Chapter 4

Sketch of a Noisy Channel Model for the translation process

Michael Carl
Renmin University of China

Moritz Schaeffer
Johannes Gutenberg University of Mainz

To advance the state of the art in translation process research, Toury (2004) requests the formulation of "probabilistic explanations in translation studies". This chapter develops these "conditioned statements" into a Noisy Channel Model of the translation process with the ultimate aim to predict "particular *modes of behavior*" by their observable traces in the user activity data (UAD). We first develop a Noisy Channel Model for the translation process and then present a number of research results that may serve as a basis for the formulation of observable behavioral units and of the latent states in a noisy translation process model. However, a large amount of research has still to be conducted before we might be able to get a complete picture of the various shades and complexities of the translation process.

1 Introduction

The Noisy Channel Model (Shannon & Weaver 1949) has been very productive for solving non-deterministic problems in communication and computational linguistics. It is a mathematical formalization of communicative processes that underlies, among many other things, speech recognition (Huang et al. 1990), statistical machine translation (Brown et al. 1993) and the translation of a text from a source language to a target language. Statistical machine translation (SMT) models translation as a process in which a source text is decoded, thereby eliminating the noise (e.g. adjusting lexical and syntactic divergences) to uncover

the intended message (i.e. the translation). In automatic speech recognition, the speech signal is segmented, analysed and mapped onto a sequence of phones, which represents possible pronunciations of the words to be recognized.

In this chapter we develop a framework of the Noisy Channel Model for the translation process. In analogy with speech recognition, the translation *process* is modeled as a probabilistic sequence of behavioral observations, such as keystrokes and eye movements, which are emitted by underlying hidden processes. The aim of the noisy channel is to anticipate and generate the behavioral user activity data (UAD) and to uncover and understand the underlying hidden translation processes that are involved in the generation of the translation.

Just as for automatic speech recognition, the segmentation and quantification of the stream of events is a precursor also for decoding translation processes. A considerable amount of work has been invested in translation process research to define and investigate various kinds of units, measures and metrics that are suited to structure and quantify processing activities.

On the one hand, the final outcome of the translation process is a text (i.e. the translation) which is defined by the spatial/sequential order of the linguistic items that it constitutes. We may thus approach the translation process from a textual angle and investigate behavioral patterns that are involved in the production of particular words or phrases. Consequently, we will deal with text-based units of investigation, spatial areas of interest (AOIs[1]) which accumulate related behavioral UAD.

On the other hand, the translation processes can be considered a temporal sequence of translational events, which may be segmented into coherent chunks or behavioral units. For instance, pauses in the translation production process (i.e. gaps in the typing activities) have been associated with cognitive meta-activity and pause analysis has been proposed as a method to detect the amount of 'cognitive effort' in translation (e.g. Immonen 2006; O'Brien 2006; Lacruz et al. 2012). However, it is unclear what exactly the cognitive processes are that take place during typing pauses and it is an unsolved problem to determine what exactly makes pauses more or less effortful. In addition, recording of gaze data is required to 'fill' the typing pauses and to identify the specific motivation of a particular pause (Kumpulainen 2015: 47).

In both cases, a distinction can be made between early, automatised translation processes and later more time-consuming processes. A number of measures exist that are suited to describe early translation processes; these include fluent typ-

---

[1]This term was coined as a tool for the investigation of eyetracking data. It can be equally used to denote a textual area to accumulate different kinds of UAD.

ing activities and early eye movement measures such as first fixation durations. However, late translation processes are more complex, more difficult to describe and their traces in the process data are more varied and difficult to classify or identify than the early processes.

As suggested by Toury (2004), translation processes and behavioral observations are probabilistic in nature. In this chapter, we suggest a probabilistic framework to assess and integrate several findings from empirical translation process research. We first lay out the general ideas of the Noisy Channel Model in §2 and apply the introduced notions to the translation process research (TPR) terminology. We show that the Noisy Channel Model provides a powerful framework to formalize "probabilistic explanations in translation studies" (Toury 2004).

One of the essential requirements for a noisy translation process model is the fragmentation and quantification of the stream of UAD into meaningful segments. §3 discusses a number of attempts to segment the UAD into meaniglful units, including production units, attention units and activity units. In the noisy channel model, these behavioral units are generated by underlying hidden translation states. Drawing on the monitor model, we make a distinction between early and later translation processes and argue that they represent different mental states. §4 provides a number of properties for these earlier and later translation states, the output of which can be measured in the behavioral UAD.

2 The noisy channel model in translation

The Noisy Channel Model conceptualizes communication as a problem of decoding (Shannon & Weaver 1949), in which a transmitter sends a message m through a noisy communication channel. The receiver perceives a signal o as a noisy encoded version of the original message. In order to reconstruct the message m, the Noisy Channel Model assumes two factors: a language model $P(m)$ which indicates the probability of the original message m and the conditional probability $P(o|m)$ which quantifies the probability of the signal o provided we know the message m. The probability $P(o|m)$ accounts for the noise that is added during the communication process. This component analysis can be achieved with the help of Bayes' theorem, which states that a conditional probability $P(m|o)$ can be decomposed as $P(m|o) = P(o|m) * P(m)/P(o)$. Decoding makes use of the right side of this equation, ignoring the common denominator $P(o)$. The noisy channel model is used to formalize a variety of different communication problems; in automatic speech recognition, o is an acoustic signal and m is the spoken message (Mylonakis et al. 2007). The model is also used for part-of-speech tagging,

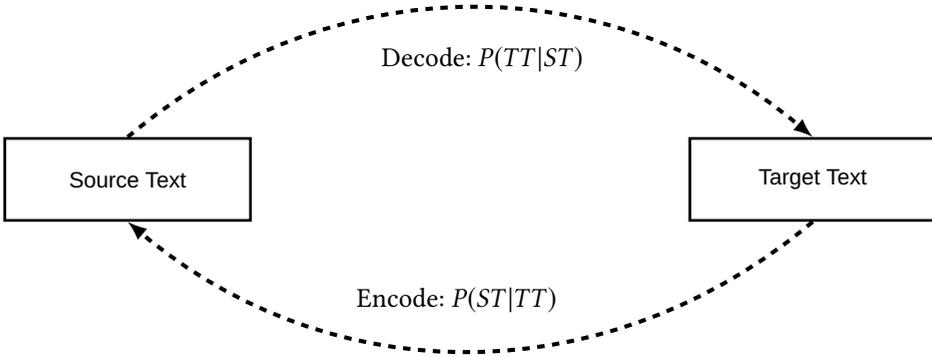in syntactical analyses, and in many other natural language processing (NLP) applications.

Figure 1: The noisy channel model for translation

In the context of SMT it is assumed that the target text (TT) corresponds to the message *m* and the source text (ST) is the signal *o* that we want to decode (see Figure 1). Given we know all the factors that are involved in the encoding process in $P(ST|TT)$ and we know the probabilities with which each single event occurs, we can reverse the encoding process based on Bayes' law as shown in equation 4.1.

$$P(TT|ST) \quad = \quad P(ST|TT) * \frac{P(TT)}{P(ST)} \tag{4.1}$$

As each of the factors that contributes to a translation (i.e. the encoding and decoding) generates a large number of hypotheses, the Noisy Channel Model makes use of a search operator *argmax* to retrieve the most probable translation among the many possible options.

$$\widehat{TT} \quad = \quad argmax \ P(TT|ST) * P(TT) \tag{4.2}$$

The *argmax* operator in equation 4.2 takes account of the fact that there can be many possible outcomes, but we are searching only for the most likely translation $\widehat{TT}$. This operator produces, under optimum circumstances, the best re-

construction of the translation $\widehat{TT}$ based on observed source text (or sentence) ST.

$$P(ST|TT) \;\; = \;\; \sum_i P(ST|TT, v_i) \tag{4.3}$$

Equation 4.3 demonstrates the possibility of including additional predictor variables v_i in the Noisy Channel. Since the total probability of a sample space always amounts to $1 = \sum_i v_i$ any number of additional variables can be introduced in this manner, so as to provide additional explanatory power to the computations.

Early approaches to SMT modelled the channel as a probabilistic translation dictionary (Brown et al. 1988). More recent SMT systems use many additional resources, such as phrase- or tree-based translation models; they encode sentences as lattices or confusion networks that enhance the noisy channel extensively. In order to integrate a large amount of features that might impact the decoding process, the Noisy Channel Model has been generalized as shown in equation 4.4, which can take into account any number of feature functions $f_k(ST, TT, v_k)$, which are weighted by a factor λ_k (Och et al. 2003). Each feature function $f_k(\cdot)$ may represent a very different aspect in the decoding process and can be trained independent of other features. Its contribution to the overall outcome of the decoding process is ranked by a factor λ_k:

$$\widehat{TT} \;\; = \;\; argmax \;\; P(TT|ST) \approx \sum_{k=1\ldots n} \lambda_k * f_k(ST, TT, v_k) \tag{4.4}$$

We propose an adaptation of the Noisy Channel Model to model human translation and post-editing processes.

2.1 Probabilistic translation processes

In an attempt to define the notion of translation universals, Toury (2004) requests the formulation of conditioned statements which would provide "probabilistic explanations in translation studies". Conditioned statements would predict "particular *modes of behavior* (or their observable results) … [based on] an array of *variables*, whose capacity to enhance (or reduce) the adoption or avoidance of a particular behavior would be verified empirically" (Toury 2004: 24). The most general format of such a conditioned statement according to Toury would be as

follows:

> If 1 and 2, and 3, and ... ∞, then there is great likelihood that X [...] where the numbers (1, 2, 3, ... ∞) stand for the different variables which may have an effect on the selection of a translational behavior. (p. 26)

In terms of the notion introduced above, Toury's conditioned translation statement can thus equivalently be expressed as a set of conditional probabilities in the form $P(x_i | v_1, v_2, \ldots, v_m)$ where x_i stands for a particular predicted translation behavior and the set $V = \{v_1, v_2, \ldots, v_m\}$ contains predictor variables which have an effect on and explain — to a certain extent — the observed behaviour x_i in a probabilistic manner. Even though there might be many different modes of translation, we assume that the translation process consists of a finite inventory of behavioral patterns $X = \{x_1, \ldots, x_n\}$ and we assume — in contrast to Toury — that for each of the observations $x_i, 1 \leq i \leq n$ there exists only a finite set of predictor variables $v_j \in V$ that has an effect on x_i. The observed translation process X can then be formalized as a sequence of possible behavioral patterns $x_1, x_2 \ldots$ that are conditioned by a number of predictor variables $v_1, v_2 \ldots$. The most likely (explanation for) translation behaviour $\widehat{X}$ can thus be computed in a similar way as the most likely translation $\widehat{TT}$.

The general idea in this model is that the value of a dependent variable (X) is related to a set of independent variables (V) through a function (F). Given the translation UAD, we learn the function (F) to minimize the error (also known as loss (L)) in prediction ($\widehat{X}$) of the variable V.

$$\text{minimize } L(X, \widehat{X}), \text{ where } \quad \widehat{X} = F(X, V) \tag{4.5}$$

2.2 Latent translation states

As an illustration of a probabilistic conditioned statement, Toury (2004) discusses a made-up example in which he illustrates a hypothetical effect of experience and fatigue on whether translational processing will be applied to small or low-level textual-linguistic entities. In this example, the level of textual-linguistic entities that a translator works with would be represented by the dependent variable X whereas the explanatory (or predictor) variables V represent the experience and fatigue of the translator which may have an effect on the choice of the translation unit.

Toury is not very consistent in his usage of the terms "translation modes" and "translation behavior". Surprisingly, for him the translator's behavior "is not really observable in any direct way" (Toury 2004: 26). He nevertheless mentions possible forms of translational behavior which include all kinds of "regularities [that] can be found on every level, from the individual act of translation [...] to the overall notion of translation", under which he also subsumes translation universals, i.e. structures in the translation product. We will come back to this issue in the conclusion.

On the one hand, Bernardini (2001: 241) points out, "an understanding of translation [processes] [...] is not derivable solely from an analysis of the final product". On the other hand keylogging and eye-tracking technologies give us today the possibility to directly observe and investigate translation behavior and empirically assess the granularity of the chunks that a translator works with. Accordingly, we conceptualize the translation process as successive intermediate versions of a text (i.e. the emerging translation), which are the direct consequences of translation behavior. Most important in this process are obviously the keystrokes which are the direct causes for text modifications.

As an extension to Toury's model, we assume that the behavioral patterns are triggered through internal (latent) states in the translator's "black-box". Using EEG and fMRI technologies we may be able to investigate and measure these latent states directly through experimental equipment in the near future (Annoni et al. 2012). Currently however, Think Aloud Protocols (TAPs) and introspection are methods used to assess the hidden (or latent) states in the translation process. Two of the main goals of TAP research are (1) to describe translation problems, and (2) to isolate strategies and translation procedures. According to Lörscher (2005: 599), the "data [that are collected through TAP] are interpreted as (observable) indicators of (unobservable, mental) translation strategies" which, for him, represent the basis for the formation of hypotheses regarding the mental translation process. Based on the collected data, Lörscher (1991) describes five basic translator types which differ with respect to how much the solution of a translation problem is automatized, whether the translator requires search, whether a translation problem is decomposed into smaller parts, and to what extent the translation problems are consciously accessible and can be verbalized. Lörscher (1991: 280) finds that "[w]hen several [translators] are faced with a problem X, many or most of them employ similar or the same types of strategy".[2] However, findings like these remain to be quantified and scrutinized for their predic-

---

[2]While this looks similar to Toury's conditioned statement, the "problem X" would here be a predictor variable, while the dependent variable "types of strategy" is a latent translation state.

tive value. For Krings (1986), TAPs have only a restricted validity. He cautions that "although verbal data do give evidence of mental processes, they cannot be claimed to be isomorphic with those processes" (Krings 1986: 264).
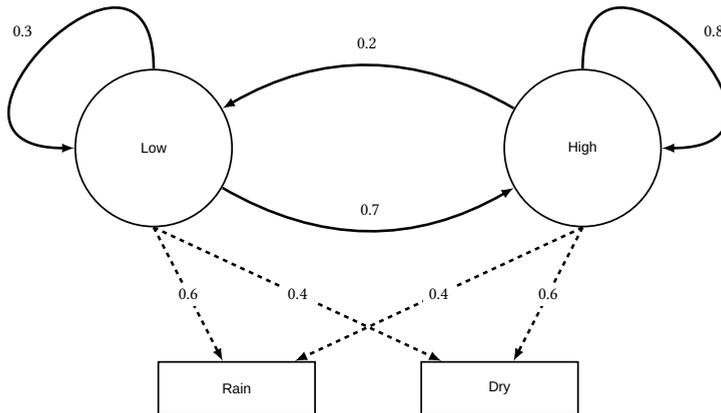

Figure 2: A hidden Markov Model

The translation model suggested by TAP analysis can be formalized as a Hidden Markov Process in which a number of interconnected "hidden" states emit observations with a certain probability. Figure 2 outlines the idea of a hidden Markov model, where the two hidden states (Low and High) emit possible observations (Rain or Dry) with a certain probability. A hidden Markov model can consist of a large number of hidden states and emit many different observations. The hidden states are organized in the form of (possibly completely connected) recursive networks and transition probabilities that indicate the likelihood with which one state follows another. A number of efficient algorithms exist to learn transition and emission probabilities from data and to compute most likely sequences of observations.

2.3 Early and late stages of translation states

de Groot (1992), Hartsuiker et al. (2004), Lopez & Resnik (2009) and Schaeffer & Carl (2013) assume that entries in the mental bilingual dictionary consist of nodes that link the lemmas, concepts, word forms and syntactical information between the two languages. The nodes are linked to all words that exhibit correspond-

ing features in a language-independent fashion. They are, however, specific to particular language combinations as not all languages always realize the same morphosyntactic and semantic aspects in the same way.

The models predict that the more nodes overlap between the source and target language words and structures, the less time it takes to retrieve associations and to generate translations. This unchallenged translation process is to a large extent a subliminal process.

The translation model by Schaeffer & Carl (2013) posits that the translation process is *recursive* given that translators often switch back and forth between source and target text in order to examine both texts for interpretive resemblance. In doing so, translators are primed by either the source or the target text, allowing them to register and analyze the resemblances in both texts. This view is in line with the monitor model (Tirkkonen-Condit 2005) according to which an automatic default translation procedure is interrupted when a problem occurs and triggers conscious translation processes.

Figure 3 visualizes an unchallenged translation process. It shows a relatively undisturbed translation progression in which an English source sentence "All of his victims were old weak woman" is translated into Danish "Alle hans ofre var aeldre svagelige kvinder" on the left and right axes respectively. Translation activities are depicted in the graph on a timescale, from ms 206.000 to 215.000. The overall translation of the eight words takes approximately 9 seconds. The figure shows keystrokes (insertions and deletions), gaze fixations on the source text (blue boxes with dots) and gaze fixations on the target text (green boxes with diamonds).
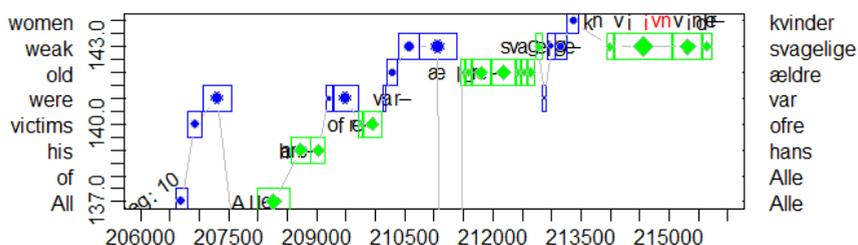

Figure 3: Example of an undisturbed translation progression

Some of the measures for the translation segment in Figure 3 are shown in Table 1 and explained as follows:

- *FFDur*: first fixation duration on the source word (blue rectangle)

- *FixS* and *FixT*: number of fixations on the source text and target words respectively; there are only few fixations (≤ 10) on each single word

- *TrtS* and *TrtT*: total reading time of source text word and its translation, respectively

- *Ins*, *Del* and *Dur*: number of counted insertions, number of counted deletions and total typing duration to produce the translation, respectively

- *Munit*: number of revisions (micro units) of a word (see page 103)

- *HTra*: word translation entropy (cf. literality criteria 3, see page 96)

Table 1: Behavioural measures for smooth translation activities.

SToken	FFDur	FixS	TrtS	FixT	TrtT	Ins	Del	Dur	Munit	HTra
All	60	2	259	1	559	0	0	622	1	0.41
of	0	0	0	1	559	5	0	622	1	0.74
his	0	0	0	3	658	5	0	462	1	0.49
victims	239	1	239	3	1177	5	0	634	1	0.49
were	478	5	1116	1	80	4	0	445	1	0.00
old	179	1	179	7	1136	6	0	1061	1	0.99
weak	159	8	1796	5	1813	10	0	1177	1	1.36
women	59	2	238	1	200	11	3	2234	1	0.24

The segment in Figure 3 is chararacterized by relatively few fixations on the source and target words and relatively short total reading times. There is a short delay between the reading of a source word and the production of the translation (i.e. the eye-key-span; see Schaeffer & Carl 2016, Schaeffer & Carl 2017 [this volume]). Only the translation of "weak" and "women" required longer reading times, perhaps due to unusual character combinations in the Danish translations.

Figure 4 shows an excerpt from an English → Chinese translation session, with much more complex patterns of ST and TT reading behavior, repeated regressions, re-reading, backtracking, deletions, revisions, etc. The production of this translation segment of 17 words took approximately 100 seconds, which is almost 5 times longer per word than the Danish translation in Figure 3. The ST segment "the extra green mile" was read at least seven times, four times during an orientation phase between seconds 210 and 240 and then again three times during translation drafting.
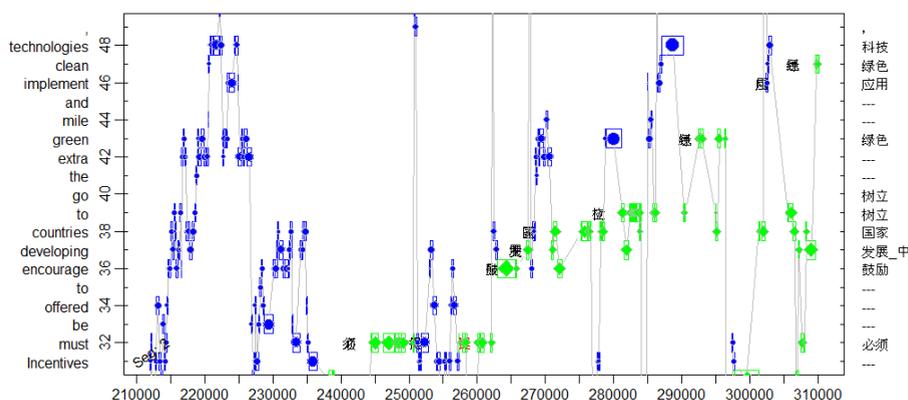
Figure 4: Progression graph with complex patterns of monitoring behavior

Table 2: Measures of challenged translation processes

SToken	FFDur	FixS	TrtS	FixT	TrtT	Ins	Del	Munit	HTra
the	183	2	316	0	0	0	0	0	1.35
extra	234	20	5133	0	0	0	0	0	1.68
green	267	24	8995	13	2683	2	0	1	0.46
mile	250	6	2515	0	0	0	0	0	1.96

Table 2 lists the behavioural measures for the translation segment which was more challenging. In total, there were 20 and 24 fixations on the words "extra" and "green", respectively. The first fixation durations *FFDur* make up less than 5% and 3%, respectively, of the total reading time for these words, indicating that most of the translation effort is related to later processes, such as source text integration or formulation of a translation hypothesis. In this example it seems that much effort was required to understand and/or formulate a first translation hypothesis for the phrase "extra green mile" since most of the reading occurred before the translation was typed.

The words "the", "extra" and "miles" remain untranslated (*Munit = 0*); only a Chinese translation of "green" was produced and aligned.[3] The relatively higher

---

[3]Dashes '—' on the right Y-axes in the translation progression graphs indicate non-translated or non-aligned words for which there are no correspondances in the translation.

HTra values indicate that translators have produced more different solutions for these words than for the translation of "green".

The available eye movement measures seem to be well suited to capture unchallenged translation processes. However, existing measures are not well suited to describe the more complex reading patterns occurring during the later stages of challenged translation, because they either describe early processes (e.g. first fixation duration or first pass reading time) or they do not capture the time course of the late processes and very few measures describe the interrelationship between reading and writing activities (see also Schaeffer & Carl 2017).

Table 3 summarizes some of the existing measures. They will be explored in more detail in §4. A distinction is made between reading measures which capture gaze activities, writing measures which describe typing processes and R&W measures which describe how reading and writing are coordinated. These measures refer to sequences of the source and/or target texts, so-called Areas of Interest (AOI). AOIs are typically single words, phrases or sentences, and are characterized by the accumulated UAD as well as their linguistic and other annotations. The first pass reading time, for instance, is the sum of fixation durations on a word (or another predefined text segment) from the first fixation before the eyes leave the AOI again. The word production time (*Dur*, cf. page 96) is the total time needed to type a word (i.e. a translation), including all its possible revisions. R&W measures shown in Table 3 can take values which may indicate early or late processes.

Table 3: Measures of the translation process

	reading measures	writing measures	R&W measures
earlier processes	first-fixation duration	keystrokes: insertions & deletions	eye-key span
⋮	first pass reading time	inter-key pauses	⋮
⋮	regression path duration	micro units	parallel R&W activities
later processes	total reading time	revisions, word production time	

2.4 A noisy translation processes model

A translation process model taking into account hidden states of early and late processes is depicted in Figure 5. The model resembles a Hidden Makov Model (HMM) in Figure 1 which consists of four levels of description. In the center is the *Translator* who is constrained by a number of factors and who produces a sequence of behavioral patterns which lead to the final translation product.

The *Predictors* are a vast number of variables which are likely to play a role in the translation process and which originate from an enormously heterogeneous field, including cognitive, linguistic, cross-linguistic, or textual, communicative and socio-cultural domains (Toury 2004). Other researchers (e.g. Risku 2014) also mention environmental conditions, including physical, geographic, economic, political and demographic aspects which might play a role in the translation process. The *Source Text* is another crucial predictor which will determine the characteristics of the target text.
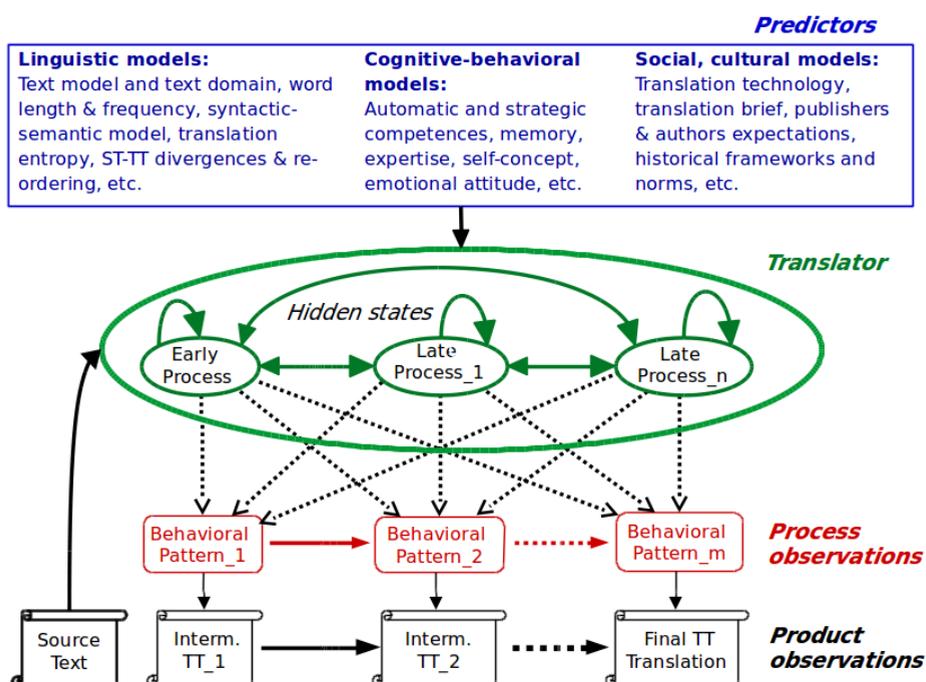

Figure 5: Observations, predictors and hidden variables in the noisy translation process model

The *Translator* is modelled as a network of hidden states which implement the actual translation processes. In contrast to earlier hierarchical-stratificational models of translation (Nida 1964; Seleskovitch 1975), it is now generally accepted that there are states of early, automatic processes and states of more deliberate, strategic processing. Hönig (1991), for instance, proposes a translation model which establishes a distinction between uncontrolled, associative translation competence (i.e. unconscious early translation processes) and a controlled workspace in which micro and macro strategies are stored. The associative translation competence corresponds to subliminal priming mechanisms, while the monitor processes occur at a later stage as they require extensive conscious effort.

The output of the model in Figure 5 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 direct consequences of translators' activities which can be traced through logging technology.

Objective UAD such as keystrokes, mouse clicks, eye movements and other behavioral data can be recorded with keyloggers, eye-trackers and other tools, but the collected UAD needs be segmented into meaningful *behavioral patterns*. However, it is neither obvious how keystrokes and gaze data should be segmented, nor is it uncontroversial what the latent states are which emit those patterns.

The HMM in Figure 5 suggests that:

- the *Translator* can be in only one state at any given time

- translation processes are driven by a large number of *Predictor* variables

- there are probabilistic transitions between successive hidden states

- each state emits exactly one *behavioral pattern* at each time

- a *behavioral pattern* produces a deterministic modification in the *interim translation*

In §3 we will be concerned with the description and analysis of the behavioral patterns. The stream of UAD can be fragmented into segments of behavioral units which are suited to describe the translation process. A Production Unit (PU), for example, is a coherent sequence of keystrokes where the lapse of time between successive keystrokes is below a given threshold, e.g., 1 sec. A PU can

thus contain a single or a large number of keystrokes irrespectively how many words are produced.

In §4 we argue that different hidden states can be related to different temporal aspects in which they are triggered. We discuss properties of earlier and later translation activities and hidden states in more detail.

3 Patterns of translational behaviour

This section discusses several approaches to fragment the UAD into sequences of behavioral patterns. Such patterns fragment the stream of translation activities on a temporal scale. We discuss units which capture gazing and typing data. In contrast to the behavioral data accumulated in textual AOIs, the UAD within the behavioral patterns may relate to several different textual items that may be at distant locations from each other. While there is a large repository of linguistic terminology to describe textual elements in AOIs — such as PoS tags, linguistic functions etc. — there are only very few approaches which fragment the translation process data and little work has beed done to describe these units.

3.1 Production units (PUs)

Carl et al. (2016); Carl & Kay (2011) define Production Units (PUs) as sequences of coherent keystrokes, where the pause between any two successive keystrokes is less than 1 second. A pause of more than 1000ms constitutes a PU boundary. PUs fragment the stream of translator activity data into sequences of coherent typing and pauses that separate them. In contrast to a micro unit (see page 103), a PU may stretch over several words, while a micro unit is defined as the flow of continuous typing that contributes to the production of one target word. A PU that stretches over m words would thus be split into m micro unit, where each produced word $1\ldots_m$ is assigned its share of keystrokes. A word can be associated with several micro and production units, depending on how often it has been revised. A PU contains, among other things, the following information:

- duration of the unit

- duration of the preceding pause

- number of insertions and deletions,

- tokens involved in the source text and target text

- average *Cross* values

- percentage of parallel source and target text reading activity during unit production

- degree of linear editing

Singla et al. (2014) investigate to what extent post-editor profiles can be identified based on the information contained in PUs. They use data from five post-editors producing together 120 translations sessions which is contained in the LS14 study[4]. They test several machine learning techniques but find that "multilayer perceptron" and "classification via regression" perform best for this task. Using 10-fold cross validation for classification, they achieve 46.48% accuracy to identify post-editors which exceeds by far the baseline accuracy of 20% which is based on guessing a post-editor by equal chance among the five participants.

Aziz et al. (2014) analyze PUs of post-edited texts, to investigate whether and how the properties of PUs are related to features of the sentences they appear in. Their investigation uses the CFT13 dataset[5] that was generated with the CAS-MACAT workbench (Alabau et al. 2013). PUs contain post-editing information about number of insertions, deletions post-editing time etc. Aziz et al. (2014) add further information to generate high dimensional feature spaces with nearly 100 features. The additional information included POS tags, named entities, chunk labels, and labels of semantic roles. The information was separated into PU level features and sentence level features "such as the number of tokens in the sentence, the number of different phrases or the number of predicates and their arguments, which could indicate that the overall sentence is complex" (179).

The authors use Principal Component Analysis (PCA) to visualize the high dimensional feature space and provide a detailed analysis of the data. Aziz et al. (2014) find that a correlation between sentence length and post-editing time can be observed mainly in cases of low post-editing activities. They find, for instance, that "PUs involving verbs are slightly more time-consuming, while PUs related to nouns require slightly more typing" (Aziz et al. 2014: 189). On the one hand, it is therefore "possible to decouple sentence length from the difficulty of each PU in terms of how time-consuming and how many edits (character level insertions and deletions) it requires." (Aziz et al. 2014: 189) On the other hand, a pause analysis becomes difficult, since "the pause prior to editing correlates very poorly to the character-level edits performed." (Aziz et al. 2014: 187) It is unclear

---

[4]The data can be downloaded from http://sourceforge.net/p/tprdb/svn/HEAD/tree/LS14/

[5]The data can be downloaded from http://sourceforge.net/p/tprdb/svn/HEAD/tree/CFT13/

why a pause occurs and whether it is related to the successive typing events, as manifested in the PUs. This is supported by their finding that "HTER correlates better with time and typing related to individual PUs than to cumulative sentence level indicators" (Aziz et al. 2014: 187)

With respect to editing times, the authors find the following relations:

- there is a stronger correlation between insertions and duration than between deletions and duration

- modal verbs, adverbs and coordinating conjunctions are more time consuming than gerunds and other non-finite verbs

- pronouns take longer to post-edit than sequences of nouns and named-entities

- consecutive NPs have a strong correlation with editing duration and the preceding pause

- the number of arguments in a sentence has an impact on its post-editing duration

(Aziz et al. 2014: 188) further find that "there is very little correlation between the length of a sentence and how time-consuming individual PUs are". In other words, post-editors process sentences in smaller units so that the post-editing duration does not necessarily depend on properties of the whole sentence, and hence sub-sentence features may provide more informative cues about actual editing effort than, for instance, sentence length. It is unclear whether and to what extent the findings for post-editing carry over to from-scratch translation.

Schaeffer & Carl (2016) attempt to predict concurrent ST reading and TT typing during from-scratch translation production. Their investigation is based on the assumption that "instances of concurrent reading and writing during translation are indicative of automatic processes and shared representations". They investigate concurrent activities in PUs using several possible predictor variables. They find that:

- the longer the PUs the more likely is concurrent reading activity

- less concurrent reading is observed towards the end of the text

- similarity of syntax in the ST and the TT facilitates concurrent activities

- more experienced translators are more likely to show concurrent R&W activities

The strong impact of the syntactic similarity on concurrent processing underpins their initial hypothesis that processes are likely to be more automatic when the ST and TT word order is similar, as in this case primed, shared syntactic representations may more easily serve as the basis for TT production.

3.2 Attention units (AUs)

In order to compare the cognitive flexibility and processing automaticity of professional and student translators, Hvelplund Jensen (2016) suggests to segment the translation activity data into attention units (AU).

Following Baddeley (2007), Hvelplund Jensen (2016) argues that, on the one hand, cognitive flexibility is linked to planning, problem solving and decision making and involves the ability to focus and switch attention, or to divide attention simultaneously into several subtasks. Hvelplund further states that a translator with good cognitive flexibility will "focus attention for precisely as long or short a period of time as is necessary only to those subtasks which are relevant to the successful execution of the translation task" (Hvelplund Jensen 2016: 153).

On the other hand, based on TAP studies (e.g. Jääskeläinen & Tirkkonen-Condit 1991) it has been suggested that professional translators rely more on automatic processing than students. Translators' automaticity is, thus, closely related to experience.

In order to assess these hypotheses on the basis of translators' UAD, Hvelplund Jensen (2016: 157) operationalizes the notion of *attention unit* (AU) in the following way:

> an AU is defined as uninterrupted processing activity allocated either to the ST (ST gaze activity), the TT (TT gaze activity and/or typing activity) or to the ST while typing (ST gaze activity and concurrent typing). Transitions to and from an AU indicate shifts in processing activity, and the point in time at which the transition occurs is used to identify the end of one AU and the beginning of the next AU.

He thus defines five AUs based on the following activities:

AU1: ST reading

AU2: ST reading + typing

AU3: TT reading

AU4: TT reading + typing

AU5: Typing

While the cognitive flexibility is computed based on the durations of the AUs, the automaticity of the process is reflected in the pupil size where smaller pupil sizes indicate relatively less cognitive load than larger ones. The pupil size for a AU was calculated as an average of all its gaze samples and a latency effect of 120 ms was factored into the calculation.

Based on an evaluation of the KTHJ08 data as shown in Table 4 (see page 97), Hvelplund Jensen (2016) finds that:

1. experienced translators spend more time on target text than less experienced translators.

2. a higher variability in AU duration by professional translators as compared to student translator indicating more flexibility and adaptability for the former group.

3. pupils are significantly larger for less experienced translators than for experienced translators.

Further, in order to assess the translation process flow, Hvelplund counts all the transitions between any two successive AU labels, separately for professional and student translators, and stores them in a 5×5 transition matrix. He compares the two matrixes and observes that experienced translators shift from AU1 (ST reading) in 65.5% of the cases to typing activity (either of AU2, AU4 or AU5) while student translators do this only in 52.2% of the cases. Student translators switch to AU3 more often than professionals, which suggests that students aim more often at confirming meaning hypotheses (reflecting some kind of uncertainty), rather than allocating the cognitive resources directly to TT typing once a meaning hypothesis has been established.

3.3 Actvity units

Not unlike Hvelplund Jensen (2016) AUs, Carl et al. (2016) suggest to fragment the activity data into seven different types of segments with the following labels:

Type1: Reading the source text (ST)

Type2: Reading the target text (TT)

Type4: Typing activity

Type5: Typing while reading ST

Type6: Typing while reading TT

Type7: Typing while reading ST and TT[6]

Type8: No activity recorded

Type1, Type2 and Type4 are basic translation activities. Type5 to Type7 take into account that source and target text reading can occur concurrently with typing, and a Type 8 is assigned to segments if no activity is logged for longer than a given threshold. Figure 6 shows the segmentation of a translation segment into AU units. The data is identical to that in Figure 3 5 but AU boundaries are marked.

Figure 6 shows a long ST reading activity (Type1, in blue) of approximately 30 seconds, between seconds 208 and 238, followed by a number of shorter pauses (Type8, in black), TT reading (Type2, in green) and typing activities (Type4, in pink) etc. These segments describe exhaustively the translation process and the properties of the sequence might be significant for certain types of translators and/or translation strategies.

In order to assess to what extent the profiles of machine-translation post-editors can be detected from the labels of the AU, Singla et al. (2014) investigate units of Type4 and Type8 (i.e. typing and pauses) of five post-editors with different amounts of experience. They use data from 120 translations sessions which are extracted from the LS14 study[7] and subdivide Type 4 and Type 8 units into five categories based on their durations. They compute a trigram language model of activity sequences for each post-editor and compute a transition matrix which is filled with the perplexity scores of each post-editor's language model on the

---

[6]In more recent work, this type of unit is decomposed into units of type 1,2,4,5,6 or 8

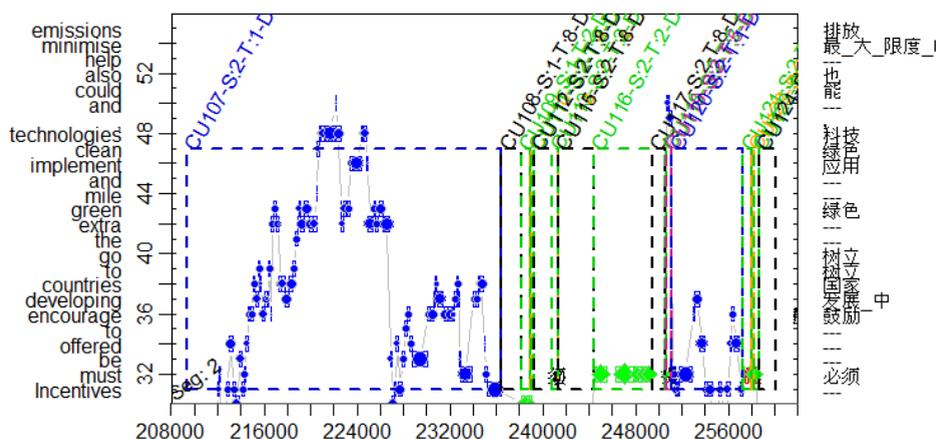[7]The data can be downloaded from http://sourceforge.net/p/tprdb/svn/HEAD/tree/LS14/

Figure 6: Translation session fragmented into CU units

other post-editor's activity sequences. A discriminative classifier is then used to cluster post-editors into two classes on the assumption that the events that make up the translation process provide enough information for the individualization of post-editor profiles.

Singla et al. (2014: 56) find that "experienced post-editors produce similar kinds of activity sequences in contrast with the activity sequences of inexperienced post-editors". They also notice that post-editors with a similarly negative attitude towards post-editing tend to have similar activity patterns.

Martínez-Gòmez, Aizawa, et al. (2014) use a subset of 204 sessions from the data shown in Table 4 that is annotated with information about translator experience and certification: 99 of the 204 sessions were produced by 47 non-certified translators, and 105 sessions were produced by 47 certified translators. They report that:

> 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. The most common translation activity was the concurrent combination of "source text reading", "target text reading" and "target text typing", which occurred around 45% of the time for non-certified translators and 65% of the time for certified translators. (Martínez-Gòmez, Aizawa, et al. 2014: n.p.)

In an extension of this experiment, Martínez-Gòmez, Minocha, et al. (2014) use the same data to recognize translator expertise, based on the assumption that "Translators have different perceptual and motor activities, depending on their level of expertise." They compare two methods to assess this hypothesis, one based on the AUs, and another one using unsupervised machine learning techniques with a view to discover regularities in the logging events and to reveal latent activities, that would otherwise not be detected.

For the unsupervised learning method, each log event (fixation and keystroke) was enriched with 31 additional features that were extracted from the immediate context, such as the number of insertions, deletion and fixations within the past and the next 10 events, together with the time offsets from the current event. The information was stored in the form of vectors which were then classified using a k-means clustering method (3 to 8 classes). Tri-gram language models were built from the sequences of cluster labels, and random forests used to predict translator expertise, such as whether the user is a certified translator or not (binary classification), his/her years of training (regression) and years of experience (regression).

Martínez-Gòmez, Minocha, et al. (2014) report an error reduction in the recognition of certified translators, and moderate but significant error reductions in the recognition of years of experience, as compared to a baseline. Best results were obtained with the unsupervised technique. They also report that CU unit of type 5 (i.e concurrent ST reading and typing) is more likely for certified translators than for non-certified translators.

3.4 OST units

Another approach to fragmenting the process data was suggested by Nitzke & Oster (2016). They manually annotate the activity data into two main categories, orientation (O) and revision (R) with five sub-categories:

- *Ost*: The participant spends time reading both source and target text

- *Os*: More than 80 % of the fixations were on the source text

- *Ot*: More than 80 % of the fixations were on the target text

- *Rl*: Every word or phrase is processed only once.

- *Rs*: The participant works on a part of the text, moves on but jumps back later to readjust the parts she already worked on.
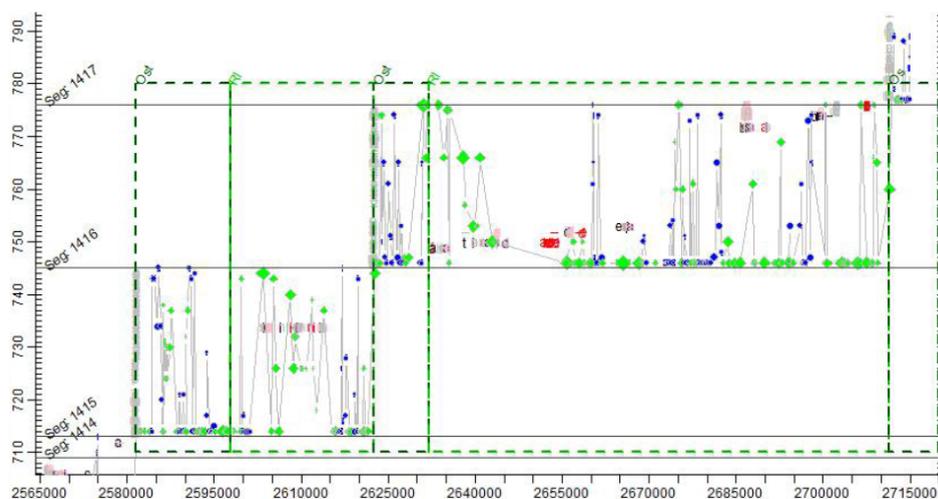
Figure 7: Annotation of OST units

With five sub-classes the annotation schema is less complex than the CU activity units - and much coarser grained. The translator activity data of 406 segments has been manually annotated into 985 segments, which is on average slightly more than two OST units per segment.

In an attempt to automatically detect OST units, Läubli & Germann (2016) segment the process data into fragments of 3 seconds, and assemble all process events (keystrokes, mouse clicks, ST fixations and TT fixations) for each segment in a vector of observations. Similar to the method used by Martínez-Gòmez, Aizawa, et al. (2014), the observation vectors are then classified with a k-means clustering method and a Hidden Markov model (HMM) is trained on the sequences of cluster labels and observation vectors. The assumption is that the cluster labels represent the underlying states of the OST annotation (orientation, revision, and, as an additional state, also pausing) where each state produces randomly an observation. The transition probabilities in the HMM and observation probability densities are then trained based on the available data. The aim of the model is to yield the most probable label for each observation, taking into account (i) the feature values (dimensions) of the current observation and (ii) the label assigned to the preceding observations. In a final step the cluster labels are mapped on the three OST labels: orientation, revision and pause. The authors show that the system reaches as high an accuracy to predict the times spent on orientation, revision and pause as some of the human annotators.

3.5 Conclusion

This section summarizes different methods of segmenting the UAD into successive chunks. Depending on the available logging data, most of the segmentation methods make use of cues in the data, such as text production pauses and/or the location of the gaze data on the source or target text to define segment boundaries. An exception is the segmentation method by Läubli & Germann (2016) who segment the UAD into chunks of 3 seconds duration. With the exception of OST units described in §3.4, all segmentation methods work fully automatically.

The reported investigations show that some segment properties are typical for different translator profiles and degrees of translator expertise. They are are also indicative of various translation problems.

The research discussed in this section can be characterized as instances of probabilistic translation modelling as discussed in §2.1 and equation 4.5 on page 76. Models are sought which predict behavioral patterns based of a number of different predictor variables. Linguistic features of the source text are investigated with respect to their effect on production times and revision behavior, patterns of reading and writing are related to cognitive models of the translator, such as translation expertise, and different translation techniques, machine translation and from-scratch translation are assessed in relation to translation effort.

Pause analysis is perhaps the most common approach for the analysis of behavioral patterns. In pause analysis it is assumed that longer pauses between successive keystrokes signal higher cognitive effort. O'Brien (2006) analyses keystroke pauses in post-editing and suggests that analyzing pauses is a useful indicator of cognitive effort in post-editing. Immonen (2006) finds that in translation, pause length is higher at word and clause boundaries. Lacruz et al. (2012) introduce average pause ratio as a metric to establish a relationship between pauses and cognitive effort in post-editing.

However, to obtain a more complete picture of the translation process, we ought to investigate the translators' "black box" in more detail. In the next section we will therefore investigate properties of the translators' hidden states, which, according to the Noisy Channel Model in §2.4 emit behavioral patterns.

4 Hidden translation states

It is unclear how many hidden translation states can or should be distinguished that participate in the translation process. However, a distinction can be made between states which are triggered through early priming mechanisms and other

more time-consuming and late(r) states which involve more cognitively demanding problem solving strategies. The TP model suggested in Figure 5 distinguishes these states into "early" and "late" states.

Priming is an unconscious mechanism that is based on the implicit memory of a first (source) stimulus which carries over to a subsequent, target stimulus and which has an impact of the execution of a following task. It has been shown that bilinguals, and therefore also human translators, use implicit memories during language production. Priming effects exist between stimuli in different modalities, such as visual and verbal. They are, however, stronger if source and target stimuli are in the same modality, e.g. within written language. Priming effects can be observed in translation and in post-editing of machine translation output (PEMT), but — as we will show — the effects are more noticeable in PEMT, presumably due to the fact that priming effects are generally stronger within one language (i.e. the MT output and final translation) than between two languages (Pickering & Ferreira 2008).

The degree of similarity between source and target items has an effect on the strength of the priming effect – the greater the similarity, the stronger the priming effect. Priming facilitates and simplifies translation. Priming effects exist for the choice of words as well as for word order. Hvelplund Jensen (2009) and Ruiz et al. (2008) report shorter ST reading times in translation if the word order in the ST is identical with the word order in the TT. Schaeffer, Dragsted, et al. (2016) report longer reading times for words with more possible choices than for words with fewer choices. This result is in accordance with Campbell's Choice Network Analysis (Campbell 2000): The more choices translators have in the selection of a translation, and the more complex the decisions are that they have to make, the more difficult the translation will be. Simpler translational decisions often lead to identical results while more variation in the traslation often implies difficult more difficult decisions.

As shown in Figures 3 and 4, translation process data encodes traces of early, automatized and later translation behavior. Automatised processes occur quickly and leave their traces early on, while later, more time-consuming processes are likely to involve more conscious problem solving activities.

A noisy channel model of translation as depicted in Figure 5 takes into account various kinds of hidden processes which ought to explain and generate the traces in the observed UAD. This section summarizes a few constraints of the hidden states, related to the observable output of early and later processes.

A large amount of research exists that investigates conscious processes in translation (e.g. Jääskeläinen & Tirkkonen-Condit 1991; Lörscher 2005). Accord-

ing to Gutt (1989), the translator's task is to recode the source text into a target language text in such a way that interpretive resemblance in regard to explicatures and implicatures of both texts is achieved. In order to examine interpretive resemblance, translators consciously apply meta-cognitive monitoring processes. Gutt's theory builds on relevance theory (RT, Sperber & Wilson 1995), which posits that linguistic forms encode semantic representations that are recovered using unconscious, automatic decoding processes. As a pragmatic theory of communication, RT seeks to explain the inference procedures that build on the automatic encode-decode mechanism and on which successful communication relies. The distinction between the process of encoding-decoding messages and the process of making inferences from evidence coincides for Blakemore (2002) with the distinction between semantics and pragmatics: linguistic semantics provides logical forms which are taken as input by pragmatic inferences constrained by the principle of relevance.

This section aims at giving empirical evidence for the existence of early and late translation processes. In §4.1 we describe the experimental material that much of the successive sections rely on. In §4.2, we investigate linguistic parameters that have an effect on the word production duration. Production duration is a possible indicator of translation difficulty and the amount of priming and more time-consuming translation strategies that went into the production of a translation. We show that, among other parameters, the number of possible translations for a word is a strong indicator of translation difficulty, which has an impact on early as well as late translation processes. In §4.3 we have a closer look at syntactic properties of lexical variation in the translation product in from-scratch translation and in post-editing. Finally in §4.4 we discuss revision behavior, which accounts probably for the latest of the translation processes.

4.1 Experimental Material, Measures and Metrics

Table 4 gives an overview of the size and number of texts. A total of 336 target texts (TTs) with a total of 48.295 target language tokens (TT Tok) were produced from six different English source texts (ST) into four target languages, Danish (da), German (de), Spanish (es) and English (en). The English TTs resulted from a copying task (C), English to English, whereas the other texts were either post-edited (P) or translated (T). The translations were produced by 95 translators over a period of 38 hours (*FDur*). The column *KDur* shows the accumulated keying time, excluding production pauses of more than one second. Note that the ratio of keying time (*KDur*) vs. total production time (*FDur*) is much smaller for

Table 4: Annotated data for the syntactical entropy study

Study	Sessions	TL	*Task*	*Texts*	*Part*	*FDur*	*KDur*	*TT Tok*
TDA	48	en	C	1-6	15	6.1	5.8	6792
KTHJ08	69	da	T	1–3	24	6.4	5.5	10571
SG12	45	de	P	1–6	23	5.6	1.9	6352
SG12	47	de	T	1–6	24	9.4	4.6	6632
BML12	64	es	P	1–6	32	2.3	0.9	9012
BML12	63	es	T	1–6	32	8.2	5.8	8936
total	336	4	3	6	150	38	24.5	48295

post-editing than for from-scratch translation, and even less in the copying task. Danish translations were only produced for three texts (1-3). The column *Part* indicates the number of different participants involved in each translation study.

From the logs of these sessions, a number of features were extracted, (cf. Carl et al. 2016), among others:

- *LenS*: length of the English source text word in characters

- *LenT*: length of the translation in characters

- *STseg*: (sequential) number of the source text segment

- *Prob1*: frequency of the English source text word (according to BNC)

- *PoS*: English source texts were part of speech (PoS) tagged. Table 5 gives an overview of the used tagset.

- *Dur*: translation duration is the amount of time needed to produce the translation of a word.

- *HTra*: word translation entropy

- *Cross*: distance between the English source text word and its translation

In order to assess the literality of a translation, Carl et al. (2016); Carl & Schaeffer (2016) introduce a literality metric which measures the similarity of a source text (ST) and its translation, the target text (TT), along the following three criteria:

- ST and TT segments have the identical word order

- ST and TT words are one-to-one translation equivalences

- ST words have one (preferred) translation in the context

Literality criterion (2) is met if each word in the ST corresponds to exactly one TT word and vice versa, while criterion (1) is realized if the translation equivalents occur in the same order in the ST and in the TT. These two criteria are represented by an integer value, referred to as *Cross* and relate to the amount of crossing word alignments (inter-lingual alignment distortion). A one-to-one correspondences results in a *Cross* value of 1, and this value grows (negatively or positively) with the distance between the aligned words. Approximately 40% of all words in the TPR-DB (English STs) have a *Cross* value greater or smaller than 1 (Schaeffer, Dragsted, et al. 2016: 190).

In order to assess literality criterion (3) we use a corpus of word-aligned, alternative translations and measure the entropy of the translation realizations. This measure is referred to as word translation entropy *HTra*. Approximately 90% of all words in the TPR-DB (English STs) have more than one translation alternative, and thus a value *HTra* > 0 (Schaeffer, Dragsted, et al. 2016: 190).

4.2 Production duration in translation and post-editing

The reduction of translation duration (the increase of productivity) is a driving force for much of the technological development of machine translation (MT) and for post-editing of machine translation (PEMT). While it has been shown in several places that PEMT is often quicker than from-scratch translation (Plitt & Masselot 2010; O'Brien et al. 2014), it has not often been investigated what the possible determining factors, and what the impact for on the translation product are.

To test which properties of the text might have an impact on the translation duration, we analysed six English texts from studies BML12 and SG12, (see Table 4). We extracted 15,313 ST and 15,568 TT words that were translated into Spanish and German by 32 and 24 translators respectively and they were also post-edited into Spanish and German by 32 and 23 translators, respectively. The

data was analyzed using R (R Development Core Team, 2014) and the lme4 (Bates et al. 2014) and languageR (Baayen 2013) packages to perform linear mixed-effects models (LMEMs). To test for significance, the R package lmerTest (Kuznetsova, Christensen, & Brockhoff, 2014) was used, which implements ANOVA for mixed-effects models using the Satterthwaite approximation to estimate degrees of freedom. The final model included participant, ST token, text and target language as random variables. The predictor variables were ST token frequency (*Prob1*), word length of the ST token in characters (*LenS*), *Cross* and *HTra*, in addition to task, i.e. post-editing and translation, as an interaction with both *Cross* and *HTra*. The dependent variable, production time per word (*Dur*) was log transformed, because it was not normally distributed. Data points which were more than 2.5 standard deviations below or above a participant's mean were excluded (3%). All effects were highly significant (all $t > 3$ and all $p < .001$).

The translation duration *Dur* indicates the production time for a translation. It is also an indicator of earlier and later processes: the more time is needed to produce a translation, the more likely will the translator be involved meta-cognitive reasoning. As shown in Figure 8 the production duration dependes on a number of additional factors.

Figure 8 shows the effect of *Cross* and word translation entropy (*HTra*) on word production time *Dur*. The Figure shows that post-editing is much quicker for words which have small *Cross* and *HTra* values. Post-editing may take as long as from-scratch translation if the MT output is modified (i.e. many different variants are produced) and/or for large *Cross* values.

It is possible that MT systems produde more acceptable translations for segments in which the word order is similar (i.e. *Cross* values are low) than for segments in which a large amount of syntactic reordering is required. In turn post-editors would need to produce less modifications for translations with low *Cross* values which would explain why post-editors take less time to produce these words as compared with words which have a very different position in the TT, in relation to the ST.

4.3 Variation in translation and post-editing

The amount of different translations that are possible for a word (*HTra*) has a strong effect on production time. In this section we investigate this phenomenon in more detail. Post-editors seem to be less creative than translators; often, they do not modify the MT output which leads to fewer variants in the translation product than during from-scratch translation.
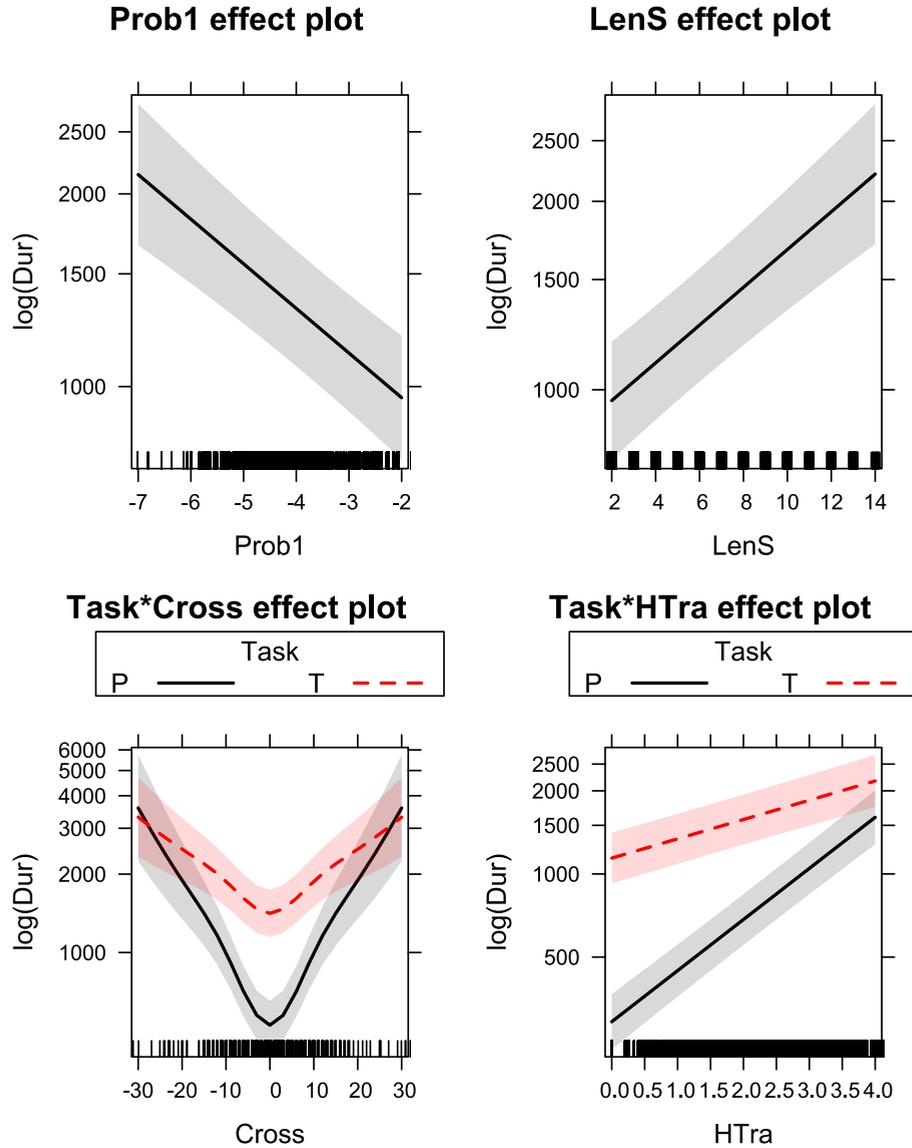
Figure 8: The effect of source text frequency (*Prob1*), ST word length in characters (*LenS*), word translation entropy (*HTra*) and word order differences (*Cross*) on ggproduction time (*Dur*) and observed translation variants for post-editing (P) and translation (T).

Čulo et al. (2014), for instance, describe a study in which 12 professional translators and 12 translation students translate or post-edit six texts from English (L2) to German (L1). Čulo et al. (2014) discuss the following MT output in detail:

EN : *In a gesture* sure to rattle the Chinese Government

DE : *In einer Geste*, die die chinesische Regierung wachrüttelt

The German translation "*In einer Geste*" is understandable but not idiomatic. It is a literal one-to-one translation – according to criteria 1 and 2 above – which was generated by an MT system but which was rarely changed by the post-editors. However, a great variation of different idiomatic versions was found in human translations of the same text segment. Human from-scratch translations for the above example include: "Als Geste", "Es ist eine Geste", "Mit der Absicht", "Als Zeichen des Widerstandes" and "Mit einer Aktion". The eight translators who translated this text produced seven different versions, while seven post-editors only came up with three different versions. The example clearly shows that translators are more creative, resulting in more diverse translation solutions and thus high *HTra* values, while post-editors are more heavily primed by and biased towards the solutions generated by the MT system which results in low *HTra* values but also faster production times. Note also that the translation *In a gesture ↔ In einer Geste* can be aligned word-by-word, which is not the case for most of the from-scratch translations.

Tightly connected to the phenomenon of interference is the amount of variation in translation solutions. Figure 9 shows word translation perplexity from English to German and English to Spanish for different word classes (PoS tags, see Table 5, below). The texts were extracted from the SG12 and BML12 studies (see Table 4) and contain approximately 800 source text words. The degree of translation variance can be measured as perplexity: an even distribution of several realised translations (e.g. all translators generate a different translation) leads to high perplexity values, while an uneven distribution (i.e. many translators generate the same translation) does not.

The values for post-editing and original translation are indicated. Some PoS tags, such as e.g. JJS (superlative e.g. "largest", "least"), NNP (proper names), CC (conjunctions) only produce a very small number of translation alternatives (low degree of perplexity). Other PoS tags, such as e.g. RP (particles), VBN (participles) exhibit more variation in the target text. In any case, the degree of word translation perplexity in post-editing is always lower than in translation from scratch. As pointed out in §4.3, this is presumably due to the fact that MT output
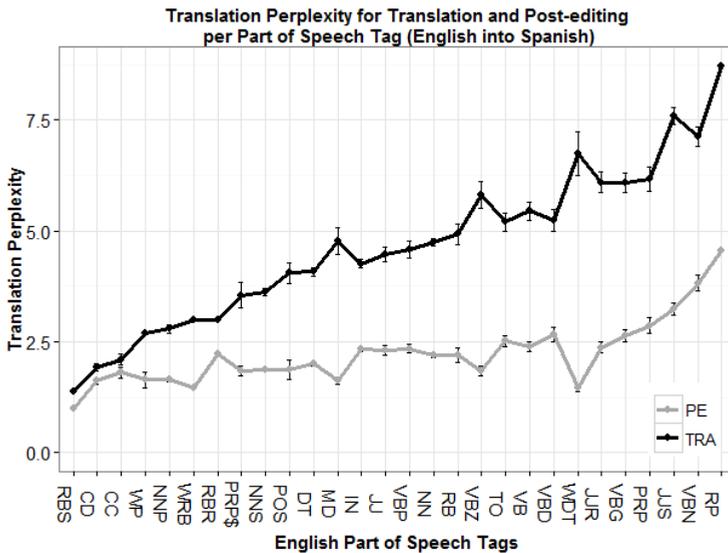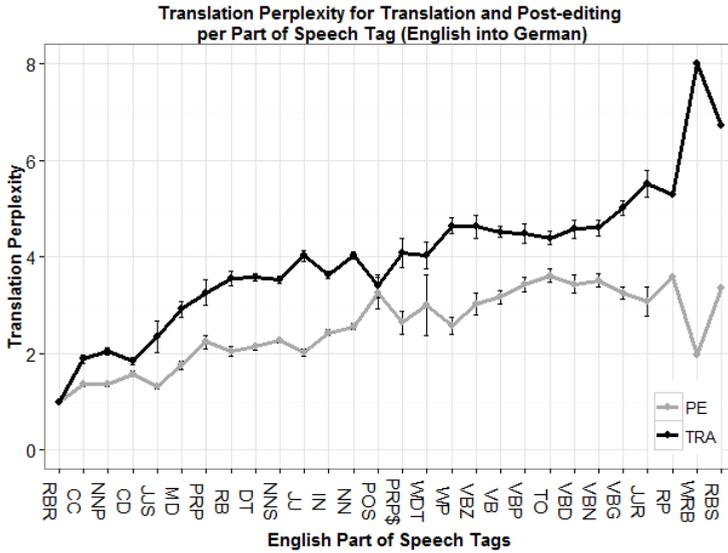
Figure 9: Priming in translation (TRA) and post-editing (PE). Perplexity in word translations exhibits the variation of generated target texts, which is always higher in translation from scratch than in post-editing.x

is often accepted without changes and all post-editors therefore often accept the same word translations.

Some PoS tags in the Spanish translation exhibit more variation than the German translation. For example, there is less variation in the translation of superlatives (JJS) in German while there is a relatively large amount of variation in the Spanish translation. Other word classes (e.g. conjunctions) seem to be translated in the same way by most translators and post-editors. The difference between post-edited texts and translations from scratch, however, are more pronounced for Spanish than for German. This suggests that Spanish post-editors accept MT output more frequently than German post-editors.

4.4 Translation revision

According to Gutt (1991) the aim of a translation is to achieve appropriate contextual effects in the target language without unnecessary effort for the reader of the target text, so that the translation corresponds to the original source text in terms of relevant aspects. In order to achieve this goal, translators consciously keep track of the possible associations between stimulus, context and interpretation, so that the resulting translations obey to the principle of cognitive and communicative relevance (Sperber & Wilson 1995: 260).

Translation revision is in many cases a compulsory activity to generate intelligible and optimal relevant translations. A distinction is made between other-revision and self-revision. Other revision is carried out by someone other than the translator, while self-revision (or checking) is done by the translator him- or herself. Self-revision of a translation is an integral part in the translators' translation process. Jakobsen (2003) distinguishes between online revision, i.e. revision during the translation drafting process, and 'end revision', which occurs after the completion of the first draft without delay. According to Mossop (2007: 109), revision may be defined as "that function of professional translators in which they identify features of the draft translation that fall short of what is acceptable and make appropriate corrections and improvements". Revisions may be due to problems in transfer, content, language and presentation (Mossop 2007) and may take place in translators' minds during the decision-making process ('internal revision') or appear on paper or the computer screen when actual changes are being made ('external revision', Künzli 2007).

Relevance Theory considers words and phrases to encode procedural components that contain instructions which control procedures that limit calculations of conceptual representations. This distinction is known as conceptual and procedural encoding. Procedural encoding thus guides the conceptual computations

and leads to processes of comprehension so that the reader may work with a conceptual representation (Blakemore 2002). In accordance with these observations, Sperber & Wilson (1993: 16) note that "conceptual representations can be brought to consciousness; procedures cannot. We have direct access neither to grammatical computations nor to the inferential computations used in comprehension."

However, according to Alves (2003), translators consciously learn how to manipulate conceptually and procedurally encoded information. They suspect that conceptually encoded information is easier to translate than procedurally encoded information as conceptual encoding exhibits a "relatively stronger interpretive resemblance between source and target texts" (p 20). Sekino (2012) reports findings based on translation data for Japanese into Portuguese. Their results corroborate Alves (2003), showing that processing effort is greater when dealing with procedural encodings in both from-scratch translations and in post-editing tasks in terms of keystrokes, fixation counts and fixation duration.

In order to assess these findings with our data, we investigate ST reading patterns and TT revision patterns on a set of UAD which included that shown in Table 4. The duration of the fixations – and also of the first fixation – signals the cognitive effort for processing a word. Fixations tend to be longer on words that require effortful processing as, for instance less frequent words, words containing spelling errors, ambiguous words, words which are inappropriate in a given context, etc. McConkie & Yang (e.g. 2003: 413).

We adopt Alves & Couto-Vale (2011) notion of micro units to quantify the amount of self-revision. A micro unit (*Munit*) is a typing burst which contributes to the translation of an ST token and which does not contain inter-keystroke pauses of more than 1 second. An *Munit* — in the way we use it here — indicates how the translation of a source word was modified. The number of *Munits* that the translation of a word is involved in is thus an indicator for its translation effort, since each revolving modification is an indicator for restructuring or reconsidering the translation a larger context.

We PoS-tagged the English source texts[8] and investigated their translations into Danish, German, Spanish, Estonian, Chinese, and Hindi (i.e. the studies ACS08, BD08, BD13, BML12, HLR13, KTHJ08,MS12, NJ12 and SG12 (cf. Carl et al. 2016)) with the hypothesis that:

1. procedurally encoded words in the English source texts would require relatively more reading time

---

[8]we used the Penn treebank PoS tagset https://www.ling.upenn.edu/courses/Fall_2003/ling001/ penn_treebank_pos.html

2. their translations into the target languages would require more revision time than conceptually encoded words

To this end, we classified the English PoS tags into 2 bins, labeled *conceptually encoding* and *procedural encodings*, according to the list shown in Table 5. We assume that the word classes in the two bins are more likely to encode their respective labels than the other one. We then investigated the distribution of effort according to the two hypothesis for these two classes.

Table 5: Penn treebank PoS tags for English source texts

Conceptual encoding		Procedural encoding	
NNP	Proper noun	IN	preposition or conjunction, subordinating
VBP	verb, present, not 3rd p. sing.	DT	determiner
NNS	noun, common, plural	PRP$	pronoun, possessive
CD	numeral, cardinal	PRP	pronoun, personal
NN	noun, common, singular or mass	MD	modal auxiliary
VBD	verb, past tense	TO	to
VBN	verb, past participle	CC	conjunction, coordinating
VBG	verb, present participle or gerund	RP	particle
JJ	adjective or numeral, ordinal	WP	WH-pronoun
VB	verb, base form	POS	genitive marker
JJS	adjective, superlative	WDT	WH-determiner
RB	adverb	WRB	Wh-adverb
VBZ	verb, present tense, 3rd p. sing.		
JJR	adjective, comparative		
RBR	adverb, comparative		
RBS	adverb, superlative		

The data for the dependent variable total reading time of the ST token (*TrtS*) was analyzed in the same way as the data for the dependent variable *Dur* described in §4.3. The dependent variable *TrtS* was log transformed because it was not normally distributed. Data points which were 2.5 standard deviations below or above a participant's mean were excluded (< 3%).

The final LMEMs had the following random variables: item, participant, text and study. The predictors were *Prob1* (ST frequency), *LenS* (word length in characters), *STseg* (sequential position of sentences in the ST), *Encoding* (see Table 5), *HTra* and *Cross*. These latter two variables implement the literality metric introduced above:

- *HTra* indicates to what extent there is a clearly preferred translation (criterion 3).

- *Cross* indicates to what extent the source and the target texts follow the same relative word order (criterion 1) and whether there is word-to-word or a phrase-to-phrase correspondence between the ST and the TT words (criterion 2).

Table 6 and Figure 10 show that translators are likely to spend more time reading conceptually encoded source text words than procedurally encoded ones. The Table shows

- Estimate: the estimated effect of the predictor variable on the dependent variable given the effect of the other predictors and the random effects.

- Std. Error: the error of the estimated effect

- t value and Pr(>|t|): the significance of the estimation. These are also given as stars (*) in the last column of the Table (three *** designate significance below the 0.001 level, two ** designate significance below the 0.01 level and one * designates significance below the 0.05 level)

Table 6: Effects of *Prob1*, *LenS*, *STseg*, *HTra*, *Cross* and kind of encoding on on total reading times of source text words.

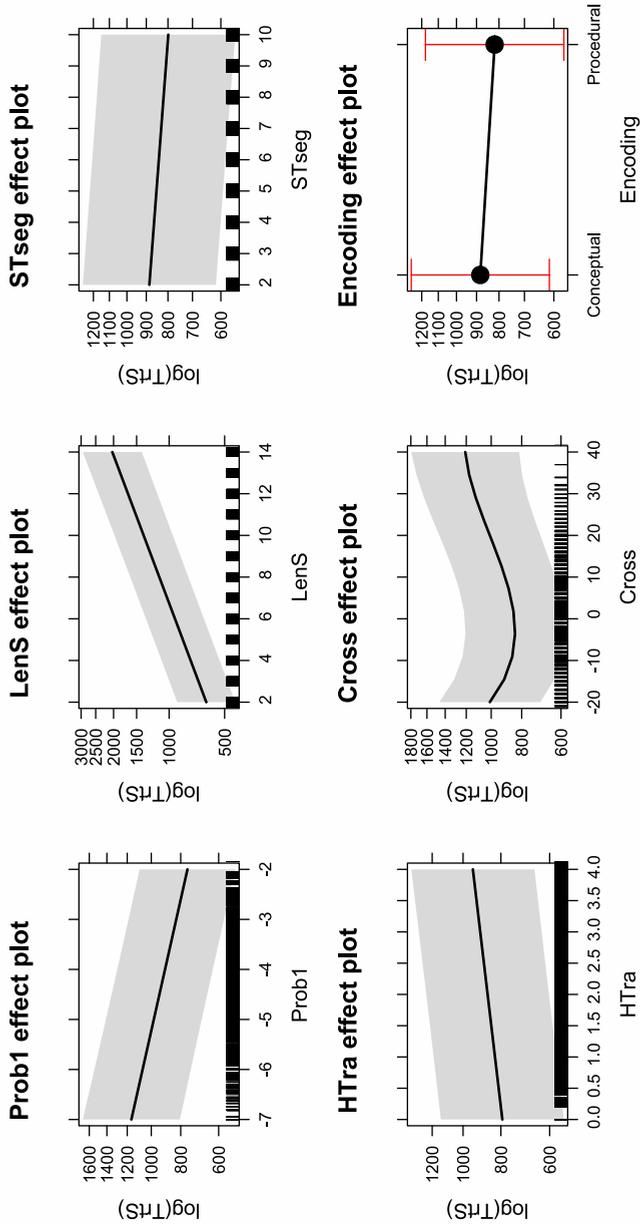|	Estimate	Std. Error	*t* value	Pr(>|t|)	
Intercept	5.94	1.89×10^{-1}	31.387	4.77×10^{-11}	***
Prob1	-8.51×10^{-2}	1.01×10^{-2}	−8.464	$< 2.00 \times 10^{-16}$	***
LenS	9.78×10^{-2}	4.14×10^{-3}	23.627	$< 2.00 \times 10^{-16}$	***
STseg	-1.29×10^{-2}	3.79×10^{-3}	−3.401	0.000681	***
HTra	4.35×10^{-2}	7.87×10^{-3}	5.525	3.55×10^{-8}	***
abs(Cross)	8.99×10^{-3}	2.13×10^{-3}	4.231	2.33×10^{-5}	***
Enc. Proc.	-7.53×10^{-2}	2.39×10^{-2}	−3.148	0.00166	**

Figure 10: Effects of *Prob1*, *LenS*, *STseg*, *HTra* and *Cross* on total reading times of ST words

Word frequency, word length, word translation entropy and relative translation distortion (i.e. *Cross*) have all a highly significant positive effect on total reading time of ST words. These findings are not surprising and have been reported elsewhere (e.g. Schaeffer & Carl 2014). Also, the facilitation effect for later (higher number) segments in the source text is well known (Schaeffer, Dragsted, et al. 2016).

The picture is different as it comes to translation revision. Table 7 and Figure 11 show that translators revise translations of procedurally encoded words more often than translations of conceptually encoded words. The dependent variable (*Munit*) indicates how often a translator revises a translation.

The analysis for the dependant variable *Munit* was carried out in the same way as previous analyses, but it was not log transformed. Data points which were 2.5 standard deviation below or above a participant's mean were excluded (< 4%). The model included the same random variables and predictors as previous analyses, with the difference that the length (in characters) of the TT word was chosen, given that this might have a more direct effect on revision than the length of the ST word. Similar to the total reading time on the ST in Figure 10, the length of the translated word, the word translation entropy and relative translation distortion (i.e. *Cross*) have all a highly significant positive effect on the number of revisions (*Munit*). This is in line with the findings that are discussed in the context of Figure 8 which show a strong effect of observed translation variants on production time.

The results of this study suggest that there is an asymmetry in the perception and in the production of conceptually and procedurally encoded information in translation. While the perception of procedurally encoded information seems to be less effortful than that of conceptually encoded information, our findings indicate the reverse relation for translation production. Taking the number of revisions as an indicator for the effort in translation production, our dataset shows that the generation of translations for procedurally encoded information is more difficult than that of that of conceptually heavy words.

5 Conclusion

Translation is an extremely challenging task that requires a translator to possess unique skills. Aside from bridging linguistic divergences between both languages, such as e.g. syntactic shifts and lexical decisions, translators must also align the author's intention with the readers' expectations while simultaneously ensuring socio-cultural interpretations of the original text in the translation. The

Table 7: Effects of *LenT*, *STseg*, *HTra* and *Cross* on translation revision

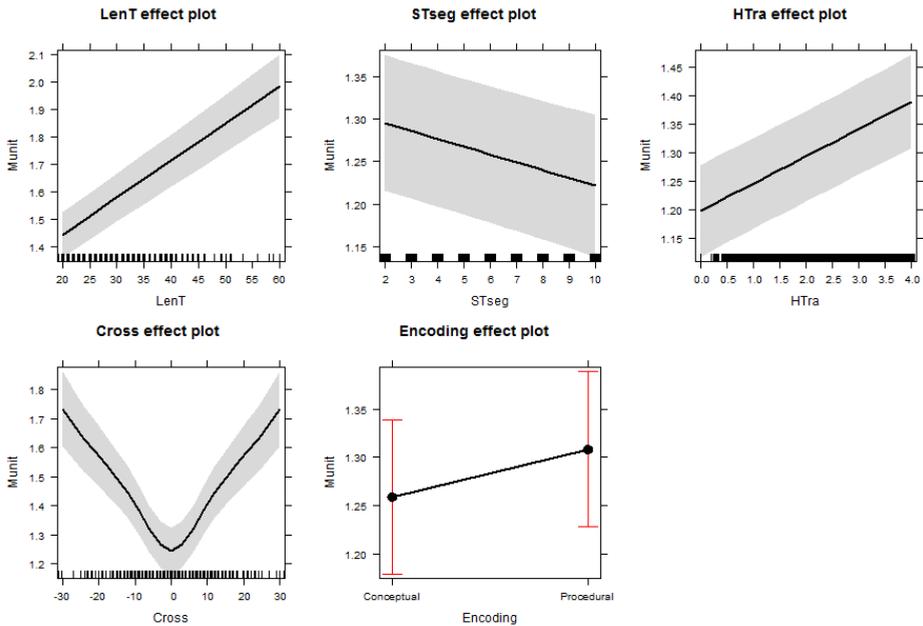|	Estimate	Std. Error	*t* value	Pr(>|t|)	
Intercept	1.09	4.23×10^{-2}	25.747	1.66×10^{-13}	***
LenT	1.35×10^{-2}	8.19×10^{-4}	16.517	$< 2.00 \times 10^{-16}$	***
STseg	-9.23×10^{-3}	2.12×10^{-3}	-4.343	1.48×10^{-5}	***
HTra	4.79×10^{-2}	4.61×10^{-3}	10.39	$< 2.00 \times 10^{-16}$	***
abs(Cross)	1.62×10^{-2}	1.84×10^{-3}	8.816	$< 2.00 \times 10^{-16}$	***
Enc. Proc.	4.93×10^{-2}	9.78×10^{-3}	5.043	4.87×10^{-7}	***

Figure 11: Effects of *LenT*, *STseg*, *HTra*, *Cross* and *Encoding* on translation revision

foundation of this activity seems to be based on unconscious memory processes: the implicit memory of source text segments primes the translator to produce a translation which is structurally and lexically similar to the target text. Subliminal priming mechanisms are the basis from which translations emerge. The first fixation is a very early behavioral measure and the word translation entropy (*HTra*) and relative translation distortion (*Cross*) have an effect on its duration such that words with small *HTra* and/or *Cross* values are easier to process than words with high *HTra* and *Cross* values.

On top of the early priming processes, translators develop a number of more consciously accessible translation strategies providing criteria to decide whether the generated translations conform to his or her expectations. This meta-linguistic knowledge is instrumental for problem-solving during the translation process. The deployment of meta-linguistic knowledge, for instance about grammatical structures or lexical translation equivalence, can be consciously directed and manipulated. For instance, repeated re-reading of a word or phrase is evidence of conscious processes. However, these processes are difficult to disentangle based on typical fixation measures such as total reading times. Each fixation on a word adds to its total reading times bt it is difficult to know which meta-linguistic strategies and problem-solving activities have been used.

Some independent variables, such as word translation entropy (*HTra*) and relative translation distortion (*Cross*) have an effect on both early and late processes, which seem to suggest that early automatized processes trigger certain later conscious ones (Schaeffer & Carl 2013). The results presented by Schaeffer, Dragsted, et al. (2016) suggest that target language-specific aspects play a role right from the beginning in the translation when reading a source text word for the first time. Words with fewer alternative translations and which do not require re-ordering in the target language require less effort than words with a higher number of alternative translations and which must be syntactically re-ordered — and this effect can be observed in early and in late measures.

The lack of appropriate late (eye movement) measures makes it difficult to assess in detail which translation strategies were deployed: a total reading time of 8 seconds, for instance, is just a conglomerate of fixation durations, but it does not tell us which translation processes were used during these 8 seconds.

The analysis of behavioral pattens is much better suited to assess translation strategies. Think Aloud Protocols (Krings 1986; Lörscher 1991; Jääskeläinen & Tirkkonen-Condit 1991) provide evidences for the existence of different translation strategies. However, the analysis of TAP data is very labor intensive and it is unclear how the identified translation strategies relate to the UAD. Alternatively,

behavioral patterns can be segmented and identified in the UAD to investigate translation strategies. Schaeffer, Carl, et al. (2016), for instance, show that translation processes are much less sequential, (sentence-by-sentence, chunk-by-chunk) and much less stratificational than predicted by earlier translation models.

To date, empirical translation process research has mainly focused on the textual product-based angle and there are some insights as to which linguistic constructions are more or less difficult to translate. However, besides some work into keystroke pause analysis (Immonen 2006; Lacruz et al. 2012), very little work is available that investigates in detail the temporal structure of the translation process and that systematically relates translation strategies to observable behavioral patterns.

In this chapter we develop a computational noisy channel model of the translation process, which can take into account a (possibly large) number of probabilistic functions that contribute to and explain the translation process. Prerequisites for the modelling of the process are measures and metrics that quantify different aspects of the observed data and that describe the various different early and late hidden translation processes in a translator's mind.

While translation process research investigates the underlying factors that lead to successive intermediate versions of a text that is to become a translation, corpus-based translation studies, including translation universal research, investigates regularities in different (final) translations, however, usually without access to the directly observable translation behavior. There are thus a number of similarities in translation product and translation process research, as both investigate the regularities in different (versions of) translations and the underlying mechanisms which may explain the observed regularities.

With the elaboration of a noisy translation process model, we hope to achieve "a scientifically sounder methodology of data collection, analysis and report" which will help in "the development of a relatively uncontroversial classification of process indicators" (Bernardini 2001: 260).

References

Alabau, Vicent, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes Garcia Martinez, Jesús González, Philipp Koehn, Luis Leiva, Bartolomé Mesa-Lao, Daniel Ortiz, Herve Saint-Amand, Germán Sanchis & Chara Tsoukala. 2013. CASMACAT: An open source workbench for advanced computer aided translation. *Prague Bulletin of Mathematical Linguistics* 100(1). 101–112.

Alves, Fabio (ed.). 2003. *Triangulating translation: Perspectives in process-oriented research*. Amsterdam: Benjamins.

Alves, Fabio & Daniel Couto-Vale. 2011. On drafting and revision in translation: A corpus linguistics oriented analysis of translation process data. *Translation: Computation, Corpora, Cognition* 1(1). 105–122.

Annoni, Jean-Marie, Hannelore Lee-Jahnke & Annegret Sturm. 2012. Neurocognitive aspects of translation. *Meta* 57(1). 96–107.

Aziz, Wilker, Maarit Koponen & Lucia Specia. 2014. Sub-sentence level analysis of machine translation post-editing effort. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard & Lucia Specia (eds.), *Post-editing of machine translation: Processes and applications*, 170–199. Newcastle: Cambridge Scholars.

Baayen, R. Harald. 2013. *languageR: Data sets and functions with analyzing linguistic data: A practical introduction to statistics*. Tech. rep. http://cran.r-project.org/package=languageR.

Baddeley, Alan David. 2007. *Working memory, thought, and action*. Oxford: Oxford University Press.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. http://cran.r-project.org/package=lme4.

Bernardini, Silvia. 2001. Think-aloud protocols in translation research: Achievements, limits, future prospects. *Target* 13(2). 241–263.

Blakemore, Diane. 2002. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Brown, Peter F., John Cocke, Stephen Della Pietra, Vincent Della Pietra, Frederick Jelinek, Robert L. Mercer & Paul S. Roossin. 1988. A statistical approach to language translation. In *Proceedings of the 12th conference on computational linguistics* (COLING), 71–76. Stroudsburg: Association for Computational Linguistics.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra & Robert L. Mercer. 1993. The mathematics of statistical Machine Translation: Parameter estimation. *Computational Linguistics* 19(2). 263–311.

Campbell, Stuart. 2000. Choice network analysis in translation research. In Maeve Olohan (ed.), *Intercultural faultlines: Research models in translations studies*, 29–42. Manchester: St. Jerome.

Carl, Michael & Martin Kay. 2011. Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta* 56(4). 952–975.

Carl, Michael & Moritz Schaeffer. 2016. Measuring literality in translation. *Translation: Corpora, Computation, Cognition* 2(1). 127–145.

Carl, Michael, Moritz Schaeffer & Srinivas Bangalore. 2016. The CRITT translation process research database. In Michael Carl, Srinivas Bangalore & Moritz Schaeffer (eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB*, 13–54. Cham: Springer.

Čulo, Oliver, Silke Gutermuth, Silvia Hansen-Schirra & Jean Nitzke. 2014. The influence of post-editing on translation strategies. In Sharon O'Brien, Laura Winter Balling, Michael Carl, Michel Simard & Lucia Specia (eds.), *Post-editing of machine translation: Processes and applications*, 200–218. Newcastle: Cambridge Scholars.

de Groot, Annette M. B. 1992. Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(5). 1001–1018.

Gutt, Ernst-August. 1989. *Translation and relevance.* University of London dissertation.

Gutt, Ernst-August. 1991. *Translation and relevance cognition and context.* Blackwell.

Hartsuiker, Robert J., Martin J. Pickering & Eline Veltkamp. 2004. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science* 15(6). 409–414.

Hönig, Hans G. 1991. Translation studies: The state of the art. In Kitty M. Leuven-Zwart & Ton Naajkens (eds.), *Proceedings of the first james s holmes symposium on translation studies*, 77–89. Amsterdam: Rodopi.

Huang, Xuedong, Mervyn Jack & Yasuo Ariki. 1990. *Hidden markov models for speech recognition.* Edinburgh University Press.

Hvelplund Jensen, Kristian Tangsgaard. 2009. Distribution of attention between source text and target text during translation. In *Proceedings of IATIS.* IATIS.

Hvelplund Jensen, Kristian Tangsgaard. 2016. Cognitive efficiency in translation. In Ricardo Muñoz Martín (ed.), *Reembedding translation process research*, 149–170. Amsterdam: Benjamins.

Immonen, Sini. 2006. Translation as a writing process: Pauses in translation versus monolingual text production. *Target* 18(2). 313–336.

Jääskeläinen, Riitta & Sonja Tirkkonen-Condit. 1991. Automatised Processes in Professional vs non-professional translation: A think-aloud protocol study. In Sonja Tirkkonen-Condit (ed.), *Empirical research in translation and intercultural studies*, 89–109. Gunter Narr.

Jakobsen, Arnt Lykke. 2003. Effects of think aloud on translation speed, revision and segmentation. In *Triangulating translation: Perspectives in process oriented research*, 69–95. Amsterdam: Benjamins.

Krings, Hans Peter. 1986. Translation problems and translation strategies of advanced German learners of French (L2). In Juliane House & Shoshana Blum-Kulka (eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*, 263–276. Tübingen: Gunter Narr.

Kumpulainen, Minna. 2015. On the operationalisation of pauses in translation process research. *Translation & Interpreting* 7(1). 47–58.

Künzli, Alexander. 2007. The ethical dimension of translation revision: An empirical study. *Journal of Specialized Translation* 5. 42–56.

Lacruz, Isabel, Gregory M. Shreve & Erik Angelone. 2012. Average pause ratio as an indicator of cognitive effort in post-editing: A case study. In Sharon O'Brien, Michel Simard & Lucia Specia (eds.), *Proceedings of the AMTA 2012 Workshop on Postediting Technology and Practice (WPTP 2012)*, 21–30.

Läubli, Samuel & Ulrich Germann. 2016. Statistical modelling and automatic tagging of human translation processes. In Michael Carl, Srinivas Bangalore & Moritz Schaeffer (eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB*, 155–181. Cham: Springer.

Lopez, Adam & Philip Resnik. 2009. Conceptual representation in the bilingual lexicon and second language vocabulary learning. In Aneta Pavlenko (ed.), *The bilingual mental lexicon: Interdiscplinary approaches*, 100–125. Bristol: Multilingual Matters.

Lörscher, Wolfgang. 1991. *Translation performance, translation process and translation strategies: A psycholinguistic investigation.* Tübingen: Günter Narr.

Lörscher, Wolfgang. 2005. The Translation Process: Methods and Problems of its Investigation. *Meta* 50(2).

Martínez-Gòmez, Pascual, Akiko Aizawa & Michael Carl. 2014. Characterization of human translation for different levels of expertise. In Akiko Aizawa, Kato Hisashi Akira & Kikui GenIchiro (eds.), *The 20th annual conference proceedings of the association for natural language processing of japan*, 153–156. Kyoto: Nakanishi Printing Company.

Martínez-Gòmez, Pascual, Akshay Minocha, Jin Huang, Michael Carl, Srinivas Bangalore & Aizawa Akiko. 2014. Recognition of translator expertise using sequences of fixations and keystrokes. proceedings of symposium on eye tracking research and applications. In Pernilla Qvarfordt & Dan Witzner Hansen

(eds.), *Proceedings of symposium on eye tracking research and applications*, 299–302. New York: Association for Computing Machinery.

McConkie, George W. & Shun-nan Yang. 2003. How cognition affects eye movements during reading. In Jukka Hyönä, Ralph Radach & Heiner Deubel (eds.), *The mind's eye*, 413–427. Amsterdam: North-Holland.

Mossop, Brian. 2007. *Revising and editing for translators*. Kinderhook: St. Jerome.

Mylonakis, Markos, Khalil Sima'an & Rebecca Hwa. 2007. Unsupervised estimation for noisy-channel models. In *In proceedings of the 24th international conference on machine learning*. Corvallis.

Nida, Eugene Albert. 1964. *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Leiden: E. J. Brill.

Nitzke, Jean & Katharina Oster. 2016. Comparing translation and post-editing: An annotation schema for activity units. In Michael Carl, Srinivas Bangalore & Moritz Schaeffer (eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB*, 293–308. Cham: Springer.

O'Brien, Sharon. 2006. Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Language and Cultures* 7(1). 1–21.

O'Brien, Sharon, Laura Winther Balling, Michael Carl, Michel Simard & Lucia Specia (eds.). 2014. *Post-editing of machine translation: Processes and applications*. Newcastle: Cambridge Scholars.

Och, Franz Josef, Richard Zens & Hermann Ney. 2003. Efficient search for interactive statistical machine translation. In *Proceedings of the 10th conference of the european chapter of the association for computational linguistics (EACL)*. Budapest.

Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427–459.

Plitt, Mirko & François Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague bulletin of mathematical linguistics* 93. 7–16.

Risku, Hanna. 2014. Translation process research as interaction research: From mental to socio-cognitive processes. *MonTI Special Issue – Minding Translation*. 331–353.

Ruiz, Carmen, Natalia Paredes, Pedro Macizo & Mª Teresa Bajo. 2008. Activation of lexical and syntactic target language properties in translation. *Acta psychologica* 128(3). 490–500.

Schaeffer, Moritz & Michael Carl. 2013. Shared representations and the translation process: A recursive model. *Translation and Interpreting Studies* 8(2). 169–190.

Schaeffer, Moritz & Michael Carl. 2014. *Measuring the cognitive effort of literal translation processes.* Stroudsburg: Association for Computational Linguistics. 29–37.

Schaeffer, Moritz & Michael Carl. 2016. Language processing and translation: Translation and non-translational language use. *Translation: Corpora, Computation, Cognition.*

Schaeffer, Moritz & Michael Carl. 2017. Language processing and translation. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting* (Translation and Multilingual Natural Language Processing 7), 117–154. Berlin: Language Science Press.

Schaeffer, Moritz, Michael Carl, Isabel Lacruz & Akiko Aizawa. 2016. Measuring cognitive translation effort with activity units. *Baltic Journal of Modern Computing* 4. 331–345.

Schaeffer, Moritz, Barbara Dragsted, Laura Winther Balling & Michael Carl. 2016. Word translation entropy: Evidence of early target language activation during reading for translation. In Michael Carl, Srinivas Bangalore & Moritz Schaeffer (eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB*, 183–210. Cham: Springer.

Sekino, Kyoko. 2012. *Investigating conceptual and procedural encodings in manual translation and in post-editing processes from Japanese to Portuguese.*

Seleskovitch, Danica. 1975. *Langage, langues et mémoire: étude de la prise de notes en interprétation consécutive.* Paris: Minard Lettres Modernes.

Shannon, Claude E. & Warren Weaver. 1949. *The mathematical theory of communication.* Urbana: University of Illinois Press.

Singla, Karan, David Orrego-Carmona, Ashleigh Rhea Gonzales, Michael Carl & Srinivas Bangalore. 2014. Predicting post-editor profiles from the translation process. In Francisco Casacuberta, Marcello Federico & Philipp Koehn (eds.), *Proceedings of the workshop on interactive and adaptive machine translation*, 51–60. Association for Machine Translation in the Americas (AMTA).

Sperber, Dan & Deirdre Wilson. 1993. Linguistic form and relevance. *Lingua* 90(1).

Sperber, Dan & Deirdre Wilson. 1995. *Postface to the second edition of Relevance: Communication and Cognition.* Oxford: Blackwell.

Tirkkonen-Condit, Sonja. 2005. The monitor model revisited: Evidence from process research. *Translators' Journal* 50(2). 405–414.

Toury, Gideon. 2004. Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? In Anna Mauranen & Pekka Kujamäkipp (eds.), *Translation Universals: Do they exist?* Amsterdam: Benjamins.