

Cognitive load in simultaneous interpreting: Model meets data

International Journal of Bilingualism

16(2) 228–242

© The Author(s) 2011

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1367006911402982

ljb.sagepub.com



Kilian G. Seeber

FTI, University of Geneva, Switzerland

Dirk Kerzel

FPSE, University of Geneva, Switzerland

Abstract

Seeber (2011) recently introduced a series of analytical cognitive load models, providing a detailed illustration of conjectured cognitive resource allocation during simultaneous interpreting. In this article, the authors set out to compare these models with data gathered in an experiment using task-evoked pupillary responses to measure online cognitive load during simultaneous interpreting when embedded in single-sentence context and discourse context. Verb-final and verb-initial constructions were analysed in terms of the load they cause to an inherently capacity-limited system when interpreted simultaneously into a verb-initial language like English. The results show larger pupil dilation with verb-final than with verb-initial constructions, suggesting higher cognitive load with asymmetrical structures. A tendency for reduced cognitive load in the discourse context compared to the sentence context was also found. These data support the models' prediction of an increase in cognitive load towards (and beyond) the end of verb-final constructions.

Keywords

cognitive load, pupillometry, simultaneous interpreting, task-evoked pupillary responses

1 Introduction

Scholars interested in the study of interpreting tend to disagree about the importance of language-specific factors in simultaneous interpreting. Some claim that factors, such as morphosyntactic asymmetry between source and target language, have repercussions on the simultaneous interpreting process as they increase cognitive workload, whereas others maintain that such factors are irrelevant on condition of sufficient linguistic proficiency in both the source and the target language. This discord is reflected in the interpreting literature, which is replete with personal accounts arguing in favour of the former (e.g., Ilg, 1959; Jörg, 1995; Kirchhoff, 1976; Riccardi, 1996; Riccardi & Snelling, 1997; Zanetti, 1999) and the latter point of view (e.g., Lederer, 1981; Seleskovitch, 1984; Willett, 1974). Furthermore, the bulk of the research addressing the issue of, and making claims about, cognitive

Corresponding author:

Kilian G. Seeber, FTI, Faculté de Traduction et d'Interprétation, Université de Genève, 40, Bd du Pont D'Arve, 1205 Genève, Switzerland

Email: Kilian.Seeber@unige.ch

processes in simultaneous interpreting has relied on analytical measures (see Gile, 1995; Seeber, 2011) and performance measures, primarily accuracy and speed (see Barik, 1973, 1975; Darò, 1989; Jörg, 1995; Kopczynski, 1980; Lee, 2002; Seeber, 2001, 2005; Shlesinger, 1995; Van Besien, 1999). Although the results show a considerable amount of syntactic restructuring in simultaneous interpreting between morphosyntactically asymmetrical languages, without a direct measure of online memory load, this research fails to substantiate the causal relation between the syntactic asymmetry of the languages involved and the overall cognitive load experienced by the interpreter. In fact, even comprehensive analyses of interpreters' mistakes (e.g., Barik, 1969; Darò, 1989; Kopczynski, 1980; Shlesinger, 1995), process breakdowns (e.g., Gile, 1995) and prosodic (e.g., Seeber, 2001; Shlesinger, 1994) and temporal features of the input and output (e.g., Lee, 2002; Seeber, 2005) still partly rely on a tenuous interpretation of data using theories that might just happen to fit the facts.

Among the noteworthy exceptions that have attempted a psychophysiological approach to the study of these phenomena are Petsche, Etlinger, and Filz (1993), who used EEG in order to measure brain activation during (covert) simultaneous interpreting from and into the interpreter's dominant language, and Rinne et al. (2000), who replicated the experiment using PET. Tommola and Niemi (1986) and Hyönä, Tommola, and Alaja (1995), on the other hand, used pupillometry to assess differences in cognitive load during word comprehension, repetition and translation. This research, along with the ongoing heated debate in the interpreting research community, about the (ir)relevance of structural features of language for the interpreting process and the desire to test the predictive power of the cognitive load models (CLMs) of simultaneous interpreting inspired our endeavour to use task-evoked pupillary responses (TEPRs) to measure local cognitive load during simultaneous interpreting between syntactically symmetrical and asymmetrical structures.

2 The models

The CLMs are the result of an analytical exercise attempting to illustrate the moment-to-moment memory load experienced by simultaneous interpreters rendering the content of a verb-final construction (e.g., a German SOV structure) into verb-initial language (e.g., English, thus an SVO structure). The models produce a cognitive load signature reflecting the total amount of cognitive load generated by the overlapping component tasks of simultaneous interpreting, i.e., comprehension, production and the storage of constituents in working memory. Using the verb-initial construction '*Wir glauben, die Delegierten treffen ihre Entscheidung nach einer langen Debatte*' ('We believe the delegates make their decision after a long debate', the verb '*treffen*' corresponds to 'make') as a baseline, the models show the cognitive load signatures of four possible interpreting scenarios for the verb-final construction '*Wir glauben, dass die Delegierten ihre Entscheidung nach einer langen Debatte treffen*', illustrating the difference in cognitive load (see vertical arrows in Figure 1). Although plotted against an abstract time scale, the models suggest a substantial increase in lag (the phase shift between the original and the interpretation) and, more importantly, cognitive load associated with some of the strategies employed by interpreters to deal with structural differences between the source and target language.

Waiting, the strategy by which interpreters halt production to wait for more input, forces them to store the information they receive in working memory while pausing. It is unclear whether interpreters keep this information activated through rehearsal until it can be encoded in the target language, or whether the information decays slowly enough not to require sub-vocal rehearsal. In both instances, the process would be sensitive to interference from concurrent language processing tasks (see Wickens, 2002). As a strategy, waiting allows interpreters to alleviate cognitive load temporarily, as the interruption of simultaneous language comprehension and production effectively transforms the process into a simple comprehension and memorization task. On the other hand, however, it causes a spillover effect, leading to a considerable increase in cognitive load downstream.

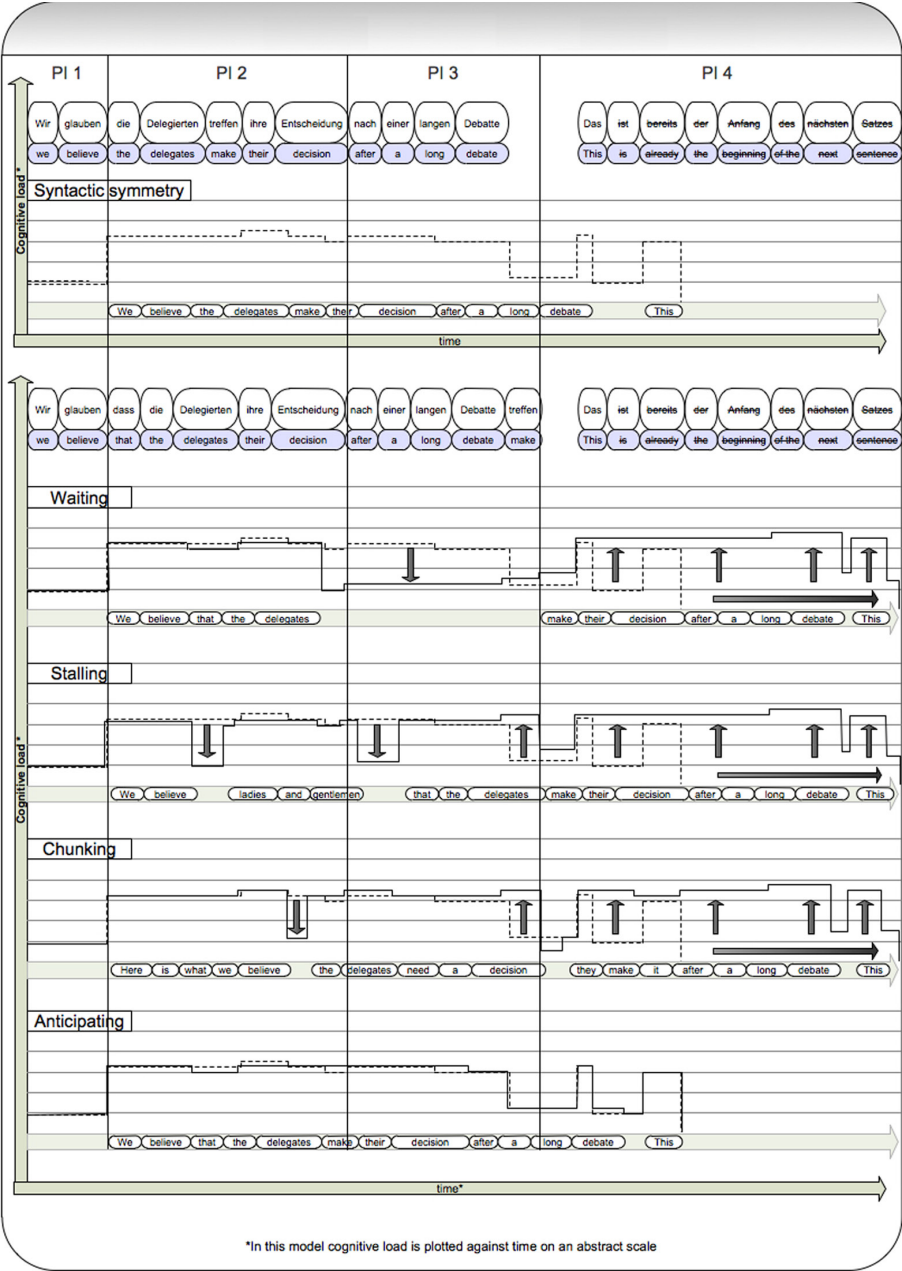


Figure 1. Cognitive load models for simultaneous interpreting. The models plot conjectured cognitive resource demands against time (on an abstract scale). The topmost graph illustrates cognitive load during simultaneous interpreting of a German verb-initial structure into English and constitutes our baseline. The remaining four graphs illustrate cognitive load during simultaneous interpreting of a German verb-final structure into English, and compare the four strategies (waiting, stalling, chunking and anticipating) to the cognitive load signature of the baseline condition. The vertical arrows indicate the variation of local cognitive load as compared to the baseline. The horizontal arrows indicate increased time lag as compared to the baseline condition. PI = period of interest.

Stalling is a strategy similar to waiting, as both aim to buy the interpreter time in order to receive more input before the integration and encoding stage. Whereas waiting normally results in a period of silence in the interpreter's output, stalling postulates the production of 'neutral padding' which fills the gap without adding any new information. As shown (see Figure 1), stalling considerably increases the interpreter's lag and adds a layer of processing complexity (and thus cognitive load) as the encoding and production of the padding material overlaps with the comprehension process.

Chunking, as a strategy in simultaneous interpreting, refers to the process whereby interpreters segment the input into smaller fragments that can be encoded without having to wait for the entire sentence to unfold. Although the input can be integrated and encoded immediately, our example illustrates how the lack of a main verb relating the arguments to each other means that the chunks often need to be strung together downstream in order to establish (or recuperate) the original meaning, causing a temporally deferred increase in cognitive load.

Anticipation is the strategy (to the extent that the process is controlled and not automatic; see Schneider & Chein, 2003) by which the interpreter attempts to guess a sentence constituent, in this case the verb, before it has been uttered in the original. This strategy allows the interpreter's overall lag to remain comparable to that of the baseline condition – at the expense of an increase in local cognitive load (see Singer, 1994) which otherwise appears to remain close to the baseline values (see Figure 1).

All in all, the models¹ therefore predict higher cognitive load for the asymmetrical condition throughout period of interest (PI) 1, PI 3 and PI 4 (see Figure 1). The most pronounced difference in cognitive load is expected in PI 4. This general prediction constituted the main hypotheses for our experiment.

3 Task-evoked pupillary responses

The pupil, the black circular opening at the centre of the human eye, regulates the amount of light that can enter the retina through its radial and circular muscles, which allow it to dilate and constrict. However, whereas the radial muscle is innervated by the sympathetic nervous system that triggers the 'fight or flight' response, the circular muscle is activated by the parasympathetic nervous system (which is primarily involved in relaxation). The process of dilation and constriction allows the pupil size to vary from 1.5 mm in bright light to up to 8–9 mm in dim light (Andreassi, 2000). If we exclude the effect of drugs, which can cause constriction (as is the case with alcohol, opioids and antipsychotics) or dilation (in the case of stimulants of the central nervous system and hallucinogens), the pupil, broadly speaking, reacts to three kinds of stimuli: luminosity, emotions and cognitive activity. First, the *pupillary light reflex* causes the pupil to constrict in response to excessive amounts of light in order to protect the retina against damage (Clarke, Zhang, & Gamlin, 2003). Similarly, the *pupillary accommodation reflex* causes the pupil size to adjust in response to an object approaching or suddenly appearing in front of the eye (Ellis, 1981). Second, average pupil size has been found to be affected by a person's general state of arousal (Hess, 1975) and to increase when subjects look at visual stimuli in which they are interested. Third, there is a sizeable amount of evidence linking problem-solving performance to the activation of the sympathetic nervous system, e.g., increased electrodermal activity, heart rate, blood pressure (see Kahneman, Tursky, Shapiro, & Crider, 1969), and thus by extension, also pupil dilation.

Pupillometry, the measurement of the size (either in terms of diameter or surface) of the pupil of the eye, was very popular in the 1960s and 1970s, yet was eventually all but replaced by modern imaging techniques (e.g., positron emission tomography, electroencephalography and magnetic resonance imaging), and only recently made a comeback as a psychophysiological technique to

assess cognitive load (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2003). After Hess and Polt's (1964) seminal work on mental arithmetic suggesting that the size of the pupil during mental activity changes as a function of task difficulty, a series of experiments have replicated these results, corroborating the usefulness of the method in a number of different paradigms, such as short-term memory load (Beatty & Kahneman, 1966), pitch discrimination (Kahneman & Beatty, 1966) and mental overload (Pook, 1973). With few exceptions (e.g., Carver, 1971) it seems to be generally accepted that the amplitude of TEPRs indicates the intensity of information processing (Just & Carpenter, 1993) and thus reflects processing demands in memory, language processing, reasoning and perception tasks (Beatty, 1982). Furthermore, mean pupil dilation during listening comprehension tasks was shown to correlate more strongly with grammatical complexity than did subject ratings of comprehensibility (Schluroff, 1982). In order to understand TEPRs, it is important to consider some of the pupil's fundamental physiological characteristics beyond those already addressed.

First, pupillary response can occur as quickly as 200 ms after stimulus presentation (Lowenstein & Loewenfeld, 1962), although dilation as a response to cognitive load usually seems to begin after 300–500 ms (Beatty, 1982; Hoeks & Levelt, 1993). Lowenstein and Loewenfeld (1962) furthermore observe that pupil diameter is largest in rested individuals, whereas it decreases with fatigue. Similarly, it has been suggested that pupillary response decreases with age, weakening the correlation between cognitive load and pupil dilation (Van Gerven et al., 2003). This phenomenon is probably related to senile miosis, i.e., the overall decrease of the pupil with age, as well as a decrease in its sensitivity, and was observed by Van Gerven et al. in subjects between 62 and 73 years of age.

Another phenomenon to be kept in mind when interpreting pupillometric data is the manifestation of cognitive overload, i.e., when the task exceeds the cognitive resources available to perform it. Although Peavler (1974) suggests that once the capacity threshold has been reached, pupil dilation will stabilize, Pook (1973) and Granholm, Asarnow, Sarkin, and Dykes (1996) found that pupil dilation decreases rapidly once cognitive overload was reached. Pook provides a tentative explanation for this finding, suggesting that, 'mental overload of the central nervous system causes it to influence the pupil to constrict and to inhibit reception of some of the input stimuli because the central nervous system is receiving information at a rate above its capacity' (1973, p. 1001).

4 Task-evoked pupillary responses in simultaneous interpreting

According to Beatty, TEPRs most likely reflect 'the cortical modulation of the reticular core during cognitive processing' (1982, p. 290), and thus provide a composite measure of task-induced processing load. What is more, TEPRs may reveal the joint demand for resources in pairs of time-shared tasks (Beatty, 1982). The examples of experiments attempting to apply TEPRs to more complex tasks, however, which might not allow the straightforward separation of individual processing stages, are scarce. Tommola and Niemi (1986) and Hyönä et al. (1995) carried out two TEPR studies on simultaneous interpreters and, while identifying some major methodological challenges, corroborated the general effectiveness and validity of the methodology. The earlier of the two papers describes a pilot study on the effect of syntactic complexity on mental load, designed to corroborate the hypothesis that simultaneous interpreting (from Finnish into English) of discourse containing left-branching structures will cause more cognitive load than discourse containing right-branching structures. Among the limitations of the study is its failure to control potential confounds in the experimental materials (e.g., lexical density, lexical frequency, grammatical complexity, abstractness). In spite of these shortcomings, the paper demonstrates the feasibility of applying TEPRs to complex processing tasks, revealing a maximum variance of 0.7 mm in pupil

dilation between the two conditions (see Tommola & Niemi, 1986). The more recent paper (Hyönä et al., 1995) heeds Tommola and Niemi's (1986) call for more systematic experimentation with two experiments designed to study the variation in processing load during simultaneous interpreting. Whereas the first experiment compares average cognitive load across three different language processing tasks at the discourse level (listening comprehension, shadowing and simultaneous interpreting) and looks at between-task differences, the second experiment replicates the first experiment at the word level (word recognition, word repetition and word translation), thus analysing cognitive load variations within a given task. The results for both experiments indicate a significant increase of mean pupil dilation, and consequently cognitive load, from listening comprehension to shadowing and simultaneous interpreting. These data suggest that this paradigm can be used to assess overall cognitive load during simultaneous interpreting. Similarly, Scheepers and Crocker (2004) and Engelhardt, Ferreira, and Patsenko (2010) successfully applied this technique to show the local manifestation of cognitive load in a comprehension task. However, to date no attempt has been made to identify *local cognitive load* during simultaneous interpreting using online psychophysiological measures.

5 The experiments

A set of experiments was designed to measure online cognitive load during simultaneous interpreting of German verb-initial (i.e., syntactically symmetrical) and verb-final (i.e., syntactically asymmetrical) constructions into English.

Given the temporal constraints imposed by the task, which requires interpreters to start with the production of their English output before the end of the German input, we expected cognitive load during the asymmetrical condition to be higher than during the symmetrical condition. As explained earlier, when confronted with verb-final structures, interpreters working into a verb-initial language avail themselves of an array of strategies allowing them to postpone the production of the (still unknown) original verb. Applying these strategies will tax the inherently capacity-limited working memory system (Just & Carpenter, 1992) and increase cognitive load more than interpreting a source-text structure where all the sentence constituents needed for the production of the target language utterance are readily available. Using the symmetrical condition as a baseline, we therefore predicted that syntactic asymmetry would engender an increase in cognitive load in PI 2, PI 3 and PI 4 (see Figure 1).

In order to allow participants to rely on inference processing, i.e., establish relations between and among individual parts of discourse as well as their own knowledge of the world, so as to complement their mental model of the information to be interpreted (see Johnson-Laird, 1983; Van Dijk & Kintsch, 1983), we decided to include the amount of available context as our second independent variable. While we do not venture to predict the extent to which discourse context will reduce load for the simultaneous interpretation of the two structures, we expect local load to be less as compared to the single-sentence context condition.

5.1 Methodology

5.1.1 Participants. Ten participants (six female, four male) were recruited from the online AIIC² database of conference interpreters with their professional domicile in Switzerland. They all had English as an A language³ and German as a C language.⁴ All participants had a minimum of seven years of professional experience and were between 32 and 63 years old (mean age = 49.7). Individual sessions were scheduled for participants according to their availability. Participation in the

experiment was not remunerated, but travel expenses for those coming from outside Geneva were reimbursed.

5.1.2 Apparatus. Pupil dilation was measured using an EyeLink II head-mounted binocular eye tracker at 250 Hz while the sound files of the recorded materials were played into the participants' ears using Bang and Olufsen A8 earphones with a total frequency range of 50–20,000 Hz and an impedance of 19 ohms.

5.1.3 Design. There were two versions of each target sentence (symmetrical and asymmetrical) and two context conditions (discourse context and sentence context). In the *sentence context* condition, target sentences were presented in clusters of three sentences: one target sentence (symmetrical or asymmetrical) embedded in two context-setting sentences (see Just & Carpenter, 1993). In the *discourse context* condition, target sentences were presented in flowing discourse, embedded in two portions of coherent context. The sequence of the context condition followed a modified Latin Square design. Each participant was exposed to all items and to all experimental conditions, but never heard more than one version of any individual item.

5.1.4 Materials. For the *sentence context* condition, 32 target sentences were created (see Table 1). The stimuli were 16 symmetrical and 16 asymmetrical constructions, and 18 syntactically unrelated fillers. Each trial consisted of a target sentence preceded and followed by a syntactically unrelated context-setting sentence. These sentences were used to ensure that the interpreter would already be interpreting simultaneously when encountering the target sentence, and to avoid 'first sentence' and 'last sentence' phenomena,⁵ which are not representative of the simultaneous interpreting task. The order of trials was randomized across lists. For the *discourse context* condition, another 32 target sentences were created. The stimuli contained 16 symmetrical and 16 asymmetrical constructions that were embedded in two transcripts of addresses given by Angela Merkel, German Federal Chancellor, at the American Council on Germany in January 2006 and by Thomas de Maizière, Head of the German Federal Chancellery, at the Federal Academy for Security Policy in January 2006. Both speeches were divided into two sections of approximately eight minutes. Each of the four resulting sections was manipulated so as to contain eight symmetrical and asymmetrical trials alternating at regular intervals of approximately 60 seconds. All materials were oralized, i.e., delivered in an attempt to preserve the key features of improvised speech, and recorded by a female native speaker of standard High German at an average presentation rate of 120 words per minute. The temporal

Table 1. Symmetrical and asymmetrical structures used in experimental materials.

	Periods of interest			
	PI 1	PI 2	PI 3	PI 4
Mean duration (in ms)	900	2500	2500	2000
Symmetrical syntax	Ich glaube	die Delegierten treffen ihre Entscheidung	nach einer langen Debatte	[pause]
[gloss]	I believe	the delegates make their decision	after a long debate	[pause]
Asymmetrical syntax	Ich glaube	dass die Delegierten ihre Entscheidung	nach einer langen Debatte treffen	[pause]
[gloss]	I believe	that the delegates their decision	after a long debate make	[pause]

evolution of the target sentences was controlled by means of digital sound manipulation (using Adobe Audition©) and the duration of the four periods of interest was kept constant across trials.

5.1.5 Procedure. Participants were seated at a desk in a soundproof room with constant artificial lighting looking at a fixation cross on a computer screen. The sequence of the experimental materials (sentence context materials, discourse context materials) was randomized using a fixed random order.

5.1.6 Confound

5.1.6.1 Word frequency. Given the increased cognitive demands during the processing of low-frequency words (see Mitchell, 2004), all words contained in our target sentences were chosen among the 10,000 most frequently used words in German (Datenbank Deutscher Wortschatz,⁶ 2003). The word frequency of the intervening discourse context was not controlled.

5.1.6.2 Cognates. An increase in processing speed for cognates (i.e., words with a common etymological origin) was observed both in bilingual speech production and speech recognition tasks (Costa, Caramazza, & Sebastien-Galles, 2000; Sherkina, 2003). To account for this phenomenon the number of cognates in the stimulus lists was controlled.

5.1.6.3 Level of abstractness. West and Holcomb (2000) suggest that participants encode and retrieve concrete words faster and more completely than abstract words. Furthermore, comprehension of concrete sentences appears to be swifter than that of abstract ones. The level of abstractness of all the nouns of the *sentence context* stimulus list was matched based on the classification in the frequency list (Datenbank Deutscher Wortschatz, 2003).

5.1.6.4 Prosody. All materials were recorded by a native speaker of standard High German (JG), with passive knowledge of French and English, as oralized texts. Unlike read text, which is hallmarked by high lexical density, complex syntactic structures and intonation patterns that are less helpful for comprehension than those of natural, improvised speech (see Déjean-Le Féal, 1978), oralized discourse contains prosodic characteristics similar to those of spontaneous speech (Shlesinger, 1994). This issue is particularly relevant as some researchers identify the lack of natural prosodic features of the materials used in experimental research as one of the main reasons for the interpreters' inability to anticipate the German verb (see Lederer, 1981). In the present experiment we controlled the materials' principal prosodic features (fundamental frequency, intensity, speed and pauses) and digitally remastered and streamlined (see Seeber, 2001) all target sentences to make them temporally identical.

5.2 Data analysis

5.2.1 Data normalization. Participants were allowed a 10-minute break between each of the four experimental sequences during which the head-mounted eye tracker was removed. Consequently, the absolute distance between the pupil and the infrared cameras (and with it the pupil's absolute size) changed from one sequence to the next. In order to account for this difference and to allow the comparison of the data gathered during different trials, a spatula with an artificial pupil was built (see Just & Carpenter, 1993). This spatula was put over the participants' dominant eye during the calibration of the eye tracker and the relative pupil size was recorded for subsequent normalization (pupil sizes were normalized to 4.00 mm).

5.2.2 Blinks. Blinks were removed from the data sets before the analysis. When blinks masked more than 50 per cent of the pupil size measurements during a target sentence, the item was not included in the analysis.

5.2.3 Lag. All instances where participants skipped a target sentence because they had accumulated an excessive lag were not included in the analysis. Overall, i.e., including trials eliminated due to blinks and lag, the final analysis was based on 513 of the 640 target sentences, i.e., approximately 80 per cent of the total number of trials.

6 Results

The mean pupil dilation in each of the four periods of interest was used to analyse main effects of, and interactions between, syntax, context and PI using three-way, repeated-measures ANOVAs: 2 context \times 2 syntax \times 4 PIs. These interactions were then further explored using contrast analyses. Finally, given the predictions of the models, separate paired *t*-tests were carried out to compare pupil dilation between the two levels of the two independent variables (context and syntax) for each PI.

Of all main effects and interactions only the main effect of PI violated the assumption of sphericity, $\chi^2(2) = 18.97$, $p < .005$, and was corrected using Greenhouse–Geisser's estimates of sphericity ($\epsilon = .40$). All effects are reported as significant at $p < .05$.

Whereas the main effects of syntax and PI on pupil dilation were not significant, $ps > .13$, the main effect of context on pupil dilation approached significance, $F(1,7) = 4.61$, $p = .069$, showing that pupil diameter was slightly smaller in the discourse context than in the sentence context (3.981 vs. 4.007 mm). The interaction effect between context and syntax was not significant, $p = .42$. However, the interaction between context and PI was nearly significant, $F(3,21) = 3.05$, $p = .051$, and the interaction effect between syntax and PI reached significance, $F(3,21) = 3.87$, $p < .05$, indicating that the effects of context and syntax grew stronger at the later PIs. The interactions between PI and context and between PI and syntax are presented in Figure 2.

To follow up on the (almost) significant interactions, we compared the two context and two syntax conditions for each PI. The difference in mean pupil dilation between sentence context and discourse context was nearly significant in PI 3 (3.973 vs. 4.02 mm, respectively), $t(7) = 2.32$, $p = .054$, and in PI 4 (4.0 vs. 4.04 mm, respectively), $t(7) = 1.99$, $p = .087$. Mean pupil dilation was significantly smaller with symmetrical syntax than with asymmetrical syntax in PI 4 (3.996 vs. 4.048 mm, respectively), $t(7) = 3.01$, $p < .05$.

7 Discussion

The discussion of the above results should be prefaced by a brief consideration on the magnitude of the pupil dilation as measured in our experiments. The differences in pupil dilation across the different experimental factors (syntax and context) are considerably smaller than those reported by Hyönä et al. (1995) between listening, shadowing and simultaneous interpreting. This is not surprising given that the latter differences reflect the increase in cognitive load generated by the combination of individual language processing tasks competing for the same cognitive resources (see Seeber, 2007; Wickens, 2002). Conversely, the magnitude of our values appears to be in line with that measured by Scheepers and Crocker (2004) for differences in sentence disambiguation.

Furthermore, it is worth noting that the raw data were not corrected to account for the latency of the pupil reactions, which, as we have seen before, can range between 200 ms and 500 ms. In spite of a 1500 ms pause between context and target sentences, we cannot exclude that the pupil dilation

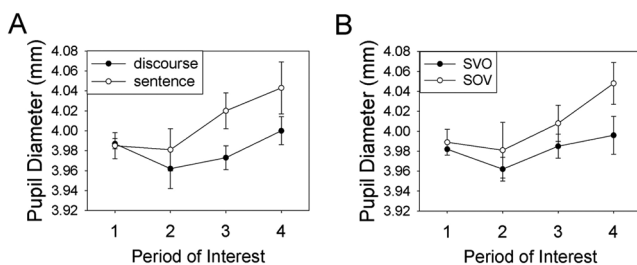


Figure 2. Mean pupil dilation as a function of period of interest and context (panel A) and as a function of period of interest and syntax (panel B). Error bars indicate the standard error of the mean (between-subjects).

during PI 1 was confounded by the load exported from the preceding sentence. This would explain why, across the four conditions, pupil dilation was relatively large in PI 1, despite the arguably easier task (at the beginning of the sentence, the interpreter merely needs to engage in a comprehension task, but not in a concurrent production task) and despite the simple syntactic structure of the input (a minimal verb phrase).

Based on the prediction of the cognitive load models, our first hypothesis was that, all other things being equal, cognitive load during simultaneous interpreting of syntactically asymmetrical structures would exceed that of symmetrical structures. The models (see Figure 1) illustrate this increase in cognitive load from PI 2 onwards, but particularly in PI 3 and PI 4. The results (see Figure 2, panel B) indicate that cognitive load indeed tends to be higher during the simultaneous interpretation of asymmetrical structures, and although the main effects of syntax and PI on load did not reach significance, the interaction between context and PI did. The follow-up analysis of this interaction revealed a significantly larger mean pupil dilation and thus significantly more cognitive load in PI 4 during the interpretation of asymmetrical structures, which confirms the predictions of the CLM. The fact that the relative maximum local load occurs in PI 4 could be evidence for the phenomenon referred to as ‘exported load’ by Gile (2008): given the dynamic nature of the simultaneous interpreting task the load experience at any given point in time might be influenced by operations upstream. In the case at hand, the interpreter would already be experiencing a considerable amount of cognitive load before even starting to hear the sentence following the target sentence.

It is important to keep in mind that our data reflect a many-to-one mapping of different strategies applied by interpreters in order to deal with syntactically asymmetrical (and symmetrical) sentence structures. This means that we are currently unable to identify which of the strategies we describe in our models (waiting, stalling, chunking, or anticipating) account for the increase of load. What we can say is that, generally speaking, the interpretation of syntactically asymmetrical structures causes more cognitive load than the interpretation of syntactically symmetrical structures towards the end of the sentence.

Unlike our first hypothesis, our second hypothesis was not directly deduced from the CLMs, which in their current version are unable to account for the notion of inference processing and context. Our prediction was based on the literature on language comprehension (Clark, 1975; Garrod, O’Brien, Morris, & Rayner, 1990; Harris & Monaco, 1978; Van den Broek, 1994), which describes the advantages of context in terms of facilitated activation of available concepts. Our data (see Figure 2, panel B) show a nearly significant interaction between context and PI, and the follow-up analysis revealed a nearly significant context effect in PI 3 and PI 4, where the lack of context led to an increase in pupil dilation and thus cognitive load. This finding corroborates the importance of context for the simultaneous interpreting process in general, and for the interpretation of syntactically asymmetrical structures in particular.

Overall, our results suggest considerable local variation in cognitive load, which in the case of verb-final sentences interpreted into a verb-initial language, is particularly high towards the end of such constructions. Given that the increase in load for verb-final constructions was only significant in PI 4, which merely constitutes approximately 25 per cent (i.e., 2000 ms) of the total time window under consideration, our data support Gile's (2008) prediction of an increase in load towards the end of sentences, but do not corroborate his 'tightrope hypothesis' (Gile, 1999), according to which simultaneous interpreters work close to the limit of their cognitive capacity most of the time. In fact, it would appear that during simultaneous interpreting of verb-final structures into a verb-initial language, such capacity limits are – if at all – only approached towards the end of the sentence. The fact that no sudden and pronounced pupil constrictions were found during the processing (i.e., the simultaneous interpreting) of our target sentences furthermore suggests that participants did not experience cognitive overload (see Granholm et al., 1996; Pook, 1973), and that the local peaks are merely relative maxima, not absolute maxima.

8 Conclusion

This experiment was designed to test the cognitive load models of simultaneous interpreting (Seeber, 2011), in other words, to gather more tangible evidence concerning the manifestation of local cognitive load during simultaneous interpreting between syntactically asymmetrical language structures. By doing so, we wanted to contribute empirical findings to the ongoing debate about the role of syntactic structure of languages in simultaneous interpreting. Our results corroborate our two working hypotheses. First, as predicted by the models, cognitive load during simultaneous interpreting of syntactically asymmetrical structures increased towards the end of the sentence (i.e., PI 4), where significant differences were found. Second, as suggested by the literature, the availability of discourse context facilitates the simultaneous interpretation of sentences and reduces the amount of cognitive load necessary to interpret individual sentences, especially towards the end of verb-final constructions.

These results are at odds with a *universalist* view of interpreting, according to which structural differences of the languages involved are irrelevant to the process. In fact, our data suggest that, at least in the case of German verb-final structures that need to be interpreted into a strictly verb-initial language like English, such differences do exist and are the cause of a significant increase of cognitive load. While our data support the predictive power of the CLMs, a more detailed analysis of the interpreters' output would be necessary to map individual strategies onto their respective cognitive load signature. This would allow a direct comparison of strategies not only in terms of quality of the output but also, and importantly, in terms of cognitive processing cost.

In the meantime, our results warrant at least a qualification of the main tenet of the so-called interpretive theory. Instead of an imperative strategy, the syntactic reshuffling of source structure constituents in the target language should be viewed as an operation executed by the simultaneous interpreter whenever the asymmetrical nature of the two languages does not allow the preservation of the linearity of the *main* sentence constituents (subject, verb and object). Although discourse context seems to allow the interpreter to attenuate some of the cognitive load caused by this operation, we have shown that such processing comes at a cost, and that it causes significantly more cognitive load than when structural symmetry allows a certain degree of linearity between the original and the interpretation to be preserved.

Acknowledgements and funding

Dirk Kerzel was supported by the Swiss National Foundation 100011-107768. We wish to thank David Souto and Caroline Dunand for helping to collect the data.

Notes

1. For a detailed discussion of the CLMs, see Seeber (2011).
2. AIIC is the International Association of Conference Interpreters (www.aiic.net).
3. According to the definition used by the AIIC, an A language is, 'The interpreter's native language (or another language strictly equivalent to a native language), into which the interpreter works from all her or his other languages in both modes of interpretation, simultaneous and consecutive.'
4. According to the definition used by the AIIC, a C language is, 'The language(s) of which the interpreter has a complete understanding and from which she or he works. Interpreters often have several C languages.'
5. When interpreting flowing discourse, the original (input) and the interpretation (output) overlap most of the time. During the first and the last sentence of an oral statement, however, one of the two streams (output and input respectively) is – at least partially – missing.
6. <http://wortschatz.uni-leipzig.de>

References

- Andreassi, J. L. (2000). *Psychophysiology: Human behavior and physiological response* (4th edn) Mahwah, NJ: Lawrence Erlbaum.
- Barik, H. C. (1969). *A study of simultaneous interpretation*. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
- Barik, H. C. (1973). Simultaneous interpretation. Temporal and quantitative data. *Language and Speech*, 16, 237–270.
- Barik, H. C. (1975). Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech*, 18, 272–297.
- Beatty, J. (1982). Task evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91, 276–292.
- Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, 5, 371–372.
- Carver, R. P. (1971). Pupil dilation and its relationship to information processing during reading and listening. *Journal of Applied Psychology*, 55, 126–134.
- Clark, H. H. (1975). Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing* (pp. 169–174). New York: Association for Computing Machinery.
- Clarke, R. J., Zhang, H., & Gamlin, P. D. R. (2003). Characteristics of the pupillary light reflex in the alert Rhesus Monkey. *Journal of Neurophysiology*, 89, 3179–3189.
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 1283–1296.
- Darò, V. (1989). The role of memory and attention in simultaneous interpretation: A neurolinguistic approach. *The Interpreters' Newsletter*, 2, 50–56.
- Déjean-Le Féal, K. (1978). *Léctures et improvisations – Incidences de la forme de l'énonciation sur la traduction simultanée*. Unpublished doctoral dissertation, Université de Paris III, Paris.
- Ellis, C. J. (1981). The pupillary light reflex in normal subjects. *British Journal of Ophthalmology*, 65, 754–759.
- Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Rapid communication: Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, 63, 639–645.
- Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 250–257.
- Gile, D. (1995). *Regards sur la recherche en interprétation de conférence*. Lille: Presses universitaires de Lille.
- Gile, D. (1999). Testing the effort models' tightrope hypothesis in simultaneous interpreting: A contribution. *Hermes*, 23, 153–172.

- Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum*, 6, 59–77.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33, 457–461.
- Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, 107, 1–22.
- Hess, E. H. (1975). *The tell-tale eye*. New York: Van Nostrand.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143, 1190–1192.
- Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments and Computers*, 25, 16–26.
- Hyönä, J., Tommola, J., & Alaja, A. (1995). Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48A, 598–612.
- Ilg, G. (1959). *L'enseignement de l'interprétation à l'école d'Interprètes de L'université de Genève*. Genève: Université de Genève.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Jörg, U. (1995). *Verb anticipation in German–English simultaneous interpreting: An empirical study*. Unpublished Master's thesis, Bradford University.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310–339.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583–1585.
- Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology*, 79, 164–167.
- Kirchhoff, H. (1976). Das Simultandolmetschen: Interdependenz der Variablen im Dolmetschprozess, Dolmetschmodelle und Dolmetschstrategien. In H. W. Drescher (Ed.), *Theorie und Praxis des Übersetzens und Dolmetschens* (pp. 59–71). Frankfurt: Lang.
- Kopczynski, A. (1980). *Conference interpreting: Some linguistic and communicative problems*. Unpublished doctoral dissertation, University of Poznan.
- Lederer, M. (1981). *La traduction simultanée: Expérience et théorie*. Paris: Minard.
- Lee, T.-H. (2002). Ear voice span in English into Korean simultaneous interpretation. *Meta*, 47, 596–606.
- Lowenstein, O., & Loewenfeld, I. E. (1962). The pupil. In H. Davson (Ed.), *The eye, Vol. 3, Muscular mechanisms* (pp. 231–267). New York: Academic Press.
- Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, EPRs and beyond* (pp. 15–32). Hove: Taylor & Francis.
- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, 11, 559–566.
- Petsche, H., Etlinger, S. C., & Filz, O. (1993). Brain electrical mechanisms of bilingual speech management: An initial investigation. *Electroencephalography and Clinical Neurophysiology*, 86, 385–394.
- Pook, G. K. (1973). Information processing vs pupil diameter. *Perceptual and Motor Skills*, 37, 1000–1002.
- Riccardi, A. (1996). Language specific strategies in simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching translation and interpreting 3* (pp. 213–222). Amsterdam: John Benjamins.
- Riccardi, A., & Snelling, C. (1997). Sintassi tedesca: vero o falso problema per l'interpretazione? In L. Gran & A. Riccardi (Eds.), *Nuovi orientamenti negli studi sull'interpretazione* (pp. 143–158). Padova: CLEUP.
- Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., et al. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, 294, 85–88.
- Scheepers, C., & Crocker, M. W. (2004). Constituent order priming for reading to listening: A visual-world study. In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, EPRs and beyond* (pp. 167–185). Hove: Taylor & Francis.

- Schlurhoff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, 17, 133–145.
- Schneider, W., & Chein, J. M. (2003). Controlled and automatic processing: Behavior, theory and biological mechanisms. *Cognitive Science*, 27, 525–559.
- Seeber, K. G. (2001). Intonation and anticipation in simultaneous interpreting. *Cahiers de Linguistique Française*, 23, 61–97.
- Seeber, K. G. (2005). Temporale Aspekte der Antizipation beim Simultandolmetschen von SOV-Strukturen aus dem Deutschen. In A. Künzli (Ed.), *Empirical research into translation and interpreting: Processes and products. Bulletin suisse de linguistique appliquée Vals-Alsa numéro 81* (pp. 123–140). Neuchâtel: Institut de linguistique de l'Université de Neuchâtel.
- Seeber, K. G. (2007). *Thinking outside the cube: Modeling language processing tasks in a multiple resource paradigm*. Conference Proceedings, Interspeech 2007, Antwerp, Belgium.
- Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories – new models. *Interpreting*, 13, 176–204.
- Seleskovitch, D. (1984). Les anticipations de la compréhension. In D. Seleskovitch & M. Lederer (Eds.), *Interpréter pour traduire* (pp. 273–283). Paris: Didier Erudition.
- Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis*. Amsterdam: John Benjamins.
- Sherkina, M. (2003). The cognate facilitation effect in bilingual speech processing. *Toronto Working Papers in Linguistics*, 21, 135–151.
- Shlesinger, M. (1994) Intonation in the production and perception of simultaneous interpreting. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 225–236). Amsterdam: John Benjamins.
- Shlesinger, M. (1995). Shifts in cohesion in simultaneous interpreting. *The Translator*, 1, 193–214.
- Singer, M. (1994). Discourse inference process. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479–515). San Diego, CA: Academic Press.
- Tommola, J., & Niemi, P. (1986). Mental load in simultaneous interpreting: An on-line pilot study. In L. S. Evensen (Ed.), *Nordic research in text linguistics and discourse analysis* (pp. 171–184). Trondheim: Tapir.
- Van Besien, F. (1999). Anticipation in simultaneous interpretation. *Meta*, 44, 251–259.
- Van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). San Diego, CA: Academic Press.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse representation*. New York: Academic Press.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2003). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41, 167–174.
- West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12, 1024–1037.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3, 159–177.
- Willett, R. (1974). Die Ausbildung zum Konferenzdolmetscher. In V. Kapp (Ed.), *Übersetzer und Dolmetscher* (pp. 96–97). Heidelberg: Quelle & Meyer.
- Zanetti, R. (1999). Relevance of anticipation and possible strategies in the simultaneous interpretation from English into Italian. *The Interpreters' Newsletter*, 9, 79–98.

About the authors

Kilian G. Seeber is assistant professor at ETI's Interpreting Department (University of Geneva). He completed his undergraduate training in translation and interpreting at the University of Vienna, did his graduate work in interpreting at the University of Geneva and his postdoctoral research in psycholinguistics at the University of York. The main focus of his research to date has been on cognitive aspects of language processing, more specifically on anticipation and working memory in simultaneous interpreting.

Dirk Kerzel is Professor of Cognitive Psychology in the Psychology Department of the University of Geneva. He did his undergraduate training in psychology and psycholinguistics at the University of Bielefeld. His graduate work was on perception and action at the Max-Planck Institute in Munich. Before his appointment in Geneva, he was a research fellow at the University of Giessen. His current research focuses on eye–hand coordination and the link between eye movements and attention.