Original Articles · Artículos originales

Literal vs. Default Translation. Challenging the Constructs with Middle Egyptian Translation as an Extreme Case in Point

Traducción literal vs. traducción por defecto. Un desafío para los dos constructos mediante la traducción del egipcio clásico como caso extremo

Christian Olalla-Soler ⬤ 0000-0001-9251-476X
Universitat Autònoma de Barcelona

ABSTRACT

This paper presents the results of a study that compares the constructs of literal translation (Schaeffer & Carl, 2014) and default translation (Halverson, 2019) by means of an observational, exploratory study with Middle Egyptian translation as an extreme case in point. Two MA students in Egyptology at Universitat Autònoma de Barcelona and three recent graduates of the same MA programme took part in the study. They translated two excerpts from two Middle Egyptian literary texts into Spanish. InputLog was used to collect translation-process data and derive word-level indicators of cognitive effort from them: typos per word, word typing speed, and within-word pause. Results showed a clear link between default translations and cognitive effort (low number of typos, low number of respites, and fast writing speed). However, the assumption that deviations from literality cause greater cognitive effort was not observed. Hence, default translation may serve as a more adequate construct to describe the regular way translators perform.

Keywords: cognitive translation studies, literal translation, default translation, observational study, keylogging, cognitive effort, Middle Egyptian

RESUMEN

En este artículo se exponen los hallazgos de un estudio observacional y exploratorio en el cual se cotejan los constructos de traducción literal, según Schaeffer y Carl (2014), y el de traducción por defecto, propuesto por Halverson (2019), mediante la traducción del egipcio clásico como un caso extremo. En la investigación participaron dos alumnos del Máster Universitario en Egiptología de la Universitat Autònoma de Barcelona junto con tres egresados del mismo programa. La tarea consistió en traducir al español dos fragmentos procedentes de dos textos literarios escritos en egipcio clásico. Se empleó InputLog como herramienta para recoger datos referentes al proceso de traducción y para derivar de ellos indicadores de esfuerzo cognitivo a nivel de palabra, tales como errores tipográficos por palabra, velocidad de tecleo y pausas intra-palabra. Los resultados evidenciaron una clara vinculación entre las traducciones por defecto y el esfuerzo cognitivo, manifestado por un bajo número de errores tipográficos, escasas pausas y una elevada velocidad de escritura. No obstante, el supuesto de que las desviaciones de la traducción literal conllevan un esfuerzo cognitivo mayor no pudo ser verificado. Así, la traducción por defecto podría presentarse como un constructo más apropiado para describir la forma habitual de traducir de los/las traductores/as.

Palabras clave: estudios cognitivos de la traducción, traducción literal, traducción por defecto, estudio observacional, registro de teclado, esfuerzo cognitivo, egipcio clásico

And he understands the difficult passages,
and he elucidates the hieroglyphic signs,
and he becomes an apprentice of the Majesty of Thoth.

The Book of Thoth, passage 338
Translated by Richard Jasnow & Karl-Theodor Zauzich

1. Introduction

This paper presents the results of a study that compares the constructs of literal translation (Schaeffer & Carl, 2014) and default translation (Halverson, 2019) by means of an observational, exploratory study with Middle Egyptian translation as an extreme case in point.

1.1. Literal and default translation

These two constructs aim at describing the regular way translators perform. The concept of literal translation (as opposed to free translation) has been around for centuries and has been discussed in translation studies since its origins as a modern, academic discipline (see, e.g., Catford, 1965; Ivir, 1981). However, more recent approaches to literal translation stem from Tirkkonen-Condit's notion of a default way of translating:

> It looks as if literal translation is a default rendering procedure, which goes on until it is interrupted by a monitor that alerts about a problem in the outcome. The monitor's function is to trigger off conscious decision-making to solve the problem. Automation also affects the monitor, so that traces of its operation are not as frequently observable in the processes and products of experts as in those of novices and non-experts. (2005: 408)

Schaeffer and Carl's literal translation construct fleshes out Tirkkonen-Condit's notion and assumes that (2014: 29): "One-to-one literal translation correspondences are easier to produce than translations that formally deviate from the source text, as the latter would require more effort, and hence will take longer for a translator to produce". Carl and Schaeffer's (2017b: 85) concept of translation literality is based on the following criteria:

1. Word order is identical in the source and target texts.
2. Source and target-text items are one-to-one translation equivalents.
3. Each ST word has only one possible translated form in a given context.

The concept of literality in translation, as per their criteria, is closely tied to the degree of similarity between the source text (ST) and the target text (TT). This similarity extends not only to syntactic structures but also to semantic representations of lexical items. The degree to which two items share structural and semantic representations may predict the strength of a priming effect that may ease the processing of a linguistic item based on its literality. The more literal an item is, the stronger the priming effect. This effect is tied to both early and late eye movement measures during the translation process. Early effects may indicate automatic cognitive mechanisms underlying the translation process, while late eye movement measures are more likely to reflect conscious behavior (Carl & Schaeffer, 2017).

Schaeffer and Carl (2014) also posit that translators often already have a translation hypothesis in mind, referring to the ST and TT to verify this hypothesis. They suggest that trans-

lations often consist of monotonous one-to-one translations, approaching an ideal literal translation. The more syntactic reordering between source and target text takes place, the more it will become non-literal.

Schaeffer et al. (2016) further elaborate on this, suggesting that a literal translation is often the first or default solution a translator applies to the source text, often only as an interim solution before a less literal translation is considered or produced.

Default translation is conceived as a "particular phase of translation production", in which:

> translators demonstrate stretches of uninterrupted production. The text produced is assumed to draw on easily accessible, routinized knowledge, including bilingual linguistic knowledge, metalinguistic knowledge (including knowledge of communication norms), and knowledge of the specific task. This knowledge is dynamic, individual and personal, but maintains a principled relationship to language patterns characteristic of relevant usage situations within the relevant linguistic communities (Halverson, 2019: 190)

Default translation is a complex process that involves the generation of fast and easy translations, which are based either on coactivation patterns established through bilingual language use (without translation) or those established through prior translation activity and subsequently reused for that purpose more or less frequently, or both (Halverson, 2019). This process can occur even when the translator is translating a particular stretch of language for the first time, as long as the linguistic material used is familiar and the cross-linguistic relationships are relatively unambiguous.

The process of default translation also involves the selection of translations that bear relatively little formal or semantic similarity to the source, if the two structures are linked often enough in a translational act. Thus, Halverson's construct makes no assumptions regarding the characteristics of the resulting relationship between the source and the target text (i.e., whether the relationship is based on a word-for-word translation). This means that default translation will show evidence of established patterns that originate in shared semantic and/or formal characteristics, as well as translational solutions that have become entrenched in spite of their lacking obvious semantic or formal similarities (Halverson, 2019).

The entrenchment process in default translation draws on experience with interpreting or translating from an early age. It is also influenced by the translator's personal history, which determines key characteristics of their linguistic knowledge. Hence, default translation is usage-based and its manifestation is intrinsically linked to the specific knowledge and experience of the individual language user. This includes but is not limited to linguistic and metalinguistic knowledge, established routines for action, and conceptualization of the specific translation task at hand (Halverson, 2019).

The constructs are also opposed in relation to their assumptions about cognition. Literal translation is grounded in the principles of computational translatology, in which reality is conceived of as external and transcendental, the human mind as different from the brain, and languages as independent from their users and based on linguistic symbols that carry stable, self-contained meaning (Muñoz, 2016: 561; see Carl, 2010). Default translation is framed within cognitive translatology, which conceives cognition as (Clark, 1996; Muñoz, 2016: 563–564): embodied ("the body substantially determines thought processes"), embedded ("the brain is nested into both a body and a physical and sociocultural environment"), enacted ("the

environment is selectively created by the subject in an emergent self-organizing process that fosters an effective use of resources and affordances"), extended ("the brain/mind actively offloads tasks and procedures into "outside" scaffolds and props so as to lower mental load"), affective ("many social activities rest upon our ability to reason about others' emotions, but also because emotions drive and fine-tune our mental processes and our behavior"), and distributed ("several cognizing and not cognizing agents conjointly perform complex tasks").

Drawing from the literature review outlined herein, numerous distinctions between literal and default translation emerge, as encapsulated in Table 1.

Table 1. A comparison between literal and default translation

	Literal translation	Default translation
Source-to-target relationship	An absolute literal translation relationship between ST and TT is one-to-one, the word order in both texts is identical, and the semantic and structural representation of a translated item in the TT is expected to be closely identical to that in the ST.	This construct does not necessarily demand a one-to-one relationship or identical word order between the ST and TT. The resulting translations might bear little formal or semantic similarity to the source, provided the two structures are frequently linked in a translational act.
Cognitive process	Framed within computational translatology, which perceives reality as external and transcendental, and languages as independent from their users, carrying stable, self-contained meaning.	Framed within cognitive translatology, which conceives cognition as embodied, embedded, enacted, extended, affective, and distributed.
Nature of process	Seen as a default rendering procedure, which continues until there is a need for a conscious decision to solve a problem in the outcome.	Seen as a dynamic process that draws upon easily accessible, routinized knowledge, including bilingual linguistic knowledge, metalinguistic knowledge, and knowledge of the specific task.
Priming effect	The construct posits a strong priming effect based on the literality of a linguistic item. The more literal an item is, the stronger the priming effect, which can ease the processing of the item.	The construct doesn't specify a priming effect. Instead, it emphasizes that default translation can occur even when the translator is translating a specific stretch of language for the first time, as long as the linguistic material is familiar.
Experience and individuality	The construct does not stress individual experiences or uniqueness. The translators often already have a translation hypothesis in mind, referring to the ST and TT to verify this hypothesis.	This construct puts a premium on the translator's personal history, linguistic knowledge, and established routines, emphasizing that translation is a usage-based and dynamic process that is intrinsically linked to the translator's specific knowledge and experience.

1.2. Empirical research into literal and default translation

Both constructs have been studied empirically. Schaeffer and Carl (2014) observed that literal translations were faster than non-literal ones, more translation choices lead to longer reading and processing time, and gaze activity rose with an increase in the variation of source and target-text syntactic structure. Schaeffer et al. (2016) observed that translation processes are partly automatic and partly conscious and willed behaviour, given that the effect of similar syntactic order in the source and the target text and a low number of possible translations for a single source-text item act as a facilitator effect which fosters automatic processes. Carl and Schaeffer (2017a) found that literal translations are less effortful to produce. In another study, Carl and Schaeffer (2017b) reached the same conclusion, i.e., producing non-literal transla-

tions is more time-consuming. In addition, the authors suggest that the number of lexical and syntactic choices a translator has to make affects translation difficulty.

In addition to two conference presentations of a pilot study in which the methodological procedures to investigate default translations were explored (Halverson & Muñoz, 2019a, 2019b), the construct of default translation has been investigated in a corpus-based study carried out by Jiménez-Crespo and Casillas (2021). The authors found out that the processing effort of literal translation renditions in a corpus did not differ from that of default translations (i.e., the time taken to start editing the translation candidate and the time taken to complete the editing of the candidate). Hence, syntactic and word order similarity and low number of possible renditions did not reduce effort.

1.3. Challenging the constructs

The study presented in this paper aimed at exploring how the two constructs differ (or coincide) when explaining the very same data on cognitive effort extracted from the process of translating two texts from Middle Egyptian as the source language by means of an observational, exploratory study. The specific aims were:

1. To estimate the effect of literality on cognitive effort of producing translation equivalents to source-text words.
2. To determine whether the defaultness of the translations and their degree of literality has an effect on cognitive effort of producing them.

The use of Middle Egyptian as a source language serves the purpose of putting both constructs to an extreme test. Middle Egyptian (2200 to 1350 BC; Cervelló Autuori, 2015) featured grammatical forms (such as particles) and syntactic structures (such as pseudo-verbal constructions, present and future participles, etc.) that do not exist in many modern languages. It was a strongly paratactic language, and the syntactic structure of the various clause types was generally fixed. Words were not separated by spaces, and no punctuation marks were used. The consequence of (combining) these elements leads to the hypothesis that the degree of literality (as operationalised by Schaeffer & Carl, 2014) when translating from Middle Egyptian to modern languages such as Spanish (as in the case of this study) should be low and hence effortful from a cognitive viewpoint since a lack of literality is assumed to force the translator to abandon the unmarked and preferred literal translation procedure and adopt a problem-solving mode.

This paper is structured as follows. The methods of the exploratory study are presented in section 2. Section 3 is devoted to the results of the study and their discussion. Section 4 presents the most relevant conclusions, discusses the study's limitations, and provides possible future lines of work.

2. Methods

This paper reports the results of an observational, exploratory study that was pre-registered in the Open Science Framework Registries on June 30, 2021.[1] Several methodological procedures described in the pre-registration had to be modified due to unforeseen methodological issues. Such deviations from the pre-registration will be discussed in the following sections.

2.1. Study plan

The study consisted in translating two Middle Egyptian texts into Spanish and conducting a semi-structured interview after completing the second translation. Data were collected remotely, so participants could use their own computers and other resources, such as dictionaries and grammar handbooks, in a physical environment in which they felt comfortable. No constraints were set, and participants were allowed to spend as much time as needed to translate each text and use any resource they deemed necessary. The decision on the order in which the two texts were translated was also left to the participants.

Data were collected from July to September 2021, and the procedure for all participants was:

1. The study description and informed consent was signed by the participant.
2. A zip file containing a document with detailed instructions, the two texts, the InputLog installer, and a link to a video tutorial to install and use InputLog was sent to the participant.
3. The participant translated both two texts with InputLog and sent both translated texts and the InputLog files to the researcher.
4. After arranging an appointment with the participant, the interview was conducted via Skype, Zoom, or telephone, depending on the participant's preferences.

2.2. Sample

Two graduating students of the MA in Egyptology at Universitat Autònoma de Barcelona (Spain) and three recent graduates of the same MA took part in the study. All participants (three male, two female) were native speakers of Spanish, and only one had experience as a professional translator.[2] All participants had completed all philological modules of the MA, which corresponded to 45 (37.5%) of the 120 ECTS of the MA. These modules were devoted to the study of the Ancient Egyptian language (with a focus on Middle Egyptian, Late Egyptian and, to a lesser degree, Coptic) and its translation into Spanish. It must be noted that, rather than as an end in itself, translation practice is used as a means to learn about and reflect on the source language, which means that translation is conceived as overt and documentary, following the terms proposed by House (1977) and Nord (1991). In these modules, no contents related to the theory and practice of translation are provided.

To recruit participants, the MA coordinator contacted students and recent graduates through email. The five participants took part in the study on a voluntary basis (i.e., voluntary response sampling), hence a non-probability sampling procedure was used. The only criterion for selecting participants was having completed (in the case of former students) or being about to complete their MA (in the case of students).

2.3. Materials

2.3.1. Texts to be translated

Two literary, Middle Egyptian text excerpts were selected for the study. They were preceded by an explicit translation brief: the prose translations of the two literary texts were to be published in a book containing literary works of Ancient Egypt for the general public.

The excerpts were extracted from *The Story of the Shipwrecked Sailor* (pHermitage 1115; written at the end of the 12th dynasty, which reigned from 1991 to 1802 BC) and *The Story of Sinuhe* (pBerlin 3022 and pAmherst n-q; written between the 10th and 17th regnal year of Senusret I, a king of the 12th dynasty who reigned from 1971 to 1926 BC). Both texts are written in hieratic and composed in narrative verse (Foster, 1980). For this study, the hieroglyphic transcription was provided to the participants.

The selection of these two literary texts was due to the fact that they are considered to be excellent representatives of the grammar and syntax of Middle Egyptian (see Parkinson, 1998). The selection of the two excerpts was based on linguistic criteria, i.e., the texts had to display a wide variety of syntactic and grammatical structures reflecting the linguistic characteristics of Middle Egyptian. The two texts were grammatically and syntactically analysed, and an excerpt was extracted for each. In *The Story of the Shipwrecked Sailor*, lines 166 to 176 and columns 177 to 179 were extracted (110 source-text words; Figure 1). In *The Story of Sinuhe*, the excerpt included column B 179 and lines B 180 to B 188, which corresponded to 107 source-text words (Figure 2).

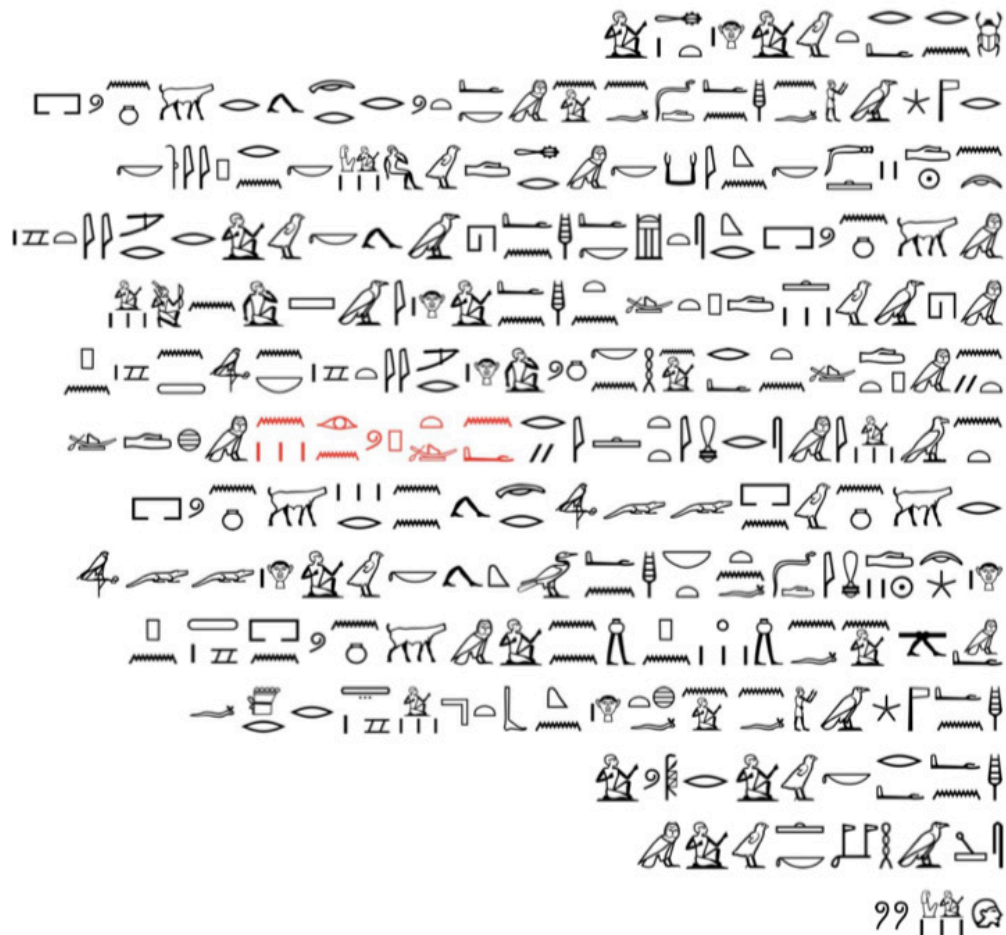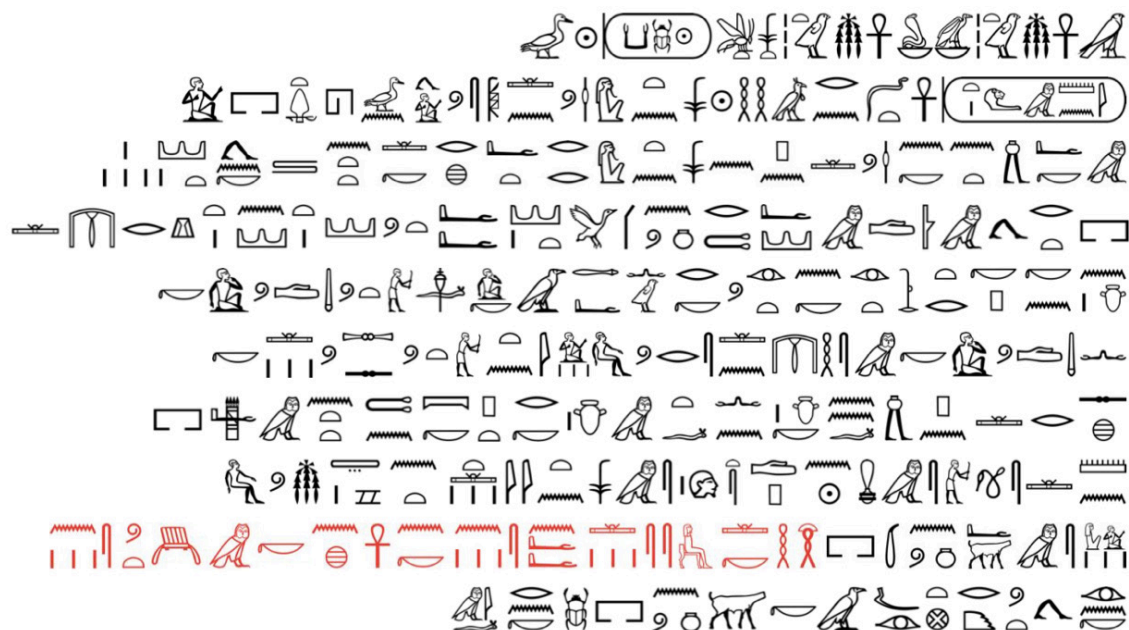**Figure 1.** Excerpt of *The Story of the Shipwrecked Sailor*

Figure 2. Excerpt of *The Story of Sinuhe*

Each text was presented in a single-page PDF document which included the hieroglyphic text preceded by the previously mentioned translation brief. The hieroglyphic transcription of the hieratic texts was extracted from Allen (2015) and rewritten using JSesh (Rosmorduc, 2014).

2.3.2. Data collection tools

The participants' translation processes were recorded with InputLog v. 8 (Leijten & Van Waes, 2013). InputLog records keyboard activity data in an ecologically valid way, as it captures keypresses performed in any programme, such as Microsoft Word, web browsers, PDF viewers, etc. Hence, the translation process becomes more realistic since participants are not limited in any way in terms of using documentation and information-seeking resources.

Since the participants were able to use their own computers for the two translation tasks with InputLog, ecological validity and the overall reliability of the data gathered increased given that they used a keyboard in which they were used to type.

Participants were given the recommendation to translate each text in a single session, which all participants followed. In all, nine InputLog log files were generated and used for subsequent analysis (see next section), given that for one participant InputLog crashed during one of the translation tasks and data were lost.

2.4. Variables and indicators

2.4.1. Literality

Carl and Schaeffer's (2017b: 85) concept of translation literality is based on the following criteria:

1. Word order is identical in the source and target texts.
2. Source and target-text items are one-to-one translation equivalents.

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

The first and second criteria are based on the assumption (a) that "[a] literal translation consists of the same number of tokens where each TT token corresponds to exactly one ST token, and tokens in both texts are ordered in the same way" (Carl & Schaeffer, 2017b: 85), while the third one is based on the assumption (b) that both source and target languages have similar semantic representation with respect to a source-text word if that word is consistently translated in the same way (ibid.: 85). The authors use two indices to measure criteria 1 and 2 (based on assumption a): CrossT and CrossS. To measure criteria 3 (based on assumption b), the authors employ the translation perplexity index and word translation entropy.

CrossS and CrossT are indexes based on the word alignment of the source and target texts. CrossS "indexes translation relations from the ST point of view, providing a metric for the syntactic distance of the TT compared to the ST" (Carl & Schaeffer, 2017b: 86), while CrossT establishes translation relations from the TT point of view. According to Carl and Schaeffer:

> from a given translation and its alignments with a source word we compute *CrossS* and *CrossT* values on the ST and the TT sides respectively, by following the alignment links and counting the distance in words between two successive alignments. We thus obtain a vector of relative alignment distortions *CrossS* and *CrossT* for ST and TT words, indicating the word order similarity of the two sentences. (2017b: 86)

Hence, to compute CrossS and CrossT, source-text and target-text sentences need to be aligned. Either the hieroglyphic or the hieratic writing systems used punctuation marks, so applying the linguistic notion of sentence to Middle Egyptian is highly problematic. Instead of sentences, in this study, alignment segments were equated to clauses. Source-text clauses were used to divide the text and align the source-text words of each clause with their translated counterparts.[3] The alignment was carried out manually.

A second methodological issue related to CrossS and CrossT that was identified when pre-registering the study is related to the implicit decisions made by the proponents of this procedure when aligning source-text and target-text words. To illustrate this, a CrossS and CrossT example of an alignment between English and Spanish provided in Carl and Schaeffer (2017a: 49) is shown below (Figure 3):

Figure 3. Word alignment example provided in Carl and Schaeffer (2017a: 49)
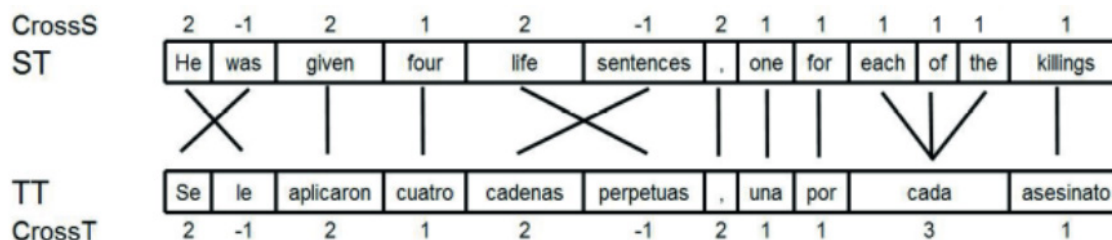
There are several alignments that do not seem to follow the authors' declared procedure. For instance:

a. The verb *was* is established as an equivalent for *Se*, a Spanish reflexive pronoun used to build the reflexive passive voice. *Se* cannot by any means be considered a translation equivalent of the third-person singular *was*, which would be either *fue* or *era* in Spanish. Word-to-word alignment does not seem to work adequately in this case since the En-

glish *was given* should be aligned with *Se aplicaron* (two passive structures with different grammatical structures). The methodological issue here is whether non-word-to-word equivalents should be aligned when an equivalence cannot be established between two words at a semantic and grammatical level.

b. The noun *life* is aligned with *perpetuas* and *sentences* with *cadenas*. It is questionable to align *life* and *sentences* as two separate units since *cadena perpetua* is the Spanish equivalence of the English collocation *life sentence* and, consequently, these two words together form a semantic unit. In this case, the methodological issue is how to align the semantics of collocations.

c. The commas are also aligned, so it is not clear if punctuation marks should be considered words and hence aligned. Middle Egyptian did not use punctuation marks, so in this case, it is not possible to align them.

Establishing equivalences at a word level does not seem to work in all cases, and the rationale behind several alignments is not clear nor transparent. In light of these issues, when computing CrossT and CrossS values, the following rules were applied:

1. Source and target-text units were aligned at a word level by default.
2. When several words had to be considered a semantic unit, they were merged into a single unit.
3. Punctuation marks were not aligned (since they did not exist in Middle Egyptian).

We now turn to perplexity, a construct used to measure the third criterion for literality (each ST word has only one possible translated form in a given context). Carl and Schaffer define perplexity in the following way:

> The perplexity of a language model is a measure that indicates how many different, equally probable words can be produced, and thus how many choices are possible at a certain point in time. The higher the perplexity, the more similar likely choices exist. Since the difficulty of picking one amongst similar possibilities grows with the number of the available options, perplexity also indicates the difficulty to make a decision. (2017b: 87)

This study applied the computation procedure of perplexity, as described in Carl, Schaeffer, and Bangalore (2016: 31) and in Carl and Schaeffer (2017b: 88). However, two major methodological problems were identified prior to pre-registering this study. First, the low number of participants influenced the results, as they reflected a small variety of translation choices and/or a low repetition of choices. Second, as indicated when discussing the CrossS and CrossT indicators, there are several alignment decisions that seem incongruent and therefore make it difficult to closely follow the authors' analytical procedures. In the example for the calculation of translation perplexity provided in Carl and Schaeffer (2017a: 47), we observe that the authors have followed different alignments for the same translation choices (Figure 4).

Figure 4. Computation example of perplexity provided in Carl and Schaeffer (2017a: 47)

s	H(he) = 1.95		H(was) = 2.88		H(given) = 3.55		H(four) = 0	
	# he	I(s→ti)	# was	I(s→ti)	# given	I(s→ti)	# four	I(s→ti)
t1	12 le	0.87	9 le	1.29	4 dieron	2.46	22 cuatro	0.00
t2	2 recibió	3.46	2 recibió	3.46	4 condenaron	2.46		
t3	2 ---	3.46	2 dieron	3.46	2 recibió	3.46		
t4	4 se	2.46	2 condenaron	3.46	2 condenó	3.46		
t5	1 se lo	4.46	1 se condenó	4.46	1 sentenciado a condenas	4.46		
t6	1 se le	4.46	1 se condemnó	4.46	1 Se le condemnó	4.46		
t7			1 lo condenó	4.46	1 los	4.46		
t8			1 han	4.46	1 impuesto	4.46		
t9			1 fue	4.46	1 han condenado	4.46		
t10			1 declarado	4.46	1 culpable	4.46		
t11			1 ---	4.46	1 conenaron a	4.46		
t12					1 condenó a	4.46		
t13					1 aplicaron	4.46		
t14					1 ---	4.46		

The personal pronoun *he* has been aligned with *le* (an indirect pronoun), *se* (a reflexive pronoun)*, se lo* (a combination of an indirect + direct pronoun)*, se le* (another combination of indirect + direct pronoun) and even with *recibió* (3rd person singular of the past tense of the verb receive). The verb *was* has been aligned with the pronoun *le* (9 times). We also find *se condenó* and *se condemnó* [sic] aligned with *was*, and even *se le condemnó* [sic] with *given*. Coherence is very low, and this influences the results (for instance, by increasing the number of translation choices when including *le* as an alignment of *was* and not of *he*, which would be more logical given that both are third person singular pronouns but with different syntactic functions). The issue of aligning source-text and target-text words has been discussed by one of the proponents of this method in Gilbert et al. (2023) and, even if the authors examine the effect of inconsistent alignments on perplexity and entropy (see below), it also affects CrossS and CrossT. In our study, once an equivalent was established, all other equivalents that are formally identical were classified in the same way.

The third criteria of literality (i.e., each ST word has only one possible translated form in a given context) has also been measured with word translation entropy, an indicator that is closely related to perplexity (since perplexity is an exponential function of word translation entropy). Schaeffer et al. define word translation entropy as follows:

> the degree of uncertainty regarding which lexical TT item(s) are chosen given the sample of alternative translations for a single ST word: if the probabilities are distributed equally over a large number of items, the word translation entropy is high and there is a large degree of uncertainty regarding the outcome of the translation process. If, however, the probability dis-

tribution falls unto just one or a few items, entropy is low and the certainty of the TT item(s) to be chosen is high. (2016: 191)

2.4.2. Default translations

Halverson suggests the following procedure to identify default translations:

in order to identify passages of default production in online data, the following procedure is proposed for piloting: (1) segment keystroke logs using the framework proposed by Muñoz and colleagues, as presented in Muñoz & Cardona (2018[2019]); (2) filter out segments including problem-solving activity; (3) using aligned gaze data, investigate whether empty task segments contain reading activity, and include all which do. This should suffice for a pilot study of the phenomenon, and the potential represented by incorporating the information given by gaze shifts and EKS [time difference between a fixation on a ST element and the first keystroke in producing its translation] can be investigated in a further development phase. (2019: 202)

The Task Segment Framework (TSF, Muñoz & Cardona Guerra, 2019; Muñoz & Apfelthaler, 2022) provides an alternative to the methodological problems regarding the analysis of translation processes as keylogged that have been discussed in many publications, such as using fixed, arbitrary pause thresholds (Muñoz & Martín de León, 2018; Muñoz & Cardona Guerra, 2019; Muñoz & Apfelthaler, 2022). The TSF identifies task segments in the keylogged translation process and distinguishes between the following subtasks: (1) adding new text to the text written so far, i.e., ADD; (2) changing the text written so far, i.e., CHANGE; (3) searching for information, i.e., SEARCH, and (4) interacting with the computer for purposes other than those immediately related to the task at hand (such as lowering the music volume, changing the font, etc.), i.e., HCI. In addition to these four subtasks, it was necessary to add a fifth one, ANALYSIS, in which participants produced a syntactic analysis of the source-text segment being translated. This subtask was only observable in the case of the students since, in their training, they were asked to produce syntactic analyses of the text being translated.

These five subtasks fit in different ways within task segments, i.e., behavioural chunks of the typing flow segments arrived at by setting an arbitrary baseline and two thresholds for inter-keystroke intervals:

a. Lags. Time spans between 1 and 199 ms. These are considered beyond reach with current knowledge and tools and, therefore, considered noise. They are not computed.

b. Delays. Time spans between the baseline of 200 ms and two times the median value for inter-stroke intervals within words. This is the lower threshold of two, and such time spans are deemed to be mainly due to mechanical disfluencies that do not stop the typing flow.

c. Pauses. Time spans above three times the median value for inter-stroke intervals between words. This is the upper threshold in the analysis. Pauses are assumed to be usually intentional, conscious stops of typing flow.

d. Respites. Time spans between the lower and upper thresholds. They are considered unintentional and related to disfluencies due to task-related attentional changes that do not involve planning and do not stop the typing flow.

The TSF was used to segment the logs of the translated texts and to identify task segments, as suggested by Halverson. However, it was not possible to use eye-tracking, so the procedure

followed in this study had to be modified. In fact, Halverson and Muñoz (2019a, 2019b) carried out two pilot tests to identify default translations with keylogged data analysed with the TSF but without eye-tracking data. The segmentation of the log file was carried out with Fácil, a piece of software created at the University of Bologna. After task segments were identified and classified into the five subtasks previously discussed, it was necessary to establish criteria to identify task segments containing default translations. They were conceptualised in the following way based on the definition provided by Halverson (2019):

a. Default translations are expected to be fluent. Hence, default translations should show fewer respites and typos than non-default translations.

b. Default translations are expected to be fast. Hence, default translations should be produced faster than non-default translations.

c. Default translations are expected to show some degree of formal similarity across translation solutions. Hence, default translations should show greater similarity than non-default translations.

d. Default translations are expected to be found in TSs of a single type. Hence, ADD-only segment types will be examined to identify default translations. Out of the 528 task segments that were produced, 67.8% were identified as ADD-only. They contained 69.5% out of the 1537 words produced in the translations.

Based on these parameters, a segment was considered a default translation when it occurred in an ADD-only segment, and at least three out of the following four conditions were met:

a. The ADD-only segment was produced with a speed higher than that of 75% of all segments, and/or

b. The ADD-only segment contained a number of respites lower than that of 75% of all segments, and/or

c. The ADD-only segment contained a number of typos lower than that of 75% of all segments, and/or

d. The ADD-only segment had a coincidence level higher than that of 75% of all segments.

To compare (and identify) non-default translations and default translations, the following indicators were computed for each ADD segment:

a. Production speed. Time spent typing an ADD segment divided by the number of keystrokes in that segment.

b. Number of respites. Number of respites in an ADD segment divided by the number of keystrokes in that segment.

c. Number of typos. Number of typos in an ADD segment divided by the number of keystrokes in that segment.

To compute the level of similarity of the task segments, the source- and target-text word alignment was used. The proportion mean of the target-text equivalents of the source-text words within a task segment was computed. Table 2 exemplifies the procedure.

Table 2. Example of computation procedure of TS coincidence levels

ST word code	ST010	ST011	ST012	ST013	ST014	ST015	ST016
ST word	*ꜥḥꜥ.n*	*ḏd.n=f*	*n=i*	*mk*	*tw*	*r*	*spr*
Participant 1	Entonces	él dijo	me	¡Mira!	Tú	regresarás	∅
Participant 2	Entonces	él dijo	me	Mira,	tú	alcanzarás	∅
Participant 3	y	él dijo	me	"Escucha,	∅	llegarás	∅
Participant 4	Entonces	él dijo	me	¡Mira!	tú estás	hacia	el llegar
Participant 5	Entonces,	él dijo	me	∅	∅	Regresarás	∅
Option 1 counts	4	5	5	3	2	2	4
Option 2 counts	1			1	2	1	1
Option 3 counts				1	1	1	
Option 4 counts						1	
Number of translations	5	5	5	5	5	5	5
Proportion of option 1	0.8	1	1	0.6	0.4	0.4	0.8
Proportion of option 2	0.2			0.2	0.4	0.2	0.2
Proportion of option 3				0.2	0.2	0.2	
Proportion of option 4						0.2	

Example 1: TS 17 for participant 1 is Eno⊠tonces_, which corresponds to *ꜥḥꜥ.n* in the source text. The proportion of this TT option is 0.8. Hence, the level of coincidence of TS 17 for participant 1 is 0.8.

Example 2: TS 26 for participant 1 is T•◊7·u_regresar◊7·as_, which corresponds to *tw r spr* in the source text. The proportion of use of the translation option for ST word *tw* is 0.4. The proportion of use of the translation option for ST word *r* is 0.4. The proportion of use of the translation option for ST word *spr* is 0.8. The mean between 0.4, 0.4, and 0.8 is 0.53. Hence, the level of coincidence of TS 26 for participant 1 is 0.5.

Note: Punctuation marks are not considered when aligning TT words. When a TS length is less than a word long (for instance, TS 7 for participant 1 is O) or it only contains a spacebar or a punctuation mark, the TS is considered as a missing data point in terms of coincidence level.

Out of the 475 ADD-only segments, 166 (35.0%) were identified as containing default translations. These 166 segments contained 215 (20.1%) out of the 1068 words produced in ADD-only segments. As reported in Table 3, these 166 ADD-only task segments containing default translations are faster and produced with less interruptions than the remaining 309 ADD-only segments without default translations.

Table 3. Indicators of *defaultness* in ADD-only segments

Task segment…	ADD-only task segments containing default translations (n = 166)	ADD-only task segments not containing default translations (n = 475)
Production speed	m = 0.72; Mdn = 0.29; SD = 1.01	m = 1.78; Mdn = 1.62; SD = 1.15
Number of respites	m = 0.01; Mdn = 0.00; SD = 0.04	m = 0.06; Mdn = 0.00; SD = 0.14
Number of typos	m = 0.00; Mdn = 0.00; SD = 0.00	m = 0.03; Mdn = 0.00; SD = 0.06

2.4.3. Cognitive effort

As discussed in Hunziker Heeb et al. (2021), cognitive load and cognitive effort have been used interchangeably on many occasions. However, the two concepts should be distinguished, as pointed out by these authors:

> We suggest differentiating between cognitive load and cognitive effort by specifying that load can be associated with the source text, the translation commission, the tools, the translation situation, the translator's psychophysical disposition, the translator's network, and so on, whereas, effort is associated with the actual responses by the task performer (see also Gile and Lei 2021). While cognitive load can theoretically be identical for different translators, cognitive effort is probably not. (2021: 56-57)

In Schaeffer and Carl (2014) and Carl and Schaeffer (2017b), eye-tracking data were used as indicators of cognitive effort. Since data were collected remotely (as the study was conducted during the COVID-19 pandemic), eye-tracking was not used. Instead, in this study, key-logging was employed as a method to record behavioural indicators of translation cognitive effort. Three indicators were derived from the keylogged data, and they focused on the word level so that they could be correlated with all literality indicators, which are also computed at a word level. The three word-level indicators of cognitive effort were:

1. Typos per word. Total number of deletions in a word divided by the number of keystrokes in that word. The assumption is that an effortful translation process will increase the number of typos per word.
2. Word typing speed. Total time spent typing a word divided by the number of keystrokes in that word. The assumption is that an effortful translation process will decrease the speed at which a word is typed.
3. Within-word pause. Total pause time between consecutive keystrokes in a word divided by the number of keystrokes in that word. In this case, the assumption is that an effortful translation process will increase within-word pause duration.

2.5. Statistical analysis

In addition to descriptive statistics (the mean, the median, and the standard deviation), two inferential procedures were used to analyse the data. When two groups are compared (i.e., the student vs. non-student condition, and the two texts being translated), the Mann-Whitney U test for independent groups was used given that the data did not follow a normal distribution. Data transformations were not effective to normalise the data due to the inclusion of values that equalled zero (in the case of the log-transformation) and the heavy skewness of the distribution (in the that of the Box-Cox transformation). Effect sizes are provided for all tests, and their interpretation is based on sensitivity analyses to detect the minimum effect size of interest, which were performed on G*Power 3.1 (Faul et al., 2007) with the following parameters: two-tailed test; alpha = 0.05; power = 0.95; sample sizes for groups 1 and 2 = their respective n. Significant results with effect sizes below the minimum effect size of interest are considered irrelevant.

To estimate the effect of literality and the effect of the *defaultness* of the translations and their degree of literality (i.e., their interaction) on the word-level indicators, two procedures were used. In the case of number of typos per word and word typing speed, generalised linear models (GLMs; Nelder & Wedderburn, 1972) with a Tweedie distribution (Tweedie, 1984) and a log link function were used. A Tweedie distribution is a case of exponential dispersion models and combines the characteristics of distributions such as the normal, the gamma, and the Poisson. This makes it useful when a distribution presents a positive mass at zero but is otherwise continuous, as is the case of the three word-level indicators of cognitive effort, which

act as outcome variables in the GLM procedures. In the case of word typing speed, GLMs with a gamma distribution and a log link function were used (Hammouri et al., 2020). Such models are suitable when dealing with continuous data that do not meet the assumptions of normality and constant variance, which are prerequisites for many traditional statistical models. The log link function is a mathematical transformation that helps to linearize relationships between variables and manage the issue of heteroscedasticity. It ensures that the predicted values of the response variable are always positive, which is particularly useful when the response variable represents quantities that cannot be negative, such as counts, times, or amounts.

All statistical analyses were performed using Python, and the corresponding scripts, along with the datasets used for the analysis, are available in the supplementary Open Data material.[4]

2.6. Ethical issues

Participants signed an informed consent prior to their participation in the study, which described: (1) the project; (2) data-collection procedures; (3) data-processing procedures; (4) data-storing procedures; (5) ways in which data was going to be disseminated and presented in research outputs; (6) risks for the participants (no risks were foreseen); (7) compensation (no economic compensation); (8) rights of the participant, and (9) how to contact the researcher who is responsible for the collection, processing and storage of data in this study.

3. Results and discussion

In this section, results on the following aspects are presented: (1) word-level indicators of effort; (2) literality indicators; (3) the crossings between word-level indicators of effort and literality indicators, and (4) the effect of the *defaultness* of the translations and their degree of literality on word-level indicators of effort.

3.1. Word-level indicators of effort

Overall, the three word-level indicators of effort show that the translation process was rather fast, uninterrupted by typos, and with short within-word pauses (Table 4). When comparing the two texts, a significant difference was detected for the three indicators (typos: $U = 294,456$; $p < 0.001$; $r = 0.055$; speed: $U = 246,922$; $p < 0.001$; $r = 0.115$; pause: $U = 245,605$; $p < 0.001$; $r = 0.120$), although the effect sizes were smaller than the minimum effect size of interest ($r = 0.193$) and therefore are considered irrelevant. Hence, no differences in effort are observed in the translation process of *The Story of Sinuhe* and *The Story of the Shipwrecked Sailor*. The same applies to the comparison between students and non-students (typos: $U = 305,148.5$; $p = 0.011$; $r = 0.040$; speed: $U = 258,515$; $p < 0.001$; $r = 0.120$; pause: $U = 284,033$; $p = 0.271$; $r = 0.033$; minimum effect size of interest: $r = 0.189$). Hence, there seems to be no difference in terms of effort between the two texts and the student vs. non-student condition.

Table 4. Word-level indicators of effort – overall results and results by text and by student vs. non-student condition

		Typos per word	Word typing speed	Within-word pause
Global (N = 1537)		m = 0.02; Mdn = 0.00; SD = 0.08	m = 0.28; Mdn = 0.16; SD = 0.48	m = 0.22; Mdn = 0.14; SD = 0.48
Text	Sailor (n = 588)	m = 0.03; Mdn = 0.00; SD = 0.09	m = 0.28; Mdn = 0.15; SD = 0.54	m = 0.21; Mdn = 0.13; SD = 0.53
	Sinuhe (n = 949)	m = 0.02; Mdn = 0.00; SD = 0.07	m = 0.28; Mdn = 0.17; SD = 0.44	m = 0.23; Mdn = 0.15; SD = 0.45
Student	yes (n = 710)	m = 0.03; Mdn = 0.00; SD = 0.08	m = 0.24; Mdn = 0.15; SD = 0.31	m = 0.19; Mdn = 0.14; SD = 0.28
	no (n = 827)	m = 0.02; Mdn = 0.00; SD = 0.08	m = 0.31; Mdn = 0.18; SD = 0.59	m = 0.25; Mdn = 0.14; SD = 0.60

3.2. Literality indicators

In all, 668 words (43.5%) of the target-text words (and more than 50% in the case of the excerpt from *The Story of Sinuhe*) could not be aligned with their source-text counterpart, as shown in the CrossS graphs included in Figure 5. This indicates that word alignment indicators are problematic, at least in this linguistic pair, due to extreme grammatical and syntactic differences. In all, 593 (38.5%) of target-text words had the same position as their source-text counterpart (CrossS in position 1), and the proportion rose to 574 (37.5%) in the case of source-text words compared to their target-text counterparts (CrossT in position 1). In CrossS, displaced target-text words were placed at a posterior position (165 words vs. 111 words in an anterior position; 59.8% vs. 40.2%) to the one occupied by their counterpart in the source text. CrossT graphs show that source-text words are heavily displaced when compared to their target-text counterparts (676 at a posterior position and 128 at an anterior one: 72,9% vs. 13.8%), which means that the syntactic distance between the two texts is high. Hence, given the assumptions of the effect of (the lack of) literality on cognitive effort, it would be expected to find very high levels of cognitive effort at a word level in the tails of the CrossS and CrossT distribution, similar to a U-shape in which the less effortful words to translate are the ones with CrossS and CrossT values closer to zero.

From the two texts, however, the alignment of the *The Story of the Shipwrecked Sailor* seems to indicate that the distance is lower than that of *The Story of Sinuhe*. However, the statistical comparison between both texts in terms of CrossS and CrossT yields statistically significant but irrelevant results in the case of CrossT given that the minimum effect size of interest of $r = 0.206$ was not reached (CrossS: $U = 96,758.5$; $p = 0.063$; $r = 0.072$; CrossT: $U = 80,977.5$; $p < 0.010$; $r = 0.103$). As for the comparison between students and non-students, non-significant results were obtained (CrossS: $U = 108,770$; $p = 0.477$; $r = 0.025$; CrossT: $U = 100,400$; $p = 0.146$; $r = 0.054$).

As for perplexity and entropy, they are significantly higher in *The Story of Sinuhe* than *The Story of the Shipwrecked Sailor* (perplexity $U = 70,079$; $p < 0.001$; $r = 0.224$; entropy: $U = 70,079$; $p < 0.001$; $r = 0.224$), but no difference was detected when comparing students and non-students (perplexity: $U = 109,096.5$; $p = 0.456$; $r = 0.028$; entropy: $U = 109.096,5$; $p = 0.456$; $r = 0.028$).

**Figure 5.** Descriptive results of the four literality indicators
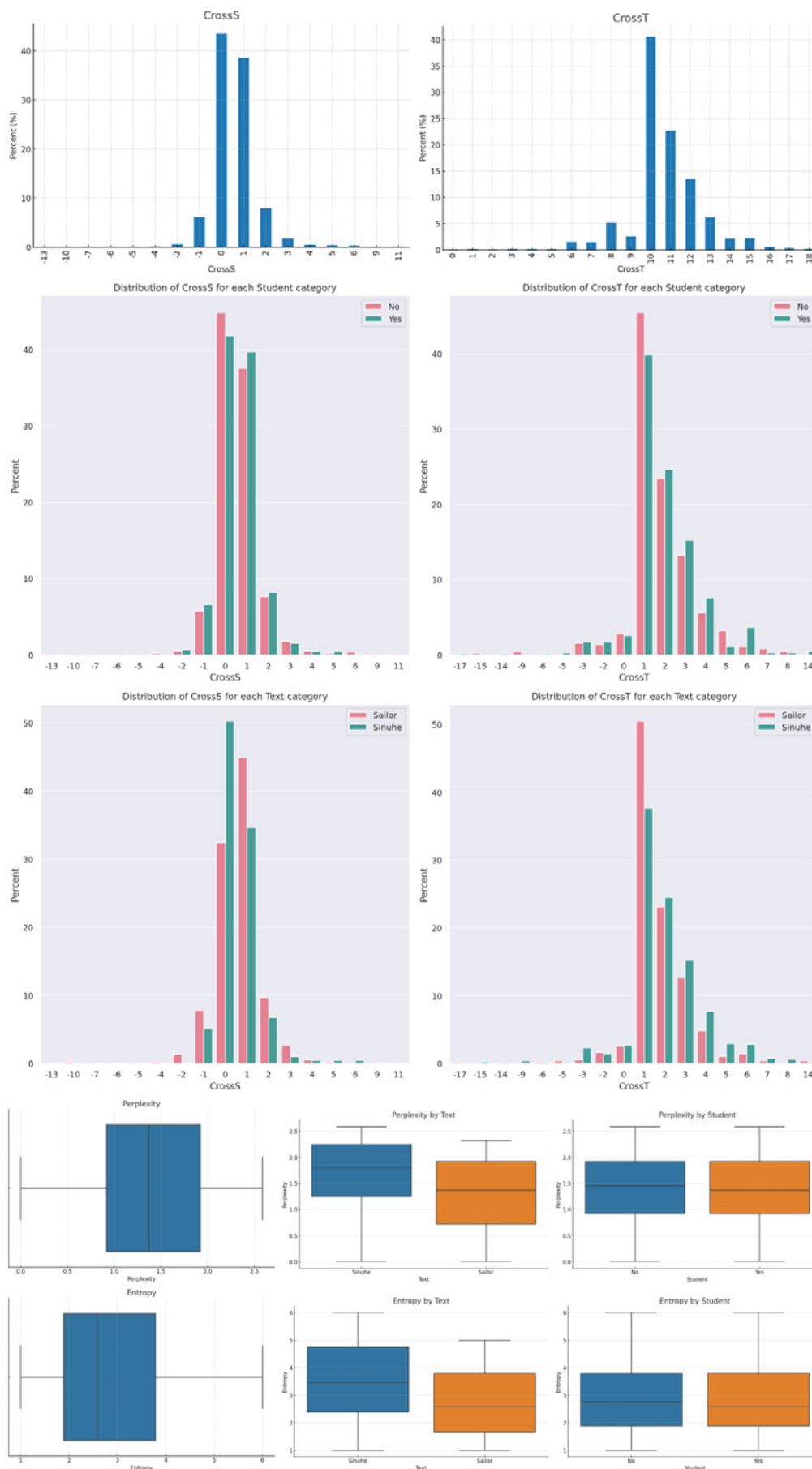
3.3. Crossings between word-level indicators of effort and literality indicators

3.3.1. Number of typos per word

The results of the GLM procedure indicate that there is no evidence in the data that the literality indicators are predictors of the number of typos per word, since no literality indicator is a significant predictor (z for intercept = 14.795; $p < 0.001$; z for CrossS = 1.521; $p = 0.128$; z for CrossT = 1.162; $p = 0.245$; z for perplexity = 1.547; $p = 0.122$; z for entropy = 1.550; $p = 0.121$).

There appears to be a slight, non-significant decrease of typos per word as CrossS and CrossT increase (Figure 6). Hence, word order differences do not increase effort in terms of number of typos per word. In the case of entropy and perplexity, results show no relationship between typos and the number of possible renditions for a given word.

Figure 6. Relationship between the number of typos per word and literality
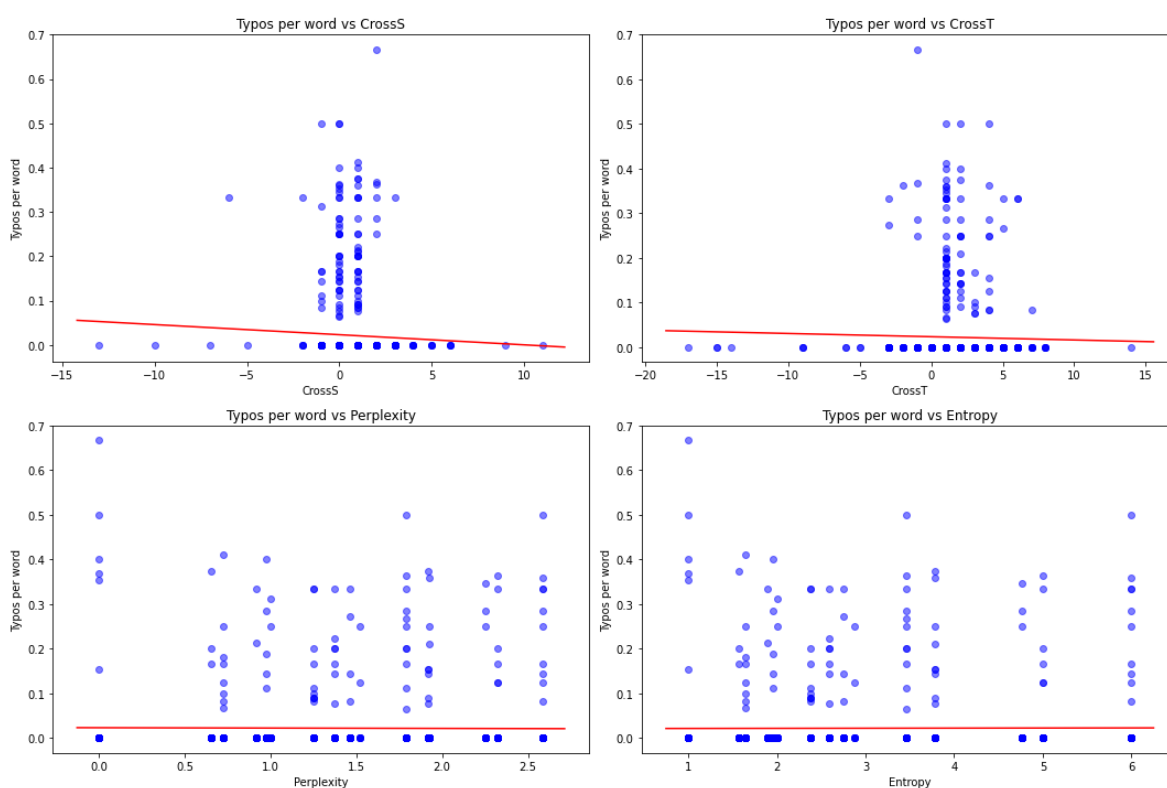
3.3.2. Word typing speed

The results obtained for word typing speed resonate with the ones obtained in the previous indicator: the literality indicators do not act as predictors of word typing speed given that none of the indicators resulted in being significant effects (z for constant = 6.797; $p < 0.001$; z for CrossS = 1.152; $p = 0.249$; z for CrossT = 0.582; $p = 0.560$; z for perplexity = 0.215; $p = 0.830$; z for entropy = 0.307; $p = 0.759$). This observation is corroborated by Figure 7, which illustrates that word order and the number of potential renditions of a specific source-text word bear no relation to word typing speed.
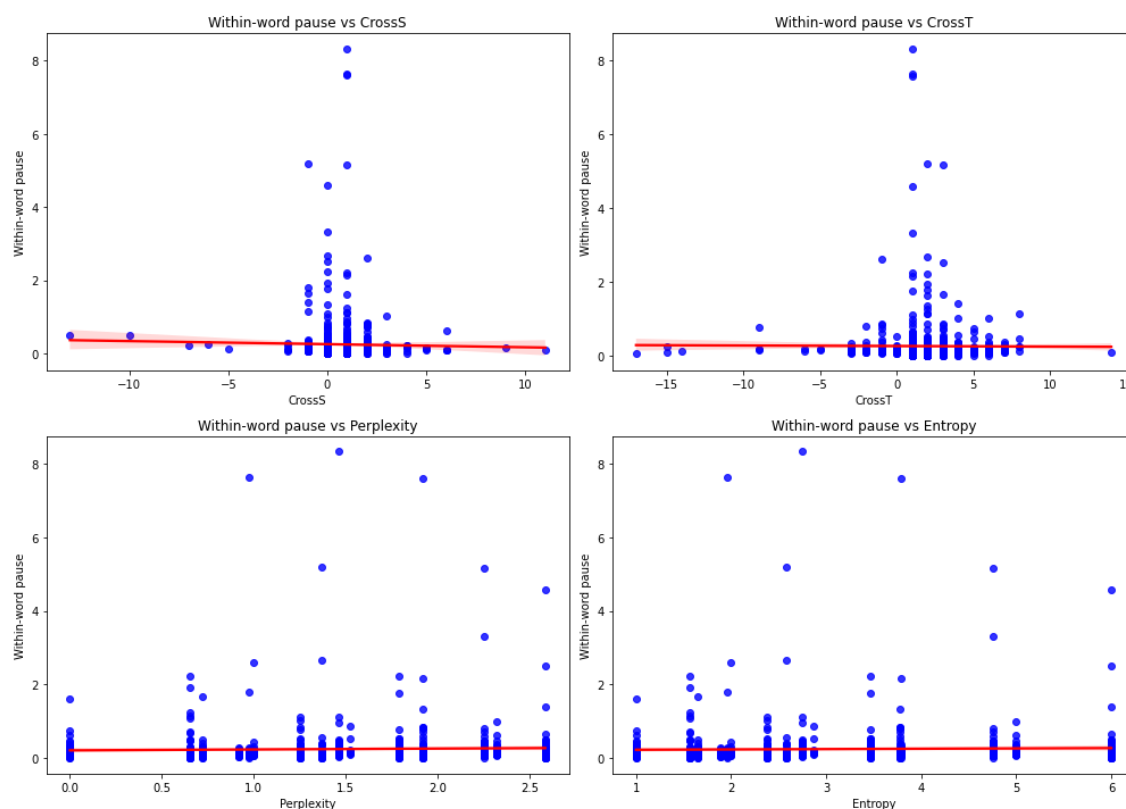
Figure 7. Relationship between word typing speed and literality
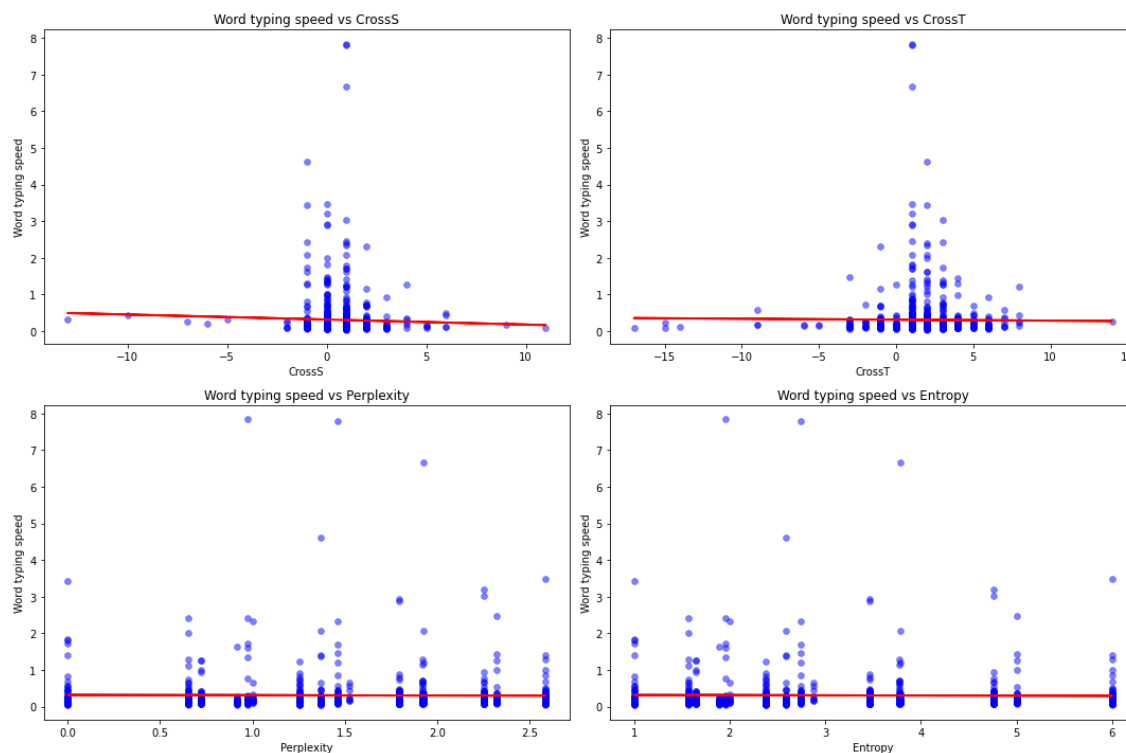
3.3.3. Within-word pause

Again, the literality indicators are not predictors of within-word pauses (z for constant = 7.431; $p < 0.001$; z for CrossS = 0.562; $p = 0.574$; z for CrossT = 0.436; $p = 0.663$; z for perplexity = 0.900; $p = 0.368$; z for entropy = 0.700; $p = 0.484$).

Figure 8. Relationship between within-word pause and literality

From a descriptive perspective (Figure 8), the relationship between within-word pause and CrossS and CrossT echoes the patterns observed in the first indicator, the number of typos per word: a slight, non-significant reduction of within-word pauses as CrossS and CrossT increase. Hence, translating words with (very) different word order in the source text compared to the target text (or the other way round) does not imply higher effort in terms of within-word pauses. As for entropy and perplexity, no relationship is observed. Again, this contradicts the assumptions of translation literality on word-level cognitive effort.

3.4. Effect of the defaultness of the translations and their degree of literality on word-level indicators of effort

As observed in the previous section, differences in word order and an increased number of possible translation renditions did not imply a higher degree of cognitive effort when translating at a word level. This section focuses on the interaction between the default translation vs. non-default translation and the four literality indicators and its effect on the three word-level indicators of effort.

### 3.4.1. Number of typos per word

The GLM results indicate that there is no significant effect of the interaction between the *defaultness* of translations and the degree of literality on the number of typos per word (Table 5).

Table 5. Test of model effects – number of typos per word

	z	*p*
(Intercept)	11.977	0.001
CrossS	1.283	0.200
CrossT	0.420	0.675
Perplexity	0.944	0.345
Entropy	0.906	0.365
Default	1.074	0.283
Default * CrossS	0.043	0.966
Default * CrossT	0.742	0.458
Default * Perplexity	1.321	0.187
Default * Entropy	1.017	0.309

Figure 9 presents a graphic representation of these interactions. In the case of default translations, none of the literality indicators shows much variability in the data, which means that, independently of how similar the syntactic structure of the source and the target texts are and the number of possible renditions, the number of typos per word is constantly low when renditions are default translations. In the case of the non-default translations, CrossS and CrossT show more variability since the ratios of typos per word are more dispersed. This is not the case with entropy and perplexity, where no relationship is found for both indicators of literality.

Figure 9. Effect of the interaction between the default vs. non-default translation condition and each literality indicator on the number of typos per word
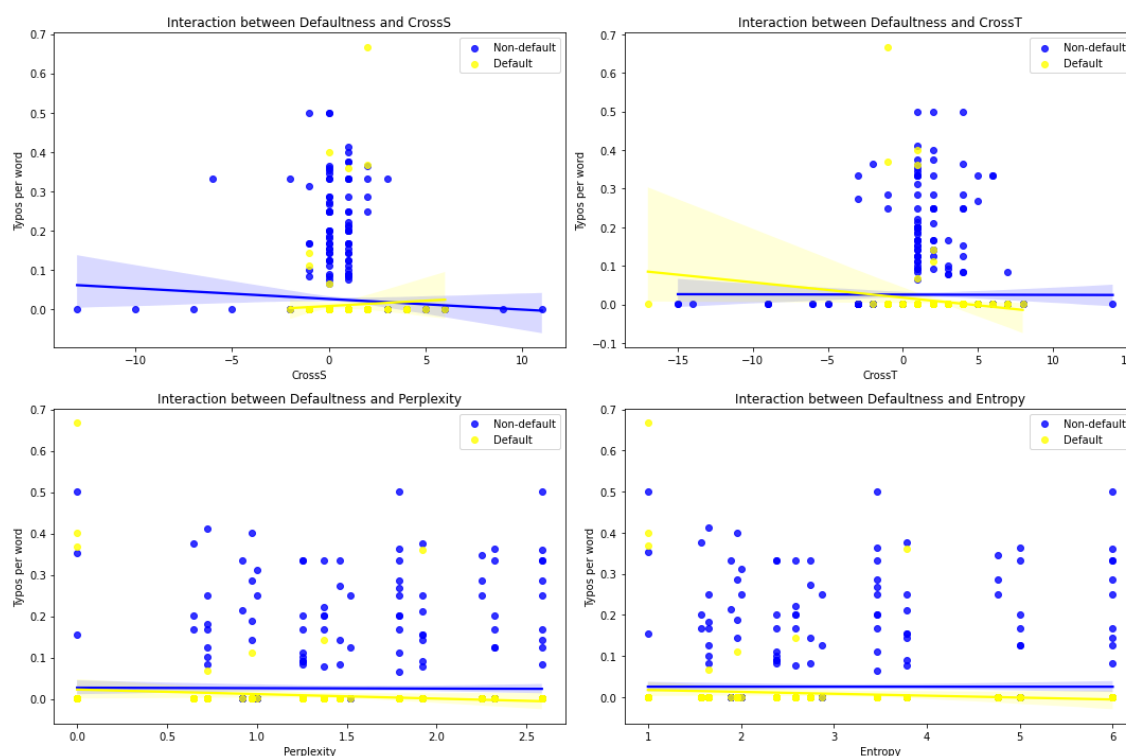
3.4.2. Word typing speed

Only a single variable (the default vs. non-default translation condition; parameter estimate for non-default vs. default: exp[B] = 1.106; 95% CI [0.402; 1.811]) in addition to the intercept was a statistically significant effect (Table 6). Hence, effort in terms of word typing speed is independent of the *defaultness* of a rendition and the interaction with its degree of literality.

Table 6. Test of model effects – word typing speed

	z	p
(Intercept)	8.423	0.001
CrossS	0.832	0.406
CrossT	0.390	0.696
Perplexity	0.822	0.411
Entropy	0.535	0.593
Default	3.078	0.002
Default * CrossS	0.705	0.481
Default * CrossT	0.318	0.750
Default * Perplexity	0.481	0.630
Default * Entropy	1.042	0.297

The word typing speed was more homogeneous in the non-default condition for all levels of both CrossS and CrossT than in the default condition (Figure 10). Word typing speed was faster in the non-default condition, even if ADD-only task segments containing default translations were faster than ADD-only task segments without them. This may indicate that

non-default translations are produced in bursts of speed (possibly caused by a longer pre-word pause), while the speed at which default translations are produced is more variable but without interrupting the typing flow so as not to break the task segment, possibly due to the ongoing planning of the process while typing.

Even if not significant, in default translations, word typing speed is more variable in all levels of syntactic similarity between the source and target text, but not so much in the non-default condition. As for the number of possible renditions, default translations perform similarly to non-default translations in terms of typing speed, but higher variation is detected in higher levels of perplexity and entropy (i.e., the less variety of possible renditions). Therefore, this would contradict the assumption that fewer possible renditions from which to choose from lowers cognitive effort.

Figure 10. Effect of the interaction between the default vs. non-default translation condition and each literality indicator on word typing speed
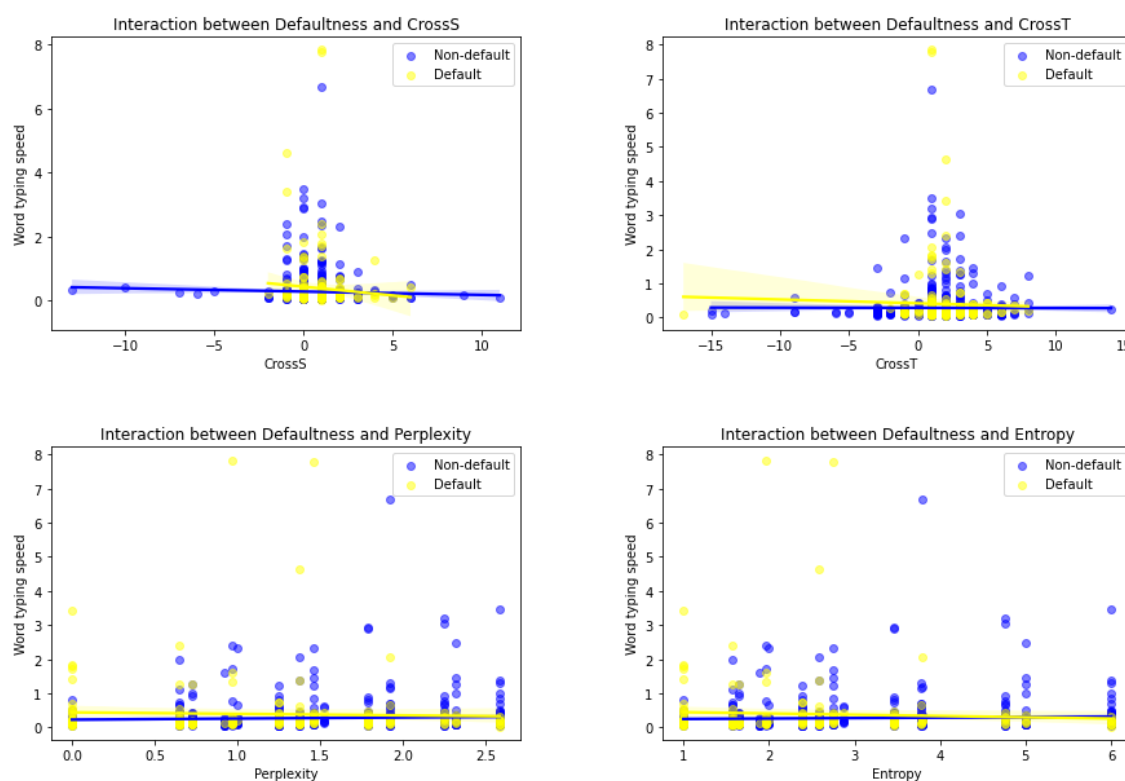

3.4.3. Within-word pause

Again, only the default vs. non-default translation condition (parameter estimate for non-default vs. default: exp[B] = 1.076; 95% CI [0.145; 2.005]) in addition to the intercept was a statistically significant effect (Table 7). Hence, effort in terms of word typing speed is independent of the *defaultness* of a rendition and the interaction with its degree of literality.

Table 7. Test of model effects – within-word pause

	z	p
(Intercept)	7.424	0.001
CrossS	0.428	0.669
CrossT	0.368	0.713
Perplexity	0.253	0.800
Entropy	0.080	0.937
Default	2.266	0.023
Default * CrossS	0.548	0.584
Default * CrossT	0.509	0.611
Default * Perplexity	1.680	0.093
Default * Entropy	1.828	0.068

Figure 11 presents two different patterns. The first one refers to CrossS and CrossT. Even if within-word pauses show more variation in default translations than in non-default ones, the effect of the defaultness of the translations and that of CrossS and CrossT on word typing speed is not observable. Hence, there is no effect on the degree of syntactic similarity on within-word pauses for both default and non-default translations. The second pattern concerns perplexity and entropy. Here the difference between default and non-default translations in relation to within-word pauses is non-existant.

Figure 11. Effect of the interaction between the default vs. non-default condition and each literality indicator on within-word pause
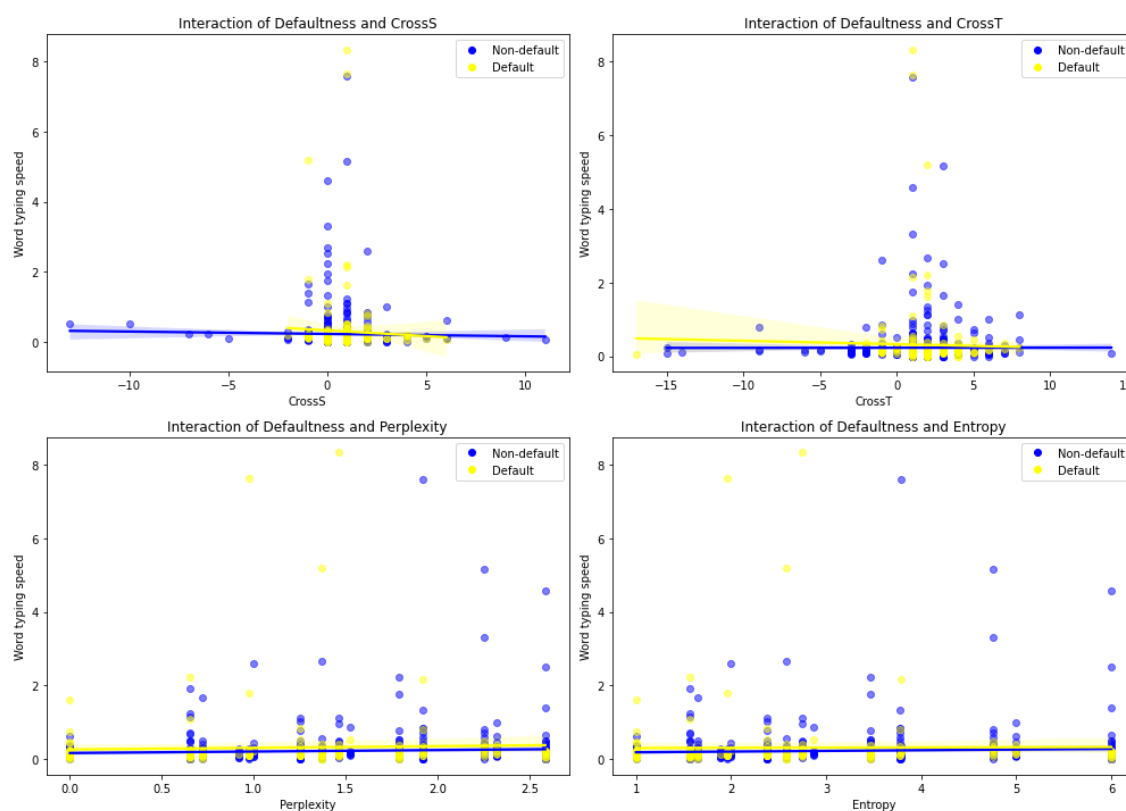
4. Conclusions

As any exploratory study is expected to do, this investigation has raised more questions than answers. Despite the heavily displaced syntactic and word order between source and target texts, word-level indicators of effort were unaltered. Translations were effortful (and effortless) in all degrees of syntactic similarity. The exploratory nature of this study prevents us from establishing definite conclusions, but these results could question the assumption that deviations from a similar word order cause higher cognitive effort. Therefore, in the limited context of our study, the findings seem to align with those of Jiménez Crespo and Casillas (2021). It may be contented that our results do not invalidate that assumption, given that the indicators of effort were different from the ones that lead to its formulation. Indeed, no eye-tracking-based indicators were used, but this behavioural measure of cognitive effort has also been questioned recently. One of the main critiques is the unquestioned acceptance of the mind-eye hypothesis (Just & Carpenter, 1980), originally formulated for reading comprehension processes, which posits that what we fixate on is closely related to what we are processing. Longer fixations may be due to the processing of a translation unit, but not necessarily, as they may be due to neurophysiological (Reichle & Reingold, 2013) and attentional constraints (Faber et al., 2020). Additionally, there have been calls to distinguish between *effort* and *arousal* measured with eye-tracking indicators (Gieshoff, 2021).

In addition to being non-significant, there is a conceptual aspect that the assumption between entropy/perplexity and cognitive effort should solve. These are product-oriented indicators computed based on the number of different renditions that a group of participants have provided. However, it is a stretch to assume that all these renditions were actively considered by each participant. Not all translators come up with the same solutions: one may come up with a single one (a default translation?), while others may come up with multiple solutions. Perplexity and entropy may be used as a product indicator, but their use as a proxy of what is going on in the translator's mind is troublesome, as it depicts the translation process as if someone was ordering at a restaurant by looking at a menu: all options are there, active, and the translator only needs to choose the more appropriate one.

This brings us back to the differentiation between cognitive load and effort (Hunziker Heeb et al., 2021). If cognitive load is associated with the source text (among others), literality, and more specifically, word order and syntactic differences, may play a relevant role in the translation process. But cognitive effort, i.e., the actual responses by the task performer, will be highly individual. Based on the results from this small sample, it seems plausible to suggest that individual responses to this parameter of cognitive load could explain why the three indicators of cognitive effort that were examined were not impacted by literality.

Individuality is one of the defining features of default translation, as suggested by Halverson (2019). Another one is the irrelevance of the characteristics of the resulting relationship between the source and the target text (i.e., whether the relationship is based on a word-for-word translation). This is exactly what the lack of effect of the interaction between the *defaultness* of the translations and their degree of literality on word-level indicators of effort showed. While the construct of literal translation and the assumption of its effects on cognitive effort were not observable, the very same data showed that the translation process can be divided into stretches of uninterrupted, (almost) effortless production and stretches of interrupted and/or effortful

production. Here, the link between default translations and cognitive effort has been clearly established (low number of typos, low number of respites, and fast writing speed). However, it was not possible to describe the translation process as literal and effortless vs. non-literal and effortful. Hence, based on our limited findings, it could be suggested that default translation may serve as a potentially more adequate description of the way translators perform.

The study presented in this paper has many limitations, many of them due to its exploratory nature. The use of Middle Egyptian as the source-text language was useful to put both constructs to an extreme test, but at the same time, it limited access to a sample of professional translators, which could have provided different results. In addition to the limited sample size (despite the more than 1500 data points included in the analysis), there are several issues related to the impossibility of using the same indicators and procedures as in the original work by Schaeffer and Carl (2014) and Carl and Schaeffer (2017b), and in that of Halverson (2019) and Halverson and Muñoz (2019a, 2019b). Additionally, it was necessary to adapt the ways in which source and target texts were aligned to compute the four literality indicators to bypass the methodological issues discussed in section 2. In summary, this study calls for a confirmatory replication study with a larger sample size and refined procedures to analyse literality and identify default translations.

References

Allen, J. P. (2015). *Middle Egyptian Literature*. Cambridge University Press.

Carl, M. (2010). A computational framework for a cognitive model of human translation processes. In *Proceedings of the 32nd International Conference on Translating and the Computer. 18–19 November 2010, London*. https://aclanthology.org/2010.tc-1.10.pdf

Carl, M. & Schaeffer, M. (2017a). Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *Hermes – Journal of Language and Communication in Business*, 56, 43–57.

Carl, M. & Schaeffer, M. (2017b). Measuring translation literality. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology* (pp. 81–106). John Benjamins.

Carl, M., Schaeffer, M. & Bangalore, S. (2016). The CRITT Translation Process Research Database. In M. Carl, S. Bangalore & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* (pp. 13–54). Springer.

Catford, J. C. (1965). *A Linguistic Theory of Translation*. Oxford University Press.

Cervelló Autuori, J. (2015). *Escrituras, lengua y cultura en el Antiguo Egipto*. Edicions UAB.

Clark, A. (1996). *Being There: Putting Brain, Body, and World Together Again*. The MIT Press.

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. Jo*urnal of Experimental Psychology: Human Perception and Performance, 46*(10), 1201–1221.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Foster, J. L. (1980). Sinuhe: The ancient Egyptian genre of narrative verse. *Journal of Near Eastern Studies, 39*(2), 89–117.

Gieshoff, A. C. (2021). Does it help to see the speaker's lip movements? An investigation of cognitive load and mental effort in simultaneous interpreting. *Translation Cognition & Behavior, 4*(1), 1-25.

Gilbert, D., Toledo-Báez, C.; Carl, M. & Espino, H. (2023). Impact of word alignment on word translation entropy and other metrics. A comparison of translation process research findings derived from different word alignment methods. In I. Lacruz (Ed.), *Translation in Transition: Human and Machine Intelligence* (pp. 203-235). John Benjamins.

Halverson, S. L. (2019). 'Default' translation. A construct for Cognitive Translation and Interpreting Studies. *Translation, Cognition & Behavior, 2*(2), 187–210.

Halverson, S. L. & Muñoz Martín, R. (2019a). Default translation in Cognitive Translation Studies. A construct and methodological issues. In *6th International Conference on Cognitive Research on Translation and Interpreting. 8-10 November 2019, Southwest University, Chongqing, China*.

Halverson, S. L. & Muñoz Martín, R. (2019b). Default translation in the wild. In *9th EST Congress, 9-13 September, Stellenbosch University, RSA*.

Hammouri, H. M.; Sabo, R. T., Alsaadawi, R. & Kheirallah K. A. (2020). Handling skewed data: A comparison of two popular methods. *Applied Sciences 10*(18): 6247.

House, J. (1977). *A Model for Translation Quality Assessment*. TBL-Verlag Narr.

Hunziker Heeb, A; Lehr, C. & Ehrensberger-Dow, M. (2021). Situated translators: Cognitive load and the role of emotions. In R. Muñoz Martín, S. Sun & D. Li (Eds.), *Advances in Cognitive Translation Studies* (pp. 47-65). Springer Singapore.

Ivir, V. (1981). Formal correspondence vs. translation equivalence revisited. In I. Even-Zohar & G. Toury (Eds.), *Theory of Translation and Intercultural Relations* (pp. 51–59). Porter Institute for Poetics and Semiotics, Tel Aviv University.

Jiménez-Crespo, M. Á. & Casillas, J. V. (2021). Literal is not always easier. Literal and default translation, cognitive effort, and comparable corpora. *Translation, Cognition & Behavior, 4*(1), 98-123.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Leijten, M. & Van Waes, L. (2013). Keystroke logging in writing research. *Written Communication, 30*(3), 358–92.

Muñoz Martín, R. (2016). Looking toward the future of Cognitive Translation Studies. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 555–572). Wiley Blackwell.

Muñoz Martín, R. & Apfelthaler, M. (2022). A Task Segment Framework to study translation processes. *Translation & Interpreting, 14*(2), 8-31.

Muñoz Martín, R. & Cardona Guerra J. M. (2019). Translating in fits and starts: Pause thresholds and roles in the research of translation processes. *Perspectives: Studies in Translation Theory and Practice, 27*(4): 525–551.

Muñoz Martín, R. & Martín de León, C. (2018). Fascinatin' rhythm – and pauses in translators' cognitive processes. *HERMES - Journal of Language and Communication in Business*, 57, 29–47.

Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*(3), 370.

Nord, C. (1991). *Text Analysis in Translation*. Rodopi.

Parkinson, R. B. (ed.) (1998). The Tale of Sinuhe *and other ancient Egyptian poems, 1940 - 1640 BC*. Oxford University Press.

Reichle, E. D. & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Front. Hum. Neurosci. 7*, 361.

Rosmorduc, S. (2014). *JSesh Documentation*. Available at: http://jseshdoc.qenherkhopeshef.org

Schaeffer, M & Carl M. (2014). Measuring the cognitive effort of literal translation processes. In U. Germann et al. (Eds.) *Proceedings of the Workshop on Humans and Computer-Assisted Translation (HaCaT)* (pp. 29–37). Association for Computational Linguistics.

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W. & Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In M. Carl, S.

Bangalore & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research* (pp. 183-210). Springer International Publishing.

Tirkkonen-Condit, S. (2005). The monitor model revisited: Evidence from process research. *Meta, 50*(2): 405–414.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In J. K. Ghosh & J. Roy (Eds.), *Statistics: Applications and New Directions—Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (pp. 579–604). Indian Statistical Institute.

Notes

1. The pre-registration and the study materials can be accessed here: https://osf.io/9d6ra/

2. The process data from a sixth participant was discarded since they initially translated the texts by hand and subsequently typed their translations into the computer using InputLog for keylogging. However, their product data were still used to generate alignments between source-text and target-text words, thereby enhancing the pool of potential renditions.

3. As stated in the pre-registration, however, the aim was to use target-text sentences as alignment units to make the analytical procedure as close as possible to the one used by Schaeffer and Carl. When data analysis began, it quickly became apparent that such a procedure was inadequate, given the large differences among the many ways in which target-text sentences were structured.

4. The Open Data material can be accessed here: https://doi.org/10.5281/zenodo.8144514