CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Stages of sight translation: Evidence from eye movements

Agnieszka Lijewska[1],* , Agnieszka Chmiel[2] and Albrecht W. Inhoff[3]

[1]Department of Psycholinguistic Studies, Adam Mickiewicz University, Poznań, Poland, [2]Department of Translation Studies, Adam Mickiewicz University, Poznań, Poland and [3]Department of Psychology, Binghamton University, Binghamton, NY, USA
*Corresponding author. Email: alijewska@amu.edu.pl

Abstract

The aim of the study was to investigate the coordination of source text comprehension and translation in a sight translation task. The study also sought to determine whether translation strategies influence sight translation performance. Two groups of conference interpreters—professionals and trainees—sight translated English sentences into Polish while their eye movements and performance were monitored. Translation demands were manipulated by the use of either high- or low-frequency critical words in the sentences. Translation experience had no effect on first-pass viewing durations, but experts used shorter re-view durations than trainees (especially in the low-frequency condition). Professionals translated more accurately and with less pausing than trainees. Translation in the high-frequency condition was more accurate and had shorter pauses than in the low-frequency condition. Critical word translation accuracy increased with the translation onset latency (TOL) for individual sentences, and pause durations were relatively short when TOLs were either relatively short or long. Together, these findings indicate that, in sight translation, the initial phase of normal reading for comprehension is followed by phases in which reading and translation co-occur, and that translation strategy and translation performance are linked.

Keywords: reading; speech production; bilingualism; lexical processing; sight translation; eye movements

Conference interpreters typically work in a simultaneous interpreting booth, where they listen to a speech in a source language and articulate a sequence of words with corresponding content in a target language. Written translation typically involves a text-to-text translation. Sight translation is a hybrid form which involves the reading of a text and the concomitant oral production of equivalent content in a target language. It is typically used when a document has to be accessed in an interpreter-mediated setting, for instance, minutes of a meeting, a letter from an absent person at a function, a medical report in a hospital, or a witness statement in a court

CrossMark

(Čenková, 2015). It progresses from the visual perception of text to the production of speech, and with the advent of eye-tracking and speech-analysis technologies, it has become possible to observe—in real time and under relatively natural task conditions—the temporal dynamics of this task. The study aimed at investigating how sight translators coordinate the viewing of source-language text with the ensuing translation leading to the production of the target language text. Furthermore, we wanted to investigate whether interpreters' decisions when to start producing the target language text (relative to the onset of source text reading) influence task performance. In other words, our goal was to test whether interpreters' individual translation onset strategies influence sight translation.

To date, a relatively small number of studies have recorded eye movements during sight translation (Chmiel & Mazur, 2013; Dragsted & Hansen, 2009; Huang, 2011; Jakobsen & Jensen, 2008; Korpal, 2012; McDonald & Carpenter, 1981; Shreve et al., 2010). Some of them focused on structural reformulation (Chmiel & Lijewska, 2019; Ma et al., 2021) and syntactic cues (Ruiz & Macizo, 2019), while other focussed on processing patterns but without using both early and late eye-tracking measures (Su & Li, 2020). Only one of them, McDonald and Carpenter's (1981) pioneering study, used a comprehensive set of measures to examine directly how the viewing of the source text is linked to the articulation of the corresponding content in a target language.

In McDonald and Carpenter's study, four German–English bilinguals sight translated English passages into German. Each passage contained an ambiguous phrase (e.g., "kick the bucket") that was preceded by context which primed either its literal or idiomatic meaning (i.e., "kicking an object" or "dying," respectively). Moreover, the phrase was followed by a disambiguating sentence that indicated whether the literal or idiomatic meaning was to be assigned. An initially selected meaning for the ambiguous was thus confirmed, when priming and disambiguating contexts matched, or disconfirmed, when they mismatched. The viewing of an ambiguous phrase and its speech production were analyzed, and its viewing pattern was related to its corresponding speech production.

The analyses revealed three distinct viewing phases: initially, translators moved the eyes along a sequence of visible words, three to five words on average. The first phase was followed by a second one during which previously read text was reviewed and during which the articulation of a corresponding translation occurred. When meaning selection was disconfirmed during the subsequent reading of the disambiguating sentence, translators often initiated a third phase. This phase consisted of an additional re-viewing of prior text and the articulation of a corrected translation.

Huang's (2011) study appears to favor McDonald and Carpenter's account according to which first-pass viewing during sight translation corresponds to first-pass viewing in reading for comprehension. In the study, eye movements of interpreting trainees were monitored in three tasks: the silent and oral reading of Chinese text and during Chinese-to-English sight translation. Huang's analysis of word viewing durations showed that initial (first-pass) viewing durations were considerably shorter during silent than during oral reading (see also Inhoff & Radach, (2014). With oral reading, the articulation of visible words thus slowed primarily the first-pass progression through the text. In striking contrast to this, first-pass

viewing during sight translation was similar to first-pass viewing during silent reading, but the translation task involved substantially more second-pass (and higher order) re-reading. Overall, Huang's findings suggest that first-pass word viewing by interpreting trainees was equivalent during sight translation and silent reading for comprehension, and that translation of words and the corresponding articulation occurred during second (and later) pass viewing.

Although Huang's task comparisons are consistent with McDonald and Carpenter's processing account, the evidence must be considered tentative, as the evidence for a key assumption, that processing during the first-pass viewing is similar during silent reading and sight translation, entails an acceptance of the null hypothesis. Moreover, other influential studies suggest that translators extract some target language information early in the processing of to-be-translated visible words (de Groot, 1992; Ruiz et al., 2008).

In Ruiz et al.'s (2008) study, professional translators were asked to sight translate Spanish to English sentences. These were displayed word-by-word, so that the presentation duration of each word was controlled by the translator through button pressing. Source-language sentences contained a member of a critical word pair that was matched on word frequency. It could occupy either the beginning or ending location in the sentence, and its corresponding English equivalent had either a high or a low frequency of occurrence. Two tasks were used: in one, the viewing of the final word of the Spanish sentence was followed by the articulation of an English translation; in the other, the viewing of the final word of the Spanish sentence was followed by a verbatim repetition of the sentence. With this approach, the word frequency of the to-be-articulated English (target language) word influenced the manually controlled presentation duration of the Spanish source word in the translation task when the target was at the end of the sentence, with a longer viewing time when the English translation of the source word had a low frequency of occurrence. The word frequency of the English translation had no effect when the critical source-language word was in the middle of the sentence and in the sentence repetition task. According to Ruiz et al. (2008), these findings imply that the translation of the critical Spanish source word occurred while it was presented; otherwise, the frequency of the English translation could not have influenced the viewing of the Spanish source word.

In de Groot's (1992) single word processing study, Dutch–English bilinguals viewed individually presented high- or low-frequency Dutch words, and their task was either to articulate a word in English with equivalent meaning or to recognize its English translation. Both tasks showed that word frequency in the source and the target language, Dutch and English, respectively, influenced responding, which was faster and more accurate for high-frequency words in each language.

While these findings (de Groot 1992, Ruiz et al., 2008) indicate that the recognition of individual source words and the accessing of equivalent lexical forms into the target language occur while they are viewed, they do not provide a fine-grained account of the processing involved in sight translation. The study by Ruiz et al. (2008) employed a self-paced reading paradigm, which made it impossible to distinguish between early and late stages of source-language processing. In the study by de Groot (1992), the translation of visible words had to occur "on the spot," that is, before the next word was shown, that is, task demands required that recognition and

translation occurred while a word was viewed. The study measured response times only and provided no chance to see the distinction between the processes occurring during first-pass viewing and the subsequent re-viewing of a word. Unlike the studies by Ruiz et al. and de Groot, the current study offers a fine-grained analysis of reading in the sight translation task, thanks to the application of eye-tracking.

In turn, the motivation for the examination of the role of participants' individual translation strategies in the current study came from Christoffels and de Groot (2005). They argued that the duration of the ear-voice span during simultaneous interpreting (i.e., the delay between hearing a source-language word and producing its translation equivalent) is determined by two competing strategies. One is to hear as much and to "wait as long as possible" (long wait) prior to translation, so that the intended meaning of the orally perceived language can be determined. The other is to "keep the lag as short as possible" (short wait) so that working memory will not be overloaded in this difficul task, and the risk of losing the thread of perceived speech is minimized. During sight translation, a long wait strategy could involve the reading of a relatively large number of words prior to translation, to minimize translation errors, and a short wait strategy could involve the reading of relatively few words prior to its translation, to simplify the translation task and to reduce working memory load.

The current study aimed at investigating the coordination of source text viewing and the articulation of translation equivalents. More specifically, sight translation performance was analyzed taking into consideration McDonald and Carpenter's account of processing stages in sight translation. We also aimed to examine whether and how interpreters' translation strategies (i.e., decisions when to start articulation of translation equivalents relative to the onset of text viewing) influenced the fluency and accuracy of their sight translation performance.

Our approach was similar to McDonald and Carpenter's (1981) study in that we recorded the viewing of to-be-translated text and the production of its spoken translation concurrently. The size of the tested population in the current study was considerably larger, however, and instead of using passages with ambiguous phrases, we used declarative sentences that contained unambiguous high- and low-frequency nouns in order to manipulate the ease of translation. Unlike in McDonald and Carpenter's (1981) study, we did not attempt to mislead translators about the meaning of to-be-translated phrases, and this was assumed to lead to the adoption of a more natural sight translation strategy. Prior work suggested that the translation of high-frequency words would be more accurate than the translation of low-frequency words (de Groot, 1992) and that professionals would translate target words and sentences more accurately than trainees (García et al., 2014).

The frequency effect constitutes one of the most robust effects in psycholinguistics (Cop et al., 2015) and is considered to arise during initial stages of reading, that is, during lexical access. Eye movement research has shown that during intensive reading of sentences (or texts) frequency effects are especially pronounced in first-pass reading measures leading to longer first fixation durations, single fixation durations, gaze durations, or even go-past times on low-frequency words than on high-frequency ones (Inhoff & Rayner, 1986; Kliegl et al., 2004; Rayner & Duffy, 1986; Slattery et al., 2007; White, 2008; but see Inhoff, 1991). The typical finding is that a large frequency effect is found in tasks requiring careful reading, for

example, during reading for comprehension (Radach et al., 2008) whereas tasks with less emphasis on careful reading (e.g., text scanning) yield reduced or no frequency effect at all (Schad et al., 2012; White et al., 2015). Consequently, the frequency effect has been used as an index of the depth of lexical processing during various reading tasks (Schotter et al., 2014; White et al., 2015). In order to gain insight into the nature of reading in sight translation, we measured the time spent viewing critical words during their initial encounter (i.e., during first pass, here indexed by first fixation duration, gaze duration, and go-past time) and when they were subsequently re-viewed (i.e., during second pass, here indexed by re-view duration and total viewing duration).

Our understanding, based on McDonald and Carpenter's account, is that sight translation begins with first-pass reading that—on a word level—is similar to regular reading for comprehension activity performed by bilinguals. Words are identified and lexical access occurs. We know from previous research from psycholinguistics and translation studies that this lexical access is language nonselective, which means that same language synonyms and other language translation equivalents are co-activated (e.g., Lauro & Schwartz, 2017; Schaeffer & Carl, 2013; Titone et al., 2011; Van Assche et al., 2020; Whitford et al., 2016). This phase is followed by second and subsequent viewing passes, during which comprehension continues and translation occurs. This is where reading in sight translation differs from reading for comprehension. In the second and subsequent viewing passes, the reader integrates lexical information for a given word with the sentence. We believe that complex translation processes leading to the articulation of translation equivalents take place here (including further activation of appropriate translation equivalents, inhibition of inaccurate previously activated words, etc., i.e., these are processing components in which translation expertise offers an advantage). Thus, the first-pass viewing of words is similar in reading for sight translation and reading for comprehension and this similarity is observable only via eye-tracking and not in self-paced reading or reaction time experiments, wherein early and later stages of word processing cannot be easily distinguished.

If first-pass measures are similar in reading for sight translation and reading for comprehension (Huang, 2011; McDonald & Carpenter, 1981), then first-pass viewing durations (i.e., first fixation duration, gaze duration, and go-past time) should be shorter for high- than for low-frequency targets, as occurred during monolinguals' and bilinguals' silent reading of text (Clifton et al., 2016; Cop et al., 2015; Rayner, 1998, 2009; Whitford & Titone, 2014, 2015). Furthermore, if the initial phase of reading for comprehension and reading for sight translation is indeed similar, then translation expertise should *not* influence first-pass word viewing durations.

Further, the first-pass viewing of words should be followed by re-viewing in line with McDonald and Carpenter's (1981) account and prior findings by Huang (2011). While the effects of word frequency emerge primarily during first-pass viewing during silent reading (Kliegl et al., 2004; Rayner & Duffy, 1986; Slattery et al., 2007; White, 2008), word frequency effects could be larger during the re-viewing (indexed by re-view duration and total viewing duration) than during the first-pass viewing (i.e., first fixation duration, gaze duration, and go-past time) in the sight translation task, as word frequency also influences processing that precedes articulation of translation equivalents. Moreover, translation expertise should influence

re-viewing (i.e., re-view duration and total viewing duration), with shorter processing times for professionals than for trainees.

The concurrent recording of translators' eye movements and of their spoken translation was used to determine their temporal ("wait") strategy prior to translation onset. A relatively rapid onset of a spoken sentence translation after the presentation of a to-be-translated sentence (referred to as translation onset latency [TOL][1]) was assumed to index an "early start" (or short wait) strategy. Conversely, looking at a sentence for a considerable amount of time prior to the onset of its translation was assumed to reveal a "late start" (long wait) strategy. In view of Christoffels and de Groot (2005), we assumed that translation strategy would influence translation performance, and that a long wait strategy (large TOLs) would generate more accurate sight translations. In addition, we expected that translation experts would prefer longer waits while trainees would prefer shorter waits.

Method

Participants

Twenty-four professional conference interpreters (13 females, 11 males) and 15 interpreting trainees (11 females, 4 males) participated in the study. The professionals, all active on the Polish market as freelance interpreters, were recruited through a translation agency. Their mean age was 38 ($SD = 8.25$), and their mean experience as conference interpreters ranged from 6 to 37 years with a mean of 13 years ($SD = 8.00$). The mean number of conference days (i.e., working days with the maximum of 6 hr of interpreting) per month in the 3 years preceding the study was 6 ($SD = 4.37$). Polish was their native language (L1) and English as their second language (L2); several were also proficient in other languages (L3s included Russian, French, German, and Italian). Their mean LexTALE score (Lemhöfer & Broersma, 2012) for English was 89.31 ($SD = 9.31$). The trainees' mean age was 23 ($SD = 0.91$). All had Polish as their L1 and English as their L2 (CEFR, Council of Europe (2001) level C1 or C2); some had L3s (German, French, Spanish, or Chinese). The trainees were familiar with strategies of sight translation which they performed during classes. All trainees were at the same stage, that is, halfway through the second semester of their four-semester conference interpreting programme and none had any professional interpreting experience. Before the experiment, all participants reported no history of hearing or neurological problems, and all of them had normal or corrected-to-normal vision. No approval from the Ethics Committee was necessary for this study. As the participants performed a sight translation task in the study, we refer to them as translators rather than interpreters in the present paper.

Apparatus

An Eye-Link 1000 Plus with a desktop mount was used to record eye movements in remote mode. The sampling rate was 500 Hz (SR-Research, Ontario, Canada). Viewing was binocular, but eye movements were recorded from the right eye only. Sentences were presented on a 21-inch LSD monitor, positioned approximately 60 cm from participants. Sentences were presented in black, 14-point Courier New font (ensuring equidistant character spacing) on a white background using Experiment

Table 1. Characteristics of English stimulus words and their Polish equivalents. Frequency counts are shown as mean Zipf scores, word length is given in the number of characters, and concreteness was estimated on a 5-point Likert scale (standard deviations in parentheses)

	High-frequency words		Low-frequency words	
	English	Polish	English	Polish
Frequency	4.49 (0.3)	3.98 (0.5)	2.37 (0.3)	2.99 (0.6)
Length	7.3 (0.7)	6.67 (1.9)	7.5 (0.2)	6.96 (2.0)
Concreteness	4.68 (0.2)		4.73 (0.1)	

Builder software (SR-Research, Ontario, Canada). Eye movements were calibrated with a 13-point grid. Speech was recorded throughout a trial with a Philips microphone and recorded input was analyzed with Audacity(*R*) recording and editing software (Audacity Team, 2017). Translation accuracy was determined offline by a professional translator.

Materials

The critical words in the to-be-translated English sentences were 30 high-frequency nouns and 30 paired low-frequency nouns. The processing of these words was of primary interest. The two members of a critical word pair were matched for length and concreteness, and their meaning was congruent with a constructed sentence frame. The frequency data for these critical words were taken from SUBTLEX-UK (van Heuven et al., 2014) and concreteness ratings from the Brysbaert, Warriner and Kuperman (2014) database. Frequency data for the Polish equivalents were taken from SUBTLEX-PL (Mandera et al., 2014). Concreteness ratings for corresponding Polish translations were not available but are not likely to significantly differ from the ratings in English. Of the 60 critical English nouns, none had a Polish cognate. Table 1 shows characteristics of the critical source-language words and their Polish equivalents. Thirty sentences frames were constructed that could accommodate the high- and the low-frequency member of a critical word pair. Also, care was taken to retain a relatively consistent syntactic structure across sentences. Each participant saw only one of the two lists.

To reduce the skipping (nonfixation) of critical words, they were preceded by a word with at least six letters (Rayner, 1998). Table 2 shows only samples of the experimental materials and the full list of stimuli is given in Supplementary Material. A MANOVA that compared frequency and length across languages revealed no significant difference in length between the critical English source words and their Polish translation in the high- and low-frequency condition ($p > .05$). The frequency manipulation, by contrast, yielded a significant difference between high- and low-frequency words in both languages ($p < .001$).

The predictability of critical source-language words was determined via a cloze test that was distributed online to 27 evaluators who were native speakers of Polish with English as their L2 at the proficiency level (C1 or C2 according to CEFR, Council of Europe, 2001). They read the English sentence frames (without target

Table 2. Sample experimental stimuli (targets in capitals)

Sentence with a low-frequency word	Sentence with a high-frequency word
Since he knew nothing about modern art, he laughed at a painting depicting URINALS covered with ribbons in the museum.	Since he knew nothing about modern art, he laughed at a painting depicting SOLDIERS covered with ribbons in the museum.
If you don't want to get too depressed, stop worrying about your problematic SINUSES all the time.	If you don't want to get too depressed, stop worrying about your problematic CHILDREN all the time.

words) and filled in fitting English words. If over 30% of respondents completed a sentence frame with the same word, the sentence was removed from the stimulus set. As a result of the first round of norming, five sentences were removed. Modified versions were normed in a second round by 35 evaluators (coming from the same population as the first group of evaluators). The mean cloze value of high- and low-frequency source words was 2.7% and 0%, respectively, indicating that none of the targets was constrained by prior context. Increases in the frequency of a target word were, however, associated with increases in the number of its acceptable English-to-Polish translations. Specifically, we calculated critical source words' translation entropy (HTra) (Carl & Schaeffer, 2017), defined as the number of "choices a translator has for a given source text word, that is how many equally likely translations may be produced in a given context" (Schaeffer et al., 2016, p. 185). The entropy values of high- and low-frequency words were 0.77 and 0.49, respectively, and the correlation between HTra values and target frequency was $r = .27$, across items ($t = 2.17$, $df = 58$, $p < .05$). Together these examinations indicate that high-frequency source words should have a general recognition and translation advantage, their recognition and translation being easier than the recognition and translation of low-frequency source words.

The materials also included 62 filler sentences so that participants would not encounter difficult words that were relatively difficult to identify and to translate on a large proportion of trials. Two lists of sentences were constructed. Each list contained the same 30 sentence frames with a different member of the critical word pair. Half the critical source words on a list were low-frequency words and the remaining critical source words were high-frequency words. Both lists also contained the identical set of filler sentences. The experimental sentences were displayed in no more than two lines of text. Importantly, the target word never occupied the beginning or ending two word locations on the screen.

Procedure

The experiment was designed using the Experiment Builder (SR-Research) program. A session started with the sight translation of 10 practice sentences, and this was followed by the translation of one list of sentences.

To record eye movements, participants faced the eye-tracker and the computer screen. We used the eye-tracker in remote mode, which meant that the participants put a target sticker on their foreheads and performed a 13-point calibration.

A recalibration was performed between trials as needed. An obligatory recalibration was performed half-way through the experimental block (after 46 sentences).

Trials began with a fixation point (black dot) displayed where the first letter of the sentence later appeared. The participants were instructed to sight translate the English sentence appearing on the screen into Polish. They were told that the usual time constraints of professional sight translation should be applied. This meant that translators should perform the translation task as fast as possible. When the participant finished the spoken translation of a visible sentence, the experimenter initiated the presentation of the next sentence (or initiated a calibration if needed). The session lasted approximately 25 min. On each trial, eye movements were recorded from the onset of a sentence until its visibility was terminated, and production of a translation was recorded from the onset of a visible sentence to the end of articulation.

Data selection and measurement specification

On 1,100 of 1,170 experimental trials, translators' viewing of a visible sentence and its spoken translation were recorded successfully. The time spent viewing critical words was examined in the first set of analyses. To be included, the critical word had to be translated accurately, which removed 188 trials, it had to be the recipient of at least one fixation, which removed another 217 trials (on which the critical word was skipped), and the duration of a critical word's first fixation duration had to be less than 1000 ms, which removed another 6 trials. This left 682 trials for the analyses of critical word viewing.

Five word viewing duration measures were computed: first fixation duration, gaze duration, go-past time, re-view duration, and total viewing duration. First fixation duration comprised the duration of the first fixation on the word, and gaze duration comprised the cumulated time spent viewing it until the eyes moved to another word[2]. Go-past time included a critical word's first-pass viewing duration and the duration of subsequent fixations, on *prior* words, until the eyes moved past the target to subsequent words. The measure is assumed to be sensitive to the ease with which a word is recognized and integrated into known sentence context during reading (Liversedge et al., 1998; Reichle et al., 2009). Since translation occurs primarily during the re-reading of words in McDonald and Carpenter's (1981) model, the re-reading of critical words was thus of particular theoretical interest. The vast majority of critical words were re-read after their first-pass viewing had been terminated (91.9%, $n = 627$), and their re-reading duration consisted of the cumulated duration of all re-reading fixations. This was supplemented with an examination of re-reading types, as it was associated with one of three saccade types. A critical word's first-pass viewing could be terminated with an outgoing regression to prior text, so that the critical word was subsequently re-read with a forward-directed saccade (this occurred on 17.6% of critical word viewing trials). More frequently, re-reading occurred when first-pass critical word viewing was terminated with a forward-directed saccade, and when it was subsequently reached with an incoming regression (50.3%). On 24% of the critical word viewing trials, re-reading followed both an outgoing regression and an incoming regression, that is, the word was re-read more than once. The final measure, total viewing duration, comprised the

cumulated duration of all fixations on a word, irrespective of whether it was re-read once or several times.

Another set of analyses investigated the role of participants' self-selected translation (wait) strategy in sight translation performance. The wait strategy measure (TOL) for each trial was defined as the interval between the presentation of a to-be-translated sentence and the onset of their spoken translation. A small number of trials ($n = 7$) were excluded because TOLs were extremely short, <300 ms, or extremely long, >18 s, as they were typically due to equipment malfunction (e.g., an improper activation of the microphone). This left 1,093 trials for analysis.

Three measures were obtained to index translation performance. Two that sought to capture the accuracy of a spoken translation, one for the critical word and the other for the full source-language sentence, and one that sought to capture the fluency of translation. When judging accuracy, we followed the procedure from Ruiz and Macizo (2019). The distinction between accurate and inaccurate critical word translations was relatively straightforward, and participants articulated an accurate Polish (equivalent) noun for the critical English source word on 905 of 1093 trials (82.8%). The distinction between accurate and inaccurate sentence translations was less clear cut. Careful evaluation of full sentence translations by a professional translator who served as referee (17 years of conference interpreting and written translation experience—with Polish as L1 and English as L2) revealed nuanced degrees of accuracy. The Polish translation of a sentence could deviate from the meaning of the visible English sentence, the translation of a particular word—the critical word or another content word—could be erroneous, pronouns could be over-used (fewer uses in Polish than in English), or some other deviation could occur. Rather than categorizing a sentence translation as either fully accurate or inaccurate, the referee thus assessed the fidelity of its translation on a 4-point scale, 1 indicating a poor translation, such as a deviant sentence meaning, and 5 indicating a fully accurate translation of a sentence. Most sentences (71%) had a high translation accuracy rating of 4 or 5.

The cumulated duration of pauses between spoken words was computed to index the fluency of a spoken translation. The EMU-SDMS tool for speech corpus creation (Koržinek et al., 2017; Winkelmann et al., 2017) was applied to segment the spoken translation of a sentence into words in conjunction with a manually entered transcription of the words of the spoken translation. An R script was then used to calculate the cumulated inter-word pause duration for each sentence. One trial had an extremely long pause duration, more than 10 *SD*s above the mean, and two trials had a pause duration of zero. These three trials were excluded from the analysis of pause durations.

Data analysis

Two sets of analyses were applied. The analysis of the viewing of critical words allowed us to investigate whether comprehension and translation are serial or co-occur during sight translation. In turn, the analysis of the quality of translation performance examined the role of participant's individual decisions of when to start translation production relative to the onset of reading (through TOL) during sight

translation. All statistical analyses were performed with the R system for statistical computing (R Core Team, 2020). Linear mixed models and generalized linear mixed models, as implemented in the lme4 library (Bates et al., 2015), were used for the analyses of trial-based numeric and binary data, respectively. A linear mixed model from library ordinal (Christensen, 2019) was used to analyze sentence accuracy ratings.

The following fixed effect structure was applied to the set of critical word viewing data. Critical word frequency, translation expertise, and the interaction of the two factors were used as fixed effects, as this sufficed for the testing of the serial processing and the co-occurrence hypothesis.

The fixed effects for the larger set of translation performance data (sentence accuracy, critical word accuracy, and cumulated pause duration) were critical word frequency (high vs. low), translation expertise (professional vs. trainee), and a linear trend for the wait strategy (TOL). Since exploratory analyses suggested that short and long TOLs could be associated with higher-quality translations, the model also included a quadratic trend for TOL. In addition, the model included fixed effects for the interaction of critical word frequency with expertise and the linear and quadratic TOL trends. Sum contrasts were applied to the factors frequency and expertise, so that effect sizes reflected the distance of a condition from the grand mean. The numeric TOL values were centered.

The binary values of the factor expertise were slightly correlated with the linear component of TOL ($r = .21$) but not with the quadratic component ($r = -.07$). Nevertheless, statistical models with these fixed effects examined potential effects of collinearity with Anova models with type 3 sums of squares. For sentence translation accuracy, we applied the Anova.clmm function from library RVAideMemoire (Hervé, 2021), and for all other models, we applied the Anova function from library car (Fox & Weisberg, 2010).

At the outset, all linear mixed models used a maximal random factor structure with random intercepts for participants and items, and random slopes for critical word frequency. Since each sentence frame contained low- and high-frequency critical words and was read by professionals and trainees, the random factor structure for items also included random slopes for expertise and the interaction of expertise with word frequency. In addition, the maximal model included the correlations of random factor components. Application of a random factor principal component analysis (with the rePCA function), as suggested by Matuschek et al. (2017), indicated that the maximal random factor structure yielded an over-parameterized fit for most statistical models. When this was the case, we followed Matuschek et al. and used a "top-down" approach to simplify the model. In a first step, we removed all random factor correlations, and when this did not eliminate random factor over-parameterization, the random factor component with the smallest amount of variance was dropped from the model until overfitting was eliminated. Across simplified models, this yielded a random factor structure with four to six components.

The frequency distributions of the to-be-analyzed numeric measures deviated from the normal distribution. Specifically, the skewing of pause durations and of the five critical word viewing duration measures was positive. In a first step, these data sets were log transformed which yielded close-to-normal distributions for all five viewing duration measures except total viewing durations. A square root
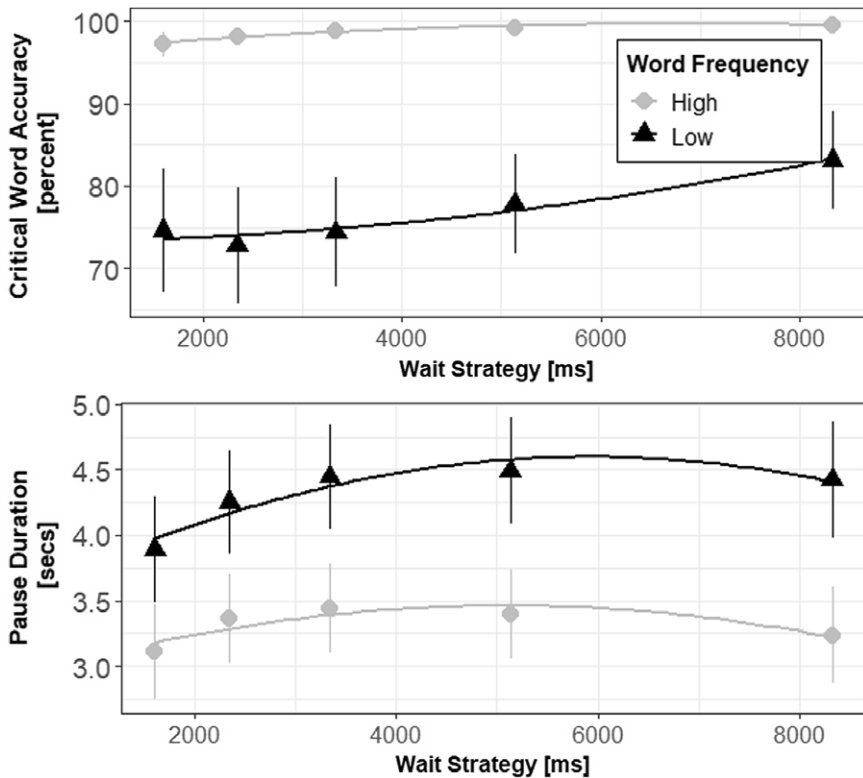
Figure 1. Critical word accuracy (upper panel) and pause duration (lower panel) with the corresponding standard error as a function of wait strategy (TOL) and target frequency. Sentence-specific TOL values were grouped into five bins of approximately equal size that corresponded to very short, short, average, long, and very long waits prior to translation onset. Standard errors were computed for each bin.

transformation yielded close-to-normal frequency distributions for this measure and also for pause durations. The reported coefficients of the statistical models are thus based on the ordinal (ranked) values for sentence translation accuracy ratings, on logits for critical word translation accuracies, on square root transformed values for pause durations and total viewing durations, and on log transformed first fixation durations, gaze durations, go-past times, and re-reading durations. Figures and tables show backtransformed condition means and standard errors that were extracted from statistical models with function emmeans from library emmeans (Lenth et al., 2020). Figure 1 was plotted with ggplot2 (Wickham, 2016).

Results

Critical word viewing

In order to determine the nature of reading in sight translation, we analyzed word viewing measures as a function of target word frequency and translation expertise.

Table 3. Means and standard errors (in parentheses) for measures of critical word viewing, re-reading, and translation quality as a function of word frequency and expertise

	Expertise			
	Professionals		Trainees	
	Critical word frequency		Critical word frequency	
	High	Low	High	Low
Critical word viewing				
First fixation duration	238 (11.5)	253 (12.9)	250 (14.8)	253 (16.9)
Gaze duration	269 (17.8)	320 (22.1)	280 (22.5)	314 (27.1)
Go past time	556 (66.3)	630 (77.9)	436 (62.9)	562 (89.1)
Re-reading duration	780 (78.7)	921 (96.3)	834 (98.5)	1293 (168.2)
Total viewing duration	1104 (110.7)	1436 (130.1)	1159 (125.8)	1752 (166.8)
Re-reading patterns				
Out-Regressions	16.5 (2.70)	16,2 (3.31)	9.26 (3.13)	15.9 (4.53)
In-Regressions	43.2 (6.70)	30.3 (7.17)	59.9 (8.39)	52.9 (9.58)
In+Out-Regressions	23.9 (4.70)	19.2 (4.36)	11.9 (3.82)	15.9 (5.34)
Translation quality				
Sentence translation (rank)	3.69 (0.31)	3.14 (0.33)	2.42 (0.32)	1.94 (0.35)
Critical word accuracy (%)	99.1 (0.47)	83.8 (5.33)	97.9 (1.15)	59.7 (10.3)
Pause duration (s)	2.95 (0.34)	3.86 (0.438)	3.95 (0.51)	5.04 (0.58)

The means and standard errors for the five critical word viewing measures are shown in Table 3 as a function of word frequency and expertise.

All five measures showed numerically shorter viewing durations for the high-frequency condition. However, the word frequency effect was not significant for first fixation duration, which yielded a relatively small numeric difference, 9 ms ($b = -0.018$, $SE = 0.020$, $t = -0.884$, $p = .347$). The frequency effect was larger, 43 ms, for gaze durations ($b = -0.063$, $SE = 0.027$, $t = -2.308$, $p = .05$) and increased to 100 ms for go-past times ($b = -0.092$, $SE = 0.043$, $t = -2.159$, $p < .05$). Importantly, the effect of expertise was negligible across these three viewing measures, amounting to 6 ms shorter first fixation durations for experts ($b = -0.011$, $SE = 0.029$, $t = -0.375$, $p = .710$), 2.5 ms longer gaze durations ($b = 0.013$, $SE = 0.036$, $t = 0.361$, $p = .720$), and 94 ms longer go-past times for experts than for trainees ($b = 0.090$, $SE = 0.072$, $t = 1.239$, $p = .222$). The interaction of the two main effects was negligible across the three measures (all $p > .30$). Together, all three show familiar benefits for high-frequency words, and none of them revealed shorter critical word viewing durations for professional interpreters; hence, no effect of expertise was found.

Re-reading durations and total viewing durations (which include re-reading time) showed, by contrast, reliable effects of expertise. The re-reading durations

of professionals were 213 ms shorter than the re-reading durations of trainees ($b = -0.102$, $SE = 0.043$, $t = -2.370$, $p < .025$). The corresponding benefit of expertise amounted to 186 ms for total viewing durations ($b = 0.078$, $SE = 0.038$, $t = 2.056$, $p < .05$). Although the numeric differences between groups were larger for low-frequency words, the interaction of the two main effects did not approach significance (both $p > .12$).

The means and standard errors for the three types of critical word re-reading are shown in Table 3. In it, regressions out of a critical word (so that it is re-read with a forward-directed saccade) are labeled Out-Regressions and backward directed regressions to the critical word are labeled In-Regressions. Instances in which both types of regressions occurred are labeled In+Out-Regressions. As can be seen, re-reading with a regression back to a critical word (In-Regressions) was more common (50.3%) than re-reading after a regression out of a critical word (17.5%). Both types of re-reading occurred on 24% of the critical word viewing trials. Across the three measures, none of the fixed effects approached significance (all $p > 0.176$).

Together, these analyses show group-specific differences, as translation expertise influenced primarily the time spent re-reading critical words. Both groups accomplished re-reading primarily through regressions that were directed back to critical words, and there were no significant differences in their types of re-reading. Translation expertise thus influenced the time spent on the re-reviewing of words but not the programming of re-reading saccades and not the first-pass viewing.

Translation quality

In order to investigate how the participants' wait strategies affected their performance in sight translation, we analyzed translation quality by looking into three measures: sentence accuracy, critical word accuracy, and cumulated pause duration (as a function of TOL coupled with word frequency and expertise).

The effects of wait strategy (TOL) on critical word translation accuracy and on pause durations are shown in Figure 1. Furthermore, Table 3 shows critical word translation accuracy, sentence translation accuracy, and pause duration as a function of word frequency and translation expertise.

The analyses of all three translation quality measures yielded a robust effect of critical word frequency, its translations being more accurate and more fluent in the high- than the low-frequency condition (an ordering effect of 0.51 for sentence translation accuracy: $b = 0.258$, $SE = 0.095$, $z = -2.706$, $p < .01$; a 27% higher critical word translation accuracy: $b = -1.653$, $SE = 0.246$, $z = -6.707$, $p < .001$; and 1.0 s shorter pause durations: $b = -0.125$, $SE = 0.024$, $t = -5.047$, $p < .001$). The main effect of translation expertise was also highly significant, professionals generating more accurate sentence translations, with an order effect of 1.25, a 23% higher critical word translation accuracy, and 1.09 s shorter pauses pause durations ($b = 0.617$, $SE = 0.152$, $z = 4.046$, $p < .005$; $b = 0.543$, $SE = 0.224$, $z = 2.432$, $p < .05$; and $b = -0.1374$, $SE = 0.068$, $t = -2.028$, $p < .05$, respectively). Most importantly, the linear effect of TOL was significant for critical word accuracy ($b = 0.732$, $SE = 0.338$, $z = 2.169$ $p < .05$), the translation of critical words being more accurate with longer TOLs (top panel of Figure 1). Furthermore, TOLs yielded a significant quadratic effect for pause duration ($b = -0.101$, $SE = 0.048$,

$t = -2.122$, $p < .05$), as pauses were shorter when TOLs were either relatively short or long (bottom panel of Figure 1). No other effect approached statistical significance (all $p > .196$).

Together, the analyses of translation performance yielded familiar effects of word frequency and expertise. Translation was more accurate and more fluent in the high- than the low-frequency condition, and professionals translated more accurately and more fluently than trainees. Most importantly, there was also a relationship between participants' TOL and performance. Critical word translations were more accurate with longer TOLs, and translation was more fluent when TOLs were short or long.

Discussion

The current study aimed at investigating the coordination of source text comprehension and translation in a sight translation task with a focus on the impact of participants' translation strategies on the quality of translation. Our approach followed McDonald and Carpenter's (1981) study, except that we used sentences with high- and low-frequency source words rather than paragraphs with ambiguous or unambiguous phrases, and that we tested substantially more participants, both trainees and professionals. We also measured the participants' translation (wait) strategy prior to the onset of the spoken translation, and we computed multiple performance measures that indexed both the accuracy and the fluency of translation. In the following, we discuss the implications of our findings for the theoretical account of sight translation and for the understanding of translation strategies.

Implications for the theoretical account of sight translation

The key feature of McDonald and Carpenter's (1981) proposed account of sight translation is two or three processing phases. An initial stage (i.e., first-pass viewing, here indexed by first fixation duration, gaze duration, and go-past time) entails reading for comprehension, including identification and comprehension of a sequence of source-language words. This is followed by re-viewing (i.e., second-pass viewing, here indexed by re-view duration and total viewing duration) when comprehension continues and translation occurs. McDonald and Carpenter's account also included a correction phase, involving third-pass viewing, when prior context misleads translators about the meaning of an ambiguous source-language expressions. Since our materials did not seek to mislead the translator, we did not examine recoveries from translation errors, and we distinguished first-pass viewing durations from all other re-view durations. In addition, we measured the accuracy and fluency of translation.

The considerably higher accuracy and fluency of translation for sentences with high-frequency critical words indicate that sight translation was easier in this condition. Word frequency also yielded numeric effects for all viewing measures, with shorter viewing durations for high- than for low-frequency words when critical words were fixated and correctly translated. The influence of word frequency on first fixation duration appeared to be relatively weak, however, and the corresponding statistical effect did not approach significance. This is not an aberrant finding.

The size of the word frequency effect varies across studies, some reporting even smaller effects of word frequency on first fixation duration that also failed to reach statistical significance (e.g., Pollatsek et al., 2008). Overall, the numeric size of the word frequency effect for the two first-pass viewing measures, first fixation duration and gaze duration, is within the range of corresponding first-pass effects in studies of silent reading (Baayen et al., 2016; Cop et al., 2015; Diependaele et al., 2013; Kliegl et al., 2004; Slattery et al., 2007; Tiffin-Richards & Schroeder, 2015).

Similar to the current study, studies of silent reading also consistently show larger word frequency effects for gaze duration than for first fixation duration. This is attributed to saccadic error and to demands of cognitive processing. Saccadic error misplaces some initial fixation on a targeted word. When this occurs, the initial viewing location is often corrected with an intra-word (first-pass) refixation, and the duration of these refixations is longer for low- than for high-frequency words. Staub et al.'s (2010) analyses of distributional properties of first fixation duration and gaze duration, for example, provide compelling evidence for weaker effects of linguistic processes demands on first fixation duration than on gaze duration.

In contrast to silent reading, where second-pass viewing of words is relatively rare, generally less than 20% (e.g., 7% for high- and 15% for low-frequency words; White et al., 2018), the current study showed that virtually all critical source words (91.9%) were re-read after their first-pass viewing. Moreover, the numeric size of the word frequency effect was considerably larger during re-reading (300 ms) than during first-pass viewing. This provides evidence for the view that a substantial amount of source word processing occurred *after* its first-pass viewing. Since the translation of low-frequency source words was slower and more error prone than the translation of high-frequency source words, it is plausible to assume that the sizable word frequency effect for re-view duration is due to the more difficult processing of low-frequency words that occur after first-pass viewing. Collectively, the effects of word frequency on critical words' first-pass viewing and on their subsequent re-viewing are consistent with McDonald and Carpenter's processing account.

The effects of expertise on critical word viewing provide converging evidence for this account. Consistent with McDonald and Carpenter's proposal, expertise did not influence the first-pass viewing of critical words, which may have been expected if this phase of processing was assumed to be similar to silent reading and entailed primarily the comprehension of words within their sentence context. Furthermore, expertise influenced re-reading duration, with numerically larger differences for low-frequency critical words, as should occur if translation occurred during the subsequent re-reading of words.

Even though the numeric effects of the word frequency effect in the current study are well within the range of first-pass viewing durations during silent reading, it could be argued that the current study should have included a silent reading for comprehension condition, so that translators' first-pass viewing of text could be compared across tasks. However, this may not provide conclusive evidence. If, as expected, a direct comparison showed that the word frequency effect was similar for the two tasks, it could be argued that the conclusion is based on an acceptance of the null hypothesis. This argument can also be leveled against the null effect of translation expertise for first-pass critical word viewing in the current study, as it

cannot be ruled out that a different manipulation of target language could influence first-pass viewing during sight translation.

The study by Ruiz et al. (2008) and its recent replication (Ruiz & Macizo, 2019) claim extensive interaction between the source and target language properties when reading text for comprehension and when it is sight translated. However, this work used the self-paced reading paradigm, which makes it impossible to differentiate between first- and second-pass viewing.

Taken together, the current study lends support to McDonald and Carpenter's (1981) account of sight translation. That is, the translation process occurs primarily during the re-reading of a corresponding sequence of source-language words. The implications of the current study for the processing account of sight translation pinpoint subtleties in the temporal distribution of the processing stages involved, that is, comprehension of words in the source text and the production of the target text. The present findings along with those from Huang's study (2011) seem to suggest that the initial reading (i.e., first fixations, gaze duration, and go-past time) in sight translation is similar to regular reading for comprehension. Studies with bilinguals that did not involve translation suggest that initial viewing may involve nonselective access of words in the target language, that is, potential translation equivalents in the other language receive activation (Duyck et al., 2007; Lauro & Schwartz, 2017; Libben & Titone, 2009; Van Assche et al., 2011). However, this activation is not sufficient for oral production of an accurate sequence of target language words since inhibition of inaccurate equivalents and further activation of the accurate equivalent are also required. Thus, further processing (the actual translation) seems to take place during re-reading, as manifested in late reading measures (i.e., in re-view duration and total viewing duration). More studies with both early and late reading measures involving direct comparisons of reading and sight translation performed by professionals are needed to shed more light on the complexity in sight translation.

Translation strategies and translation expertise

Sight translation requires the comprehension of a sentence in the source language and the overt articulation of its meaning in the target language. The eye-voice span data reported in prior research have indicated that the viewing of one part of the source sentence typically occurs while an earlier part of the sentence is translated (Chmiel et al., 2020). Therefore, the coordination of functionally distinct processes that are executed at different points in time could make sight translation also amenable to the use of translation strategies that are applied even before the spoken translation is started. A novel aspect of the current study was the a priori specification of a temporal strategy index, TOL, that indexed the self-selected "wait time" prior to the onset of a sentence translation. TOL has a fixed temporal starting point, the onset of a sentence, prior to which no sentence processing can occur. Furthermore, TOL appears to capture general preparatory processes that typically include the viewing of several words prior to translation onset.

Following Christoffels and de Groot (2005), we assumed that longer TOLs would yield more accurate and more fluent translations. This prediction received only partial support. Critical word translation accuracy increased with TOL duration, but

the numeric size of the effect was relatively small. For instance, the accuracy difference between the TOL grouping with the shortest and longest values (bottom panel of Figure 1) is approximately 4%, and this is substantially smaller than the difference between trainees and experts, 13%. In addition, sentence translation accuracy yielded a robust effect for translation expertise but no significant effect for TOL.

Why was the influence of TOL duration not consistent across the two performance measures? One possibility is that TOL influenced only word-level processes, hence the effect for critical word accuracy but not for sentence accuracy. This account appears unattractive, however, because the relatively long duration of TOLs suggests precisely the opposite: that is, TOL should have been influenced by sentence-level processes. A more attractive alternative account is that the self-selection of TOL leveled its influence on translation performance. This leveling occurred with the translation of critical words and—even more so—with the translation of full sentences. For instance, prior to translation onset, translators could have waited until their understanding of the source sentence reached some criterion and/or when they could gauge the success of its translation. Even if this performance criterion changed somewhat from sentence to sentence, it would have a leveling effect, and relatively large TOL differences could be associated with relatively small differences in translation performance.

Self-selection may also provide a tentative account for the quadratic trend of TOL when pause durations were analyzed. Relatively short TOLs may occur when the success of sentence comprehension is estimated after the viewing of a few words and this input is considered sufficient for the translator to feel confident enough to start the production of the translation. This may result in a relatively fluent sentence production. Conversely, relatively long TOLs may occur when a relatively large part of the sentence needs to be comprehended perhaps because it is difficult to translate or because the translator needs more input to feel certain to start a successful translation, which may also be followed by a relatively fluent production.

Future work, in which TOL and the translation quality criterion are under experimental control, could be used to test our ad-hoc accounts for the effects of TOL in the current study. Experimental control over the onset of a translation could be obtained by signaling the translator when to start with the speaking of a translation, and the interval between the presentation of the source sentence and the go-signal for the translation could range from short ($\sim$2 s) to long ($\sim$8 s). The criteria for the quality of translation could be manipulated via instruction. With this approach, longer signal intervals could result in more accurate sentence translations and interval duration may not yield a quadratic trend for pause duration.

Conclusions

The present study provided novel evidence for stages in sight translation performance, wherein the initial phase of normal reading for comprehension (i.e., first-pass viewing, indexed by first fixation duration, gaze duration, and go-past time) is followed by phases in which reading and translation occur (i.e., second-pass viewing, indexed by re-view duration and total viewing duration). Thereby, the study lent support to McDonald and Carpenter's (1981) account. Furthermore,

the analysis of the role of interpreters' strategic decisions when to start producing the translation (relative to the onset of source text reading) revealed a nuanced picture. Critical word translations (but not sentence translations) were more accurate when interpreters tended to wait longer (longer TOLs), and translation was more fluent when TOLs were short or long. In contrast, expertise profoundly affected not only the quality of translation at the word level and at the sentence level, but it also strongly influenced how the source text was viewed during re-reading. Our study also shows that real-life task conditions and fine-grained eye movement analyses should be employed in order to empirically test the dynamics of processes involved in sight translation as other research methods and unnatural tasks may lead to task-specific effects.

Notes

1 We differentiate between TOL (translation onset latency) and eye-voice span. The latter is the interval between reading and producing oral translation measured for any source text word. The former has a fixed temporal starting point, the onset of a sentence, prior to which no sentence processing can occur.

2 We thus used the term "gaze" as it is used in reading research while a different definition is typically used in translation research. There "gaze" often means the total time spent viewing a word, that is, this definition of gaze includes the initial (first pass) viewing of words and their subsequent re-viewing (using the terminology of reading research, we refer this measure as total viewing duration).

References

Audacity Team. (2017). *Audacity(R): Free Audio Editor and Recorder [Computer application]* (2.2.1) [Computer software]. https://audacityteam.org/

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, **30**(11), 1174–1220. https://doi.org/10.1080/02687038.2016.1147767

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, **46**(3), 904–911.

Carl, M., & Schaeffer, M. J. (2017). Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *HERMES - Journal of Language and Communication in Business*, **56**, 43. https://doi.org/10.7146/hjlcb.v0i56.97201

Čenková, I. (2015). *Routledge Encyclopaedia of Interpreting Studies* (F. Pöchhacker) (Ed.). Routledge.

Chmiel, A., & Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters: Professionals are more time-efficient while trainees view the source text less. *Target*, **31**(3), 378–397. https://doi.org/10.1075/target.18091.chm

Chmiel, A., & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartłomiejczyk (Eds.), *Tracks and Treks in Translation Studies* (Vol. **108**, pp. 189–205). John Benjamins Publishing Company. https://doi.org/10.1075/btl.108.10chm

Chmiel, A., Janikowski, P., & Cieślewicz, A. (2020). The eye or the ear?: Source language interference in sight translation and simultaneous interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*, **22**(2), 187–210. https://doi.org/10.1075/intp.00043.chm

Christensen, R. H. B. (2019). *ordinal—Regression Models for Ordinal Data*.

Christoffels, I. K., & Groot, A. M. B. de. (2005). Simultaneous Interpreting: A Cognitive Perspective. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). Oxford University Press.

Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40year legacy. *Journal of Memory and Language*, **86**, 1–19. https://doi.org/10.1016/j.jml.2015.07.004

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, **22**(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Press Syndicate of the University of Cambridge.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**(5). http://psycnet.apa.org/psycinfo/1993-04299-001

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, **66**(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dragsted, B., & Hansen, I. G. (2009). Exploring translation and interpreting hybrids. The case of sight translation. *Meta: Journal des traducteurs*, **54**(3), 588. https://doi.org/10.7202/038317ar

Duyck, W., Van Assche, E., & Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**(4), 663–679.

Fox, J., & Weisberg, S. (2010). *An R companion to applied regression (Second edition).* SAGE Publications, Inc.

García, A. M., Ibanez, A., Huepe, D., Houck, A. L., Michon, M., Lezama, C. G., Chadha, S., & Rivera-Rei, A. (2014). Word reading and translation in bilinguals: The impact of formal and informal translation expertise. *Frontiers in Psychology*, **5**. https://doi.org/10.3389/fpsyg.2014.01302

Hervé, M. (2021). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics.* https://CRAN.R-project.org/package=RVAideMemoire

Huang, C.-C. (2011). *Tracking eye movements in sight translation_* [Unpublished M.A. thesis, National Taiwan Normal University]. http://rportal.lib.ntnu.edu.tw/.

Inhoff, A. W. (1991). Word frequency during copytyping. *Journal of Experimental Psychology. Human Perception and Performance*, **17**(2), 478–487.

Inhoff, A. W., & Radach, R. (2014). Parafoveal preview benefits during silent and oral reading: Testing the parafoveal information extraction hypothesis. *Visual Cognition*, **22**(3–4), 354–376. https://doi.org/10.1080/13506285.2013.879630

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, **40**(6), 431–439.

Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jacobsen, & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (Vol. **36**, pp. 103–124). Samfundslitteratur.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, **16**(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Korpal, P. (2012). On language-pair specificity in sight translation: An eye-tracking study. *Übersetzen in Die Zukunft Dolmetscher Und Übersetzer: Experten Für Internationale Fachkommunikation*. 2. Internationale Fachkonferenz des BDÜ „Übersetzen in die Zukunft", Berlin.

Koržinek, D., Marasek, K., Brocki, Ł., & Wołk, K. (2017). Polish Read Speech Corpus for Speech Tools and Services. *ArXiv:1706.00245 [Cs]*. http://arxiv.org/abs/1706.00245

Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, **92**, 217–233. https://doi.org/10.1016/j.jml.2016.06.010

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, **44**(2), 325–343.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.4.8) [Computer software]. https://CRAN.R-project.org/package=emmeans

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**(2), 381–390.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Elsevier. https://doi.org/10.1016/B978-008043361-5/50004-3

Ma, X., Li, D., & Hsu, Y.-Y. (2021). Exploring the impact of word order asymmetry on cognitive load during Chinese–English sight translation: Evidence from eye-movement data. *Target. International Journal of Translation Studies*, **33**(1), 103–131. https://doi.org/10.1075/target.19052.ma

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2014). Subtlex-pl: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, 1–13. https://doi.org/10.3758/s13428-014-0489-4

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, **94**, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McDonald, J. L., & Carpenter, P. A. (1981). Simultaneous translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior*, **20**(2), 231–247. https://doi.org/10.1016/S0022-5371(81)90397-2

Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology. Human Perception and Performance*, **34**(3), 726–750. https://doi.org/10.1037/0096-1523.34.3.726

R Core Team. (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, **72**(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, **62**(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, **14**(3), 191–201.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, **13**(3), 115–119. https://doi.org/10.1016/j.tics.2008.12.002

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica*, **128**(3), 490–500. https://doi.org/10.1016/j.actpsy.2007.08.004

Ruiz, J. O., & Macizo, P. (2019). Lexical and syntactic target language interactions in translation. *Acta Psychologica*, **199**, 102924. https://doi.org/10.1016/j.actpsy.2019.102924

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, **125**(2), 179–194. https://doi.org/10.1016/j.cognition.2012.07.004

Schaeffer, M., & Carl, M. (2013). Shared representations and the translation process: A recursive model. *Translation and Interpreting Studies*, **8**(2), 169–190. https://doi.org/10.1075/tis.8.2.03sch

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., & Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB* (pp. 183–210). Springer International Publishing. https://doi.org/10.1007/978-3-319-20358-4_9

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, **131**(1), 1–27. https://doi.org/10.1016/j.cognition.2013.11.018

Shreve, G. M., Lacruz, I., & Angelone, E. (2010). Cognitive effort, syntactic disruption, and visual interference in a sight translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition: Vol. XV* (pp. 63–84). John Benjamins Publishing Company. https://doi.org/10.1075/ata.xv

Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, **35**(6), 1283–1292. https://doi.org/10.3758/BF03193601

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, **36**(5), 1280–1293. https://doi.org/10.1037/a0016896

Su, W., & Li, D. (2020). Exploring processing patterns of Chinese-English sight translation: An eye-tracking study. *Babel. Revue Internationale de La Traduction/International Journal of Translation*, **66**(6), 999–1024. https://doi.org/10.1075/babel.00192.su

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, **113**, 33–43. https://doi.org/10.1016/j.visres.2015.05.008

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **37**, 1412–1431. https://doi.org/10.1037/a0024492

Van Assche, E., Brysbaert, M., & Duyck, W. (2020). Bilingual lexical access. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual lexical ambiguity resolution* (1st ed., pp. 42–68). Cambridge University Press. https://doi.org/10.1017/9781316535967.004

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, **64**(1), 88–107. https://doi.org/10.1016/j.jml.2010.08.006

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, **67**(6), 1176–1190.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, **34**(1), 205–223. https://doi.org/10.1037/0096-1523.34.1.205

White, S. J., Drieghe, D., Liversedge, S. P., & Staub, A. (2018). The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *Quarterly Journal of Experimental Psychology*. https://journals.sagepub.com/doi/10.1080/17470218.2016.1240813

White, S. J., Warrington, K. L., McGowan, V. A., & Paterson, K. B. (2015). Eye movements during reading and topic scanning: Effects of word frequency. *Journal of Experimental Psychology: Human Perception and Performance*, **41**(1), 233–248. https://doi.org/10.1037/xhp0000020

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology*, **67**(6), 1151–1165. https://doi.org/10.1080/17470218.2013.848216

Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**(4), 1118–1129. https://doi.org/10.1037/xlm0000093

Whitford, V., Pivneva, I., & Titone, D. (2016). Eye movement methods to investigate bilingual reading. In R. R. Heredia, J. Altarriba, & A. B. Cieślicka (Eds.), *Methods in bilingual reading comprehension research* (pp. 183–211). Springer New York. https://doi.org/10.1007/978-1-4939-2993-1_8

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis (Second edition)*. Springer.

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, **45**, 392–410. https://doi.org/10.1016/j.csl.2017.01.002