14

Translation, effort and cognition

Daniel Gile and Victoria Lei

14.1 Introduction: The emergence of the “effort” topic in the history of Translation and Interpreting Studies (TIS)

In translation and interpreting (henceforth “Translation”), some direct source-Text to target-Text transfers seem effortless and immediate. Such is the case with proper names, well-established technical terms and standard phrases when they are known to the translator or interpreter. However, these short Text segments (Text stands for written texts and spoken and signed speeches) are generally embedded in Texts that require reflection, problem solving and decision making as to what the author or speaker actually means and intends to achieve, and how the Translator should reformulate the message in the target language so as to meet applicable professional and social requirements. Such analyses and decision-making processes involve effort. In particular, translation generally requires the search for ad hoc information, both linguistic and extra-linguistic, to identify target-appropriate lexical and phraseological usage, as well as reflection and decisions on how to address the tension arising from differences in information-explicitness requirements between the source and the target language (“linguistically and culturally induced information”—see Gile, 2009, Ch. 3). In interpreting, considerable added pressure is associated with stringent time constraints.

Interestingly, at the beginning of scientific investigation into translation, starting in the 1950s, this aspect of the translator’s work was given little attention. The initial focus was on lexical and stylistic contrasts between languages (e.g. Mounin and Vinay and Darbelnet’s work) and on the related issue of equivalence (e.g. Nida’s innovative distinction between formal and dynamic equivalence), and then moved on to philosophical, cultural and sociological issues (e.g. Toury’s norms). One exception was Jiří Levý’s (1967) formulation of the game-theory-based minimax principle, according to which translators invest as little effort as possible to achieve maximum effect.

Investigators of translation became more interested in the topic with the advent of translation process research, starting in the mid-1980s, and the integration of concepts from cognitive science. Interestingly, Gutt (1991/2000) introduced the notion of (cognitive) processing effort into translation theory via Sperber and Wilson’s relevance theory (1986), which basically also postulates optimization of human cognitive investment as a function of expected
gain. Gutt quotes Levý, but only refers to the minimax principle once (in a footnote, on page 20), interpreting it and criticizing it in terms of pragmatics. Levý’s use of the term seems to refer not to intense cognitive effort but to effort in the more mundane sense of the word, similar to Pym’s (2004) effort measurable in “time, hardship, technology costs, interpersonal exchanges”.

In cognitive psychology, a key reference discipline for cognitive research into Translation, cognitive effort is viewed as the engaged proportion of limited-capacity central processing (see Tyler et al., 1979). The link between the abstract concept of cognitive effort and the more mundane meaning of effort in everyday life becomes clear when considering the distinction between so-called automatic operations, which are very fast and require little attention, and non-automatic operations, which are far slower and do require attentional resources—and are associated with a feeling of effort (Kahneman, 2011). According to Kahneman, effort is basically unpleasant, and “System 2”, the symbolic name he gives to a set of mechanisms that govern non-automatic operations—those which require significant attention—is “lazy”, feels uncomfortable when exerting effort and would like nothing more than to hand the control of cognition back to “System 1”, which relies on heuristics and performs automatic operations. When humans are forced to deploy effortful activities, as a result of a phenomenon called “ego depletion”, they are less inclined to take on new effortful activities later. This intrinsic “laziness” of human cognition, which ties in with Zipf’s principle of least effort in human behaviour (1949), is not unrelated to relevance theory. It also draws attention to the potentially important role of motivation, be it under a personal ethical philosophy, under a code of ethics or under the threat of sanctions if sufficient effort is not devoted to achieving quality.

To pioneer psychologists and psycholinguists who showed interest in the cognition of simultaneous interpreting (SI) in the 1960s and early 1970s, the relevance of limitations of cognitive resources in interpreting was clear. Actually, to a large extent, their interest in SI was triggered by curiosity as to how interpreters managed to conduct multiple operations with limited attentional resources under cognitive pressure, for instance under high-speed delivery conditions (see Gerver, 1969). Gerver talks about an information-handling system, which is subject to overload if required to carry out complex processes at too fast a rate, and of evidence that within this system, attention is shared between the input message, processes involved in translating a previous message, and the monitoring of current output. He adds that while under normal conditions attention can be shared between these processes, when the total capacity of the system is exceeded, less attention can be paid to either input or output if interpretation is to proceed at all. Hence, less material is available for recall for reformulation, and more omissions and errors in the output will occur. A similar analysis, initially based on introspection, was developed by Gile a little over a decade later with his Effort models. Kirchhoff (1976), a translation and interpreting teacher with an interest in bilingualism, drew on psycholinguistics to attribute the fundamental challenge that arises when interpreting between structurally dissimilar languages to the need to process larger segments, presumably because of the need to store more information before being able to reformulate the source-speech message. She talks about “cognitive load” and the need to choose “appropriate strategies” as a function of available capacity.

Investigation of interpreting by psychologists and psycholinguists in this early history of interpreting studies was short-lived. Seleskovitch, Lederer and their followers, who developed interpretive theory in the 1970s and 1980s and gained a virtually exclusive influence in the field for close to two decades, were conference interpreters. As practitioners, they were aware of the existence of such cognitive pressure (as evidenced by tactical advice given to students in the classroom on how to handle various difficulties and on the need to focus one’s attention
on listening rather than note-taking in consecutive interpretation, and on listening rather than reading in simultaneous interpretation with text) but did not incorporate it into their theory, probably in order to avoid giving salience to language issues and what they considered the excessive influence of theories and methods from cognitive psychology. Chernov, who developed his own theory in the USSR in parallel, did so, albeit indirectly, through his idea that anticipation was what made SI possible (Chernov, 1994). Starting in the early 1980s, Gile brought the focus back to cognitive effort with the Effort models in an attempt to account for a large proportion of errors, omissions and infelicities observed and measured in interpreting performance. Other interpreting studies researchers followed, including Tommola (Tommola & Hyönä, 1990; Tommola & Niemi, 1986) and Setton (1998), who adopted relevance theory as one of the pillars of his own model of interpreting, and many others in the following years.

Over the past two decades or so, with the emergence of cognitive translation and interpreting studies (CTIS), the concept of effort, especially cognitive effort, with its implications and measurements, has gained much visibility in research into written translation as well, making it one of the topics where investigators of translation and interpreting meet.

14.2 Effort and Translation performance

14.2.1 The relevance and importance of effort in Translation performance

The most obvious reason for which the topic of effort deserves attention in TIS is the link between the amount of effort invested by the Translator and the quality of his/her performance. The relation is not straightforward. The correlation can be assumed to be strong for low effort investment: a translation done carelessly will most probably exhibit weaknesses, linguistic infelicities if not inaccuracies, and outright errors. As the amount of effort invested in ad hoc information collection, in analytical thinking about the source Text and implications of the Translation brief, and in revision increases, infelicities and errors are likely to recede. What is more difficult to determine is how much quality improvement will ensue from incremental additional efforts. For instance, what is the return of an increasing number of iterations of self-revision and revision by others or of consulting more documentary sources when retrieving thematic, terminological and phraseological information? Beyond a certain point, further effort may not contribute much or may even become counterproductive, though where this point lies can vary greatly.

Another reason for the importance of the topic is that effort, in the sense of the use of one’s attentional resources to perform demanding tasks, has become and continues to be an important tool in the investigation of various translation and interpreting phenomena, as illustrated later in this chapter.

Finally, in the light of recent interest in the place of risk and risk analysis in Translation behaviour, and in particular in the links between risk and effort as discussed by Pym (2015), it will be argued here that effort is also an ethical issue (see also Pym, this volume, for a discussion of translation, risk management and cognition).

14.2.2 How much effort investment?

Cognitive psychologists tend to focus on “effort” as it arises in the form of non-automatic operations drawing upon limited available attentional resources. When engaged in such operations, people may only become aware of associated discomfort when it reaches high intensity, as happens regularly in SI. A feeling of discomfort also arises when low-intensity effort
is prolonged, as in the preparation of glossaries, in repeated self-revision in translation, or in repeated search for terminological information in external sources.

Both high-intensity and low-intensity effort are found in translation and interpreting, but in TIS, authors do not generally make the distinction. And while “general theoreticians”, in particular Levý and Pym, seem to focus on low-intensity effort, investigators interested in Translation cognition focus more on high-intensity effort, i.e. on intensive use of attentional resources (see also Hvelplund, this volume, for a discussion of translation, attention and cognition).

According to what could be termed universals in professional ethics and codes of conduct, in view of the link between effort and quality, it would seem natural for translators and interpreters to invest considerable effort in their work. Levý, however, believes that translators seek to optimize the effort/gain ratio. Taken at face value and without qualifications, such behaviour is ethically debatable. Pym (2015) advocates a principle whereby so-called high-risk translation problems deserve more effort than low-risk translation problems. His use of the term “risk” is problematic: in Pym (2011), he defines it as the probability of an undesired outcome of an action, while in Pym (2015), the term seems to denote stakes. Setting aside this terminological ambiguity, the idea that more effort should be invested in high-stake Text segments makes sense—provided as much effort as is required to achieve acceptable quality is invested in lower-stake segments as well.

Gile’s sequential model of translation (Gile, 2009) includes iterative tests of source-text comprehension and of fidelity and acceptability on translation units and aggregates until satisfactory solutions are found. According to this model, which specifically refers to risk and recommends the adoption of solutions with the best combination of loss/gain associated with their respective probabilities of occurrence, effort is only discontinued when satisfactory solutions are found. It was developed for didactic purposes and cannot be taken as a good reflection of reality, not only because in real life, operations it describes as a sequence are often performed with overlapping and backtracking, but also because entirely satisfactory solutions to translation problems are not always found.

When, and how much, do translators and interpreters invest effort as they address Translation challenges in real life? Risk management has the potential to account for some effort-related behaviour, especially in a business context (for instance as regards investment in equipment and translation software, or in human resources management), and to some extent in individual Translation behaviour—in interpreting, minimizing the risk of cognitive interference can be a determinant of coping tactic selection (see Section 14.3.1). However, other factors may be far more important. Professionalism, personal motivation and resistance to fatigue-induced ego depletion as well as socially induced emulation between peers could be stronger determinants, as suggested, for instance, by a recent survey among conference interpreters by Zwischenberger (2017), who found that the strongest reason for satisfaction/dissatisfaction was succeeding/failing to fulfil the respondents’ own standards. This point is taken up again later in this chapter.

14.3 TIS research on cognitive effort

14.3.1 Cognitive effort in TIS literature

With the spectacular growth of cognitive investigations into translation over the past two decades, from general considerations of “effort” in the everyday sense of the word, a new focus has been crystallizing around cognitive effort, which is associated with longer information processing time, more pauses, filled or unfilled (e.g. Plevoets & Defrancq, 2016; Shreve et al., 2011), slower target-text production, which includes longer gaze time and longer reading time (e.g. Dragsted,
Translation, effort and cognition

2012; Schaeffer & Carl, 2014), and physiological changes. In particular, Jakobsen et al. (2007) used task time as a metric to compare the cognitive effort required to render idioms and non-idioms in written translation and sight translation, and found that idiomatic expressions slowed down production in both modes.

Patterns of segmentation have also been correlated with cognitive effort in written translation. In a study using eye tracking and keyloggings, Jakobsen (2011) observes that comprehension, reformulation, and the actual typing and monitoring of the translation product overlap at times, as some of the chunks recorded are too long to be possibly stored in short-term memory, and proposes a six-step “micro-cycle” of motor and cognitive efforts, stressing that some of the steps can be skipped, while some may be repeated again and again.

Relevance theory was integrated into experimental translation process research, as it “postulates an optimal cognitive relation between the minimum processing effort necessary for the generation of the maximum cognitive effects” and “offers a productive way to investigate the role of effort in translation task resolution” (Alves & Gonçalves, 2013, p.108). The cross-linguistic priming effect in written translation has also been taken up in research into translation cognition. Various studies have sought to find out whether translation involving larger cross-linguistic differences takes up more cognitive effort (e.g. Bangalore et al., 2016; Schaeffer et al., 2016; Vandepitte & Hartsuiker, 2011). Schaeffer & Carl (2014) put forward a recursive model of translation consisting of an early priming process based on shared semantic and syntactic representations between the source text (ST) and the target text (TT) (“horizontal translation” based on translinguistic equivalences, which is automatic, and “vertical translation”, which is conceptually mediated and requires conscious effort).

One phenomenon that clearly distinguishes interpreting from translation (except for the case of sight translation, a hybrid) is that while in translation, the intensity of expended cognitive effort is relatively low most of the time, and the feeling of discomfort associated with effort mostly only arises if it is prolonged, in interpreting, it can be very high, e.g. when the speaker is fast or speaks with a strong, unfamiliar accent, or when the speech is informationally dense (a list of “problem triggers” is drawn up in Gile, 2009 and corroborated empirically inter alia in Mankauskienė, 2018).

When discussing high-intensity effort in interpreting, it is convenient to use the cognitive psychological construct of working memory (WM), as defined by Baddeley and Hitch (1974) and later elaborated on through different avatars, e.g. by Miyake and Shah, Ericsson, Kintsch, Cowan, and Barrouillet and Camos (for a clear and thorough presentation, see Gieshoff, 2018). This concept refers to some virtual cognitive space where information is stored for up to a few seconds and processed. WM is postulated to have very limited storage space of just a few “chunks” of information, which disappear rapidly if not “refreshed” with some effort, and is associated with “executive functions”: inter alia, attention allocation and attention switching.

Intense cognitive effort arises naturally from the simultaneous processing of information coming from the speaker and production of the interpreter’s target speech (in the simultaneous mode) or notes (in the consecutive mode). When interpreters feel that they are close to saturation or anticipate the imminent arrival of such a state (when under high cognitive load), they may resort to various coping tactics to relieve the pressure, e.g. buying time to better understand an utterance or decide how to reformulate the message (stalling) or making sure the recipients understand a reference or idea (paraphrasing, explicating and explaining). Many have a cognitive cost insofar as they take up time and further effort, with the associated risk of overloading WM and jeopardizing the fragile balance of attention allocation between Efforts to the detriment of the processing of neighbouring speech segments through a cascade effect (Gile, 2009, Ch. 8; Pointurier-Pournin & Gile, 2012 for signed language interpreting). The specific dynamics of
such phenomena, including measurements of the level of cognitive effort expended and reaction chains, are clearly a key to better insight into interpreting cognition.

Gile (2009) considers that while interpreting, the combined attentional requirements of all functional Efforts (reception, production, short-term memory operations and attention management) tend to be close to maximum available capacity, which makes interpreters vulnerable to sudden increases of requirements (through problem triggers) and to attention management errors, as well as to attention fluctuations. This assumption is referred to as the “tightrope hypothesis”.

Interpreters are aware of this and, as explained earlier, use various coping tactics (actions with an immediate objective, as opposed to strategies, actions with a wider, longer-term objective) to alleviate cognitive pressure and optimize their target speech. Ideally, they seek to reformulate the maximum amount of information meant by the speaker to be transmitted to the addressees subject to compatibility with the speaker’s assumed intention and interests (in the default value of speaker-loyalty—see Gile, 2009), including compliance with applicable social norms.

The selection and implementation of such coping tactics can require much deliberate effort on the part of the interpreter, who may be tempted to opt for the (effortless) omission tactic or for another effort-sparing but sub-optimal tactic in terms of communication efficiency, such as simply repeating a term as it was pronounced by the speaker without translating it.

In other words, in the interpreter’s tactical behaviour, conscientiousness (or “professionalism” in the everyday sense of the word) and personal motivation can be determinants of effort investment.

Actually, this applies to low-intensity effort as well, and thus to interpreting strategies such as conference preparation: by acquiring new relevant information about the future meeting to which they have been assigned (e.g. the participants, the stakes, relevant concepts and ideas, and relevant terminological and phraseological information) or by refreshing it in their long-term memory, interpreters expect to prime the information and therefore to need less cognitive effort to understand the speakers’ utterances and to reformulate them in the target language. Preparatory investment in low-intensity effort is thus assumed to reduce high-intensity effort requirements when the meeting starts.

Besides these low- and high-intensity efforts related to information, socio-affective effort can also be an important part of interpreting work in public-service settings (e.g. healthcare, courts or asylums), where, at least on one side, stakes are high and principals may be in emotional states that require special attention. Interpreters need to get attuned to the specific needs in the situation at hand and make the right decisions with respect to interventionism, sometimes beyond the standard boundaries defined by professional codes of conduct (this has become a central topic in community interpreting and signed language interpreting circles in the past two decades or so—see e.g. Janzen, 2005).

Since the 1990s, interpreting researchers have been following more closely developments in cognitive science in their investigation of cognitive load and cognitive effort. In particular, referring to Gile’s Effort models, Seeber (2011) sought to fine-tune the analysis by drawing on Wickens’ multiple-resource model (2002). According to Wickens, contrary to a single-resource model postulated by Kahneman (1973), cognitive operations draw on a shared pool of resources to a small extent but rely mostly on different resource pools depending on the processing stage (i.e. central processing and response processes), the type of process (manual/spatial and vocal/verbal), the processing modality (aural or visual), and within visual processing, on focal vs. ambient vision. In his view, such a model gives a better account of the relative amount of interference between tasks involving WM to store or transform information. Seeber uses this model to calculate the relative (theoretical) amount of cognitive interference from various interpreting-related tasks, namely shadowing, sight translation and SI. Note that Barrouillet & Camos (2012) argue
for a different time-based resource-sharing model (TBRS), according to which maintenance of memory traces depends on attentional focusing within a single resource-limited central system.

14.3.2 Quantifying cognitive effort

Being a construct, cognitive effort can only be measured indirectly through indicators. This section lists the main indicators and measurement methods found in research into TIS over the years, with a special focus on recent research.

14.3.2.1 Rating scales

They involve post-hoc questionnaire surveys asking participants to rate the level of cognitive load they experienced when performing a task, under the assumption that increased capacity invested in a task is associated with “subjective feelings of effort or exertion” (O’Donnell & Eggemeier, 1986, 42/7), on which people can report retrospectively (Paas et al., 2003). A study of translation difficulty by Sun and Shreve (2014) is an attempt to apply subjective rating to TIS. In the study, a four-category rating scale (Mental Demand, Effort, Frustration level and Performance) adapted from the NASA-TLX (Task Load Index) developed by Hart and Staveland (1988) was used for evaluation of post-hoc subjective workload. Sun (2018) believes the method can also be employed to measure cognitive load in post-editing of machine translation. Chen (2017a) points out that the rating scale techniques have been widely used in research related to Cognitive Load Theory, and their potential in TIS is still to be tapped.

14.3.2.2 Behavioural indicators

14.3.2.2.1 Hesitation pauses

Lacruz and Shreve (2014), which explores “the promise of simple pause metrics as tools for measuring cognitive effort during post-editing” (p. 266), puts forward the average pause ratio (average time per pause ÷ average time per word) as a potential measure for cognitive effort in post-editing, suggesting that high pause densities during post-editing, which tend to produce low average pause times, indicate high levels of cognitive effort. The measure is tested out in a few case studies, which demonstrate a significant negative correlation between the average pause ratio and the event-to-word ratio.

In a large corpus of interpreted and non-interpreted texts, Plevoets & Defrancq (2016) operationalize informational load, which generates cognitive load, in terms of delivery rate, lexical density, percentage of numerals and average sentence length. Interpreted texts were analysed based on the interpreter’s output and compared with the input of non-interpreted texts, and the effect of source-Text features was measured by counting the occurrence rate of the speech disfluency u(h)m. Interpreters produced significantly more uh(m)s than non-interpreters in original speeches. The authors attribute this difference mainly to the effect of lexical density on the output side. The main source predictor of uh(m)s in the target Text was shown to be the delivery rate of the source Text.

14.3.2.2 Dependency distance and lexical simplification

Liang et al. (2017) used reduction of dependency distance as an indicator of cognitive effort. Dependency distance (DD), the distance (the number of words) between two syntactically related words in a sentence, is assumed to be correlated with cognitive load during speech production, because to produce a correct sentence, traces of the first word in the relevant syntactic relation have to be kept in WM until the last syntactically dependent word is produced. By virtue of the law of least effort, speech producers are assumed to tend to minimize DD. Liang et al.
compared mean DDs from SI, consecutive interpreting (CI) and written translations and found them to be largest for translated texts, second largest for SI and smallest for CI. They conclude that CI may entail heavier cognitive load than SI. This inference is problematic, since under the heavy cognitive pressure of simultaneous interpretation, interpreters are not as free to reorganize their information into sentences as they are in consecutive interpretation, because such reorganization would entail much waiting and possibly saturate WM. A similar comment applies to a related study by Lv and Liang (2018) using lexical simplification (choosing more frequent words, shorter words, etc.) as an indicator of cognitive effort, which concludes that CI may be associated with higher cognitive effort based on the finding that lexical (and associated) simplification was larger for CI than for SI.

14.3.2.2 Ear-voice span and eye-key span

Ear-voice span, the time lag between source-speech segments and corresponding target-speech segments, was a focus of interest of early investigators working on SI (e.g. Oléron & Nanpon, 1965; Gerver, 1969; Goldman-Eisler, 1967), perhaps partly because it actually measured the “simultaneity” of SI, but also because it was related to questions about the storage capacity of WM. Also note that in cognitive psychology and psycholinguistics, reaction times are a central dependent variable in experiments, and time lag was a likely candidate to be used as an indicator of cognitive effort. Lee (2003) studied a corpus of authentic English to Korean simultaneously interpreted speeches and found that longer source-language sentences induced longer lags, reduced accuracy and longer intra-sentence pauses in the target speech, all of which suggest that the lag indicates more processing effort. However, longer lags can also generate higher cognitive load insofar as they force the interpreter to store more information from the incoming speech before reformulation becomes possible. This ambiguity is also found in Chen’s use of ear-pen span (Chen, 2017b), the lag between the source speech and the notes being taken by the interpreter.

Combining keylogging and eye-tracking technologies has facilitated accurate measurement of time lag in translation. Dragsted and Hansen (2008) introduced the concept of eye-key span, the time lag between the first fixation on source-text words and the beginning of the typing of the corresponding target-text words. Translation and interpreting students were recruited to translate from English into Danish. The authors claim that longer eye-key span is associated with “problem words”, such as metaphorical expressions, that require more cognitive effort, as these words also elicit more regressive eye movements and longer accumulated fixation durations.

14.3.2.2.4 Eye movement patterns

When reading is involved, the number of fixation counts, the duration of fixations and the number of regressive saccades are considered as indicating cognitive effort (e.g. Just & Carpenter, 1980). Sjørup (2008) found that the processing of metaphors was associated with longer fixations than the processing of literary expressions. Jakobsen and Jensen (2008) investigated cognitive processing by professional translators and translation students in four different types of reading tasks—reading for comprehension, reading in preparation for translation, sight translating, and reading while typing a translation. For both groups of subjects, sight translation and written translation were associated with significantly longer task time, more fixation counts and longer total gaze time (total duration of all fixations) than the two reading tasks that did not involve actual translation. Eye movement patterns and task time also suggest that all four tasks were more effortful for the students than for the professionals, and in written translation, that students allocated more effort to reading the source text, while the professionals invested more visual attention in the target text. A study by Pavlović and Jensen (2009), with professional translators and translation students with native Danish and English as their L2, investigated directionality...
Translation, effort and cognition

in translation using eye movement patterns as well as pupillometry to measure cognitive effort. The results corroborated the hypothesis that in both translation directions, processing the target text demands more cognitive effort than processing the source text: it was associated with longer gaze time, longer average fixation duration and larger pupil dilation. However, only pupil dilation values and task time supported the hypothesis that translating from L1 to L2 demands more cognitive effort than translating from L2 to L1. In both groups, L2-to-L1 translation was associated with longer gaze time. Chang (2009) used similar eye-tracking metrics to compare cognitive effort in L1-to-L2 and L2-to-L1 translation. The study, which includes two language combinations, recruited two groups of translation students, one with Mandarin and English, and one with Spanish and English. They were asked to type-copy a text in L1, type-copy a text in L2, type-translate from L2 to L1 and type-translate from L1 to L2. The results suggest that for both language combinations, L1-to-L2 translation requires more cognitive effort, as it is associated with significantly larger pupil dilation, a significantly larger number of fixations, higher fixation frequency, higher blink frequency and longer task time. Eye movement patterns also indicate that for both translation directions, more visual attention was allocated to the TT.

14.3.2.3 Brain imaging and associated techniques

14.3.2.3.1 Electroencephalography (EEG)

Electroencephalography (EEG) records electrical cerebral activities from the scalp and has been identified as a physiological index for measuring online cognitive effort (e.g. Antonenko et al., 2010). Compared with neuroimaging techniques based on blood-flow measurement as listed later, EEG provides high temporal resolution but low spatial resolution. It is therefore suitable for assessing online task-evoked cognitive effort, but not for drawing inferences on the locations of brain activation.

Petsche et al. (1993) investigated directionality in SI by comparing EEG changes in three professional interpreters interpreting mentally (to prevent physiological artefacts from speech utterance) into their L1 and L2 and compared coherence of the EEG signals as an indicator of cognitive effort, which turned out to be larger when interpreting into L2. Lachaud (2011) combined EEG (for measuring time-frequency power), eye tracking (for measuring fixation duration and pupillary responses) and keylogging (for measuring response time) to look at the cognitive effort expended when transcoding between English and Norwegian at the word level when Deceptive Cognates (false friends), True Cognates and Non-cognates are involved. Translation involving Deceptive Cognates was found to be the most effortful: it was associated with significantly larger time-frequency power, longer total fixation duration, bigger pupil size amplitude variation and longer reaction times. EEG and keylogging data also suggest that transcoding Non-cognates requires significantly more cognitive effort than transcoding True Cognates. Eye-tracking data did not demonstrate significant differences between the two.

An ERP study by Koshkin et al. (2018) is the first published work testing Gile’s Effort models in a “naturalistic setting requiring the participants to interpret continuous prose overtly” using neuroimaging techniques. Event-related potentials (ERPs), which reflect brain responses to events, are calculated by averaging a continuous EEG signal over a large number of trials (e.g. Antonenko et al., 2010). The study tried to find out whether more resources given to WM would result in reduced availability of attention for listening. Based on previous functional magnetic resonance imaging (fMRI) studies (e.g. Corbetta & Shulman, 2002; Mayer et al., 2007) claiming that attention and WM are subserved by overlapping brain regions, it was assumed that changes in listening effort and WM load would be reflected by changes in the amplitude of ERPs (N1 and P1). Interpreters were asked to interpret speeches from English into Russian
and vice versa. The source and target speeches were transcribed and manually time-stamped. The WM load of each unit was defined by ear–voice span and estimated using the number of content words in the unit, the number of content words in the unit weighted by their frequency, and all the words in the unit weighted by their respective syllabic lengths. A larger negativity in the P1 and N1 amplitude was found when the WM load was lower and vice versa, suggesting that the listening effort and the WM effort competed for cognitive resources.

14.3.2.3.2 Positron emission tomography (PET)
PET looks at blood flow and oxygen consumption in different parts of the body, including the brain. Increased blood flow and metabolism imply increased activities and effort. It is a relatively invasive method compared with EEG, functional near-infrared spectroscopy (fNIRS) and fMRI, if only because it involves injecting a tracer into a vein.

Rinne et al. (2000) used PET to measure brain activation in professional interpreters during SI between Finnish (L1) and English (L2) in both directions. Interpreting into L1 elicited left frontal activation, while interpreting into L2 elicited much more extensive left fronto-temporal activation. The results indicate that SI activates predominantly left-hemispheric structures and imply that translating into L2 demands more cognitive effort.

14.3.2.3.3 Functional magnetic resonance imaging (fMRI)
fMRI is a non-invasive neuroimaging technology using magnetic resonance to measure brain activity by detecting changes associated with blood flow. The technology provides high spatial resolution and allows precise localization of activated areas of the brain. Larger amplitude of activations may imply enhanced cognitive effort. Ahrens et al. (2010) conducted an fMRI experiment with student interpreters to compare brain activity between two conditions—SI from Spanish (L2) to German (L1) and free speech production in German. SI and free speech production elicited different activations in the brain. Chang (2009) used fMRI to investigate novice interpreters’ cognitive effort during L1-to-L2 and L2-to-L1 interpreting. Interpreters with Mandarin as L1 and English as L2 were recruited to perform sight interpreting from L1 to L2 and vice versa during the scan. The results suggest that L1-to-L2 translation was more effortful, as it involved more cortical areas than L2-to-L1 translation.

14.3.2.3.4 Functional near-infrared spectroscopy (fNIRS)
fNIRS is a portable neuroimaging technique that makes non-invasive optical measurements of blood flow in the brain. As fMRI does, it detects brain activation by measuring blood oxygenation changes. However, compared with fMRI and PET, which require the subject to lie still during the scan, fNIRS has the potential to allow more ecologically valid experiments for translation and interpreting studies. Lin et al. (2018) combined behavioural measures and fNIRS technology to examine the cognitive effort associated with pairing, transphrasing and non-translation. Students of interpreting were asked to perform three tasks—a pairing task, a transphrasing task and a non-translation task. The stimuli were visually presented two-character cultural-specific items in Chinese (L1), which subjects were required to interpret into English (L2). The area of interest was the left prefrontal cortex (PFC), which includes Broca’s area and has been associated with lexical search, semantic processing, bilingual processing, production of speech and cognitive control. Pairing elicited the most intense activation, which was localized in Broca’s area, while transphrasing induced the longest, most extensive activation overall in the left PFC, suggesting that the latter is most likely to lead to cognitive overload in SI. However, the ecological validity of the study is limited, as the stimuli used in the study were at word level. And as the study only investigated one part of the brain, the results do not give a full picture of the cognitive effort involved in the task.
14.3.2.4 Pupillometry

It is widely accepted that the pupil reacts to cognitive activity (e.g. Kahneman et al., 1969). Pupil dilation, which is usually measured with an eye tracker, can therefore be seen as a response to increased cognitive load (e.g. Holmqvist et al., 2011; Rayner et al., 2006). Pupillometry is less invasive and more convenient than some of the other physiological measures presented here. Tommola and Niemi (1986) used pupillometry to investigate the impact of source-text syntactical complexity on cognitive load. Tommola and Hyönä (1990) and Hyönä et al. (1995) further tested the method by comparing pupil dilation levels associated with three language processing tasks, namely listening, shadowing and SI, which clearly require different levels of mental effort. Relative pupillary responses to the three tasks were consistent with the hypothesized differences in task difficulty. O’Brien (2006) recruited professional translators to translate an English text into their L1s using a translation memory tool. Their pupillary responses were recorded and analysed in combination with other measures, including processing speeds and retrospective protocols to compare the cognitive effort required to process four types of translation memory matches into German or French, namely “No Match”, “Fuzzy Match”, “Machine Translation Match” and “Exact Match”. The results indicate that No Match, which means the translator needs to translate from scratch, required the largest cognitive effort, as suggested by the largest average pupil dilation and the lowest processing speed. Exact Match required the least effort, with the smallest average pupil dilation and the fastest processing speed. Pavlović and Jensen (2009) and Chang (2009) applied pupillometry as one of the indicators of cognitive effort in their studies of translation directionality. Seeber and Kerzel (2011) used pupillometry to measure cognitive effort during SI from German into English. Pupil dilation elicited by interpreting German verb–final structures into English was significantly larger than that induced by interpreting verb–initial structures, suggesting that translating between non–parallel structures takes up more cognitive effort than translating between parallel structures. Hvelplund’s (2011) study, which combined keylogging and eye-tracking technologies to investigate the allocation of cognitive resources in the translation process, applied an array of measures, including task time, fixation duration and retrospection, but used pupillometry as the sole indicator to measure cognitive effort. Professional translators and student translators participated in a series of experiments involving translating English (L2) texts into Danish (L1) under two conditions—with or without time constraints. The results confirmed that under all conditions, translation was more effortful for the students than for the professionals, and for both groups, translating under time pressure generated more cognitive effort than translating without time pressure. However, no significant difference in pupil dilation was found between translating easier texts and more difficult texts, though the subjects did confirm that one text was more difficult than the other in post–hoc interviews. This suggests that pupillary responses might not always be sensitive enough to reflect cognitive load differences and the effort incurred by translating texts of different levels of complexity. As pointed out by Gieshoff (2018), pupil dilation also responds to sources of arousal other than cognitive effort. Hvelplund also found that more cognitive effort was involved in target-text processing than in source-text processing or source-text/target-text parallel processing.

14.4 Methodological and other challenges in the use of the concept of effort in TIS

Scientific research is effective when it discovers/uncovers facts, when it produces theories that help explain them and predict what is yet uncovered, and when it develops concepts and methods that help uncover more facts and develop and test theories. The feeling of effortfulness in translation and interpreting is part of reality. The concept of cognitive effort has demonstrated
its usefulness in developing theories that explain phenomena observed in translation and interpreting. How predictive these theories are depends on the level of accuracy and detail sought. For instance, the Effort models and the associated tightrope hypothesis have been shown to successfully predict holistic phenomena such as increases in the rate of errors, omissions and infelicities in interpreting under certain working conditions (in particular high delivery speed) and in the presence of well-identified problem triggers. However, the precision of their predictions is limited because of their very nature as a holistic conceptual framework. The tightrope hypothesis did not quantify the interpreters’ “closeness” to saturation. It only indicated functionally that they tended to work close enough to saturation to become vulnerable to unexpected hikes in attention requirements and to attention management errors. More specific cognitive theories could do better, provided researchers find a way to offset the high variability arising from the large number of highly influential parameters, which are determinants of performance and are difficult to control in ecologically valid experiments.

Measurements of cognitive effort per se are also of potential use in uncovering interesting correlations; for example, if it could be shown, beyond general principles as postulated by existing theories, that certain tactical behaviour patterns and certain types of Translation quality deteriorations tend to occur at specific thresholds of cognitive load or cognitive effort. Many indicators provide relative data but not interval scales against which particular Translational phenomena can be identified. What is the level of effort at which errors of specific types occur, and what is the level of effort at which Translators give up trying hard and adopt less effortful but also less efficient strategies and tactics as regards the quality of the final product? Answers may come from triangulation studies measuring Translation behaviour and cognitive effort while eliciting retrospective reports (as in Gumul, 2018, in which many retrospective comments on 240 interpretations by students turned out to point to processes described by the Effort models).

Yet another limitation of many quantitative cognitive effort indicators such as pupillometry, gaze duration and hesitation pauses in their use to test ideas and theories about interpreting cognition (the problem may be less conspicuous in translation, where source-text reading is easier to separate from target-text writing) is that while they may be able to detect variations in cognitive effort intensity, they are not specific enough to point to the particular sub-processes or process components that caused them. How does one know whether increased cognitive effort is caused by a comprehension difficulty, a production difficulty or by the retrieval of information from memory, and what is the triggering speech segment? When considering ear-voice span, or ear-pen span in the case of CI, as in Chen (2017b), lag time can be taken as an indicator of cognitive effort possibly arising from processing difficulty associated with a previously heard source-speech segment. However, it can also indicate a tactical pause, which is not linked to any such difficulty, for instance when the interpreter reviews previous notes to make sure s/he will be able to read them when reformulating the speech.

Brain imaging techniques may provide one solution to the problem, if they show that distinct processes tend to mobilize specific areas of the brain or generate specific activity. However, in many cases, a far simpler solution can also help, again in the form of retrospective reports. Cued retrospection, as already used with keylogging techniques a long time ago, inaccurate and incomplete as it is, could well help interpret more specifically causes of increased cognitive effort as detected by objective indicators.

14.5 Concluding remarks

In human Translation, effort cannot be reduced to one parameter in a productivity or profitability equation. On the professional side and on the didactic side, it is a determinant of quality,
and as such, an ethical obligation—up to a certain point. It is also, for at least a decade or so in every Translator’s early career, an investment in future expertise (Ericsson, 1996).

On the research side, it is also a useful concept for the investigation of Translation behaviour, including tactics, strategies and sub-optimal performance (e.g. errors, omissions and infelicities) and also socio-affective behaviour determinants.

It is particularly useful in research into Translation cognition, especially with the help of theories from cognitive sciences and with relevant technology that provides physiological evidence of mental activity. Pupillometry and brain imaging techniques, which are becoming less and less invasive, hold much promise, but when aiming for the best results, it is important to take on board the specific technical and social environment of Translation and to consider triangulating these “objective” techniques with more qualitative techniques, in particular retrospection, which can be of much help when seeking to interpret quantitative data correctly.

**Note**

1 In this chapter, “cognitive load” is used to denote the cognitive pressure that a process imposes by virtue of environmental and task-specific factors, while “cognitive effort” refers to the effort actually expended by the Translator when performing the task.

**Further reading**


A good overview and discussion of cognitive load in interpreting and beyond.


This doctoral dissertation includes very well-written discussions of cognitive effort and associated theories.


This book, which is largely devoted to interpreting and translation cognition, offers further simple-language explanations on points mentioned in this chapter.


A very fundamental book, which explains many facets of less than rational decisions in humans, but as regards this chapter more specifically, it is an excellent introduction to effortful and less effortful thinking and acting in humans.


A good overview of existing research and technologies available, including technical details, and reflections about future directions.

**References**


Translation, effort and cognition


Gile and Lei


