Temporal eye-voice span as a dynamic indicator for cognitive effort during speech processing:
A comparative study of reading aloud and sight translation

Abstract: This chapter examines the dynamic latency between human translators' reading input and speaking output during reading aloud and sight translation. It aims to determine whether the temporal *eye-voice span* (EVS) at sentence level could work as a dynamitic indicator of cognitive effort during speech processing. Thirty participants performed both the reading aloud and sight translation tasks with either English or Chinese texts. Their eye movements and speech outputs were recorded by an eye-tracker and an audio recorder, respectively. EVS at sentence initial and sentence terminal positions in the reading aloud and sight translation tasks were analyzed. The results show that the lengths of both sentence-initial and sentence-terminal EVS in sight translation tasks are significantly longer than those in reading aloud tasks. This is in line with results of total gaze fixation duration and fixation count, which are closely related to cognitive effort. Further correlation tests show that both initial and terminal EVS yield a positive although weak correlation with the fixation indexes in the sight translation tasks, while discrepant results emerge in the reading aloud tasks. Hence, we suggest that temporal EVS can be used to discriminate different types of reading-speaking tasks and has the potential to serve as a dynamic indicator of cognitive effort during sight translation.

Keywords: eye-voice span; eye-tracking; sight translation; reading aloud; cognitive effort.

1. Introduction

Eye-voice span (EVS) is "a measure of the amount of material or time by which the voice lags behind the eyes in oral reading" (Morton 1964, p. 347). The material span is referred to as *spatial* EVS (De Luca et al. 2013) or *eye-voice distance* (Inhoff et al. 2011); while the time span is called *temporal* EVS (Geyer 1966) or *eye-voice latency* (Holmqvist et al. 2011). A considerable body of psycholinguistic evidence suggests that both spatial and temporal EVS are common phenomena when reading aloud. In addition, EVS is claimed to be indicative of reading competence. For example, Buswell (1921) proposed that people's spatial EVS expand as their reading competence develops. Inhoff et al. (2011) found that the duration of temporal EVS can be routinely down-regulated to make the eyes closer to the spoken word through two complementary regulation strategies: increasing individual fixation duration; and programming regressions. Laubrock & Kliegl (2015) support their argument by pointing out that both programming regressions and refixations can be taken as ways of regulating the duration of temporal EVS, if it becomes too long. Thus, whereas a substantially larger spatial EVS signifies a higher level in one's reading competence, an overly long temporal EVS could instead account for a lower level in reading competence. It follows that spatial and temporal EVS may represent different or even opposite competence indications, and thus they should be dealt with separately.

The calculation of spatial EVS began earlier than the calculation of temporal EVS. Spatial EVS can be measured using the light-out approach, a report-based off-line method.[1] In contrast, temporal EVS is more frequently measured using modern on-line

[1] When a subject read to a predetermined position, the researcher would switch off the lights or the screen

methods, including eye-tracking and audio recording. Laubrock & Kliegl (2015) point out that the report-based procedures may ignore the effects of readers' parafoveal preview, the guessing effect, and some task-dependent strategies. As a result, spatial EVS measured by such an off-line method could be "grossly overestimated" (Laubrock & Kliegl 2015, p. 2). The temporal EVS measured by modern on-line methods, however, could effectively address these problems and yield a more precise picture. Thus, the latter method has been widely adopted for reading aloud research in recent years, and will be the approach adopted in the present research. From now on, EVS will only refer to temporal EVS in our presentation.

Apart from being commonly seen in reading aloud research, EVS occurs ubiquitously during sight translation as well (e.g., Dragsted et al., 2009; Zheng & Zhou 2018). In a sight translation task, translators are presented with a written source text (ST), and an immediate oral reformulation of the ST into the target language is required. Thus, sight translation is "a specific type of written translation as well as a variant of oral interpretation" (Lambert 2004, p. 298; but see Ho et al., 2020). By coordinating the tasks of reading/receiving new information and orally delivering the translated information during sight translation, the translator's eye movements and speech outputs have to work together to perform the task smoothly.

With reference to Gile's (2009, this volume) Effort Model of sight translation (= Reading + Short-term Memory + Production + Coordination), it can be assumed that: (a) managing reading input and oral output demands extra effort than carrying out the two activities separately; (b) coordinating reading and speaking involves attention shifts; (c) devoting cognitive effort to short-term memory and coordination is, to some extent, an automatic mental activity. In short, having EVS during sight translation indicates that cognitive effort is devoted to and divided among Reading, Short-term Memory, Production, and Coordination, the four elements suggested by Gile (2009).

By contrast, reading aloud involves less cognitive processing compared with those in sight translation. During reading aloud, the "lower level processes are usually enabled as soon as the word is encoded" (Carpenter & Just 1983, p. 304), whereas the sight translation task requires a higher-level integration such that more cognitive processing is involved. This explains why sight translation is generally considered to demand more cognitive effort than reading aloud. Nonetheless, the underlying mechanism accounting for such a difference in EVS between different types of reading-speaking tasks is worth investigating.

Sandrelli (2003), Agrifoglio (2004) and Gile (2009) agree that translators have to read ahead to identify keywords, re-structure the linguistic elements in advance, and produce smooth oral renditions in the target language, when they are performing a sight translation task. In this sense, concerning its dynamic characteristics, an EVS between and within sentences could be the result of both macro- and micro-planning activities. While macro-planning involves the semantic and conceptual preparation between sentences, micro-planning is concerned with the cognitive process within sentence segments (Dragsted & Hansen 2009). Huang (2011) examined sentence boundaries in the English-Chinese sight translation task and found that reading input was ahead of oral output in 99.4% cases. This means that the voice was ahead of the fixations in only 0.6% of cases, where "the interpreter employed anticipation skills during interpreting and was able to predict from the context the content was coming up in the speech" (Huang, 2011, p. 64). This finding indicates how common it is to show a positive EVS in sight translation and that, although the oral output overwhelmingly occurs behind the

---

and count the number of words that the reader could produce orally. The number of words was then accounted as a spatial distance.

reading input, an overlap between reading and oral production during sight translation could happen. Zheng & Zhou (2018) show that the time for reading ahead into a metaphorical expression is normally longer than that beyond a metaphorical expression. This finding substantiates that the length of EVS is affected by the local processing difficulty caused by the ST. Moreover, compared to reading aloud tasks, the *end of sentence effects* in sight translation tasks could cause extra processing burdens at the sentence terminals, because translators need to finish the integration that was not completed earlier (Carpenter & Just 1983, p. 301). Hence, EVS may have a close relationship with the cognitive effort expended on the dynamically changing processing unit during the reading-speaking activities.

Although EVS during reading aloud has received substantial attention over the last century, there is scarce research on EVS during sight translation, and even fewer comparative studies of EVS between sight translation and reading aloud processes. To fill this gap, the present chapter aims to determine whether EVS (temporal), which represents the buffering system during the reading-speaking process, could serve as a dynamic cognitive effort indicator in reading aloud and sight translation. By measuring the dynamic EVS between human translators' reading input and speaking output and examining their eye movements, we will test the following hypotheses: (1) there will be less cognitive effort (i.e., total fixation duration and fixation count) invested in the reading aloud task than in the sight translation task; (2) EVS in the reading aloud task will be shorter than that in the sight translation task; (3) EVS at different positions of a sentence could vary in a consistent manner due to the changing processing difficulty; (4) EVS is positively correlated with the representative fixation indexes of cognitive effort.

2. Methodology

2.1. *Tools*

A Tobii TX300 eye-tracker integrated with a 23" LCD monitor at 1280×1024 pixels was used in this research. Participants were asked to sit in a chair at around 60-65 cm from the PC screen. The eye-tracker was calibrated with regular 5-point calibration at the beginning of all the warm-up exercises and formal tasks. To keep the lighting consistent, the window-less room was equipped with an overhead fluorescent light. Due to the need for "standardization of the different filter settings" in experiments (Alves et al. 2009, p. 274), all experiments in this study used the same Velocity-Threshold Identification (I-VT) Filter, which has its velocity threshold set at 30 degrees/second. The maximum time between fixations was set at 75 ms; the maximum angle between fixations was set at 0.5 degrees; and the fixation filter was set to include fixation samples that fell within a time window of at least 60 ms.

2.2. *Participants*

After a process of data screening (see criteria in Section 3.1), data from thirty (out of thirty-one) participants were included in the study. The average age of the participants is 24.29 years (*SD*=1.87 years, range=22-27). They were master's students majoring in Translation Studies at Durham University (all with 40 teaching hours' exposure to the interpreting classes throughout the academic year), and they were recruited on a voluntary basis. They were all native Chinese speakers with English as their second language, and were considered representative of advanced learners of English-Chinese translation. These late bilinguals were ranked as highly proficient in English, with a mean IELTS score of 7.31 (*SD* = 0.33, range=7-8). All participants had normal or corrected-to-normal vision, did not wear bifocal lenses, mascara or other eye-makeup, and had had no alcohol or caffeinated drink on the day. Participants were told that anonymity and confidentiality would be ensured, and they all signed a consent form before each experiment. Each participant received a £10 book voucher as a reward for their work. Experiments comply with the June 1964 Declaration of Helsinki, and have been approved by the research ethics committee of Durham University.

2.3. *Materials*

There are two materials for this research (see Appendix). Material A (200 English words, 10 sentences) was used as the reading material for the reading aloud task in English (RA-E), as well as the ST for the English-Chinese sight translation task (STR E→C). This excerpt from the 42nd US President Bill Clinton's farewell address was employed as the ST in Zheng and Xiang (2013; 2014). Material B (426 Chinese characters, 10 sentences) was used as the reading material for the reading aloud task in Chinese (RA-C) and the ST for the Chinese-English sight translation task (STR C→E). It is the translation of another excerpt from the same farewell address, published on Soho Chinese on-line news, one of the largest Chinese online news websites.[2] The original English text consists of 9 sentences and 193 words, which is similar to Material A in terms of linguistic features.

2.4. *Procedure*

All trials were carried out at the university's eye-tracking laboratory, with each participant being scheduled for an individual session to perform both reading aloud and sight translation tasks. The participants were first briefed on the procedure and the background of the ST. Half of the participants were assigned a RA-E warm-up exercise followed by the formal RA-E task; for the other half, a RA-C warm-up exercise was followed by the formal RA-C task. Participants were then given three minutes to prepare for either STR E→C or C→E with Material A or B (see Figure 1 for the procedure). To avoid excessive or noisy latency caused by unknown words in their L2 (English) during sight translation, they were also given a glossary in preparation for the STR E→C task. Employing the same text in reading aloud and the following sight translation tasks was to ensure the comparability of the source materials and eliminate the difference in effort invested into the two tasks induced by a change of text. Furthermore, the respective reading aloud processes—always ahead of the sight translation tasks—constituted part of the preparation for sight translation, an apparently more difficult task than reading aloud. This was to reduce the enormous trade-off effect in effort expended in the reading aloud task if both tasks had been performed in a reverse order. The source material was presented in multiple slides and the slide change was automatic. If the participant finished reading aloud or translating the content on one slide before the predefined timeframe, their eye data after task completion was discarded, as it was deemed irrelevant to the task. After the sight translation task, each participant was asked a couple of questions in retrospection while watching the replayed recordings of their performance on the two tasks. The total session for each participant lasted around one hour.

[2] http://news.sohu.com

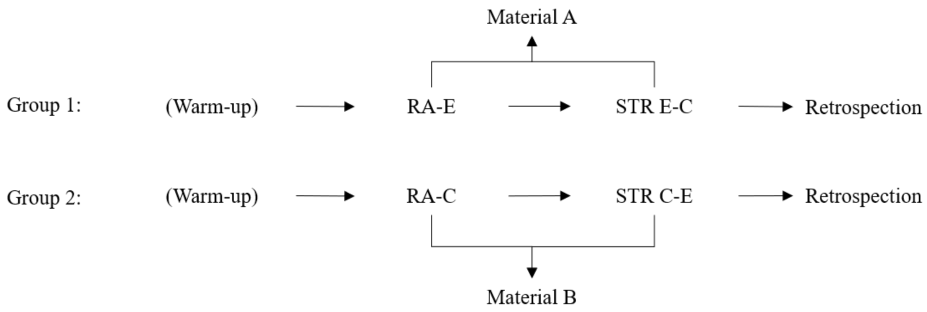

Figure 1. Experimental procedure

2.5. Data acquisition

Rayner & Pollatsek (1989) suspected that “the only really satisfactory way to measure the eye-voice span is by actually making a record of the eye movements and relating the eye-movement record to a record of the vocal output” (p. 181). To this end, EVS is often calculated by onset-to-onset or offset-to-offset algorithms in reading aloud (Laubrock & Kliegl 2015). In brief, onset-to-onset EVS is measured by the time of the articulation onset of the syllable minus the time of the fixation onset. Offset-to-offset EVS is measured by the time of the articulation offset of the syllable minus the time of the fixation offset. These two approaches for EVS calculation might yield different results, and should not be mix-matched. Figure 2 is an example of onset-to-onset calculation approach that measures an EVS from a fixation onset to an articulation onset of a syllable/word. EVS in the present study is calculated by this approach only.

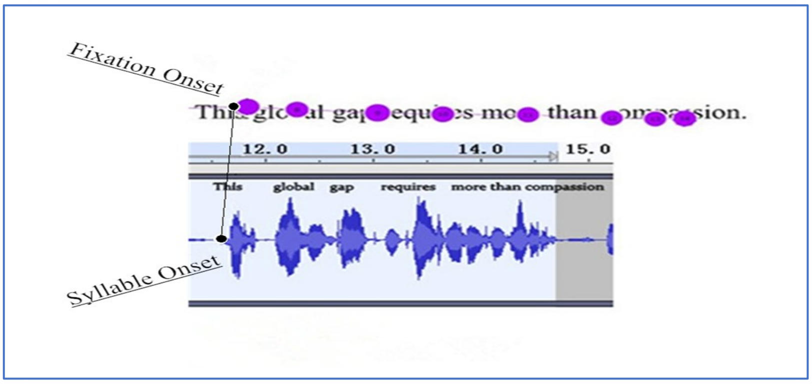

Figure 2. Onset-to-onset approach for EVS calculation

3. Data Processing

3.1. *Quality assessment of eye-tracking data*

Hvelplund (2011; 2014), Sjørup (2013) and Schmaltz et al. (2016) have applied some quality assessment criteria on eye-tracking data in their written translation studies. In contrast, stimuli for reading aloud and sight translation situated in the centre of the screen do not lose any purposeful gaze data to the keyboard. Hence, occasional fixations away from the valid area were considered noise in our experiments, with the quality assessment criteria on eye-tracking data being adjusted as follows.

Four criteria were applied to assess the quality of the eye-tracking data obtained in reading aloud and sight translation tasks: Fixation Time on Source text (FTS) as a percentage of total task time [(total fixation duration ÷ total task time)×100]; Time of Unclassified Sample (USP) as a Percentage of total task time [(unclassified data sample duration ÷ total task time) ×100%]; Mean Fixation Duration (MFD), which is total fixation duration ÷ fixation count; and the standard range of the Saccade duration as a Percentage of pure Gaze activities (SPG) which is the sum of fixation and saccadic durations [(total saccade duration ÷ the sum of fixation and saccadic durations) ×100].

Following recommendations from researchers in statistical computing (e.g., Crawley 2002) and researchers in translation studies (e.g., Macizo & Bajo 2004; Jensen et al. 2009), the data quality thresholds for FTS and USP are set at mean ± 2.5 standard deviations in the present study. MFD threshold is set at 200 ms as suggested by Pavlović & Jensen (2009) and Hvelplund (2014). The SPG threshold is set to be between 5% and 15%, in accordance with Zheng & Zhou (2018), in which the range of SPG mean ± 2.5 standard deviations was between 4.76% and 15.16%. The data that met the requirements of at least three out of the above four criteria was included for further analysis. As a result, the data from one participant was deemed invalid, with the data discard rate being 3.23%.

3.2. *EVS calculation at sentence level*

Following Laubrock & Kliegl (2015), the present research calculates the EVS at sentence boundaries using the onset-to-onset algorithm. This calculation method has also been adopted in the majority of eye-tracking studies, as suggested by Holmqvist et al. (2011, p. 443). The first fixation (F_1) on the first reading/translation unit and the first fixation (F_x) on the last reading/translation unit are defined as two critical fixations for a sentence under investigation. These two positions are named *sentence initial* and *sentence terminal*. The EVS measured at these two positions based on F_1 and F_x, therefore, are called *initial* EVS and *terminal* EVS. Initial and terminal EVS are calculated by subtracting the fixation time on the onset of F_1/F_x from the articulation of the onset, which is marked using the software Audacity 2.2.2. Figure 3 shows that the initial EVS=39.968-38.921=1.047 (s), and the terminal EVS=45.338-44.677=0.661 (s).

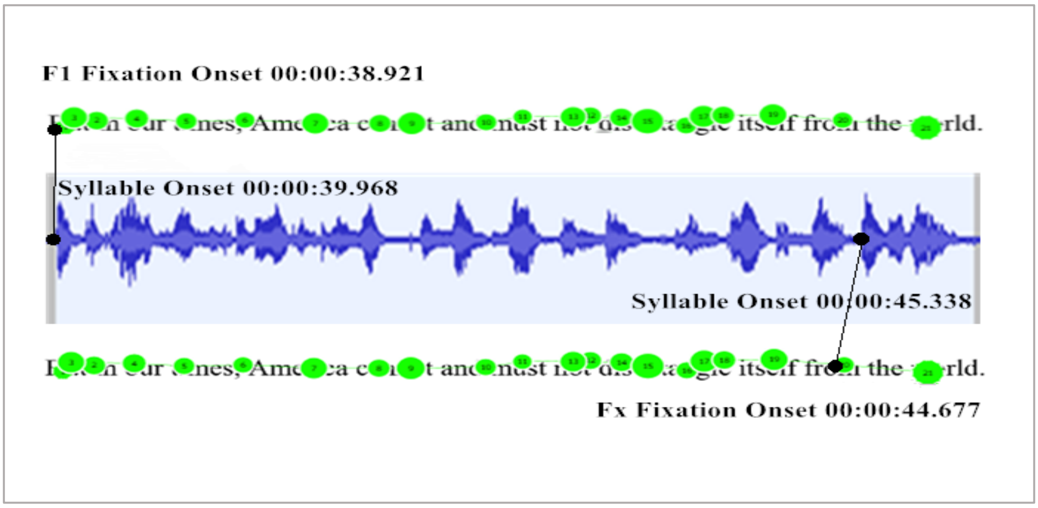

Figure 3. EVS calculated at sentence initial and terminal

3.3. *Fixation indexes of cognitive processing*

Total fixation duration (TFD) and fixation count (FC), as the two most frequently used fixation indexes related with cognitive processing, have been adopted in this research to indicate cognitive effort allocated to reading aloud and sight translation. Sentences as the natural linguistic unit during reading aloud and sight translation were drawn as the Area of Interest to elicit the two fixation indexes, i.e., TFD and FC.

4. Results

4.1. *Cognitive effort allocated on reading aloud and sight translation*

The t-tests were performed to compare the means of the two selected fixation indexes (TFD and FC) and the two EVS indexes (initial and terminal) in the two types of tasks (reading aloud and sight translation), respectively (see Table 1).

Table 1. T-tests of TFD, FC, I-EVS and T-EVS between reading aloud and sight translation tasks

Indexes	RA-E		STR E→C		t-test	RA-C		STR C→E		t-test
	Mean	*SD*	*Mean*	*SD*	*p*	*Mean*	*SD*	*Mean*	*SD*	*p*
TFD (s)	76.42	9.82	118.34	17.97	<.01	74.27	8.83	121.78	15.26	<.01
FC (n)	304.23	32.31	491.87	86.84	<.01	283.83	27.25	470.37	70.41	<.01
I-EVS (ms)	1181.64	119.35	2166.50	623.72	<.01	1187.14	142.11	1909.12	521.52	<.01
T-EVS (ms)	861.62	78.19	1559.32	320.64	<.01	724.70	103.28	1213.69	273.67	<.01

Note: The assumptions of a t-test have been met.

Table 1 shows that the mean values of TFD and FC in sight translation tasks are significantly higher than those in reading aloud tasks ($p<.01$, $d>0.8$), meaning that for the same text (stimulus) being presented in English or Chinese, the amount of cognitive effort invested on sight translation is significantly higher than that on reading aloud.

Following the examination of the fixation indexes, the results of both initial and terminal EVS on the reading aloud and sight translation tasks are in line with what we have found with TFD and FC: the mean values of EVS in sight translation tasks are significantly longer ($p<.01$, $d>0.8$) than those in reading aloud tasks, both at initial and terminal positions. Table 1 also reveals that, whether in reading aloud (E and C) or sight translation (E→C and C→E) tasks, the mean values of initial EVS are consistently much higher than those of terminal EVS, and with statistical significance ($p<.01$, $d>0.8$). This indicates that the EVS presents a certain consistency at sentence-initial and sentence-terminal positions in both tasks.

4.2. *Correlation between EVS and fixation indexes*

In order to further ascertain whether EVS can serve as a dynamic indicator of cognitive effort, we further conducted correlation tests between EVS and the two fixation indexes (i.e., TFD and FC) in reading aloud (E and C) and sight translation (E→C and C→E) tasks. Since normality was violated, we chose the Spearman correlation coefficient as the calculation method. The results are presented in Table 2.

Table 2. Spearman correlation coefficient (*r*) between EVS and fixation indexes in reading aloud and sight translation tasks.

	RA-E						RA-C					
	TFD			FC			TFD			FC		
	N	*r*	Sig.	N	*r*	Sig.	N	*r*	Sig.	N	*r*	Sig.
I-EVS	300	0.083	.150	300	0.068	.243	300	0.007	.907	300	0.024	.678
T-EVS	299	-0.164	**.004**	299	-0.175	**.002**	300	0.211	**<.001**	300	0.225	**<.001**
	STR E→C						STR C→E					
	TFD			FC			TFD			FC		

	N	r	Sig.	N	r	Sig.	N	r	Sig.	N	r	Sig.
I-EVS	297	0.210	**<.001**	297	0.236	**<.001**	298	0.218	**<.001**	298	0.199	**.001**
T-EVS	279	0.264	**<.001**	279	0.278	**<.001**	281	0.133	**.026**	281	0.146	**.015**

Note: Although the data recorded from all the 300 sentences were analysed, not all sentences were successfully read out or translated, i.e., without an initial EVS or terminal EVS. So, in some cases the number of sentences is below 300.

Table 2 shows that all the correlations between initial or terminal EVS and the two fixation indexes (TFD and FC) in all the reading aloud and sight translation tasks are weak ($|r| < 0.3$). Initial EVS exhibits positive correlations with the two fixation indexes in all the tasks. Terminal EVS shows negative correlations with TFD and FC in the RA-E task, and positive correlations with the two fixation indexes in all the other tasks. Although the correlation coefficients are not strong enough, the results generally indicate that, for the task of reading aloud in English, longer total fixation duration and more fixation counts in a sentence tend to signal slightly shorter EVS at the sentence terminals; however, when the participants were reading aloud in Chinese and sight translating both the English and Chinese materials, more intense fixation in a sentence signals slightly longer EVS at both the sentence initials and sentence terminals.

5. Discussion

This chapter aims to determine whether EVS at sentence level could serve as a dynamic cognitive effort indicator in reading-speaking tasks, by examining how it is related to the amount of cognitive effort invested to reading aloud and sight translation tasks. The results show that EVS at both sentence-initial and sentence-terminal positions are longer in sight translation than in reading aloud tasks, which is in line with the results of the representative fixation indexes of cognitive effort. Thus, both initial EVS and terminal EVS have successfully detected the differences in effort invested into the two tasks, suggesting that reading aloud is a relatively less cognitively demanding activity than sight translation. Based on these results, the first two hypotheses are corroborated. Initial EVS was also consistently longer than terminal EVS in the two types of reading + speaking tasks, which could indicate potentially different attributes of these two EVS indexes. Therefore, the third hypothesis has also been corroborated. Furthermore, the correlation tests displayed a significant, but weak positive relationship between both EVS indexes (initial and terminal) and the fixation indexes (TFD and FC) in the sight translation tasks (C→E and E-C). For the two reading aloud tasks, initial EVS yielded a weak positive correlation with the two fixation indexes, whereas terminal EVS showed a weak positive correlation with the fixation indexes only in the RA-C task. Consequently, the last hypothesis can only be partially corroborated. Implications of these findings are discussed below.

As mentioned, in Gile's (2009) Effort Model, sight translation involves extra effort in coordination and short-term memory to manage reading input and oral output. Thus, sight translation is apparently a task demanding more effort than reading aloud, which involves almost no extra effort in short-term memory and coordination. As the major fixation indexes of cognitive processing show, i.e., TFD and FC, the present research has validated this account in that sight translation came with significantly more intensive fixations than reading aloud. In line with this result, the two EVS indexes have also depicted the same picture, with sight translation possessing significantly longer initial and terminal EVS than those of reading aloud. Even considering that there could be a trade-off effect on the sight translation task as to exerting effort due to the prior reading aloud task on the same source material, the difference in effort indicated by the fixation and EVS indexes between the reading aloud and sight translation task is still significant. Thus, the difference in effort induced by the two tasks was enormous and it was successfully detected by both initial and terminal EVS indexes.

A closer inspection of the initial and terminal EVS indexes revealed an evident and consistent difference: sentence initials have significantly longer EVS than sentence terminals, regardless of task type. Typically, the reading process during both reading aloud and sight translation follows a sequential scan to ensure a smooth and uninterrupted oral production. In general, the mind integrates all the information collected from successive fixations and forms a stable and coherent oral representation of the text during both reading aloud and sight translation. O'Brien (1926), Tinker (1965) and Vernon (2014) agree that the size of spatial EVS has a strong correlation with the position of meaning units in a sentence. Quantz (1897) and Buswell (1920) also argued that the spatial EVS is longer at sentence-initial positions than at sentence-terminal positions. Like its spatial counterpart, the results in our research confirm the account that sentence initials tend to yield longer EVS than the sentence terminals. This major difference in the length of EVS at the bipolar positions of a sentence shows a high degree of adaptability and flexibility of EVS. On the basis of this, together with the successful detection of the difference in effort investment between reading aloud and sight translation, the application of EVS as a dynamic indicator of cognitive processing can be proposed.

Regarding the differences of EVS between sentence positions, we claim that, during the reading-speaking process, both the visual and cognitive sources of information might be guiding the eyes (McConkie 1983, p. 75). Information such as the length of the emerging words and punctuation could be obtained within the perceptual span to the right of a fixation (Rayner 1983, p. 97). Hence, when a comma or period becomes visible in one's peripheral vision, the likelihood of it being recognised as the sentence boundary increases. Moreover, one's knowledge about the language and the text could be the second guidance. The reader might guess that the sentence will end soon, based on its content and grammatical structure. In this way, the two types of guidance work together to help coordinate the eyes and voice, although we do not know which one of them dominates in every case. This mechanism constitutes an explicit explanation for why terminal EVS was significantly shorter than initial EVS in the reading aloud and sight translation tasks.

Another finding of the present research is initial and terminal EVS, on the one hand, and the two fixation indexes (TFD and FC), on the other, had weak positive correlation coefficients in the sight translation tasks (both E→C and C→E). This indicates that both EVS indexes represented, though to a limited extent, the cognitive investment during sight translation. For the reading aloud tasks, however, it appears that initial EVS is positively but weakly correlated with the fixation indexes, representing the cognitive effort invested into the reading aloud tasks of both directions; while terminal EVS is only positively (and again, weakly) correlated with the fixation indexes in reading aloud the material in the participants' L1. Although all the correlation coefficients are not strong, this discrepant result may be caused by the different nature of the two types of indexes (i.e., fixation and EVS) functioning on the two types of tasks.

On the one hand, TFD and FC are entirely based on eye-tracking data, while EVS does not depend solely on the eye-movements. Since EVS is not a conventional eye movement-based indicator, we should not only focus on the *E (eye)* of EVS but also look at the role of the *V (voice),* because "articulatory output of a word presumably tells us that it no longer needs to be buffered in working memory" (Laubrock & Kliegl 2015,

p. 2). Thus, it can be considered that the fixation indexes are more macro-level indicators signifying the depth or the intensity of cognitive processing, whereas EVS is a micro-level measurement indicating the eye-voice coordination effort. Therefore, we suggest that EVS and the fixation indexes are representatives of different sub-categories of cognitive effort devoted to completing a reading aloud or sight translation task. To be more specific, EVS might correspond primarily to the participants' ability to coordinate the input processing and the output execution. Thus, it can be inferred that there might be a difference in the relationship between the micro-level eye-voice coordination and the macro-level intensity of processing when participants performed the reading aloud task in English (their L2) and in Chinese (their L1). However, this discrepancy (i.e., the direction effect) is not pronounced in the sight translation tasks. This could be caused by the nature of these sight translation tasks: that sight translation demands substantially more intensive processing as well as enormous coordination effort (as compared to the reading aloud tasks) regardless of translation direction. Thus, the difference in the relationship between micro-level coordination and macro-level intensity of effort in the sight translation tasks (of different translation directions) may not be evident. Nonetheless, the correlations are very weak in the present study, and further systematic investigations in this regard could be conducted to validate the present results.

On the other hand, having an EVS during the reading-speaking process coincides with what Carpenter & Just (1983) called applying "a moving bin strategy" which means "collecting input from several words before processing any one of them" (p. 303). Readers integrate thoughts into the prior context held in the processing buffer as soon as they decode a piece of information. The eyes, then, tend to proceed to the next fixation without waiting for the *bin* to be empty through oral production. In Rayner & Pollatsek's (1989, p. 181) words, "if the eyes moved further ahead, there would be a lot of undigested material" before the reader can produce it orally. When this happens, perhaps the "eyes need to wait for the voice because the size of the working memory buffer is limited" (Laubrock & Kliegl 2015, p. 17). As such, it is natural that participants may pronounce the words at the sentence terminals in their second language quickly with the help of visual and cognitive guidance, but they may also automatically spend more effort to process these sentence terminals even after the reading aloud. In other words, the quicker they read aloud the words, the more effort may be required for them to further process the meaning of the words afterwards. Such an effect could be much smaller in processing during reading aloud in their native language. Thus, this may provide supportive evidence for the slight negative correlation coefficient found between erminal EVS and the fixation indexes in the RA-E task.

However, if the cognitive processing and the voice output can catch up with the eyes, as in the task of reading aloud in the participants' L1 (RA-C), less information would remain undigested and the eyes do not need to slow down. This situation may also apply to the task of sight translation (in both directions), but in an opposite, i.e., slowed-down, manner. Since the nature of sight translating materials from one language to another language would require an extra process of coordination, the voice output (i.e., the translation) should rarely be generated without full digestion of the input.

The present study argues that both the visual and cognitive sources of information act as a basis for eye movement control at sentence initial and terminal positions, and that they become joint forces to facilitate the maintenance of an EVS during reading aloud and sight translation. The general patterns of EVS within and across sentences' boundaries during reading aloud and sight translation show that thought determines the dynamics of EVS, and cognitive processing modulates the duration of EVS. The successful detection of differences between different task types and sentence positions reflects the relative stability of EVS in measuring cognitive effort in general reading+speaking activities. Determined by its nature of involving a complicated interaction mechanism between eye and voice coordination, this more elaborate interpretation could be applied in different cases such as comparing different types of reading-speaking task. Nevertheless, the discrepant results mixing positive and negative correlations between EVS and the fixation indexes in the reading aloud tasks, and the fact that only weak correlations were captured in this study, could also imply the inability of either the fixation or the EVS measures to properly capture the variation of cognitive effort at more fine-grained levels. Thus, further study involving more types of cognitive effort indicators can be conducted to explore such differences systematically.

6. Conclusion

This chapter aimed to investigate whether temporal EVS at sentence level can serve as a valid indicator of cognitive effort in reading-speaking activities. Combining the fixation indexes of cognitive processing, the length of EVS at different sentence positions (initials and terminals) in different reading-speaking tasks (reading aloud and sight translation) was examined. The results show that EVS in sight translation tasks is significantly longer than that in reading aloud tasks, which is in line with the results obtained from the two fixation indexes. Furthermore, it was found that sentence initials possess consistently longer EVS than sentence terminals in the two types of tasks. The detected differences of EVS between reading aloud and sight translation as well as between sentence initials and terminals reveal the relative stability of EVS. The subsequent correlation tests further confirm the validity of both sentence-initial and sentence-terminal EVS as dynamic cognitive effort indicators during the sight translation tasks, because of their significant, although weak, positive correlations with the fixation indexes. For the reading aloud tasks, only initial EVS was found to be in a positive and weak correlation with the fixation indexes; while terminal EVS was found to be positively correlated with the fixation indexes in the RA-C task only. This discrepancy may be caused by the pronounced direction effect, interacting with the complicated nature of EVS involving the coordination between eye and voice. Overall, we propose that EVS can be used to discriminate different reading-speaking tasks in terms of cognitive effort and has the potential to serve as a dynamic indicator of cognitive effort, especially during sight translation.

This study can provide some reference for future research regarding the application of EVS in different reading-speaking activities. However, we are aware of two limitations in the study. Firstly, the limited sample size and absence of a professional interpreter group may weaken the statistical power and make the findings not representative enough for the population. The fact that no strong correlations between fixation and EVS indexes were captured in this study may also be caused by the limited sample and data collected. Secondly, the fixed task order, i.e., reading aloud followed by sight translation, for all the participants could lead to a fatigue effect which may influence the results as well. Therefore, it would be beneficial for future research to enlarge the sample size, diversify the participants, and counterbalance the tasks, in order to further validate these findings.

Acknowledgement

The authors would like to thank the two anonymous reviewers for providing their valuable comments and constructive suggestions. The work was supported by Project of Technology Consulting Expert, Shanxi Province.

Reference

Agrifoglio, M. 2004. Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting* 6(1): 43-67.

Alves, F., A. Pagano and I. Silva. 2009. A new window on translators' cognitive activity: Methodological issues in the combined use of eye tracking, key logging and retrospective protocols. In *Methodology technology and innovation in translation process research,* eds. I. M. Mees, F. Alves, and S. Göpferich, 268-269. Copenhagen: Samfundslitteratur.

Buswell, G. T. 1920. *An experimental study of the eye-voice span in reading*. Chicago: University of Chicago Press.

Buswell, G. T. 1921. The relationship between eye-perception and voice-response in reading. *Journal of Educational Psychology* 12(4): 217-227.

Carpenter, P. A., and M. A. Just. 1983. What your eyes do while your mind is reading. In *Eye movements in reading,* ed. K. Rayner, 275-307. New York and London: Academic Press.

Crawley, M. J. 2002. *Statistical computing: An introduction to data analysis using S-Plus*. Chichester: John Wiley and Sons.

De Luca, M., M. Pontillo, S. Primativo, D. Spinelli, and P. Zoccolotti. 2013. The eye-voice lead during oral reading in developmental dyslexia. *Frontiers in Human Neuroscience,* 7: 696.

Dragsted, B., and I. G. Hansen. 2009. Exploring translation and interpreting hybrids. The case of sight translation. *Meta,* 54(3): 588-604.

Dragsted, B., I. G. Hansen, and H. S. Sørensen. 2009. Experts exposed. In *Methodology technology and innovation in translation process research,* eds. I. M. Mees, F. Alves, and S. Göpferich, 294-317. Copenhagen: Samfundslitteratur.

Geyer, J. J.1966. *Perceptual systems in reading: A temporal eye-voice span constant*. PhD Thesis. Berkeley: University of California.

Gile, D. 2009. *Basic concepts and models for interpreter and translator training*. Amsterdam and Philadelphia: John Benjamins.

Holmqvist, K., M. Nystrom, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Ho, C., T. Chen, and J. Tsai. 2020. How does training shape English-Chinese sight translation behaviour? An eyetracking study. *Translation, Cognition and Behavior* 3(1): 1-24.

Huang, C. 2011. *Tracking eye movements in sight translation: The comprehending process in interpreting*. Master Thesis. Taibei: National Taiwan Normal University.

Hvelplund, K. T. 2011. *Allocation of cognitive resources in translation: An eye-tracking and key-logging study.* PhD thesis. Copenhagen: Copenhagen Business School.

Hvelplund, K. T. 2014. Eye tracking and the translation process: Reflections on the analysis and interpretation of eye tracking data. *MonTI Special Issue—Minding Translation* 1: 201-223.

Inhoff, A. W., M. Solomon, R. Radach, and B. A. Seymour. 2011. Temporal dynamics of the eye voice span and eye movement control during oral reading. *Journal of Cognitive Psychology* 23(5): 543-558.

Irwin, D. E. 2004. Fixation location and fixation duration as indices of cognitive processing. In *The interface of language, vision, and action: Eye movements and visual world,* eds. J. Henderson and F. Ferreira,105-134. New York: Psychology Press.

Jensen, K. T. H., A. C. Sjørup, and L. W. Balling. 2009. Effects of L1 syntax on L2 translation. In *Methodology technology and innovation in translation process research,* eds. I. M. Mees, F. Alves, and S. Göpferich, 319-336. Copenhagen: Samfundslitteratur.

Lambert, S. 2004. Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta,* 49(2): 249-306.

Laubrock, J., and R. Kliegl. 2015. The eye-voice span during reading aloud. *Frontiers in Psychology,* 6, 1432.

Liu, Y., Zheng, B., and Zhou, H. 2019. Measuring the difficulty of text translation. *Target* 31(1): 125-149.

Macizo, P., and M. T. Bajo. 2004. When translation makes the difference: Sentence processing in reading and translation. *Psicológica* 2 (1): 181-205.
McConkie, G. W. 1983. Eye movements and perception during reading. In *Eye movements in reading,* ed. K. Rayner, 65-95. New York and London: Academic Press.
Mishra, A., P. Bhattacharyya, and M. Carl.2013. Automatically predicting sentence translation difficulty. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (Volume 2: Short Papers), 346-351. Sofia.
Morton, J. 1964. The effects of context upon speed of reading, eye movements and eye-voice span. *Quarterly Journal of Experimental Psychology* 16(4): 340-354.
O'Brien, J. A. 1926. *Reading, its psychology and pedagogy*. New York: The Century Company.
Pavlović, N., and K. T. H. Jensen. 2009. Eye tracking translation directionality. *Translation research projects 2*, ed. A. Pym and A. Perekrestenko, 93-109. Tarragona: Intercultural Studies Group.
Quantz, J. O. 1897. *Problems in the psychology of reading.* New York: Macmillan.
Rayner, K. 1983. *Eye movements in reading*. New York and London: Academic Press.
Rayner, K., and A. Pollatsek. 1989. *The psychology of reading*. Englewood Cliffs. N. J.: Prentice Hall.
Sandrelli, A. 2003. New technologies in interpreter training: CAIT. In *Textologie und Translation*, ed. H. Gerzymisch-Arbogast, 261-294. Tübingen: Gunter Narr Verlag.
Schmaltz, M., I. A. L. da Silva, A. Pagano, F. Alves, A. L. V. Leal, D. F. Wong, L. S. Chao, and P. Quaresma. 2016. Cohesive relations in text comprehension and production: An exploratory study comparing translation and post-editing. In *New directions in empirical translation process research,* eds. M. Carl, S. Bangalore and M. Schaeffer, 239-263. New York: Springer.
Sjørup, A. C. 2013. *Cognitive effort in metaphor translation: An eye-tracking and key-logging study*. PhD thesis. Copenhagen: Copenhagen Business School.
Tinker, M. A. 1965. *Bases for effective reading*. Minneapolis: University of Minnesota Press.
Vernon, M. D. 2014. *The experimental study of reading*. Cambridge: Cambridge University Press.
Zheng, B., and H. Zhou. 2018. Revisiting processing time for metaphorical expressions: An eye-tracking study on eye-voice span during sight translation. *Foreign Language Teaching and Research*, 50(5): 744-759.
Zheng, B., and X. Xiang. 2013. Processing metaphorical expressions in sight translation: An empirical-experimental research. *Babel*, 59(2): 160-183.
Zheng, B., and X. Xiang. 2014. The impact of cultural background knowledge in the processing of metaphorical expressions: An empirical study of English-Chinese sight translation. *Translation and Interpreting Studies*, 9(1): 5-24.

Appendices

Material A: excerpt from Clinton's farewell speech (2001)

Slide page 1.
The expansion of trade hasn't fully closed the gap between those of us who live on the cutting edge of the global economy and the billions around the world who live on the knife's edge of survival. This global gap requires more than compassion. It requires action. Global poverty is a powder keg that could be ignited by our indifference.

Slide page 2.
In his first inaugural address, Thomas Jefferson warned of entangling alliances. But in our times, America cannot and must not disentangle itself from the world. If we want the world to embody our shared values, then we must assume a shared responsibility.

Slide page 3.
We must embrace boldly and resolutely that duty to lead, to stand with our allies in word and deed, and to put a human face on the global economy so that expanded trade benefits all people in all nations, lifting lives and hopes all across the world.

Slide page 4.
Third, we must remember that America cannot lead in the world unless here at home we weave the threads of our coat of many colors into the fabric of one America. As we become ever more diverse, we must work harder to unite around our common values and our common humanity.

Material B: excerpt from the Chinese version of Clinton's farewell speech (2001)

(Note: only Chinese texts were presented on screen for reading aloud and sight translation tasks)

Slide page 1.
机会属于所有的美国公民；责任源自全体美国人民；所有美国人民组成了一个大家庭。我一直在为寻求一个更小、更现代化、更有效率、面对新时代的挑战充满创意和思想、永远把人民的利益放在第一位、永远面向未来的新型的美国政府而努力。
(Opportunity for all, responsibility from all, a community of all Americans. I have sought to give America a new kind of government, smaller, more modern, more effective, full of ideas and policies appropriate to this new time, always putting people first, always focusing on the future.)

Slide page 2.
我们要加倍努力地工作，克服生活中存在的种种分歧。于情于法，我们都要让我们的人民受到公正的待遇，不论他是哪一个民族、信仰何种宗教、什么性别或性倾向，或者何时来到这个国家。我们时时刻刻都要为了实现先辈们建立高度团结的美利坚合众国的梦想而奋斗。
(We must work harder to overcome our differences, in our hearts and in our laws. We must treat all our people with fairness and dignity, regardless of their race, religion, gender, or sexual orientation, and regardless of when they arrived in our country—always moving toward the more perfect Union of our Founders' dreams.)

Slide page 3.
对我来说，当我离开总统宝座时，我充满更多的理想，比初进白宫时更加充满希望，并且坚信美国的好日子还在后面。我的总统任期就要结束了，但是我希望我为美国人民服务的日子永远不会结束。
(As for me, I'll leave the presidency more idealistic, more full of hope than the day I arrived and more confident than ever that America's best days lie ahead. My days in this office are nearly through, but my days of service, I hope, are not.)

Slide page 4.

在我未来的岁月里，我再也不会担任一个能比美利坚合众国总统更高的职位、签订一个比美利坚合众国总统所能签署的更为神圣的契约了。当然，没有任何一个头衔能让我比作为一个美国公民更为自豪的了。谢谢你们，愿上帝保佑你们，愿上帝保佑美国！
(In the years ahead, I will never hold a position higher or a covenant more sacred than that of president of the United States. But there is no title I will wear more proudly than that of citizen. Thank you, God bless you, and God bless America!)

Authors' Brief Bios

Dr Hao Zhou obtained her BA in English Studies from Southwest Jiaotong University, two MAs (TESOL and Translation Studies) and a PhD (Translation Studies) from Durham University. Her research focuses on Cognitive Translation and Interpreting Studies, and had published journal papers in *Target*, *Foreign Language Teaching & Research*, *Asia-pacific Translation and Interdisciplinary Studies*.

Dr Binghan Zheng (corresponding author) is Distinguished Professor at Shanxi University and Associate Professor of Translation Studies at Durham University. His research interests include cognitive translation and interpreting studies, neuroscience of translation, and comparative translation and interpreting studies. His recent publications appeared in journals such as *Target*, *the Interpreter and Translator Trainer*, *Across Languages & Cultures*, *Journal of Pragmatics, Brain & Cognition*, *Perspectives*, *LANS-TTS, Babel*, *Translation & Interpreting Studies*, *Foreign Language Teaching & Research*, and *Journal of Foreign Languages*.

Miss Yu Weng is a PhD candidate in Translation Studies at Durham University. She obtained her MSc. in Scientific, Technical and Medical Translation with Translation Technology from UCL in 2014, a BA in Linguistics and Sociology (First class) from the University of Manchester and a BA in English (Translation) from Sun Yat-Sen University in 2013. Her research interest lies in the psychology of translation, cognitive translation studies and linguistics, and had published journal articles in *Target* and *LANS-TTS*.

Address for correspondence
Binghan Zheng (ORCID: 0000-0001-5302-4709)
School of Foreign Languages, Shanxi University, 030006, P. R. China
School of Modern Languages and Cultures, Durham University, DH1 3JT, UK
Email: binghan.zheng@durham.ac.uk