ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 15, No. 10, pp. 3325-3335, October 2025
DOI: https://doi.org/10.17507/tpls.1510.19

Recognition of Neologisms During Translation: An Eye-Tracking and Key-Logging Study

Majdi Alzahrani
Department of English, College of Science and Arts, King Abdulaziz University, Rabigh, Saudi Arabia

***Abstract*—This study explores whether neologisms are recognized as unfamiliar words or treated as standard words during translation by examining the cognitive effort they require. Previous research has highlighted the challenges of recognizing, comprehending, and translating newly coined terms. Focusing specifically on recognition, this study hypothesizes that translating neologisms requires greater cognitive effort than translating standard words. To test this, an experiment was conducted with six participants using three short source texts, each containing a single neologism. Eye-tracking and keystroke logging were used to measure fixation and pause durations. A Linear Mixed Model was applied to determine whether neologisms required significantly more cognitive effort. The eye-tracking results showed a statistically significant difference, suggesting that neologisms were recognized as unfamiliar words during translation. In contrast, keylogging data revealed no significant difference, indicating that participants did not spend additional time typing neologisms. Methodological factors—such as the merging of pause types (PDB and PDW) and participants' unfamiliarity with physical keyboards—likely influenced the keylogging results. These findings highlight the need to refine pause-type analysis and account for typing behavior to better distinguish cognitive from non-cognitive influences in translation.**

***Index Terms*—cognitive effort, translating neologisms, eye-tracking, keystroke logging**

I. Introduction

The translation of neologisms is considered one of the translator's greatest challenges (Newmark, 1988), mainly because of the difficulties in comprehending and reproducing them in the target language (Lahlali, 2014). Even native speakers, who tend to have knowledge of the source language and culture, can experience difficulties in understanding neologisms, particularly when the text provides minimal or no contextual clues to discern the meaning of such unfamiliar terms (Lehrer, 2003). Furthermore, when consulting sources (e.g., a dictionary) to establish the denotative reference of a newly coined word, it is common to find that no definition is available (Hanaqtah, 2019). Consequently, neologisms are expected to require more time and more mental effort to translate, regardless of whether the result is accurate or inaccurate. Additionally, and perhaps most importantly, the nature of neologisms – the way they are created and formed – can increase the ambiguity of exactly what a neologism is (and is not), potentially leading to misinterpretations or the term being ignored altogether in the final translation.

Therefore, the identification of neologisms precedes the challenge of translating them. The structure of neologisms can sometimes be misleading, both visually and auditorily, particularly with blends – the most common type of neologism (Lehrer, 2003) – where part of the word that was used to create the neologism is absent. This compels translators to infer the missing portion, a process that can lead to incorrect or incomplete results (Beason, 2010). For instance, the term 'infomercial', a merger of 'information' and 'commercial' that refers to a marketing approach with hard-sell promotion that originated in television advertisements (Hope & Johnson, 2004), might be confused in translation with merely 'information' or 'informational', thus overlooking the marketing and sales dimensions.

According to a study conducted by Al-Jarf (2010) involving twenty translation students, a significant majority (76%) failed to recognize neologisms as such. While Al-Jarf's findings highlight challenges in identifying neologisms, her study relied primarily on questionnaire and test-based approaches. In contrast, this paper investigates the same phenomenon – whether neologisms are recognized or perceived as standard words – through a process-oriented lens, using two data-elicitation methods: eye-tracking and keylogging (Hvelplund, 2011). Specifically, the study examines the cognitive effort involved in translating neologisms by measuring fixation duration for each source-text word and pause duration for each corresponding target-text word. A linear mixed model is then used to test for significant differences between neologisms and standard words.

II. Literature Review

The term neologism refers to "newly coined lexical units or existing lexical units that acquire a new sense" (Newmark, 1988, p. 140). Such expressions are coined for cultural, artistic, scientific, or technological reasons. For example, the term 'twindemic', a merger of 'twin' and 'pandemic', refers to the flu season alongside COVID-19 (Schmied & Ivanova, 2023). Another widely used neologism is Brexit, a portmanteau combining 'British' and 'exit', which was popularized in media discourse not only in English and European languages, but also globally, following the United Kingdom's 2016 departure from the European Union.

Neologisms are categorized according to their method of creation and assigned types. Some are created entirely from scratch, although such occurrences are exceedingly rare. More commonly, neologisms are created through standard linguistic formation methods such as compounding, blending, derivation, extending the meaning of existing words, or loan translation. Moreover, neologisms are classified into distinct types, including form-based (e.g., compounds), functional (e.g., conversions), semantic (involving shifts in meaning), and borrowed types (direct borrowings from other languages) (see, e.g., Rogers, 2015).

However, neologisms differ in several respects: (1) their functional role, for instance to accommodate cultural shifts and technical inventions, (2) their dynamics with synonyms, that is, how they interact with existing synonyms in replacing or coexisting with older terms, (3) the resources used in their formation (i.e., the linguistic formation methods mentioned above), (4) the longevity of the established term (i.e., some neologisms last for a long time, while others disappear quite quickly), (5) their integration into existing linguistic systems (i.e., to what extent they gain acceptance in terms of semantics and/or syntax), and (6) their interaction with different linguistic systems – some neologisms are mainstreamed in other languages, whether by translation or other means (Cabré, 1999, pp. 205–207).

As part of their functionality – specifically, their ability to capture the reader's attention (Llopart-Saumell, 2023) – neologisms can serve aesthetic and ornamental purposes rather than being merely informative (Lehrer, 2003; Thomsen, 2006). Authors, especially in media discourse, often employ neologisms for creative purposes when crafting product names, slogans, and foregrounded expressions through wordplay and puns and other literary techniques (see, e.g., Kalugina et al., 2019). Moreover, regardless of their intended use, neologisms, alongside other linguistic and non-linguistic resources available to users, contribute to the expansion of the language's lexicon, thus, enhancing the expressive and communicative modes of speakers' thoughts and emotions (Steinmetz & Kipfer, 2012).

In relation to translation, Peter Newmark suggested eleven procedures for translating neologisms: "transference (with inverted commas), TL [target language] neologism (with composites), TL derived word, naturalization, recognized TL translation, functional term, descriptive term, literal translation, translation procedure combinations (couplets etc.), through-translation, internationalism" (1988, p. 150). Newmark (1988) believes that, given the use of various methods, achieving a definitive translation of a neologism is both difficult and rare. This highlights the extent to which translating neologisms is a laborious process, requiring cognitive effort and meticulous attention from the translator to render the original into the target language (Gile & Lei, 2021).

The source-text context – that is, the overall framework of the text interpretation – plays a crucial, but not necessarily determining, role in understanding the total meaning of a neologism (Lehrer, 2003). Therefore, an examination of context is often the first attempt to make sense of a source-text neologism. Chen (2021) noticed that, in the absence of context when translating neologisms, the translator's cognitive effort increases.

Measurement of Effort

Within Translation Process Research (TPR) cognitive effort refers to the mental exertion associated with several key tasks: comprehending texts, actively thinking about translation strategies, correcting errors, and critically evaluating translation outcomes (Lacruz, 2017). Given the difficulty of directly observing and measuring cognitive processes, TPR researchers draw upon theories from neighboring disciplines, such as psycholinguistics and cognitive psychology, as gaining insights into the translation process has become a critical area within Translation Studies. Various methodologies and experimental equipment, including eye-tracking from psychology and keylogging from human-computer interaction, have been employed in TPR studies. Therefore, cognitive aspects such as effort (e.g., Dragsted, 2012), attention (e.g., Hvelplund, 2011), emotion (e.g., Hubscher-Davidson, 2018), and metaphors (e.g., Massey, 2016) have been explored.

Hans Krings (2001) developed a framework for analyzing translators' cognitive effort involving three categories: temporal, cognitive, and technical, as illustrated in Figure 1. The temporal category measures the total time required to complete a translation task, including reading, comprehending, and translating. The cognitive category includes mental processes such as problem-solving and decision-making related to strategies and error correction. The technical category covers physical aspects, including typing actions (e.g., insertion, deletion) and mouse operations (e.g., text highlighting). Initially designed to examine post-editing (PE) of machine-translated outputs, Krings' framework has also been used to explore human translation (see, e.g., Koglin, 2015).

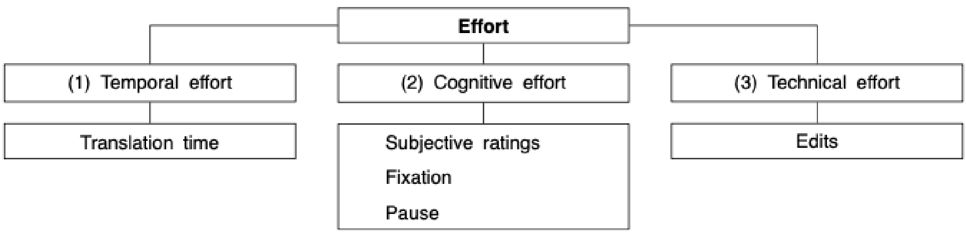

Figure 1. Krings Framework of Effort

In TPR, a variety of eye-movement and user-activity measures are employed to understand cognitive processing. The eye-movement measures are fixation-based (e.g., fixation duration), pupil-based (e.g., pupil dilation), saccade-based (e.g., saccade length), and transition-based (e.g., transition frequency). Fixation measures, particularly fixation duration and count, are the most frequently used in TPR studies. Pupil-size measurements are also employed by many researchers because pupil dilation reflects cognitive load (Hvelplund, 2014).

The user-activity measures identify pause events. Pause duration and count are used to calculate pause ratio, average pause ratio, and pause-to-word ratio. Another important measure is processing time—the time spent completing a translation task (Lacruz, 2017). According to research, there is a correlation between source-text difficulty and pauses during translation (Koponen, 2016).

III. Methodology

This section outlines the research methods used to investigate whether neologisms are recognized as unfamiliar terms or perceived as standard words during translation, based on the cognitive effort involved. It describes the participant selection and demographics, the instruments for data collection, the texts used in the experiment, the cognitive effort metrics, and the procedures followed.

A. Participants

Six participants were selected for the eye-tracking and keystroke logging experiments. They were all undergraduate translation students at the same level of study and from the same college, sharing a common linguistic background: Arabic as their native and target language (L1), and English as their second and source language (L2). The participants were male, right-handed, and aged between 21 and 23 years ($M = 22.17$, $SD = 0.50$). Four had normal vision, and two wore corrective lenses. None reported neurological or psychological impairments or medication use. This small homogeneous sample was intentionally chosen to reduce individual variability in cognitive and linguistic background, thereby allowing for a more focused investigation of the effect of neologisms on cognitive effort. While the sample is not intended to represent a broader population, it addresses the exploratory aim of the study: to test whether neologisms are recognized as distinct from standard words during translation, as indicated by fixation and pause durations. Small-scale studies like this remain valuable in Translation Process Research (TPR), particularly for exploring cognitive phenomena and refining data elicitation methods (O'Brien, 2009). However, generalizability is necessarily limited, and future studies should include larger and more diverse participant samples to validate and extend the findings reported here (Balling & Hvelplund, 2015).

Recruiting participants for the study involved several challenges. Although convenience sampling was used, participation depended on volunteer willingness – a common limitation in social science research. Despite offering a token of appreciation, only 7 out of 14 invited students agreed to participate, and one was excluded due to an eye-tracking calibration failure. Such logistical issues are well-documented in TPR studies (Korpal, 2015). Participants were fully briefed on the study's purpose and procedures, gave informed consent, and were assured of data confidentiality and the right to withdraw at any time. Each participant received an Apple Store gift card.

B. Instruments

Two instruments were used to collect data: an eye-tracking device and keystroke logging software. The former captures eye position and movement, while the latter records keyboard activity. In TPR, both tools offer empirical data, such as fixation and pause durations, that provide insights into cognitive processes during translation.

The Gazepoint GP3 HD Desktop Eye-Tracker was used to collect participants' gaze data. As a binocular system with reliable performance (Cuve et al., 2022), it operates at 60–150 Hz and offers a viewing angle resolution between 0.5° and 1°. It captures eye movements at 6.7-millisecond intervals. While higher sampling rates (e.g., 250 Hz and above) are preferred, 150 Hz is suitable for fixation-based studies (Cummins, 2017). However, it may introduce errors in fixation categorization, as Gazepoint does not distinguish between blinks and missing data. This limits certain quality checks, such as calculating gaze sample-to-fixation percentage (GSF). Furthermore, Gazepoint Analysis UX Edition was used to extract fixation data. The software includes tools for visualizing and analyzing eye-tracking data, including heat maps and fixation paths, and supports the assignment of areas of interest (AOIs).

Translog-II was used to record keyboard activity during translation tasks (Carl, 2012). Designed for TPR studies, the software has two components: Supervisor and User. The Supervisor is used by the investigator to manage experiments and extract log data, while the User generates log files of participant activity. Each file includes time-stamped keystrokes from which pause durations are extracted. Although Translog-II can integrate with eye-trackers like Tobii and Eyelink 1000, it does not natively support Gazepoint (CRITT, 2024). To address this, the 'catch screen' function in Gazepoint Analysis UX Edition was used to enable simultaneous recording, a workaround suggested by Cui et al. (2023).

C. Metrics

Two key metrics were considered in this study, along with their associated thresholds. First, total fixation duration (FD) is the eye-tracking metric used to measure cognitive effort. FD refers to "the cumulative amount of time that a specific word was gazed during the translating process" (Rydning & Lachaud, 2010, p. 94). As a fixation-based metric, it positively correlates with cognitive effort (see, e.g., Saldanha & O'Brien, 2014). A threshold of 200 milliseconds was applied, which falls within the average fixation range of 100–250 milliseconds and is consistent with common practice in eye-tracking research (Alves et al., 2009). This threshold also fits within the broader 100–500 millisecond range observed across readers (Rayner, 1998). While some TPR studies have used lower thresholds (e.g., 100 milliseconds; Pavlovic & Jensen, 2009), this was considered inappropriate for the present participants—university-level translation students—who may require longer reading times.

Second, total pause duration (PD) is the keylogging metric employed for measuring cognitive effort. PD refers to the length of time a participant pauses during translation. As a metric, PD has been extensively applied in TPR due to the correlation between source text difficulty and pauses during translation (Koponen, 2016). It provides insights into the translator's cognitive effort (Cui et al., 2023), especially when the source text contains difficult components such as neologisms. A threshold of 1000 milliseconds was applied to identify significant pauses during the experiments, which is within the range commonly used in TPR studies (see, e.g., Jakobsen, 2017).

D. Materials

The neologisms used in the experiment (bikelash, gramping, and Zoombombing) were taken from a list published by Grant Pearson Brown Consulting Ltd (GPB, 2023), a London-based communication advisory firm. GPB has been collecting and publishing English neologisms in cooperation with London's *Evening Standard* and *The Daily Telegraph* since 2014. The GPB's list includes over 200 neologisms with their meanings, sources, and first-appearance dates as observed by GPB.

According to GPB, bikelash is a merger of 'bike' and 'backlash' and first appeared in February 2021 in Cycle Magazine. It refers to the public response to the temporary bicycle lanes established during the COVID-19 pandemic. The term, however, seems to have been coined and popularized before the pandemic, possibly around 2014 during New York's bike lane rollout. The neologism gramping is a blend of 'grandparents' and 'camping' and simply refers to going camping with one's grandparents. A Google News search reveals that this term appeared in a *Daily Mail* headline in 2012 (Mail, 2012). The neologism Zoombombing is composed of Zoom (the digital platform for online meetings) and the word 'bombing'. It is believed to have been coined during the COVID-19 pandemic when Zoom became a popular platform for virtual meetings but initially suffered from low security, allowing uninvited people to join and disrupt meetings.

Next, three source-text examples – each containing one neologism from the above – were inspired by online news portals and modified to manage length and context.

I. Politicians recognize bikelash. In cities worldwide, shifting roads from cars to cycle use upsets some residents.
II. Families ski with grandparents this season. Gramping is becoming a British tradition.
III. Zoom will pay $85 million due to privacy issues and Zoombombing trends online.

The decision to use three sentences and treat each as a separate recording was based on two methodological considerations: participant profile and the need to reduce data loss. As students, participants were more likely to stay engaged with shorter texts. Separating sentences also simplified the task and allowed for breaks (if needed). Finally, this setup minimized the risk of losing all data at once—any loss was limited to a single sentence (or recording).

E. Procedure

The experiment consisted of three parts: the instruction period, the lab training session, and the experimental tasks. During the instruction period, participants were informed about the purpose and procedures of the study, including potential risks and benefits. They were asked to sign an informed consent form and complete a demographic survey. This was followed by a 30-minute training session designed to familiarize participants with the keyboard and seating positions. The ideal distance from the eye-tracker (65 cm) was maintained, as recommended by the Gazepoint Control user manual (GP3, 2024). Participants then performed an initial eye calibration using the Gazepoint Control software. Before beginning the experimental tasks, the eye-tracker was activated, and the Gazepoint Control interface was prepared for calibration. Two applications were launched in the background: Gazepoint Analysis UX Edition to record gaze data, and Translog-II to log keystroke activity. The Translog-II interface included two horizontally divided windows: one displaying the source text and the other left blank for participants to type their translations.

During the experimental tasks, each participant performed an eye calibration and then began translating the source text from scratch, while both Gazepoint Analysis UX Edition and Translog-II recorded the session. After each task, the researcher saved the recorded files. Once the experimental phase was complete, eye-tracking and keylogging data were extracted and prepared for analysis. This included processing fixation and pause data, conducting a follow-up survey with participants, and performing statistical tests. The follow-up survey was not part of the original methodology but became necessary during result interpretation.

IV. DATA COLLECTION

A. Eye-Tracking Data

After completion of the experiment, a preliminary review of the recordings was conducted primarily as an initial quality indicator to confirm the presence and precision of saccades during the experiments. Using the software's built-in feature, Areas of Interest (AOIs) were assigned to each source-text word (e.g., W1, W2, W3). Consequently, the fixation duration for each word in each translation experiment became accessible to the investigator and ready for analysis.

Next, fixation data were extracted from the recordings using the eye-tracker software, including AOI-related data. The extraction produced three types of Excel files containing thousands of fixation samples, stored automatically in the same file as each participant's recording. These files included: 'Data Summary Report' (DSR), 'All Gaze Samples' (GS), and 'Fixation Samples' (FS). The DSR file provided an AOI summary and statistics, such as total and average time views in seconds and fixation counts. The GS file comprised all fixation samples, both short and long. The FS file contained fixation data specific to the screen.

Since the Gazepoint software has no option to modify the threshold, nor does its Analysis User Manual specify the threshold used for producing the DSR data, it was necessary to process the raw data in the GS and FS files to modify the threshold to 200 milliseconds, and to conduct other types of analysis, including data quality checks. Although determining the exact built-in threshold is beyond the scope of this study, early analysis of GS and FS datasets suggested that the software is extracting fixation data at a 100-millisecond threshold.

(a). Mean Fixation Duration

To distinguish between acceptable and unacceptable fixation data, the mean fixation duration (MFD) was adopted as a quality measure. It was calculated by comparing the total fixation duration of the AOI to its fixation counts at a 200-millisecond threshold (Hvelplund, 2011). The MFD analysis for the first recording (R1) shows that 92.71% of the fixation data are acceptable, making it suitable for generating meaningful insights related to the hypothesis. The seven instances where an AOI mean fixation fell below the 200-millisecond threshold are as follows: W9, W13, W14, and W15 for Participant 5; W10 for Participant 6; W15 for Participant 2; and W16 for Participant 3. For the second recording (R2), 80.95% of the fixation data are acceptable. The segments with below-threshold fixations include: W10 and W14 for Participant 1; W12 for Participant 2; W5 and W14 for Participant 3; W5, W10, and W14 for Participants 4 and 5; and W10 and W14 for Participant 6. The third recording (R3) shows the highest rate of acceptable data, with 94.64%. Only three data points fell below the threshold: W7 for Participant 3 and W11 for Participants 5 and 6. Therefore, these instances where the mean fixation values fell below the threshold were dismissed from further analysis. It is important to note, however, that none of these dismissed data points corresponded to neologisms.

(b). Total Fixation Duration

After the data quality check, total fixation duration for each area of interest (AOI) was calculated using a 200-millisecond threshold, as shown in Table 1. The first recording (R1) included 89 fixation samples; the second (R2), 68; and the third (R3), 81—totaling 238 fixation duration sample per word (FDW).

TABLE 1
TOTAL FIXATION DURATION (IN SECONDS)

Recording	AOI	P2	P3	P4	P5	P6	P7
R1	W1	31.525	5.622	37.633	19.223	6.538	3.616
R1	W2	23.839	6.653	16.219	9.553	8.866	3.495
R1	W3	25.044	13.899	10.786	5.486	8.367	3.892
R1	W4	6.894	0.916	12.769	0.442	1.741	1.769
R1	W5	3.999	3.831	3.335	1.870	0.951	1.192
R1	W6	7.233	4.642	1.922	3.926	5.122	2.170
R1	W7	7.778	2.633	7.850	7.590	7.523	0.683
R1	W8	3.061	2.044	4.568	6.532	4.212	0.985
R1	W9	0.328	2.860	0.208	1.521	-	-
R1	W10	10.875	0.389	1.146	2.553	1.600	0.442
R1	W11	2.585	0.857	1.333	0.536	0.450	0.576
R1	W12	5.934	3.048	5.186	1.493	0.649	0.656
R1	W13	4.306	0.857	0.267	6.624	-	2.136
R1	W14	6.746	2.787	3.235	9.369	-	1.755
R1	W15	2.613	-	0.215	10.534	-	0.348
R1	W16	3.476	1.795	-	4.105	1.225	0.281
R2	W1	0.208	3.047	24.291	1.308	1.272	7.121
R2	W2	4.292	1.004	4.937	1.585	0.818	1.541
R2	W3	1.674	2.659	10.889	0.214	0.576	3.160
R2	W4	6.564	10.432	2.772	0.657	0.756	2.792
R2	W5	1.849	1.105	-	-	-	0.976
R2	W6	3.255	3.378	1.132	0.938	0.275	1.909
R2	W7	5.139	6.745	15.607	36.931	3.308	2.526
R2	W8	1.152	3.216	6.813	0.744	0.208	0.228
R2	W9	2.812	3.068	7.414	2.364	1.239	-
R2	W10	-	0.235	0.904	-	-	-
R2	W11	3.564	2.834	1.574	3.148	-	0.234
R2	W12	3.704	-	0.448	1.286	0.422	3.256
R2	W13	-	7.994	1.332	3.316	-	0.449
R2	W14	2.070	11.221	-	-	-	-
R3	W1	0.268	0.254	24.041	1.218	0.804	2.036
R3	W2	1.266	0.970	13.354	1.146	3.736	3.967
R3	W3	1.868	2.137	7.595	1.293	1.829	2.163
R3	W4	2.016	7.299	19.954	1.365	3.070	0.596
R3	W5	4.078	5.853	10.809	1.970	3.570	1.762
R3	W6	2.196	0.985	0.261	2.437	3.791	1.193
R3	W7	1.628	1.318	-	0.898	1.467	2.230
R3	W8	0.810	1.855	0.221	2.960	3.797	2.758
R3	W9	2.324	18.174	4.140	1.346	2.256	1.815
R3	W10	3.416	4.475	6.925	3.717	1.622	1.380
R3	W11	0.261	3.203	1.287	0.469	-	-
R3	W12	5.806	5.414	12.111	3.463	2.028	22.712
R3	W13	5.476	2.123	1.627	4.482	0.536	1.494
R3	W14	1.474	3.741	1.975	3.980	0.241	0.576

B. Keystroke Data

As previously mentioned, keystroke data were extracted using Translog-II, which generated an XML file for each participant in each recording. These files included session details (e.g., date and time), text information (source and translated texts), event logs (text insertions, deletions, pauses, mouse actions), and timing data (e.g., event timestamps).

The keylogging data from the XML files was analyzed to identify the pause duration (PD) for each translated word (PDW), using a 1000-millisecond threshold. PDW is defined as the pause while typing a word, plus the pause immediately before the next word (PDB). In other words, it includes the pause at the word and the one between words (PDW + PDB). For example, the total pause for the second target word is calculated as: *PDW2 = PDW2 + PDB [W1, W2]*. Figure 2 illustrates this process.

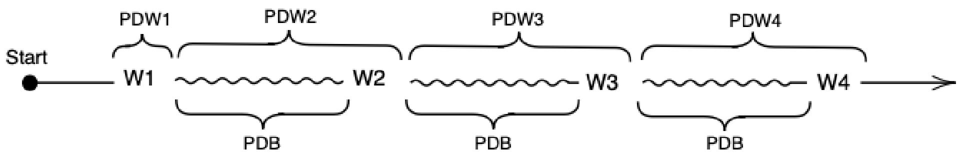

Figure 2. Total Pause Duration

Including the PDB as part of the PDW assumes that pauses between translated words reflect cognitive processing involved in creating the word. For example, a translator finishes typing a word (e.g., W2), then pauses—likely to read the source text or consider the next word—before typing W3.

The total number of pause duration (PD) samples from all three recordings is 177, as shown in Table 2. Of these, 21 were linked to the translation of neologisms. Since participants produced varying numbers of words, word counts differed across translations. In Table 2, an empty cell indicates where a translation ended. A symbol indicates the PD sample did not exceed the 1000-millisecond threshold and was excluded from analysis.

TABLE 2
TOTAL PAUSE DURATION (IN SECONDS)

Recording	AOI	P1	P2	P3	P4	P5	P6
R1	W1	56.327	2.688	2.516	16.625	-	-
R1	W2	33.829	2.531	2.578	27.484	15.891	2.062
R1	W3	20.719	1.703	93.766	1.078	1.015	1.11
R1	W4	1.281	2.296	-	2.828	4.516	-
R1	W5	-	11.234	-	24.749	-	1.093
R1	W6	2.249	3.329	14.907	-	16.969	-
R1	W7	13.703	12.39	-	-	1.125	-
R1	W8	1.734	1.125	3.702	-	2.985	-
R1	W9	1.094	1.078	1.594	1.297	-	-
R1	W10	-	1.36	-	1.172	-	-
R1	W11		-	11.921	-	5.906	1.515
R1	W12		4.922	2.39	21.156	3.469	-
R1	W13		2.484	13.329	-	-	-
R1	W14		9.953	11.781	-	9.735	-
R1	W15		-	18.47	2.641	2.141	6.609
R1	W16		10.032		-	-	-
R1	W17		-		3.75	-	1.203
R1	W18		5.125		-	3.016	-
R1	W19		-		3.844	-	
R1	W20				-		
R1	W21				5.203		
R2	W1	1.094	6.625	31.062	5.281	1.203	5.813
R2	W2	1	11.625	1.765	5.828	-	-
R2	W3	-	1.094	1.953	1.375	-	-
R2	W4	9.281	-	1.344	1.875	-	-
R2	W5	1.641	1.781	2.063	1.062	3.844	-
R2	W6	1.235	2.937	1.016	45.75	-	-
R2	W7	-	-	3.156	-	1.031	1.063
R2	W8	7.188	5.171	2.344	11.219	-	1.61
R2	W9	-	-	-	-	-	-
R2	W10	2.953	5.171	4.298	4.031	-	2.484
R2	W11	1.046	1.781	-	-	2.062	-
R2	W12	2.985	9.391	3.625	-	1.875	1.094
R2	W13		-	7.734	-	1.172	-
R2	W14		-	1.266	1.312	-	-
R2	W15			-	13.751	6.609	
R2	W16			1.531	12.048	2.313	
R2	W17			57.516	-	2.563	
R3	W1	1.203	1.063	2.11	1.453	-	-
R3	W2	1.625	-	1.188	-	-	2.062
R3	W3	10.249	1.125	1.781	-	1.047	-
R3	W4	-	16.188	1.047	-	-	-
R3	W5	11.391	3.312	6.469	1.297	-	-
R3	W6	-	6.813	-	2.062	1.984	1.046
R3	W7	-	2.624	5.016	16.171	-	1.235
R3	W8	-	1.454	2.563	2.953	-	-
R3	W9	10.656	-	32.766	-	1.828	9
R3	W10	-	6.031	-	-	-	-
R3	W11	-	5.031	1.844	6.344	4.406	2.547
R3	W12	-	3.453	-	-	-	1.281
R3	W13	2.531	-	-	-	-	-
R3	W14	1.062	2.625	8.344	-	-	-
R3	W15	-	1.203		-	-	-
R3	W16		1.422		2.531	-	2.453
R3	W17		3.907		4.922	1.875	
R3	W18		3.234		-		
R3	W19		18.734		-		

V. ANALYSIS AND RESULTS

Before the main statistical analysis, a preliminary descriptive comparison was carried out to assess the size of the differences between neologisms and non-neologisms. Specifically, Cohen's d values were calculated using the means and standard deviations of raw fixation and pause durations across the two groups. These values were not part of the inferential analysis but served to evaluate the practical magnitude of the differences prior to applying a Linear Mixed-Effects Model (LMM), which better suits the hierarchical structure of the data.

As shown in Table 3, the effect size for fixation duration (FD) was large (d = 1.22), suggesting a meaningful difference in cognitive effort between neologisms and non-neologisms. In contrast, the effect size for pause duration (PD) was small (d = 0.18), reflecting a minor difference. This pattern is consistent with the nature of the task: FD reflects immediate visual processing during reading, while PD may be influenced by later stages of processing, which vary more across individuals and trials. These descriptive results provided initial support for proceeding with model-based analysis.

After identifying the fixation duration and pause duration data from the recordings, an LMM analysis was conducted to determine whether neologisms required greater cognitive effort than standard words during the translation process. LMM was chosen to account for both participant-level variability and the nested structure of AOIs within participants. Unlike simpler linear regression models, LMMs incorporate random effects to account for individual differences while estimating the impact of fixed factors.

In the analysis, the fixed effect was defined as the role of the word (neologism vs. non-neologism), while random intercepts were included to account for variability in fixation duration (FD) and pause duration (PD) across participants. This model structure captured data complexity and provided robust estimates of the cognitive effort linked to neologisms.

TABLE 3
EFFECT SIZE RESULTS FOR FD AND PD

Metric	FD	PD
Mean neologisms	10.51	8.44
Mean non-neologisms	3.80	6.41
Std neologisms	9.39	20.32
Std non-neologisms	5.10	9.22
Cohen's d	1.22	0.18

Due to the small sample size (n = 6), the random-effects structure was limited to random intercepts to avoid overfitting. To enhance statistical power, data from all three recordings were combined into a single LMM test. This approach yielded a comprehensive result for the dataset, allowing for more reliable conclusions about the relationship between word type and cognitive effort.

To address the natural imbalance in the dataset and assess the stability of the findings, a subsampling approach was employed. For both eye-tracking and keylogging data, non-neologism AOIs were randomly sampled to match the number of neologisms, and this process was repeated 30 times. In each iteration, an LMM was fitted, and key statistics (estimates, standard errors, z-values, p-values, and confidence intervals) were recorded. These results were not treated as independent hypothesis tests; rather, the subsampling served as a robustness check to examine the consistency of the effect across different random subsets of balanced data. The outputs were then aggregated descriptively by calculating the mean, standard deviation, and range of each parameter to summarize variation across iterations. While this exploratory strategy does not correct for multiple comparisons, it was not intended to generate additional p-values, but to assess whether the observed effect remained stable under repeated resampling.

A. Eye-Tracking Result

The LMM analysis of the combined dataset (FD) revealed a significant effect of word type on cognitive effort, as measured by fixation duration. The intercept for non-neologisms was estimated at 3.75 (SE = 0.75, z = 5.01, $p < 0.001$, 95% CI [2.28, 5.21]), representing the baseline effort for familiar words. The effect of neologisms was estimated at 6.77 (SE = 1.31, z = 5.18, $p < 0.001$, 95% CI = [4.21, 9.33]), indicating a substantial increase in fixation duration and cognitive effort. The variance attributed to random intercepts (differences between participants) was 2.58, reflecting variability in baseline fixation times.

To confirm the robustness of these effects, a subsampling analysis was conducted across 30 iterations, with each model fitted to a randomly balanced subset of the data (equal number of neologism and non-neologism AOIs). The intercept had a mean estimate of 3.69 (SD = 1.10), ranging from 1.74 to 6.23. The average standard error was 1.82 (range: 1.08–3.00), with z-values between 1.10 and 2.81, and p-values from 0.005 to 0.27. The neologism effect had a mean estimate of 6.82 (SD = 1.10), ranging from 4.28 to 8.78. The mean standard error was 2.50 (range: 1.91–3.50), with z-values between 1.27 and 4.37. P-values ranged from 1.23×10^{-5} to 0.203, mostly below 0.05.

These results demonstrate that the observed increase in cognitive effort for neologisms is both stable and statistically significant in most iterations, providing strong support for the reliability of the main effect.

B. Keylogging Result

The LMM analysis of the combined keylogging data (PD) revealed no significant effect of word type on cognitive effort. The intercept, representing the baseline pause duration for non-neologisms, was estimated at 6.21 (SE = 1.29, z = 4.82, $p < 0.001$, 95% CI = [3.68, 8.73]). This reflects the average pause duration for familiar words. The effect of neologisms was estimated at 2.50 (SE = 2.56, z = 0.97, p = 0.33, 95% CI = [–2.52, 7.52]), indicating a slightly longer pause, though not statistically significant. The variance attributed to random effects (participant-level differences) was estimated at 5.02, indicating substantial individual variability in pause duration.

To assess the robustness of these findings, a subsampling analysis was conducted using 30 balanced iterations. In each run, a random subset of non-neologisms was matched to the number of neologisms, and an LMM was fitted to the resulting dataset. The intercept across iterations had a mean estimate of 5.87 (SD = 2.16), ranging from 2.21 to 12.32. The mean standard error was 4.14 (SD = 0.74), with z-values from 0.60 to 2.03 and p-values between 0.042 and 0.55. The effect of neologisms had a mean estimate of 3.06 (SD = 2.01), ranging from –1.51 to 6.67. The mean standard error was 4.96 (SD = 0.50), with z-values from –0.32 to 1.57 and p-values from 0.12 to 0.99.

These results reveal high variability across iterations, with most models failing to detect a statistically significant effect of neologisms. This suggests that, unlike fixation duration, pause duration may be a less stable or sensitive indicator of cognitive effort in this context. The inconsistency may reflect participant-specific behaviors or task-related complexity not fully captured by the model. Overall, these findings suggest that pause duration may be an unreliable indicator of cognitive effort, reflecting external factors such as motor behavior or typing habits more than cognitive load.

VI. Discussion

The fixation duration (FD) findings support the hypothesis set out at the beginning: that newly coined terms require greater cognitive effort to process (Lehrer, 2003). Participants spent more time fixating on neologisms than on standard words, mirroring the recognition of neologisms during translation.

In contrast, the pause duration (PD) findings showed no significant difference between neologisms and standard words, suggesting that participants did not spend extra time translating neologisms. Therefore, one might infer that neologisms were recognised (as evidenced by the FD results) and comprehended during the reading of the source text, so no additional pauses were required during translation. Although assessing translation quality is beyond the scope of this study, it is unlikely that comprehending and translating neologisms was entirely straightforward—particularly for university-level students still in training and non-native speakers of the source language.

Taken together, these findings suggest that pause duration may not reliably reflect cognitive effort in this context, in contrast to fixation duration, which yielded stable and significant effects. While fixation duration appears to capture early processing effort during reading, pause duration seems more susceptible to interference from other factors during the production phase.

As explained earlier, PDW and PDB were combined due to the way pauses were captured and processed from Translog-II's XML files. While it would have been helpful to pinpoint PDB occurrences more accurately using integrated keylogging and eye-tracking in this study, future research may benefit from a clearer distinction between PDW and PDB, supported by such integration. Additionally, the analysis showed that the high pause-to-word ratio (1.29) within words is atypical, potentially indicating non-cognitive factors related to typing behaviour (Kumpulainen, 2015). For instance, participants' unfamiliarity with physical keyboards or limited typing experience may have contributed to these interruptions, rather than increased cognitive effort associated with translating neologisms. It is reasonable to assume that some typists—particularly among the younger generation, including college students—are more accustomed to touchscreen keyboards than to traditional ones. Typing tasks such as emails and assignments are now more frequently completed on virtual keyboards (e.g., on smartphones or tablets). Thus, in eye-tracking and keylogging experiments where physical keyboards are required, participants of this kind may pause frequently both within and between words.

To better understand the unexpectedly high number of within-word pauses (PDW), a short follow-up survey was conducted after data analysis. This survey targeted participants from the experiment and aimed to explore their typing habits and familiarity with physical keyboards—factors that may have influenced the keylogging results. According to the survey, 40% of participants reported feeling uncomfortable typing on a physical keyboard, while the remaining 60% were split among "very comfortable," "comfortable," and "neutral." In contrast, 80% reported feeling comfortable or very comfortable typing on a smartphone, suggesting a preference for virtual keyboards. When asked about their primary device for academic work, 60% selected a smartphone and 40% selected a computer. Notably, 80% indicated that they need to look at the keyboard "always" or "often" while typing, which may explain the high incidence of within-word pauses. Although this survey was not part of the initial experimental design, its findings offer valuable context for interpreting the keylogging data and underscore the importance of accounting for typing habits in future TPR studies.

Another factor that may have influenced the PDW results is the nature of the translated texts produced by participants. Unlike the fixed ratio of neologisms and familiar words in the source text used for FD analysis, participants' translations varied—particularly in their treatment of neologisms. The translated texts had to be reviewed to confirm whether neologisms were retained. In some cases, participants omitted neologisms entirely; in others, it was unclear whether a given word corresponded. Although evaluating translation quality lies beyond the scope of this paper, these checks were necessary to ensure accurate interpretation of the PDW data.

In light of these methodological and practical factors, the reliability of PDW findings should be approached with caution. Challenges in attributing pauses, variations in neologism translation, and limitations in data collection all underscore the need for careful analysis of PDW data. Although various pause types have been explored in TPR studies, PDW has not been widely examined. For example, Immonen and Mäkisalo (2017) briefly touch on word-medial pauses—occurring between keystrokes within a word—but focus instead on pauses at higher linguistic levels, such as between words, phrases, and clauses. Similarly, Rosa et al. (2018) identify within-word pauses as common but offer limited insight into their cognitive implications. These studies highlight a gap in our understanding of within-word pauses and their role in keylogging analysis. Taken together, these findings suggest that pause duration may be an unreliable indicator of cognitive effort in this context, likely reflecting external factors such as motor behaviour or typing habits rather than cognitive load.

VII. Conclusion

This study investigated whether neologisms were recognized as unfamiliar terms or processed similarly to standard words by measuring the cognitive effort involved in translating three texts, each containing one neologism. As expected, and as shown by the eye-tracking results, neologisms elicited greater cognitive effort compared to standard words, indicating that participants did recognize them as unfamiliar during translation. However, the keylogging data showed no significant difference, suggesting that participants did not spend additional time translating neologisms. This discrepancy may stem from methodological factors, such as the merging of pause types (PDB and PDW) and participants' unfamiliarity with physical keyboards, both of which influenced the keylogging outcomes. These findings underscore the importance of refining data collection and analysis methods—particularly by separating pause types and accounting for non-cognitive influences—to better capture cognitive effort in translation. Additionally, the study highlights the potential impact of growing reliance on virtual keyboards, which may affect both eye-tracking and keylogging data and should be explored further in future TPR research.

Acknowledgements

This research received grant no. (307/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

References

[1] Al-Jarf, R. (2010). Translation students' difficulties with English neologisms. *Analele Universităţii "Dunărea De Jos" Din Galaţi Fascicula*, *24*(2), 431–437.
[2] Alves, F., Pagano, A., & da Silva, I. (2009). A new window on translators' cognitive activity. In I. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, Technology and Innovation in Translation Process Research* (pp. 267–291). Samfundslitteratur.
[3] Balling, L. W., & Hvelplund, K. T. (2015). Design and statistics in quantitative translation (process) research. *Translation Spaces*, *4*(1), 169–184.
[4] Beason, L. (2010). *Eyes Before Ease*. McGraw Hill Professional.
[5] Brand, J., Diamond, S. G., Thomas, N., & Gilbert-Diamond, D. (2021). Evaluating the data quality of the Gazepoint GP3 low-cost eye tracker when used independently by study participants. *Behavior Research Methods*, *53*, 1502–1514.
[6] Cabré, M. T. (1999). *Terminology: Theory, methods, and applications* (Vol. 1). John Benjamins Publishing.
[7] Carl, M. (2012). Translog-II: A program for recording user activity data for empirical reading and writing research. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 4108–4112). European Language Resources Association (ELRA). Istanbul, Turkey.
[8] Carl, M. (2012). Translog-II: A program for recording user activity data for empirical translation process research. *International Journal of Computational Linguistics and Applications*, *3*(1), 153–162.
[9] Chen, J. (2021). Translating Chinese neologisms without knowledge of context: An exploratory analysis of an eye-tracking and key-logging experiment. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (Vol. 3, pp. 315–337). Springer.
[10] CRITT. (2024). *Translog-II*. Retrieved December 9, 2024, from https://sites.google.com/site/centretranslationinnovation/translog-ii
[11] Cui, Y., Liu, X., & Cheng, Y. (2023). A comparative study on the effort of human translation and post-editing in relation to text types: An eye-tracking and key-logging experiment. *SAGE Open*, *13*(1), 1–15.
[12] Cummins, R. G. (2017). Eye tracking. In *The International Encyclopedia of Communication Research Methods* (pp. 1–9). Texas Tech University. https://doi.org/10.1002/9781118901731.iecrm0099
[13] Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*, *54*(2), 1027–1049.
[14] Daily Mail. (2012). *Grandparents take children on holiday as busy parents take a break in new 'gramping' trend*. Retrieved October 18, 2024, from https://www.dailymail.co.uk/news/article-2191369/Grandparents-children-holiday-busy-parents-break-new-gramping-trend.html
[15] Dragsted, B. (2012). Indicators of difficulty in translation—correlating product and process data. *Across Languages and Cultures*, *13*(1), 81–98.
[16] El Mustapha Lahlali. (2014). *Advanced English-Arabic Translation: A Practical Guide*. Edinburgh University Press.
[17] Gile, D., & Lei, V. (2021). Translation, Effort and Cognition. In R. Muñoz, C. Alvstad, A. Wadensjö, & A. Assis Rosa (Eds.), *The Routledge Handbook of Translation and Cognition* (pp. 263–278). Routledge.
[18] GP3. (2024). *GP3 Eye Tracker Placement and Positioning*. https://gazept.com/dl/GP3_Placement_and_Positioning.pdf
[19] GPB. (2023). *Neologisms and New Words*. Retrieved May 15, 2023, from, https://www.gpb.eu/2023/05/neologisms-new-words.html

[20] Hanaqtah, M. F. (2019). Translation of Political Neologisms Coined by Politicians; Issues and Strategies. *Journal of Social Sciences (COES&RJ-JSS)*, *8*(1), 157–168.
[21] Hope, W., & Johnson, R. (2004). What is an Infomercial? *Advertising & Society Review*, *5*(2). https://doi.org/10.1353/asr.2004.0006
[22] Hubert, M. D. (2016). Teaching translation to foreign-language majors. In *Teaching Translation* (pp. 55–62). Routledge.
[23] Hubscher-Davidson, S. (2018). *Translation and Emotion: A Psychological Perspective*. Routledge.
[24] Hvelplund, K. T. (2011). *Allocation of Cognitive Resources in Translation: An Eye-Tracking and Key-Logging Study*. Copenhagen Business School.
[25] 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*, *Special Issue 1*, 201–223. https://doi.org/10.6035/monti.v0i0.292854
[26] Immonen, S., & Mäkisalo, J. (2017). Pauses reflecting the processing of syntactic units in monolingual text production and translation. *HERMES–Journal of Language and Communication in Business*, *44*, 45–61.
[27] Jakobsen, A. L. (2017). Translation Process Research. In *The Handbook of Translation and Cognition* (pp. 19–49).
[28] Kalugina, O. A., Vasbieva, D. G., Valeeva, R. A., Maystrovich, E. V., & Pashanova, O. V. (2019). Peculiarities of using creative neologisms in publicistic texts. *XLinguae*, *12*(4), 119–129.
[29] Koglin, A. (2015). An empirical investigation of cognitive effort required to postedit machine translated metaphors compared to the translation of metaphors. *Translation & Interpreting*, *7*, 126–141.
[30] Koponen, M. (2016). *Machine translation post-editing and effort: Empirical studies on the post-editing process* [Unpublished PhD dissertation]. University of Helsinki, Faculty of Arts.
[31] Korpal, P. (2015). Eye-tracking in Translation and Interpreting Studies: The growing popularity and methodological problems. In Ł. Bogucki & M. Deckert (Eds.), *Accessing Audiovisual Translation* (pp. 199–212). Peter Lang.
[32] Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes* (G. S. Koby, Ed.). Kent State University Press.
[33] Kumpulainen, M. (2015). On the operationalisation of 'pauses' in translation process research. *Translation & Interpreting: The International Journal of Translation and Interpreting Research*, *7*(1), 47–58.
[34] Lacruz, I. (2017). Cognitive effort in translation, editing, and post-editing. In *The Handbook of Translation and Cognition* (pp. 386–401). Wiley-Blackwell.
[35] Lehrer, A. (2003). Understanding trendy neologisms. *Italian Journal of Linguistics*, *15*, 369–382.
[36] Llopart-Saumell, E. (2023). "Learn the rules like a pro, so you can break them like an artist": On the emotional effects of breaking word-formation rules. *Current Perspectives in Spanish Lexical Development*, *68*, 189–214.
[37] Massey, G. (2016). Remapping Meaning: Exploring the Products and Processes of Translating Conceptual Metaphor. In Ł. Bogucki, B. Lewandowska-Tomaszczyk, & M. Thelen (Eds.), *Translation and Meaning* (Vol. 2, pp. 67–83). Peter Lang.
[38] Mellinger, C., & Hanson, T. (2016). *Quantitative Research Methods in Translation and Interpreting Studies*. Routledge.
[39] Newmark, P. (1988). *A Textbook of Translation*. Prentice-Hall.
[40] O'Brien, S. (2009). Eye tracking in translation process research: Methodological challenges and solutions. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Methodology, Technology and Innovation in Translation Process Research* (Vol. 38, pp. 251–266). Samfundslitteratur.
[41] Pavlovic, N., & Hvelplund Jensen, K. T. (2009). Eye tracking translation directionality. In A. Pym & A. Perekrestenko (Eds.), *Translation Research Projects 2* (pp. 93–109). Intercultural Studies Group.
[42] Płużyczka, M. (2020). Tracking mental processes in sight translation: Neurobiological determinants of selected eye-tracking parameters. *Translation, Cognition & Behavior*, *3*(2), 209–232.
[43] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.
[44] Rogers, M. (2015). *Specialised Translation: Shedding the 'non-literary' tag*. Springer.
[45] Rosa, R. N., Sinar, T. S., Ibrahim-Bell, Z., & Setia, E. (2018). Pauses by student and professional translators in translation process. *International Journal of Comparative Literature and Translation Studies*, *6*(1), 18–28.
[46] Rydning, A. F., & Lachaud, C. M. (2010). The Reformulation Challenge in Translation. In *Translation and Cognition* (pp. 85–108).
[47] Saldanha, G., & O'Brien, S. (2014). *Research Methodologies in Translation Studies*. Routledge.
[48] Schmied, J., & Ivanova, M. (2023). *Comparing Confidence and Trust Online and Offline*. Cuvillier Verlag.
[49] Steinmetz, S., & Kipfer, B. A. (2012). *The Life of Language: The Fascinating Ways Words are Born, Live & Die*. Diversified Publishing.
[50] Thomsen, O. N. (Ed.). (2006). *Competing models of linguistic change: Evolution and beyond* (Vol. 279). John Benjamins.

Majdi Alzahrani is an Assistant Professor in the Department of English, College of Science and Arts, King Abdulaziz University, Rabigh, Saudi Arabia. He earned his PhD in Translation Studies from the University of Leeds in 2022. His scholarly interests lie in Literary Translation and Translation Process Research.