc hypotheses about the underlying translation processes that could be quickly checked in the data. A further investigation could lead to a *substantive grounded theory* which would explain why procedural encoding are more time-consuming to postedit than conceptual encoding. For such a theory to emerge, further theoretical modelling could go on in various directions:

–   investigate in more detail the types of MT errors that are particularly time-consuming to translate and postedit. Theoretical codes could be borrowed from the distinction between procedural and conceptual coding (e.g., Alves et al. 2016).
–   assess whether similar behavior patterns are also observed for different types of texts for different translation purposes, different languages or translators with different degrees of expertise.

The analysis, as discussed here, is an iterative process of regression analyses using existing or new codes and categories, which will eventually lead to further theoretical codes, further targeted (theoretical) sampling and may result in a *formal grounded theory* (FGT). According to Glaser (2007, 100),

> […] an FGT is simply a conceptual extension […] nothing more than extending the general implications of a core variable by sampling more widely in the original substantive area and in other substantive areas and then constantly comparing with the purpose to conceptualize the general implications.

Additional information, such as usage of external resources during the translation process (consultation of dictionaries, collocation tools, web search, etc), more fine-grained quality annotations of the translations, or other information that relates to the linguistic or behavioral units, can be merged into the TPR-DB tables and be taken into account during the analysis, which will eventually lead to a more detailed FGT. We sense that this direction of research has still much potential until reaching *theoretical saturation*.

5.   Open coding: Activity Units

A less explored strand of research is the investigation of activity units (AU). AUs represent a category in the TPR-DB which fragment behavioral data (keystrokes and gaze) into snippets of coherent activity. Jensen (2011) points out that "the activity is either ST reading, TT reading/production or overlapping activity" (i.e., reading and writing at the same time), and Carl, Schaeffer & Bangalore (2016) fragment the activity data into seven different types of AUs with the following type labels:

–   1 Reading the ST
–   2 Reading the TT
–   4 Writing activity (with no recorded concurrent gaze data)
–   5 Writing while reading ST (touch typing)
–   6 Writing while reading TT (translation monitoring)

–   8 No activity recorded for more than 2.5 seconds

Types 1, 2 and 4 are basic translation activities. Types 5 and 6 take into account that reading can occur concurrently (in parallel) with writing, and a type 8 is assigned to segments if no activity is logged for longer than a given threshold.

Figure 7 shows a segment of a translation progression graph in which a translator produces the translation of English *the awareness of other hospital staff* into Spanish *la atención de otros empleados del hospital*. The graph shows the distribution of the keystrokes within a sequence of approximately 6.7 seconds (time stamp 434.400ms to 441.600ms) in which the Spanish translation *otros empleados de* ['other staff of'] is typed. There is a long pause of approximately 3.8 seconds between the writing of *t* and *r* which separates two production units (marked in horizontal gray stripes) in which respectively *ot* and *ros empleados de* are typed. The graph shows also how eyes move back and forth between the ST (blue dots) and the TT window (green diamonds). For instance, when writing *ot,* the translator first monitors the production of *o* in the TT and then gazes at the ST while writing *t*.

Figure 8 shows the segmentation of the same data into 13 AUs. The first AU in Figure 8 is of type 6, which consists of concurrent TT reading—a fixation is detected on the translation for ST word 104—while at the same time writing *o*. The following AU (type 2) consists of two TT fixations on the translation of ST words 105 (the *o* that has just been typed) and translation of word 103 *(atención)*. It is followed by ST reading (type 1) in which words 106 *(hospital)* and 105 *(other)* are fixated. The gaze then remains on the ST while *t* is typed (AU of type 5). After this, the long pause of almost 3.8 seconds is structured into five successive AUs of type 2-1-2-1-2 in which the translator reads and rereads ST words 103 to 104 and the translations of words 105 to 108; he seems to develop a translation strategy for *hospital staff* until around time stamp 439200 in an activity of concurrent writing and TT monitoring (type 6), *ros empleados* is produced without much hesitation.

There is no doubt that such kind of narrative may be insightful to uncover the underlying cognitive processes that take place during translation production—as it puts into context the input and the output of the translator's black box. A large-scale analysis of the data may reveal interesting details about the frequencies and relations of the observed patterns. However, how those narratives (and the data) should be coded is basically an open issue. Given the amount of data (hundreds of hours), the codes would have to be generated automatically in such a way that meaningful categories can emerge and research themes developed. The investigation of AUs has, to date, evolved in different categ ories and themes, but mostly avoided the rich coding that might be possible with the data.
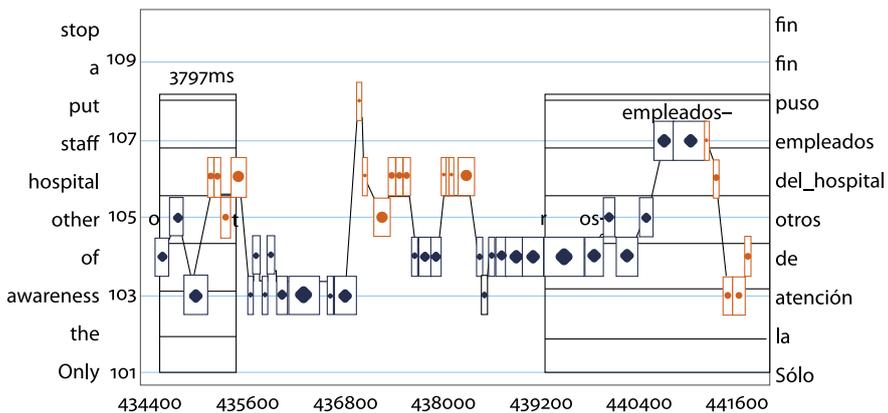
Figure 7.   Translation progression graph visualizing ST and TT fixations, keystrokes and two production units
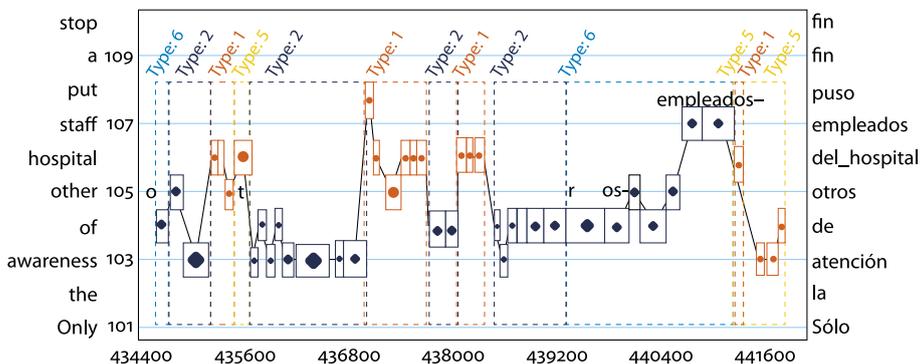
Figure 8.   Translation progression graph visualizing the same sequence of gaze and keystroke patterns as AUs of different type

Jensen (2011) is interested in the allocation of cognitive resources by student and experienced translators in 72 English-to-Danish translations from scratch. He reports that:

– Professionals and translation trainees spend most time on the TT, less on the ST.
– Translation trainees spend more time on the ST than professionals. TT Time is identical for both.
– Professionals shift less frequently between the ST and the TT.
– Duration per AU was shorter for professionals than for translation trainees.

- Professionals and translation trainees process in a mix of concurrent and sequential translation production, where professionals show more concurrent activity (aus of type 5 and 6) than translation trainees.
- More difficult texts require more shifts between ST and TT, longer duration and less concurrent activity for both translation trainees and professionals.

In a follow-up study, Martínez et al. (2014) corroborate these findings in the larger *multiLing* dataset. They use a subset of 204 sessions from the TPR-DB, annotated with information about translator experience and certification. They report that (2014, 154):

> […] translators engage 14% of their time in source text reading, between 17% to 37% in target text reading, between 35% to 42% inserting characters and 4% deleting characters. Certified translators spent significantly larger proportions of time in target text reading and target text typing than non-certified translators.

Hvelplund (2016) investigates the transitions between two successive AUs separately for professionals and trainees and counts them in a transition matrix. He compares the two matrices and observes that experienced translators shift from ST reading in 65.5% of the cases to writing activity (AUs of type 4, 5 or 6) while translation trainees do this only in 52.2% of the cases. Trainees switch to ST reading more often than professionals, which suggests that trainees aim more often at confirming meaning hypotheses (reflecting some kind of uncertainty), rather than allocating the cognitive resources directly to TT writing once a meaning hypothesis has been established.

Schaeffer & Carl (2017) investigate the average duration within different types of AUs (i.e., translation states) in postediting and translation from scratch. They find that approximately the same percentage of time is spent on ST reading in translation and postediting, 26% and 29% for postediting and translation from scratch, respectively. Posteditors spent, on average, 40% of the time on TT reading as compared to 19% for translators. In contrast, translators spent 44% of the time on writing activities (W)—AUs of type 4, 5 or 6—while this is only the case for 19% of the time for posteditors.

With regard to transitions between successive AUs (i.e., successive translation states), Schaeffer & Carl (2017) find that posteditors perform many more transitions between ST reading and TT reading (81% ST → TT and 56% TT → ST) than translators (52% ST → TT and 42% TT → ST). Most surprising is perhaps that posteditors start writing preferably after TT reading (41%) and only exceptionally after ST reading (16%), while for translators the writing probability is more balanced (54% TT → W and 42% ST → W).

In order to obtain a more fine-grained understanding of the reasons that trigger translators and posteditors to engage in writing, Schaeffer & Carl (2017) analyze the inner structure of the AUs, and find that the probability of a writing activity to happen depends on the duration of the preceding reading activity on the ST or TT. Schaeffer & Carl (2017) report that:

– Increased ST reading duration *decreases* the probability of successive writing, for both postediting and translation.
– Long ST reading might indicate translation difficulties. Potentially emerging translation hypotheses might first need to be crosschecked with the TT (i.e., triggering a transition $ST \rightarrow TT$), before writing the translation solution. That is, after gathering a translation hypothesis during long(er) ST reading the TT context needs first to be refreshed before writing is possible.
– Longer TT reading activities in postediting *increases* the probability of successive writing, while longer TT reading during translation *decreases* the probability of successive writing. A possible explanation of this observation could be that longer reading of the TT during translation makes it more likely that translators need to retrieve new (or refresh) information from the ST in order to continue writing. Posteditors, on the other hand, are already reading a TT that was produced by an MT system, where writing following a long stretch of TT reading might indicate the correction of a fluency error which does not require crosschecking with the ST.
– The more complex (the less literal) a translation is, the less likely posteditors and translators type a translation solution immediately after reading the TT. Both of them are more likely to refer back to the ST. In other words, the problem of a complex translation problems—if a particular ST item shares little syntactic or semantic properties with its TT item(s)—is discovered during TT reading, and resolved while referring back to the ST.

As the examples show, the research documented here is rather of *abductive* nature where the development and analysis of subcategories—e.g., types of AUs and their relations—and themes (translation mode, translator expertise) emerge through an iterative investigation and assessment of the data. Many research questions emerge from a dialogue with the data e.g., progression graphs, as was the case regarding the *Cross* feature, but also by iterating regression analyses, refinement of the coding schemes via selective or axial coding. In this process, the dialogic nature of research question development and confirmation or rejection cannot be stressed enough.

The objective is to understand the processes that have generated the data, to refine the categories and assess their relation in different contexts, rather than *quantifying* a pre-existing research hypothesis. Of course, the results could have

been presented as the outcome of a *quantitative* piece of research, by stating an initial hypothesis and then formulating the findings accordingly in order to confirm or reject the hypothesis. However, the research was *not* conducted in this way.

The research on AUs is an example of a rather immature GTM which led to developing a few themes, such as (un)certainty in translation, probability of writing, etc. However, a theoretical saturation of categories is far away, a systematic initial coding of the inner structure of the AUs has not been undertaken until now and the development of a *grounded theory,* e.g., an integration with the monitor model (see above), a connection to problem solving theories or a detailed analysis of translation strategies has to date not been attempted.[8]

A possible reason of this shortcoming might be the lacking adequacy of the currently deployed analysis tools. Regression analysis might not be powerful enough for handling time series and sequences of AUs. While we suspect that a large number of themes and categories may emerge through future coding and trough progressive identification and integration of categories, more powerful analysis methods might be required. Schaeffer & Carl (2017) suggest to model transitions of AUs as (Hidden) Markov Models which will be further elaborated in the next section.

6.   A formal translation processes model

As the preceding sections show, various categories, themes, hypotheses and theories have emerged through the investigation of the translation process data. In order to relate and integrate them into a comprehensive model and to assess their importance and validity, we might need a more formal framework which allows us to test the hypotheses and to evaluate their predictive power in a more systematic way. Figure 9 depicts a formal translation process model taking into account some of the processes discussed in this paper. The model resembles a Hidden Makov Model which consists of four levels of description.

In the center is the *Translator,* who is constrained by a number of *predictors* and who produces a sequence of *behavioral patterns* that lead to the final translation product. The *translator* is modelled as a network of (early and late) translation states which implement the actual translation processes. The early states represent automatic processing while later states represent more deliberate, strategic processing. Hönig (1991), for instance, proposes a model that distinguishes between uncontrolled, associative translation competence (i.e., unconscious early transla-

---

8.   However, Hvelplund (2011) makes a connection between AUs and memory models (Baddeley 2007).

tion processes) and a controlled workspace in which micro- and macrostrategies are stored. These workspaces may represent sequences of states of the translation network and produce particular sequences of AUs, as discussed above. The states are linked through transition probabilities which can be trained through machine learning techniques (e.g., Rabiner 1989).

There is a large number of heterogeneous *predictors,* which are likely to have an impact on the translation process, including cognitive, linguistic, cross-linguistic, and textual, communicative and socio-cultural domains (Toury 2004). Risku (2014) also mentions environmental conditions—including physical, geographic, economic, political and demographic aspects—which might play a role in the translation process. The *source text* is another predictor that will determine the characteristics of the target text.

The output of the model in Figure 9 has two levels of observations: *product observations* capture the changes in the translation product, i.e., the sequence of intermediate texts that are produced during the translation process. The final translation product is the final outcome in a series of successive intermediate text snapshots that emerge during the translation process and the translation process can be approximated by comparing the successive intermediate text snapshots. These observable textual changes are consequences of translators' keyboard activities, which can be traced through logging technology. *Behavioral patterns* include keystrokes, mouse clicks, eye movements etc. which may be segmented into activity units, as discussed above. The Hidden Makov Model in Figure 9 suggests that:

– translation processes are driven by a large number of *predictor*s
– the *translator* can be in only one state at any given time
– there are probabilistic transitions between successive hidden states
– each state emits exactly one behavioral pattern at a time

We believe that the implementation of such a network has the potential to put into relation and evaluate many of the codes, categories and themes discussed in this paper. It will likely uncover bottlenecks in the TPR-DB and lead to new codes which may in turn lead to new themes, etc. and finally to a substantial grounded theory of the translation process. We feel that necessary conditions are met for such an endeavor.

7.   Conclusion

This paper traces the development of the TPR-DB in the light of the Grounded Theory Method (GTM). This method was originally developed as an alternative to deductive modeling of sociological research, to overcome the restrictions of
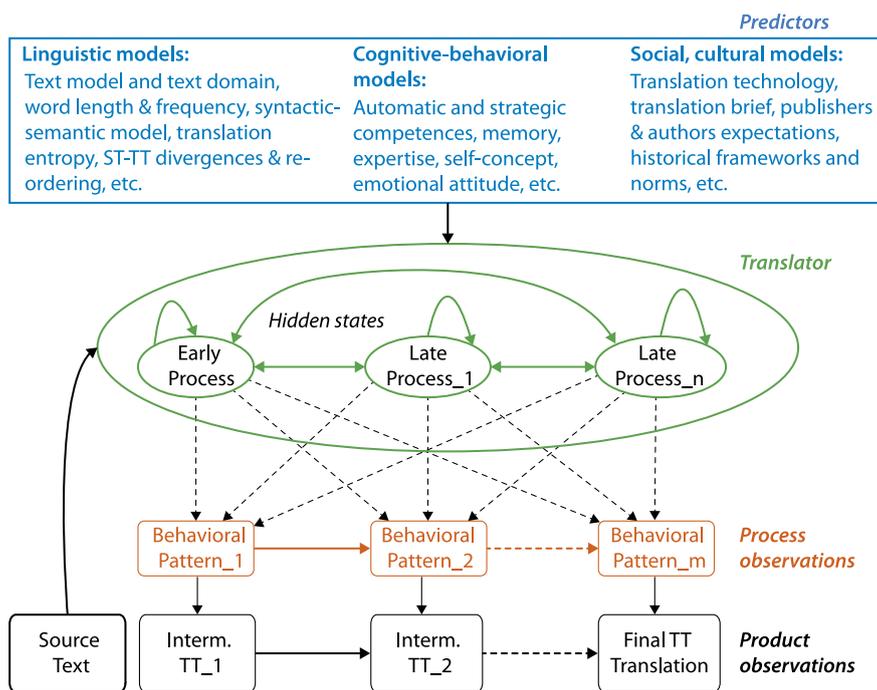
Figure 9.   Formal model of the translation process

empirical *quantitative* research to do merely hypothesis testing. GTM does not start with precise hypotheses—rather, research themes, hypotheses and categories are generated in the process of investigating the data. The development of themes from empirical data depends on a constant comparative analysis of cases with each other and the availability of adequate theoretical categories (Kelle 2007, 206). As a guiding principle, in order to make these GTM steps, the first question to ask is: What is happening here? (Glaser & Strauss 1967). According to Reichertz (2007, 215) GTM "falls within the realm of abductive logic research […] [it] was to a very small extent abductive from the start and became more and more abductive in its later stage."

In this paper, we have shown how the development of the TPR-DB can be seen as guided by similar principles. The first studies that led to compiling the TPR-DB in 2012 were clearly produced as a basis for *quantitative* research. Logs of keystroke data and gaze records were evaluated in line with a precise research hypothesis in mind. However, the incorporation of numerous data-sets from different studies, language pairs and translation modes allowed for exploratory, *abductive* and *qualitative* research, to gain an understanding of underlying processes, that had not been anticipated in the original studies. A rich set of codes and cat-

egories—compatible across all studies in the TPR-DB—supports this research, which makes it possible to quickly test and evaluate ad-hoc hypotheses and develop new categories and substantial theories on various aspects of translation processes.

The outstanding property of translation, as compared to many other cognitive activities, is the equivalence between the input and the output patterns with respect to semantic content. It allows to compare and interpret output and input at a certain level of structure for which the TPR-DB provides a set of codes. These codes facilitate the inference of underlying cognitive processes of comprehension and production that have generated the patterns in the observed behavioral data. Gazing behavior opens a door to understanding the input to the cognitive system which outputs at the other end a sequence of keystrokes in time, while the coordination of the input and the output reflects properties of the cognitive system.

The paper highlights some of the themes and categories that have emerged in the development of the TPR-DB and ends with open issues and with the proposal of a formal framework that could help understand human translation processes in more detail. Writing this paper was an instance of abductive reasoning and thus itself part of "The Logic of Discovery of Grounded Theory" (Reichertz 2007): in the onset, there was the intuition that the development of the TPR-DB could be argued to be a GTM. Given that the authors of this paper had almost a decade of experiences in the field of TPR and 5 years of active development of the TPR-DB, we had ample bottom-up "material" (data) that would need to connect with the top-down concepts of GTM. One of the reviewers of an earlier draft version found this to be "a bold undertaking," which we hope to have accomplished to some degree using the inductive/deductive method inherent to GTM.

References

Alves, Fabio, Karina Sarto Szpak, José Luiz Gonçalves, Kyoko Sekino, Marceli Aquino, Rodrigo Araújo e Castro, Arlene Koglin, Norma B. de Lima Fonseca & Bartolomé Mesa Lao. 2016. "Investigating Cognitive Effort in Postediting: A Relevance-Theoretical Approach." In *Eyetracking and Applied Linguistics*. Edited by S. Hansen-Schirra and S. Grucza, 109–142. Berlin: Language Science Press. doi: 10.17169/langsci.b108.296

Alves, Fábio, José Luiz V. R. Gonçalves, and Karina S. Szpak. 2014. "Some Thoughts about the Conceptual / Procedural Distinction in Translation: A Key-Logging and Eye-Tracking Study of Processing Effort." In *Minding Translation*. Edited by R. Muñoz, 151–175. *MonTI*, special issue 1. Alicante: Universidad de Alicante.

Baddeley, Alan D. 2007. *Working Memory, Thought, and Action*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198528012.001.0001

Bangalore, Srinivas, Bergljot Behrens, Michael Carl, Maheshwar Ghankot, Arndt Heilmann, Jean Nitzke, Moritz J. Schaeffer, and Annegret Sturm. 2016. "Syntactic Variance and Priming Effects in Translation." In *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Edited by M. Carl, S. Bangalore, and M. Schaeffer, 211–38. Cham: Springer. doi: 10.1007/978-3-319-20358-4_10

Birks, Melanie and Jane Mills. 2011. *Grounded Theory: A practical guide*. London: Sage Publications.

Blakemore, Diane. 2002. *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486456

Carl, Michael, and Martin Kay. 2011. "Gazing and Typing Activities during Translation: A Comparative Study of Translation Units of Professional and Student Translators." *Meta* 56 (4): 952–75. doi: 10.7202/1011262ar.

Carl, Michael, Barbara Dragsted, and Arnt Lykke Jakobsen. 2011. "A Taxonomy of Human Translation Styles." *Translation Journal* 16 (2). Available at http://translationjournal.net/journal/56taxonomy.htm

Carl, Michael, and Barbara Dragsted. 2012. "Inside the Monitor Model: Processes of Default and Challenged Translation Production." *TC3, Translation: Computation, Corpora, Cognition*, 2 (1): 127–145.

Carl, Michael, Moritz J. Schaeffer, and Srinivas Bangalore. 2016. "The CRITT Translation Process Research Database." In *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* edited by M. Carl, S. Bangalore, and M. J. Schaeffer, Chapter 2 (pp. 13–54). Cham: Springer. doi: 10.1007/978-3-319-20358-4

Carl, Michael, and Moritz J. Schaeffer. 2017. "Why Translation Is Difficult : A Corpus-Based Study of Non-Literality in Post-Editing and From-Scratch Translation." *Hermes—Journal of Language and Communication Studies* 56: 43–57.

Carl, Michael. In print. "Literal Translation, Default Translation and the Similarity of Language Systems." *Target*

Charmaz, Kathy. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: Sage.

Diamond, Bruce J. and Gregory M. Shreve. 2017. "Deliberate Practice and Neurocognitive Optimization of Translation Expertise." In *The Handbook of Translation and Cognition*. Edited by J. W. Schwieter and A. Ferreira. Malden, MA: Wiley-Blackwell. doi: 10.1002/9781119241485.ch26.

Dragsted, Barbara, and Michael Carl. 2013. "Towards a Classification of Translation Styles Based on Eye-Tracking and Keylogging Data." *Journal of Writing Research* 5 (1): 133–158. doi: 10.17239/jowr-2013.05.01.6

Ericsson, K. Anders. 2000. "Expertise in Interpreting: An Expert-Performance Perspective." *Interpreting* 5 (2): 189–222. doi: 10.1075/intp.5.2.08eri

Gasson, Susan. 2004. "Rigor in Grounded Theory Research: An Interpretive Perspective on Generating Theory from Qualitative Field Studies." In *The Handbook of Information Systems Research*. Edited by M. Whitman and A. Woszczynsk. 79–102. IGI Global. doi: 10.4018/978-1-59140-144-5.ch006

Glaser, Barney G. 2008. "Qualitative and Quantitative Research." Gounded Theory Review 7 (2). Available at http://groundedtheoryreview.com/2008/06/30/1052/ Originally published in Glaser. Barmey G. 2003. *The Grounded Theory Perspective II: Description's remodeling of grounded theory methodology*, chapter 7 (99–113). Mill Valley, CA: Sociology Press.

Glaser, Barney G. 1978. *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley, CA: Sociology Press.

Glaser, Barney G. and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Hayes, John R., and Linda Flower. 1980. "Identifying the organization of writing processes." In *Cognitive Processes in Writing*. Edited by L. Gregg and E. R. Steinberg, 3–30. Hillsdale, NJ: Lawrence Erlbaum.

Heist, Brian. 2012. *From Thematic Analysis to Grounded Theory*. Lecture notes available at http://www.mermaid.pitt.edu/documents/MERMAID%20GrndedThry%20talk%20HEIST.pdf

Hood, Jane. 2007. "Orthodoxy vs. Power: The defining Traits of Grounded Theory." In *The SAGE Handbook of Grounded Theory*. Edited by A. Bryant and K. Charmaz, 151–164. London: Sage.

Hvelplund, Kristian Tangsgaard. 2016. "Cognitive Efficiency in Translation." In *Reembedding Translation Process Research*. Edited by R. Muñoz, 149–170. Amsterdam: John Benjamins.

Hvelplund, Kristian Tangsgaard. 2011. *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging Study* (PhD Series, No. 10.2011). Frederiksberg: Samfundslitteratur.

Hönig, Hans G. 1991. "Holmes' 'Mapping Theory' and the Landscape of Mental Translation Processes." In *Translation Studies: The State of the Art*. Edited by K. van Leuven-Zwart and T. Naajkens, 77–89. Amsterdam: Rodopi.

Immonen, Sini. 2006. "Translation as a Writing Process: Pauses in Translation versus Monolingual Text Production." *Target* 18 (2): 313–336. doi: 10.1075/target.18.2.06imm.

Jakobsen, Arnt Lykke. 2002. "Orientation, Segmentation, and Revision in Translation." In *Empirical Translation Studies: Process and Product*. Edited by G. Hansen, 9–20. Copenhagen: Samfundslitteratur.

Jensen, Kristian T. H. 2009. "Indicators of Text Complexity." In *Behind the Mind: Methods, Models and Results in Translation Process Research*. Edited by S. Göpferich, A. L. Jakobsen, and I. M. Mees, 61–80. Copenhagen: Samfundslitteratur.

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, 215–236. London: Bloomsbury Academic.

Kelle, Udo. 2007. "The Development of Categories: Different Approaches in Grounded Theory." In *The SAGE Handbook of Grounded Theory*. Edited by A. Bryant and K. Charmaz, 191–213. London: Sage

Krings, Hans Peter. 1986. *Was in Den Köpfen von Übersetzern Vorgeht: Eine Empirische Untersuchung Zur Struktur Des Übersetzungsprozesses an Fortgeschrittenen Französischlernern*. Tübingen: Günter Narr.

Kumpulainen, Minna. 2015. "On the Operationalisation of 'Pauses' in Translation Process Research." *Translation and Interpreting* 7 (1): 47–58. doi: ti.106201.2015.a04.

Lacruz, Isabel, Gregory M. Shreve, and Erik Angelone. 2012. "Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study." *AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*, 21–30.

Martínez Gómez, Pascual, Akiko Aizawa, and Michael Carl. 2014. "Characterization of Human Translation for Different Levels of Expertise." *The 20th Annual Conference Proceedings of the Association for Natural Language Processing of Japan*, 153–156.

O'Brien, Sharon. 2006. "Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output." *Across Languages and Cultures* 7 (1): 1–21. doi: 10.1556/Acr.7.2006.1.1.

Peirce, Charles Sanders. 1929. "Guessing." *Hound and Horn*, 2 (3): 267–282.

Rabiner, Lawrence R. 1989. "A tutorial on Hidden Markov Models and selected applications in speech recognition." *Proceedings of the IEEE*. 77 (2): 257–286. doi: 10.1109/5.18626

Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124 (3): 372–422. doi: 10.1037/0033-2909.124.3.372.

Reichertz, Jo, 2007. "Abduction: The Logic of Discovery of Grounded Theory." In *The SAGE Handbook of Grounded Theory*. Edited by B. Anthony and K. Charmaz, 214–228. Available at http://www.qualitative-research.net/index.php/fqs/article/view/1412/2902

Risku, Hanna. 2014. "Translation Process Research As Interaction Research: From Mental To Socio-Cognitive Processes." In *Minding Translation*. Edited by R. Muñoz, 331–353. *MonTI*, special issue 1. Alicante: Universidad de Alicante.

Schaeffer, Moritz, and Michael Carl. 2013. "Shared Representations and the Translation Process: A Recursive Model." *Translation and Interpreting Studies*, 8 (2): 169–190. doi: 10.1075/tis.8.2.03sch

Schaeffer, Moritz J., and Michael Carl. 2014. "Measuring the Cognitive Effort of Literal Translation Processes." In *Proceedings of the Workshop on Humans and Computer-Assisted Translation (HaCaT)*. Edited by U. Germann, M. Carl, P. Koehn, G. Sanchis Trilles, F. Casacuberta, R. Hill, and S. O'Brien, 29–37. Stroudsburg, PA: Association for Computational Linguistics.

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

Schaeffer, Moritz, and Michael Carl. 2017. "A Minimal Cognitive Model for Translating and Post-Editing." In *Proceedings of MT Summit XVI*. Edited by S. Kurohashi and P. Fung, vol. 1, 144–55. Japan: International Association for Machine Translation.

Schaeffer, Moritz J., and Michael Carl. 2017. "Language Processing and Translation." In *Empirical Modelling of Translation and Interpreting*. Edited by S. Hansen-Schirra, O. Čulo, S. Hofmann, and B. Meyer. Berlin: Language Science Press.

Strauss, Anselm L. and Juliet M. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.

Strauss, Anselm L. and Juliet M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Tirkkonen-Condit, Sonja. 2005. "The Monitor Model Revisited: Evidence from Process Research." *Meta* 50 (2): 405–14. doi: 10.7202/010990ar.

Toury, Gideon. 2004. "Probabilistic Explanations in Translation Studies: Welcome as They Are, Would They Qualify as Universals?" In *Translation Universals: Do They exist?*. Edited by A. Mauranen and P. Kujamäki, 15–32. Amsterdam: John Benjamins. doi: 10.1075/btl.48.03tou.

Vieira, Lucas Nunes. 2016. "Cognitive Effort in Post-Editing of Machine Translation: Evidence from Eye Movements, Subjective Ratings, and Think-Aloud Protocols." PhD dissertation, Newcastle University. Available at https://theses.ncl.ac.uk/dspace/bitstream/10443/3130/1/Nunes%20Vieira%2C%20L%202016.pdf

Walsham, Geoff. 1993. *Interpreting Information Systems in Organizations*. Chichester: Wiley.

Authors' addresses

Michael Carl
Renmin University of China
School of Foreign Languages
No. 59 Zhongguancun Street
Haidian District Beijing, 100872
P.R. China

mc.ibc@cbs.dk

Moritz Schaeffer
Johannes Gutenberg-Universität Mainz
Postfach 1150
D-76711 Germersheim
Germany

moritzschaeffer@gmail.com

Biographical notes

Michael Carl is a professor at Renmin University of China, Beijing, and director of the Center for Research and Innovation in Translation and Translation Technology at the Copenhagen Business School. His current research interest is related to the investigation of human translation processes and interactive machine translation.

Moritz Schaeffer is a research assistant at the Center of Research and Innovation in Translation and Translation Technology (Copenhagen Business School), the Institute for Language, Cognition and Computation (University of Edinburgh) and the National Institute of Informatics (Tokyo). He is currently a research associate at the school of translation studies, linguistics and cultural studies at the Johannes Gutenberg University of Mainz in Germersheim.