Entropy as a measurement of cognitive load in translation

Yuxiang Wei   yuxiang.wei3@mail.dcu.ie
Centre for Translation and Textual Studies, Dublin City University, Ireland

Abstract

In view of the "predictive turn" in translation studies, empirical investigations of the translation process have shown increasing interest in studying features of the text which can predict translation efficiency and effort, especially using large-scale experimental data and rigorous statistical means. In this regard, a novel metric based on entropy (i.e., HTra) has been proposed and experimentally studied as a predictor variable. On the one hand, empirical studies show that HTra as a *product*-based metric can predict effort, and on the other, some conceptual analyses have provided theoretical justifications of entropy or entropy reduction as a description of translation from a *process* perspective. This paper continues the investigation of entropy, conceptually examining two ways of quantifying cognitive load, namely, shift of resource allocation and reduction of entropy, and argues that the former is represented by surprisal and ITra while the latter is represented by HTra. Both can be approximated via corpus-based means and used as potential predictors of effort. Empirical analyses were also conducted comparing the two metrics (i.e., HTra and ITra) in terms of their prediction of effort, which showed that ITra is a stronger predictor for TT production time while HTra is a stronger predictor for ST reading time. It is hoped that this would contribute to the exploration of dependable, theoretically justifiable means of predicting the effort involved in translation.

1. Introduction

In recent years, process-oriented translation studies which investigate the "black box" of the translator's mind have been prolific and less of a speculative nature, due to the emergence of new methodologies for collecting, processing, and analysing behavioural data. While early research depends heavily on think-aloud protocol, more recent ones tend to adopt relatively sophisticated techniques including eye tracking, electroencephalography (EEG), functional magnetic resonance imaging (fMRI), etc. Such experimental tools have largely enabled translation process research (TPR) to become increasingly predictive (Schaeffer et al., 2019). Large-scale, multilingual, and comparable behavioural data collected via these tools (e.g., the CRITT TPR-DB; see Carl, Schaeffer et al., 2016), and analysed through rigorous statistical approaches, have provided a necessary means for building models of human translation "which makes specific, falsifiable predictions regarding the process and the product of translation" (Carl, Bangalore et al., 2016, p. 4). This allows for systematic investigations beyond the *description* of translation, taking a step further towards *explaining*, and especially *predicting*, translation phenomena from empirical observations.[1]

---

[1] When Holmes (1972) argued for an independent academic status for translation studies, it was described as an empirical discipline in nature, where there are two main objectives of inquiry: "(1) to describe the phenomena of translating and translation(s) as they manifest themselves in the world of our experience, and (2) to establish general principles by means of which these phenomena can be explained and predicted." (Quoted from the republished version of Holmes' paper in Venuti, 2000, p. 176)

Not surprisingly, it has been argued that a "predictive turn" is now being triggered in translation studies, constituting a new paradigm where predictive methods and models, driven by large-scale empirical data, are adapted to the cognitive processes of translation (Schaeffer et al., 2019).

This is the result of two aspects of technological development, namely, the machine learning approaches to translation (e.g., Neural Machine Translation) and the computational techniques that facilitate the empirical modelling of the human translation process (ibid). For the latter, the fact that many aspects of behaviour and cognition have become increasingly measurable and quantifiable (e.g., translators' strategies, typical translation patterns, and cognitive effort), and the use of rigorous statistical and computational tools, seem to have made it possible "for the first time to empirically model the translation process" (ibid, p. 5).

1.1 Entropy as a predictor variable

In view of this predictive turn, there has been increasing interest in investigating, especially by statistical means, particular features of the text that can predict the efficiency and cognitive load/effort [2] of translation, post-editing, interpreting, and other modes of translation production. These studies examine the translation product in relation to those aspects of the process which can be used as measurements of translation efficiency or difficulty. For example, eye-key span has been shown to be predicted by the number of translation alternatives for the ST word in question (Dragsted, 2010; Dragsted and Hansen, 2008), and reading time can be predicted by the change of word order between the ST and TT, the number of occurrences of the word in previous context, the length of phrases, etc. (Jensen et al., 2009)

Another novel metric which has been recently proposed and empirically examined is word translation entropy (see, e.g., Carl, Schaeffer, et al., 2016 p. 29-33). This entropy-based predictor variable, often denoted as HTra, is typically considered a statistical measure of the translation product which represents variance, literality, and translation ambiguity (Carl, 2021b; Carl, Bangalore, et al., 2016), and is used in many empirical investigations to analyse its correlation with effort, to find evidence for early priming processes, and to discuss ways of quantifying translation difficulty. It has also been considered a better measure for the variation of the translation alternatives than simply counting the number of these alternatives (Bangalore et al., 2016). Further studies on word translation entropy show a positive and statistically significant effect on different measures of effort, including, among others, first fixation duration, word production duration, the probability of a fixation, and total reading time (e.g., Carl and Schaeffer, 2017; Schaeffer et al., 2016). In other words, HTra predicts effort. On the basis of such empirical findings, words with higher HTra values have often been considered more difficult to translate (Carl et al., 2019).

1.2 Entropy as a mental process

For such and many other studies, the concept of entropy seems to be consistently used as a measure of the *product*, rather than as a representation of specific aspects of mental states during the process, nor as a way of describing the process of transition between one mental

---

[2] Although the terms "cognitive load" and "cognitive effort" can sometimes be confusing and are often used interchangeably, this paper considers cognitive *load* as the difficulty that is posed by a task or process (i.e., the required amount of cognitive effort), and considers cognitive *effort* as the actual effort expended in the process or task, where this effort is realised by optimising the allocation of limited cognitive resources.

state and another. An exception, however, is the "systems theory perspective" (Carl et al., 2019), where the human translation process is considered "a hierarchy of interacting word and phrase translations systems which organize and integrate as dissipative structures" (p. 211), and where entropy is defined as the internal order of these word translation systems. Expenditure of cognitive effort to arrive at a translation solution — where this effort is described as "average energy" (ibid) — decreases the internal entropy (i.e., disorder) of the system. In this regard, the definition of entropy is apparently from a systems theory perspective.

In terms of the conceptual investigations of entropy in relation to the mental states, Wei (2021) analyses translation entropy from a different perspective, focussing more on the probabilistic nature of this concept (as Kullback-Leibler divergence, see Kullback, 1959), the dynamic change of probability distribution, the uncertainty of choice, its representation of cognitive resource allocation in the activation, suppression, competition, and selection of candidates when multiple options are available (i.e., when the ST is translation-ambiguous), and the specific processes in which entropy is reduced through the transition of mental states. The process of lexical translation selection is analysed in close detail through the lens of entropy and entropy reduction. This brings the concept into the assumed mental states, using entropy to describe and explain cognitive activities when mental states transition between one another during lexical activation and selection. Following these conceptual explorations, Wei's (2021) study also examines the behavioural manifestations of this process through detailed observation of eye movements in a large database (i.e., the CRITT TPR-DB).

In Wei's (2021) analyses, the mental processes in translation are conceptualised under the assumption of non-selective co-activation of both source and target languages, similar to most studies that draw inferences from bilingualism. Upon encounter of a particular ST item, possible translations for this item would be subliminally co-activated, and the translator is assumed to "engage in an activation pattern where the activated items receive different degrees of priority for resource allocation" (p. 170). This pattern would then be dynamically updated during lexical selection, where there is continual shift of cognitive resource allocation as mental states transition from one towards another. The shift of resource allocation results in reduction of entropy and expenditure of cognitive effort. In this view, the amount of cognitive effort needed in the process (i.e., the cognitive load imposed) can thus be quantified via two means — either the shift of cognitive resource allocation, or the reduction of entropy (ibid).

The present paper examines these two ways of quantification, and argues that the shift of resource allocation can be represented by surprisal of the item selected (i.e., ITra, see below), and that the reduction of entropy can be represented by HTra (as formulated in Carl, Schaeffer, et al., 2016).

Section 2 briefly reviews the concept of surprisal, focusing on its conceptualisation as a means of quantifying cognitive load in psycholinguistics. This lays the foundation for the discussion on relative entropy in the subsequent section 3, where surprisal (also described as ITra in recent studies) will be shown to be equivalent to the *relative entropy* between the final and initial mental states of the translation choice. This means that the required amount of cognitive effort in the transition between these mental states can be determined by surprisal (ITra), if one adopts the formulation in resource-allocation processing difficulty.

Section 4 demonstrates that if one adopts another means for quantifying effort (i.e., reduction of entropy value), this effort would be represented by HTra.

Section 5 provides further discussion on HTra and ITra, leading to an empirical investigation in Section 6 where the two metrics are compared in terms of their prediction of translation effort. Section 7 ends the paper with concluding remarks.

2. Surprisal and ITra

In psycholinguistics, surprisal (i.e., negative logarithm of probability) is often used as an important quantification of cognitive load (Attneave, 1959; Hale, 2001; Levy, 2008, 2013; Levy and Gibson, 2013), especially in the context of structural disambiguation. The surprisal of a word in its context is considered a useful quantification of the cognitive effort required to process this word during online sentence processing (see Hale, 2001). This is because, from that view, incremental sentence comprehension is a step-by-step disconfirmation of possible phrase-structural analyses for the sentence, which means that cognitive load can be interpreted as the combined difficulty of disconfirming the disconfirmable structures at a particular point of the sentence (i.e., at a given word).

This quantification of cognitive load also raises "a unified treatment of structural ambiguity resolution and prediction-derived processing benefits" (Levy, 2013 p. 158). Both Hale (2001) and Levy (2008) illustrate much successful use of the surprisal framework for explaining a variety of psycholinguistic phenomena, many of which are closely relevant to garden-path sentences (i.e., temporary ambiguity). In addition, theoretical justifications for surprisal as a metric for cognitive processing difficulty has not been lacking (see e.g., Levy, 2013), especially within the frameworks of rational cognitive models (Shepard, 1987; Tenenbaum and Griffiths, 2001). Difficulty, or measurable disruption, in real-time sentence processing can arise either from an overload in memory (i.e., an overload in the cognitive resources for the storage and retrieval of the representational units which are used to analyse the linguistic input), or from a sufficiently unexpected input which causes a shift in cognitive resource allocation "to various alternatives in the face of uncertainty" (Levy, 2013 p. 144). Although theories based on the former (i.e., resource-limitation theories) have been a dominant paradigm for studies of differential processing difficulty, the latter (i.e., resource-allocation approach) has been a line of investigation which largely has ambiguity resolution as a primary concern (Levy, 2008).

In the latter approach (i.e., resource-allocation), the *size* of the shift in cognitive resource allocation which is induced by a word is indicative of the difficulty in processing this word, and the size of this shift is equivalent to the *change* (i.e., update) in the conditional probability distribution over all interpretations before and after the word (Levy, 2013). Mathematically, this change would be measured in terms of entropy (e.g., Cover and Thomas, 1991) — specifically, the *relative entropy* of the conditional distributions before and after encountering the word.

This seems largely consistent with the use of word translation entropy to measure the difficulty of a translation choice in the face of uncertainty (at the lexical, rather than syntactic, level), where this difficulty can be represented by the conditional probability distribution over TT alternatives.

Of particular note is that in sentence comprehension, the *relative entropy* mentioned above has been shown to be equivalent to the *surprisal* of the word in question (Levy, 2008 pp. 1131-1132), which Levy views as the reranking cost in incremental disambiguation where cognitive resources are re-allocated to the possible analyses of the sentence.

Here, it is worth mention that the concept of *surprisal* is also known — in different contexts — as information, self-information, or Shannon information content, all referring to essentially the same mathematical equation (i.e., the negative logarithm of probability). In some recent papers, the surprisal regarding a particular translation item is specifically called *word translation information*, and denoted by ITra (see e.g., Carl, 2021a; Heilmann and Llorca-Bofí, 2021). These terms, although focusing on quite different aspects, are in fact mathematically expressed in the same manner as the surprisal discussed here (i.e., the

negative logarithm of probability, or equivalently, the logarithm of the inverse of the probability).

3. ITra and relative entropy

As mentioned in 1.3, the cognitive effort that is required in the word translation selection process (i.e., the cognitive load imposed by this process) is proposed to be quantifiable by either the shift in resource allocation, or the reduction of entropy (see Wei, 2021 for details). The size of the shift in cognitive resource allocation would be represented mathematically by relative entropy (i.e., Kullback-Leibler convergence), whereas the reduction entropy would simply be the absolute difference of entropy values, regarding the initial and final stages of the process.

In other words, there are two ways of representing cognitive load via entropy — relative entropy and decrease of entropy. Here, the relative entropy of the mental state at the end of the process, with respect to the initial stage of activation, will be shown below as being equal to the surprisal (i.e., ITra) of the TT item eventually chosen by the translator.

At the end of the selection process (i.e., when the mental processing has arrived at a decision as to which particular target item is to be selected), the distribution of cognitive resources in the mental state can be reasonably assumed to have, after a series of continual update (or shift) which incurs cognitive effort, eventually concentrated on one single item (i.e., the item chosen by the translator) whose probability therefore equals 1 given this mental state. According to the definition of Kullback-Leibler divergence (i.e., relative entropy), the divergence of the updated distribution $Q(x)$ from the original distribution $P(x)$ equals the expectation of the logarithmic difference between $Q(x)$ and $P(x)$, with the expectation taken using $Q(x)$. Suppose there are n possible items in the mental lexicon (i.e., n values for x in $x \in \chi$), among which the item chosen by the translator is W, then the above description would mean that $Q(W)=1$, that $Q(x)=0$ when $x \neq W$, and that $P(x)$ represents the probabilities in the initial activation pattern for both $x=W$ and $x \neq W$. In this case, the divergence $D_{KL}(Q \parallel P)$ would be:

$$D_{KL}(Q \parallel P) = \sum_{x \in \chi} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= \sum_{i=1}^{n} Q(x_i) \log\left(\frac{Q(x_i)}{P(x_i)}\right)$$

$$= Q(W) \log\left(\frac{Q(W)}{P(W)}\right) + (n-1) \lim_{Q(x)\to 0+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= \log \frac{1}{P(W)} + (n-1) \lim_{Q(x)\to 0+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= -\log P(W) + (n-1) \lim_{Q(x)\to 0+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= -\log P(W) + (n-1) \lim_{Q(x)\to 0+} [Q(x) \log Q(x) - Q(x) \log P(x)]$$

$$= -\log P(W) + (n-1) \left[\lim_{Q(x)\to 0+} Q(x) \log Q(x) - \lim_{Q(x)\to 0+} Q(x) \log P(x) \right]$$

As $\lim_{Q(x)\to 0+} Q(x) \log P(x) = 0$ and

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas, Orlando, USA, September 12-16, 2022. Workshop 1: Empirical Translation Process Research

Page 79

$$\lim_{Q(x)\to 0^+} Q(x) \log Q(x) = \lim_{Q(x)\to 0^+} \frac{\log Q(x)}{\frac{1}{Q(x)}}$$

$$= \lim_{Q(x)\to 0^+} \frac{\frac{d}{d(Q(x))} \log Q(x)}{\frac{d}{d(Q(x))} \frac{1}{Q(x)}}$$

$$= \lim_{Q(x)\to 0^+} \frac{\frac{1}{Q(x) \ln 10}}{-\frac{1}{[Q(x)]^2}}$$

$$= -\lim_{Q(x)\to 0^+} \frac{Q(x)}{\ln 10}$$

$$= 0$$

,

it then follows that

$$D_{KL}(Q \parallel P) = -\log P(W)$$

In other words, the Kullback-Leibler divergence of these two distributions (i.e., the relative entropy between initial activation and final selection) equals the surprisal of the item that is eventually chosen by the translator, i.e., -log$P(W)$.

As the $P(W)$ in the surprisal equation here represents the probability of W in the initial activation pattern (i.e., when W is first activated together with all other items), this surprisal should in theory refer to the surprisal in the corresponding mental state at the initial stage, rather than the surprisal of the item in the textual material.

However, if the activation of lexical items is modulated by context and the frequency of the different meanings/translations (e.g., in the re-ordered access model, see, e.g., Duffy et al., 2001), the $P(x)$ which describes the mental state of activation would be the same as the probabilities that can be observed in the text. This means that the initial surprisal for this item W in the mental state, i.e., -log$P(W)$, can be approximated by, if not equivalent to, its surprisal in the text.

In this manner, the relative entropy with respect to the above mental process would be, albeit arguably, equal to the corresponding surprisal in the text. Cognitive load can thus be represented by this surprisal (consistent with Levy's formulation), and in turn approximated by corpus-based analyses. As mentioned in Section 2, this surprisal is the same as word translation information (ITra).[3]

4. HTra and decrease of entropy

Similarly, the initial entropy value in the mental state would be equal to the entropy value that is observed in the text (i.e., HTra). If the *decrease* of entropy value, i.e., the absolute difference between the two respective entropy values regarding the initial and final mental states, is used as a measurement of cognitive effort in the selection process, then at the point when the translation choice is made, this decrease would equal the initial entropy when all the

---

[3] It is important to note that the CRITT TPR-DB estimates this value on the basis of the translation choices made by all participants in each experiment. However, the surprisal here can in fact be approximated in other ways as well, using different corpus data, and would result in different ITra values than those in the CRITT TPR-DB. This is the same for HTra.

TT candidates are activated given the ST item (i.e., the entropy in the mental state between activation and selection), and in turn equal the HTra value. This will be shown below in detail.

Specifically, when the choice is made, the entropy in the mental state refers to the entropy for distribution $Q(x)$, which equals zero:

$$H_1(x) = -\sum_{x \in \chi} Q(x) \log Q(x)$$

$$= -\sum_{i=1}^{n} Q(x_i) \log Q(x_i)$$

$$= -Q(W) \log Q(W) - (n-1) \lim_{Q(x) \to 0^+} Q(x) \log Q(x)$$

$$= -Q(W) \log Q(W)$$

$$= 0$$

The initial entropy associated with the pattern of activated lexical items, i.e., the entropy in the initial mental state, is as follows:

$$H_0(x) = -\sum_{x \in \chi} P(x) \log P(x) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

where $P(x_i)$ refers to the conditional probability with which x_i is to be selected, given the mental state at the initial stage of activation.

Accordingly, the decrease of entropy between these two points, i.e., from $P(x)$ to $Q(x)$, or from initial activation to final selection, would be simply:

$$H(x) = H_0(x) - H_1(x) = H_0(x) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

Here, if the activation of lexical items is modulated by context and the frequency of meanings/translations, as mentioned in Section 3, the $P(x)$ in this equation can be considered equal to the probabilities observed in the text. This means that the $H(x)$ here would be the same as the entropy equation which is formulated in Carl et al. (2016), i.e., that which is calculated from the probabilities in the text and approximated from the sample. In other words, the decrease of entropy value in the mental state is perhaps equal to the HTra value.[4]

5. HTra and ITra

The above sections have shown that between the two ways of quantifying cognitive load in lexical translation choice, namely, shift of resource allocation and reduction of entropy (see Wei, 2021), the former is equal to surprisal of the chosen item and the latter is equal to the entropy generalising over all alternative options. Both can be approximated from the text (as ITra and HTra), and can perhaps be used as theoretically justifiable ways of quantifying cognitive load. This means that these two formulations can provide useful means for predicting the effort of translation at the lexical level.

---

[4] See previous footnote.

So two metrics, HTra and ITra, merit further discussion. If they are considered from a *product* perspective, the difference between them seems straightforward — HTra generalises over different translation items while ITra indicates the unpredictability of a specific translation item given a particular ST token (see also Carl, 2021a, p. 122; Heilmann & Llorca-Bofí, 2021, pp. 213-214). From a *process* perspective, the above sections have shown that between the mental state of initial activation and that of final selection, HTra represents the reduction of entropy while ITra indicates the size of the shift in cognitive resource allocation.

In terms of their mathematical expression, HTra represents the initial $P(x)$ distribution when alternative options are activated, whereas ITra indicates the surprisal of the final choice. HTra is equivalent to the *absolute difference* of entropy between the two mental states, while ITra is equivalent to the *relative* entropy of the final mental state with respect to the initial mental state.

In this regard, it is worth asking — which metric is a better predictor of translation behaviour, if we examine the empirical data? To answer this, a few smaller questions need to be addressed: Does HTra still predict effort if we control for the effects of ITra, and vice versa? If so, which one has a larger strength of prediction? When HTra is controlled, does ITra make an additional contribution in explaining variance in effort (and vice versa)?

6. Prediction of effort

A subset of the CRITT TPR-DB[5] was used to examine these two predictors (i.e., HTra and ITra) in terms of their significance and strength in predicting production time and ST/TT reading time. This data is within the multiLing dataset, where six English texts are translated into various languages. In total, the data used for analysis includes 500 experimental sessions from six studies (AR19, BML12, ENJA15, KTHJ08, RUC17, and ST12).

Production time is represented by Dur and refers to the duration of TT production for each ST token. For reading time, early measures of eye movement include first fixation duration on the ST token (FFDur), first pass duration on the ST token (FPDurS), and first pass duration on the TT token (FPDurT). Late measures are total reading time on the ST (TrtS) as well as on the TT (TrtT). All these were regarded as response variables in the analysis and examined in relation to HTra and ITra.

For each of these response variables, outliers were removed by 2.5 standard deviations per participant, and a sequential multiple regression analysis was conducted. In the regression analysis of each response variable, HTra was first entered as a predictor, then ITra is added. A comparison between the base model (with HTra only) and the full model (with both HTra and ITra) can show the contribution of ITra in explaining the variance in the response variables.

A set of base models with ITra entered was also examined in relation to the full model, shedding light on the contribution of HTra in explaining the variance in production time and reading time.

Through an examination of the full models in greater detail, the strength and significance of each predictor (HTra and ITra), when controlling for the other predictor, was also analysed.[6]

---

[5] This is a publicly available database. For details, see, e.g., Carl, Bangalore, et al. (2016). A description of the up-to-date public studies is also available on the CRITT@kent website: https://sites.google.com/site/centretranslationinnovation/tpr-db/public-studies?authuser=0

[6] VIF scores in the full models are all between 1.9 and 2.1.

6.1 Production time (Dur)

For the prediction of word production time, results are shown in Tables 1 and 2.

The two base models (Dur 1 and Dur 2), for HTra and ITra respectively, are both significant. For HTra, $R^2 = .03$, $F(1, 30104) = 957.42$, $p < .001$, and the model explained 3% of the variance in production time. For ITra, $R^2 = .05$, $F(1, 30314) = 1453.19$, $p < .001$. Here, the model with ITra explained a higher percentage (5%) of the variance than that with HTra.

The full model where both predictor variables were entered (Dur 3) was also significant, $R^2 = .05$, $F(2, 30103) = 748.75$, $p < .001$. With the two predictors combined, this model explained 5% of the variance in production time.

Here, although the impact of both ITra and HTra was strong and significant in the full model, it is apparent that ITra ($\beta = 689.58$) was more than three times as a stronger predictor than HTra ($\beta = 196.05$).

After controlling ITra, adding HTra to the base model did not lead to any R^2 change (see Dur 2 and Dur 3). This means that with ITra controlled, no additional variance was explained by HTra. In contrast, when ITra was added after controlling HTra, the model significantly explained an additional 2% of the variance (see Dur 1 and Dur 3). In other words, while controlling for the other predictor variable, ITra made an additional contribution in explaining the variance in production time, whereas HTra did not.

	Dur 1	Dur 2	Dur 3
(Intercept)	2434.25 ***	2434.25 ***	2434.25 ***
HTra	675.76 ***		196.05 ***
ITra		825.96 ***	689.58 ***
N	30106	30106	30106
R2	0.03	0.05	0.05

All continuous predictors are mean-centered and scaled by 1 standard deviation.
*** p < 0.001; ** p < 0.01; * p < 0.05.

Table1. Prediction of production time (Dur)

6.2 ST Reading time (FFDur, FPDurS, TrtS)

Table 2 illustrates the results for the prediction of FFDur, FPDurS, and TrtS. Similar to the results for production time, all impacts in all models here were significant, for all measures of ST reading time. However, for both early and late measures of eye movement on the ST, HTra seemed to be a much stronger predictor than ITra, in contrast to the results for production time (see Section 6.1).

This is notable for all response variables regarding ST reading, where, for FFDur, HTra ($\beta = 45.75$) was more than three times as a strong predictor as ITra ($\beta = 13.03$), and for FPDurS, HTra ($\beta = 42.76$) was more than four times as strong as ITra ($\beta = 9.49$). For the late measure of eye movement on the ST (TrtS), HTra ($\beta = 222.22$) was twice as strong as ITra ($\beta = 103.51$).

For early measures (FFDur & FPDurS), HTra explained an additional 1% of the variance only in FPDurS. For late measures, no additional variance was explained by either variable.

	FFDur 1	FFDur 2	FFDur 3	FPDurS 1	FPDurS 2	FPDurS 3	TrtS 1	TrtS 2	TrtS 3
(Intercept)	188.54 ***	188.54 ***	188.54 ***	180.72 ***	180.72 ***	180.72 ***	960.87 ***	960.87 ***	960.87 ***
HTra	55.10 ***		45.75 ***	49.57 ***		42.76 ***	296.33 ***		222.22 ***
ITra		45.84 ***	13.03 ***		40.15 ***	9.49 ***		262.61 ***	103.51 ***
N	69191	69191	69191	69364	69364	69364	69256	69256	69256
R2	0.01	0.01	0.01	0.04	0.03	0.04	0.03	0.03	0.03

All continuous predictors are mean-centered and scaled by 1 standard deviation.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 2. Prediction of ST reading time

6.3 TT Reading time (FPDurT, TrtT)

For both early and late measures of eye movement on the TT, HTra and ITra did not show a large difference in their strength of prediction, at least not as large as the difference shown above regarding ST reading (see Section 6.2), although all predictions are significant. These results are shown in Table 3.

	FPDurT 1	FPDurT 2	FPDurT 3	TrtT 1	TrtT 2	TrtT 3
(Intercept)	468.76 ***	468.76 ***	468.76 ***	2317.80 ***	2317.80 ***	2317.80 ***
HTra	195.00 ***		125.58 ***	706.94 ***		490.14 ***
ITra		186.87 ***	96.94 ***		653.80 ***	302.30 ***
N	68795	68795	68795	69025	69025	69025
R2	0.07	0.07	0.08	0.03	0.03	0.04

All continuous predictors are mean-centered and scaled by 1 standard deviation.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 3. Prediction of TT reading time

7. Concluding remarks

The above sections have analysed, both theoretically and empirically, two ways of quantifying cognitive load in translation choice, namely, shift of resource allocation and reduction of entropy. Both can be approximated via corpus-based means. At a conceptual level, the paper argues that HTra approximates the reduction of entropy in the mental state and that ITra approximates the size of shift in cognitive resource allocation, providing theoretical justifications for both HTra and ITra as potential means of quantifying cognitive load. Empirical analyses on the CRITT TPR-DB showed that although both metrics had significant and strong impact on effort, ITra was a much stronger predictor for word production time while HTra was a stronger predictor for ST reading time. The difference between the two for prediction of TT reading was found to be relatively small. It is hoped that this would contribute to the search for a dependable means of predicting effort in translation.

References

Attneave, Fred. (1959). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*. Oxford, England: Henry Holt.

Bangalore, Srinivas, Bergljot Behrens, Michael Carl, Maheshwar Ghankot, Arndt Heilmann, Jean Nitzke, . . . Annegret Sturm. (2016). Syntactic variance and priming effects in translation. In Michael Carl, Srinivas Bangalore, and Moritz Schaeffer (Eds.), *New Directions in Empirical Translation Process Research* (pp. 211-238). Cham, Switzerland: Springer.

Carl, Michael. (2021a). Information and entropy measures of rendered literal translation. In Michael Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 113-140). Cham, Switzerland: Springer.

Carl, Michael (Ed.) (2021b). *Explorations in Empirical Translation Process Research*. Cham, Switzerland: Springer.

Carl, Michael, Srinivas Bangalore, and Moritz Schaeffer (Eds.). (2016). *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Cham: Springer.

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

Carl, Michael, Moritz Schaeffer, and Srinivas Bangalore. (2016). The CRITT translation process research database. In Michael Carl, Srinivas Bangalore, and Moritz Schaeffer (Eds.), *New Directions in Empirical Translation Process Research* (pp. 13-54). Cham, Switzerland: Springer.

Carl, Michael, Andrew Tonge, and Isabel Lacruz. (2019). A systems theory perspective on the translation process. *Translation, Cognition & Behavior, 2*(2), 211-232. doi:10.1075/tcb.00026.car

Cover, Thomas M, and Joy A Thomas. (1991). *Elements of information theory*. New York: John Wiley & Sons.

Dragsted, Barbara. (2010). Coordination of reading and writing processes in translation. *Translation and Cognition, 15*, 41.

Dragsted, Barbara, and Inge Gorm Hansen. (2008). Comprehension and production in translation: a pilot study on segmentation and the coordination of reading and writing processes. *Copenhagen Studies in Language*(36), 9-29.

Hale, John. (2001, June). *A probabilistic Earley parser as a psycholinguistic model.* Paper presented at the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, Pittsburgh, Pennsylvania, 1- 7 June 2001. Association for Computational Linguistics.

Heilmann, Arndt, and Carme Llorca-Bofí. (2021). Analyzing the effects of lexical cognates on translation properties: A multivariate product and process based approach. In Michael Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 203-229). Cham, Switzerland: Springer.

Holmes, James S. (1972). *The Name and Nature of Translation Studies*. Paper presented at the Translation Section of the Third International Congress of Applied Linguistics, Copenhagen.

Jensen, Kristian Tangsgaard Hvelplund, Annette C Sjørup, and Laura Winther Balling. (2009). Effects of L1 syntax on L2 translation. *Methodology, Technology and Innovation in Translation Process Research. Copenhagen: Samfundslitteratur*, 319-338.

Kullback, S. (1959). *Information Theory and Statistics*. Hoboken, NJ: John Wiley & Sons.

Levy, Roger. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Levy, Roger. (2013). Memory and surprisal in human sentence comprehension. In Roger van Gompel (Ed.), *Sentence Processing* (pp. 90-126). London and New York: Psychology Press.

Levy, Roger, and Edward Gibson. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Front Psychol, 4*, 229. doi:10.3389/fpsyg.2013.00229

Schaeffer, Moritz, Barbara Dragsted, Kristian Tangsgaard Hvelplund, Laura Winther Balling, and Michael Carl. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In Michael Carl, Srinivas Bangalore, and Moritz Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* (pp. 183-210). Cham, Switzerland: Springer.

Schaeffer, Moritz, Jean Nitzke, and Silvia Hansen-Schirra. (2019). Predictive turn in translation studies: Review and prospects. In Stanley Brunn and Roland Kehrein (Eds.), *Handbook of the Changing World Language Map*. Cham, Switzerland: Springer.

Shepard, Roger N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323.

Tenenbaum, Joshua B, and Thomas L Griffiths. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629-640.

Venuti, Lawrence (Ed.) (2000). *The Translation Studies Reader*. London: Routledge.

Wei, Yuxiang. (2021). Entropy and eye movement: A micro-analysis of information processing in activity units during the translation process. In Michael Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 165-202). Cham, Switzerland: Springer.