SKASE Journal of Translation and Interpretation, 2025; 18(2): 4–37
doi: 10.33542/JTI2025-S-2

Directionality in translation: Throwing new light on an old question

Zoë Miljanović[*], Fabio Alves[§], Celina Brost[*], and Stella Neumann[*]
[*]RWTH Aachen University [§]Universidade Federal de Minas Gerais (UFMG)

Abstract
Investigating directionality in translation from an empirical-experimental perspective has long occupied translation studies scholars. Most empirical studies about translating from L1 into L2 or from L2 into L1 use the time spent on the task as an indicator of processing effort. Whyatt (2019) is a notable exception which investigates not only time spent on the task but also the impact of text type on translation directionality. However, a combined focus on translation directionality and editing patterns observed over the course of the translation process as an indicator of processing effort has been rarely analysed (Alves & Gonçalves 2013). Against this background, two research questions guide the rationale of this experimental study: (1) Is there a difference in translation task execution when translating from L2 into L1 than when translating from L1 into L2?, (2) Does it differ in terms of text register, editing patterns, pause length or time spent on the task? To answer these questions, we investigated L1>L2 and L2>L1 translation task execution in the language pair English/German by means of keystroke-logged data. The results show that translations into the L1 in our data set involve a higher proportion of editing procedures as opposed to unchallenged translations than translations into the L2. Furthermore, results for inter-word pauses are register-specific. These findings suggest that the effect of directionality on translation behaviour is complex and requires a research design that accounts for this complexity.

Keywords: *translation process research; directionality in translation; editing patterns; processing effort; register effect*

1. Introduction

Investigating directionality in translation from an empirical-experimental perspective has long occupied translation studies scholars. While many empirical studies about translating from the first language (henceforth L1) into a second language (henceforth L2) or from an L2 into the L1 use the time spent on the task as an indicator of processing effort (Pavlović & Hvelplund Jensen 2009; Ferreira et al. 2016; Schmaltz et al. 2016), we propose that a combined focus on time-based *and* text editing patterns observed in the course of the translation process may provide a more robust account of the impact of directionality on task execution. Even though translation into the L2 is neither uncommon in many speech communities nor necessarily ‘worse’ than translation into the L1 (Pokorn 2005; Pavlović 2013), it is not the preferred translation direction in many cultures including Germany, where translation programmes focus primarily on teaching L2>L1 translation. Translation competence in both directions is, therefore, not necessarily comparable. That is why there is a need to understand effects of translation direction on the translation process more comprehensively. This paper offers a new perspective on the issue by answering the following primary research question: (1) Is there a difference in translation students’ task execution when translating from an L2 into the L1 than when translating from the L1 into an L2? To specify this guiding question, we further ask:

(2) Does it differ in terms of text register, editing patterns, pause length and/or time spent on the task?

Research on the effect of translation direction has gained popularity in recent years. Yet the terminology is still not entirely standardised within the field. Directionality is typically used to denote whether a translation task demands translating into the first language, or into a second language and this is also how it will be used in this study. This already displays two of the main approaches to this issue. Translations into the second language are often called "inverse" or "reverse", which, however, implies some sort of judgment on what would be the "right" direction, namely *direct* translation into the first language. To avoid assumptions yet to be proven empirically we will adopt the more neutral terminology also employed by others referring to the L1 as the first, or most frequently used language, and L2 for the second, or lesser used language (cf. Pavlović 2007; Stewart 1999). Note that Evert & Neumann (2017) refer to a different type of directionality in their investigation of asymmetries between the languages involved in translation in their role as source or target language. What they investigate is independent of translating into the L1 or L2 and might be called *language pair directionality* to distinguish it from the L1-L2 directionality we are interested in here, which is in turn independent of relationships between specific languages.

In the remainder of the paper, we will first review the literature and then describe our innovative remote elicitation procedure and data processing in the method section. The subsequent discussion and conclusion of our results offer potential explanations for our findings while remaining critical of their overall implications for the question of translation directionality and processing behaviour.

2. Literature review

Empirical research on the effect of translation direction is rather scarce and presents a somewhat divided field of observations, especially regarding the translation process (Whyatt 2019: 80). Findings also diverge as far as the translation product is concerned. While some studies do not report on any textual differences regarding directionality (Rodríguez-Inés 2013), other studies have found directionality to have an effect on translation quality in grammar, punctuation or even readability of the target text in both directions (Whyatt 2019: 89). Furthermore, Whyatt also reports a modulating effect of text type on said directionality effect, which suggests a significant influence of the source text register on translation depending on the direction. Specific typographical differences corresponding to translation direction have further been reported by Rodríguez-Inés (2017: 257), who found a general tendency of translators to use more brackets to add information or provide equivalent terms in translations into the L2 than into the L1.

Differences in syntactic and lexical complexity in L2 translations from French into English have been found by Penha-Marion et al. (2024). They report a tendency toward lower lexical density due to a reduced vocabulary range with more high-frequency words and increased syntactic complexity by means of longer sentences in L1>L2 translations than in L2>L1 translations and in original English texts. Penha-Marion et al. (2024) attribute this to the effect of translation experience, or rather the lack thereof, as their participants were translation students. An influence of the source texts as well as idiosyncratic translation behaviour was also observed.

More generally, many studies elicit data from translation students or inexperienced translators (Pavlović 2007, 2010, 2017; Cao et al. 2023), and it has been shown that the translators' L2 proficiency and general linguistic knowledge might have more of an impact on translation quality than the direction of the translation task (Elston-Güttler et al. 2005; Pokorn et al. 2019). This suggests that translator training should include second language competence training, but also that studies on translation directionality need to be aware of the individual capabilities of their participants.

These findings concerning the effect of language proficiency as well as the prominence of specific features depending on the translation direction have informed the discussion about the supposed difference in difficulty and cognitive demand of translating into the L1 or into the L2. Difficulty in either translation direction is suggested to be linked to time spent on the task or the cognitive effort needed to complete it. For this, translators' behavioural patterns during the translation task are frequently consulted. This often involves the use of eye-tracking software and/or keystroke logging tools in combination with self-reports after or even during the translation task, to enable a more comprehensive understanding of the cognitive processes involved in translation (Alves et al. 2009).

Previous studies investigating said difference in cognitive effort respective of the translation direction have produced inconclusive results. Investigations of gaze and pause behaviour during translating show that translations into the L2 demand more cognitive effort, hence corroborating the so-called *translation asymmetry* (Pavlović & Hvelplund Jensen 2009; Chang & Chen 2023; Jia et al. 2023). This effect also holds true when language specific features are investigated (Jensen et al. 2009). Furthermore, the total time spent on the translation task has been shown to be indicative of the cognitive effort applied. Yet even though participants sometimes spend perceivably more time on translations into their L2, this difference was not proven to be statistically significant (Ferreira et al. 2021: 115). In other studies, no difference in terms of time spent is found at all (Whyatt 2019: 88). Fonseca (2015: 123) further claims that the supposed higher cognitive demand might rather be the result of strategic translation behaviour than a sign of increased difficulty. Ferreira et al. (2018), in contrast to Whyatt (2019), find higher metalinguistic self-monitoring in L2>L1 translation with more explicit critical appraisal of the choices made in the participants' L1. These two studies also contrast in terms of the difficulties in L1>L2 translation, which may be of lexical (Ferreira et al. 2018) or grammatical nature (Whyatt 2019). This indicates a lack of conclusive evidence concerning cognitive effort in translation depending on L1-L2 directionality, as multiple factors are at play (cf. Pavlović & Hvelplund Jensen 2009).

Pavlović (2007: 13) also shows that only 44% of translators indicate L1>L2 translation to be more difficult, while 33% claim it to be the other way around, although these preferences could very likely be due to differences in training and proficiency (Pietryga 2022: 105). These findings suggest that language proficiency, individual differences and especially the source text intricacy might impact the perceived difficulty and general outcome of the translation task even more than the directionality of the translation, in terms of both process and product (Ferreira et al. 2016: 61; Ferreira et al. 2018: 111–112). Concerning revision behaviour as well, individual preferences between translators seem to have a larger impact than L1-L2 directionality (Alves & Vale 2011/2017; Alves & Gonçalves 2013), although this remains to be confirmed by studies with larger numbers of participants.

As shown above, translation research has employed many different methods over the years. The most promising and meaningful approach appears to be a multi-methods research design that combines the investigation of the translation process and product. There have been a few attempts to take on this task on a larger scale through compiling a translation process corpus. The project TransComp (Göpferich et al. 2008) conducted a longitudinal study that investigated the development of translation competence of translation students in comparison to professional translators. By using the keystroke logger Translog, and a screen recording tool, they provide insight into the translation process as well as the translation product of their comparably small corpus. A similar approach has been taken by a team at the Zürich University of Applied Sciences with the Capturing Translation Processes (CTP) project (Ehrensberger-Dow 2013: 7). They collected translation process data of translation students and professional translators at different points in their careers, employing a multi-methods research design. Their data shows that especially students employ different strategies when translating into the L2 like, for example, longer contemplation times with fewer revisions.

Keystroke logging the translation process involves analysing latencies between the specific computer activities such as all kinds of keystrokes, mouse clicks etc. and segmentation patterns, which have been used as operationalisations of cognitive processing effort. According to Alves & Vale (2011/2017), editing patterns can be mapped onto micro and macro translation units (TUs). Micro TUs correspond to translational text production segments delimited by pauses of a pre-defined length. In this, the concept is comparable to the concept of the production unit by Carl et al. (2016). However, we are interested in the editing operations on these units (see below). Note that Alves & Vale's micro TU is not equivalent to Carl's (2021) notion, which he also calls micro unit, although he defines it differently.

Once a translator revisits an initial micro TU and edits it, it will become part of a macro unit, which thus consists of all micro TUs (still separated by pauses) that involve edits of the initial one and make changes to some of its wording (Alves & Vale 2009). This includes all interim changes such as deletions, insertions and revisions that lead to the final translation solution. Alves & Vale (2011/2017) differentiate between different types of macro units depending on whether text production and edits occur in the drafting and/or end revision phase. A P1 is produced and possibly changed across multiple micro units during the drafting phase only, while P2 exhibits straightforward text production during the drafting phase and is then only edited during the end revision phase. The last type of macro unit, P3, consists of multiple micro units in the drafting and at least one editing pattern in the end revision phase, therefore having undergone multiple editing procedures across translation phases. To those macro units, Alves & Gonçalves (2013: 113) add a P0 category which equals a micro unit that does not undergo any editing procedures whatsoever. They also annotate the distance between the micro units that together make up one macro unit. Table 1 summarises the different editing patterns characterising translation units proposed by Alves & Gonçalves (2013).

Table 1: Overview of the editing procedures described by Alves & Gonçalves (2013)

Procedure	**Characterisation**
P0	continuous text production only, no edits
P1	edits of the produced unit during drafting phase only
P2	edits of the produced unit only during the end revision phase
P3	edits of the produced unit during drafting and end revision phase

Against the background of the conflicting findings of previous studies, for example in relation to time spent on the translation task or different foci of cognitive effort (Ferreira et al. 2021; Whyatt 2019), we tentatively formulate the following expectations. Since we are working with translation students, and language proficiency was shown to be a significant factor, participants might need time to look up more words when translating into the L2 than into the L1, resulting in longer pauses in this direction. Furthermore, a difference in editing behaviour might become visible, as translation into the L1 has been shown to exhibit more complex editing procedures. Moreover, the strong influence of register on linguistic choice has been widely reported generally in linguistics. In particular, Whyatt (2019) also identifies an interaction between register and directionality.

3. Method

3.1. Experimental design

We designed an experiment in which translation students (L1 speakers of German) translated four texts, two from English into their L1 German and two from German into English, their L2. To account for the expected influence of register on directionality we included texts from the registers of popular scientific writing and product review respectively. The source texts thus cover both translation directions as well as two different registers. In order to delimit the amount of translation units to be included in the subsequent manual analysis (see below), we identified cohesive chains in the four texts. Throughout the text, different items refer back to an antecedent, thus establishing a cohesive chain. Drawing on Halliday & Hasan (1976), they consist of elements linked through lexical cohesion or cohesive reference. These items naturally occur in any text and therefore require little manipulation. Their analysis in the translations grants insight into the participants' translation and editing strategies. While they work through the source text and encounter more items belonging to the cohesive chain, the field of sense relations unfolds and reveals a clearer meaning of each individual item. This means that the translators might be prompted to go back in their target text production to revise choices they initially made or to adapt their subsequent translation choices to produce a more cohesive target text. We identified one cohesive chain in each text, starting from the antecedent representing the text's topical focus in the title or, in the case of the German product review, in the first line. The chains consist of sense related items, such as repetitions, synonyms, hyponyms or meronyms, as well as lexically related items belonging to the same semantic space and pronominal reference items. These pronouns are co-referential with the antecedent as well as with the other cohesive items in the chain.

The texts (see Appendix A) were edited such that they were comparable in length and in number of the different cohesive items and proofread by native speakers of English and German. Moreover, the experimental design (see below) was tested in two pilot studies with students from RWTH Aachen University. The second pilot study was necessary because of issues with the first set of source texts. Table 2 provides a comparative characterisation of the final set of source texts including the distribution of cohesive items in the source texts. In order to access translational behaviour, we analysed linear representations of the unfolding target texts produced in Translog-II (Carl 2012) rather than the source or target text products (see below for further detail on preprocessing Translog's log files).

Table 2: Overview of length and cohesive items of the source texts

	DE_POP	EN_POP	DE_REV	EN_REV
Words	237	253	242	241
Characters	1,804	1,651	1,629	1,233
Lexical cohesion	16	9	15	18
Reference	3	1	5	3
Total number of cohesive items	19	10	20	21

Previous experimental studies of directionality in translation used small cohorts of participants with process data elicited on-site in a laboratory. To allow for a larger cohort of participants and consequently more statistical power, we collected data remotely. The experimenter met the participant via video call using the video chat environment Zoom (Version 5.17.7) and gave the participant remote access to their computer on which the translation task was displayed and recorded in the keystroke logging environment Translog-II. The fact that keystroke logging was recorded remotely is likely to have an unavoidable impact on the accuracy of the recordings. In the worst case, accuracy varies between participants due to differences in the hardware and/or internet speed. Keller et al. (2009) report a replication study for self-paced reading, which also involves keystroke logging. Their web-based replication of a lab experiment yielded a close match between the original and the web-based results. While these findings cannot simply be applied to our setting involving a different software, it at least gives us some confidence in the validity of the recordings. Needless to say that future work will have to include a replication study for Translog-II similar to the study reported by Keller et al. (2009).

During the experiment, cameras and microphones were switched off on both ends of the call. All participants were given an individual ID to pseudonymise the elicited data and ensure confidentiality. Prior to the experiment, the participants gave their informed consent to participating voluntarily in the study, and they were debriefed afterwards. Participants received 45 Euros as remuneration.

On the screen next to the Translog window, there was an internet browser with four open tabs, including two bilingual online dictionaries (https://www.dict.cc/ and https://www.leo.org/) as well as two monolingual thesauri (https://www.merriam-webster.com/ for English and https://www.dwds.de/ for German), which the participants could switch between. The participants were asked to only use these resources and neither other resources on their own devices nor print dictionaries. In some cases, we encountered Zoom or internet-related issues, but they could all be resolved rather quickly and were noted in the experiment protocols. Some participants were unable to use the umlauts (ä, ö, ü) on their keyboards for the German translations, which could not always be fixed during the experiment but was noted in the protocols as well.

After they completed each translation, participants notified the experimenter, who saved the recording and opened the next source text to be translated in Translog. This experimental setup ensures sufficient experimental control while increasing ecological validity in comparison to a laboratory setting by allowing the participants to work at their own pace in a familiar physical environment. Nevertheless, awareness of being recorded as well as the constraint imposed on access to resources clearly make this an experimental setting with consequences on the participants' behaviour. Participants first performed a copy task and then translated all four texts. The order in which the texts were presented was randomised. After completing all translation tasks, participants were asked to fill in a survey adapted from the Translation and Interpreting Competence Questionnaire (Schaeffer et al. 2020). As part of the survey, we asked participants to take the English and German LexTALE tests (Lemhöfer & Broersma 2012) and enter their results.

A total of 38 students (29 female, 8 male, 1 diverse; ages 19–30, SD 2.736534; 18 undergraduate, 20 graduate students) enrolled in translation degree programmes at four German universities[1] who had successfully completed at least two translation modules participated in the study. They were required to have German as (one of) their L1(s) and English as their first L2. The mean scores for the German and English LexTALE tests were 85.62% (SD 8.808537) and 81.46% (SD 10.76115) respectively.

3.2. Mapping translation units (annotation)

The resulting 152 keystroke log files (38 participants x 4 texts) were subsequently processed. In a first step, the XML files generated by Translog-II were automatically segmented into micro units using individual pause thresholds and represented as text files with the help of a Python script. The thresholds were determined individually per participant based on the median pause duration between pressing alphanumeric keys (as these are more likely to be used when typing words). To this median, we added two standard deviations of this duration in order to capture 95 percent of the variance in inter-key pause duration (see Martín & Apfelthaler (2022) for a similar approach using two times the median value as an arbitrary threshold to identify segments). In this way, we obtained an individual measure of a participant's typical typing speed that captures individual sensorimotor specifics (e.g. whether someone is a touch typist or not) and helps mitigate the influence of any challenges posed by the technical infrastructure. Any pause duration longer than this thus represents an extreme value that cannot be explained by sensorimotor or technical considerations and is likely to be caused by processing-related factors.[2] In addition to symbols representing mouse clicks or keystroke operations such as backspacing or deletion, the linear representations produced by our Python script represent each micro unit in one line, thus tracking the sequential order of text production. They also include an indication of the micro unit's position in the translation product: Each line starts with the cursor position in the unfolding target text at the current stage. The pause length dividing the current from the next micro unit is given in milliseconds at the end of each line.

These pre-segmented files representing the processes of target text production were uploaded to the annotation platform INCEpTION (Klie et al. 2018), where seven students enrolled in the MA programme Cognitive, Digital and Empirical English Studies at RWTH Aachen University annotated the files for course credits under the supervision of the first author on the basis of detailed annotation guidelines. The students were first trained in the annotation and practised it in a trial round, after which potential issues were addressed and the annotation guidelines refined. The student annotators were made familiar with the source texts and their cohesive chains and additionally had access to Translog in order to be able to replay the

---

[1] We gratefully acknowledge support from Oliver Czulo (University of Leipzig), Silvia Hansen-Schirra (University of Mainz/Germersheim), Kerstin Kunz (Heidelberg University) and Ekaterina Lapshinova-Koltunski (University of Hildesheim) in finding participants.
[2] We are indebted to Florian Frenken for this reasoning and for writing the Python scripts required for the automatic segmentation of the Translog-II logfiles and for exporting the INCEpTION annotation to a data table for further analysis.

translation process. All annotations were performed on the linear representations of the target texts, which were not aligned with the source texts.

The annotation proceeded in three steps. First, translations of the cohesive items in the source texts were identified and labelled in the target texts, including 1) the cohesive relation to their antecedent and 2) the type of text production operation involved (that is, full or partial production or deletion, or a mixed category for both production and deletion operations within one micro unit). Furthermore, any subsequent revision, such as an addition to or editing of these items, during the course of translation was also categorised. In the case where an item belonging to a cohesive chain was edited but that edit did not establish a (different) cohesive relation to the target antecedent, we added an escape label for the category 'cohesive relation', as it would not apply. Figure 1 shows a linear representation of the target text in INCEpTION annotated for cohesive items in step 1.

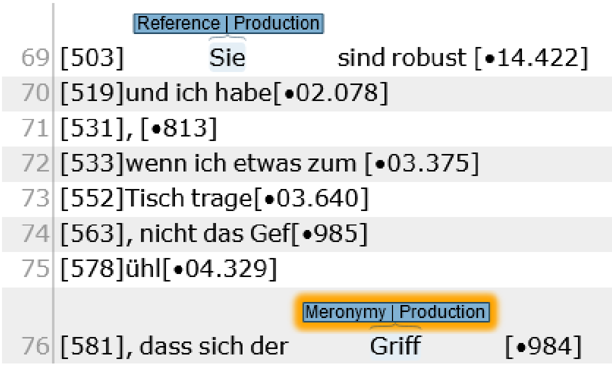


Figure 1: Example of an annotation of the linear representation in INCEpTION, including the pronominal reference *sie* ('it') and the meronym *Griff* ('handle') referring to the antecedent *Pfanne* ('pan')

Next, each production segment containing a cohesive item identified in step 1 and any subsequent micro units where each such item was edited were linked and labelled as macro units. This involves a first annotation of the segment containing the item as a micro unit representing (in the case of P0) or belonging to a macro unit as well as a link of that segment to its related micro units, which together form the macro unit (P1-3). These comprise the initial production of the segment containing a cohesive item and any subsequent segments including an editing operation on that item, so that all micro units belonging to one macro unit are connected. The links between the micro units are labelled with a running number. The number of the link between the last two micro units of the macro unit is thus also the number of times the item was picked up again and worked on during the translation process. Figure 2 shows an example of a macro unit consisting of three micro units, in which the item initially produced in segment number 164 was immediately deleted in the following segment number 165 and produced again later in the translation process in segment 170. This example represents a P1 type of translation unit, in which the item was produced and edited only during the drafting phase. The numbers attached to the links between the micro units count the number of revisions of the initially produced item. In this case, the initial item *extension* was deleted in the first linked revision and produced again in the second linked revision. The number in square brackets at the start of each line represents the cursor position, and the number in square brackets at the end of each line is the pause length in ms.

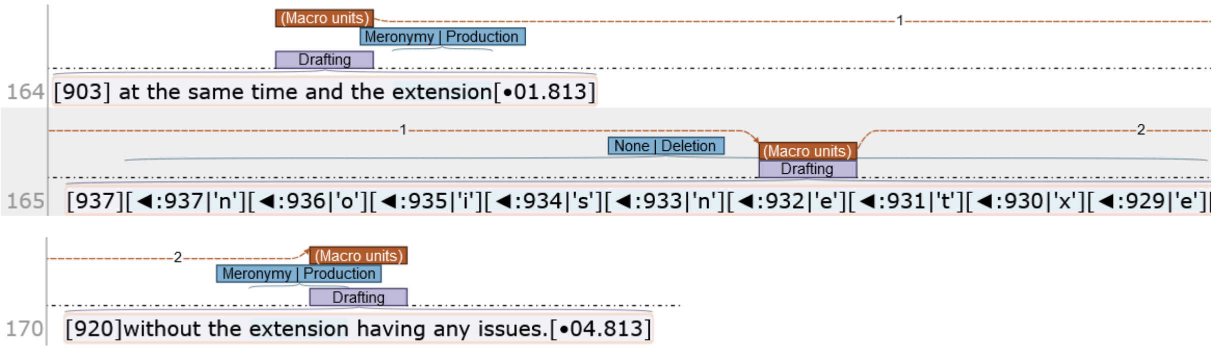


Figure 2: Example of a meronym produced and edited across several micro units in the drafting phase

In the third and last step, the translation phases orientation, drafting and end revision (Jakobsen 2002) were determined with another span layer in INCEpTION. The orientation phase describes the phase before any translating in terms of actual text production takes place. Possible keystrokes in that phase may for example include noting down vocabulary in the Translog editor during source text reading in preparation of translating. Drafting includes all text production from the first key press to produce the target text until a first version of the entire source text has been transferred into the target language. After that final full stop, the end revision phase begins (Jakobsen 2002: 91–92).

After each step, the students' annotations were checked and corrected first by the students themselves and finally by the first author before proceeding to the next step of the annotation. These annotations were exported from the INCEpTION platform in the form of XML files, which were then processed to result in a data table containing each annotated item in a micro unit, its position in the text approximated by the cursor position at the beginning of the processing segment, its position in the translation process as the number of the line of the segment in the linear representation, and the annotated information, that is the cohesive relation of the item in the segment, the text production operation, and the translation phase the segment occurs in. Based on the macro units established in the second step of the annotation and the translation phase, the macro units were determined and classified into one of the editing procedures presented in Table 1 (Alves & Vale 2011/2017; Alves & Gonçalves 2013). The data table furthermore contains the register, translation direction, source text and participant ID as metadata information.

On the basis of the data thus obtained, we analysed two indicators claimed to be linked to directionality: inter-word pause length as a time-related indicator of processing effort (potentially linked to lexical retrieval) and the distribution of editing procedures across micro and macro units as an indicator of the strain of text production.

4. Results

4.1. Pause length

As to pause length, we focus on all inter-word pauses as these are more likely to be linked to lexical retrieval. These are extracted automatically as latencies from a non-alphanumeric key such as white space or a punctuation mark (representing the end of one word) to the next alphanumeric key (representing the start of the next word). This information was automatically extracted for the entire data set and can therefore be analysed in more statistical detail, whereas the examination of editing procedures required manual annotation and will therefore only be characterised with the help of descriptive statistics (see below).

We fitted a mixed linear regression model to analyse the relationship between translation direction and inter-word pause length. Table 3 gives an overview of the variables included in the analysis. In addition to our response variable total inter-word pause length ("interword_pauses_sec_c") and our main predictor of interest, translation direction ("direction_sum"), we included register as a predictor, as this has been shown in multiple studies to be an important factor influencing linguistic choice ("register_sum"). Since Whyatt (2019) also discussed an interaction with directionality, we include an interaction term between these two variables in our model.

We assumed above that proficiency in the L2 could influence the process. We therefore included the score of the English LexTale test as the variable "LexTale_EN_c". Pause behaviour could be influenced by the participants' typing abilities. To account for this, we include a measure of typing speed during the copy task as the individual baseline ("keystroke_time_copy_sec_c"). The last predictor is "no_activity_sec_c": The length of pauses during translating could be moderated by whether the participants took time to familiarise themselves with the source text at the beginning of the task and reviewed the product for a last time before ending it. This information is captured by the variable "no_activity_sec_c".

As each participant contributes four data points, these are not independent. Each person might introduce idiosyncratic patterns across tasks. We therefore include a random intercept for "participantID". Note that we do not include a random effect for source text, although all participants translated the same four texts, as the levels map entirely on the combined fixed effects of translation direction and register.

Table 3: Overview of the variables included in the regression model

Type	**Variable**	**Scaling**	**Levels**	**Description**
Response	interword_pauses_sec_c	Continuous	-	The overall time without computer activity between words (centred)
Predictor	direction_sum	Binary	L1>L2, L2>L1	Direction of translation into the L2 or the L1 (sum-coded)
	register_sum	Binary	REV, POP	The two registers from which the source texts were taken (sum-coded)
	LexTale_EN_c	Continuous	-	The score of the English LexTale test (centred)
	keystroke_time_copy_sec_c	Continuous	-	The number of characters typed per second in the copy task (centred)
	no_activity_sec_c	Continuous	-	Time spent on the task before the first and after the last computer activity (centred)
Random effect	participantID	Categorical	[participant ID]	The unique identifier of each participant

Table 4 provides descriptive statistics of the pause-related variables for all 152 data points. The variable "interword_pauses_sec_c" is computed by subtracting the total word-internal pause lengths (used in the calculation of typing speed, see above) from the sum of all pause lengths per text in seconds. As can be seen in Table 4, the median sum of inter-word pause lengths per text is 1,935.1 seconds (the median overall task time is 2,333.2 seconds). The measure of typing speed in the copy task per participant was obtained by dividing the length of the copy text in characters by the time participants needed for the task (given in seconds). This gives us the average number of letters a participant types per second. Recordings of the copy task were missing for four participants. The median number of letters typed per second across the remaining 34 participants is 2.757.[3] The variable "no_activity_sec_c" is the result of subtracting the time elapsed between the first and the last computer activity from the overall time spent on the task. One extreme data point (1521.3 seconds of no activity) was more than double the length of the second-longest value. We removed this data point, as there may have been an unrelated interruption in task completion.[4] The median time without activity is 85.125 seconds (SD 103.6124), that is, roughly 4 percent of the overall task time. Descriptive statistics for "no_activity_sec_c" are given with and without the extreme data point in Table 4. We took the decision to continue without this extreme data point (leaving us with 151 data points for the regression analysis).

Table 4: Descriptive summary of pause-related variables

	tasktime_ sec	**interword_ pause_sec**[5]	**LexTale _EN_c**	**keystroke _time_ copy_sec**[6]	**no_activity _sec**	**no_activity_ sec**[7]
n	152	152	152	34	152	151
Min.	952.9	733.6	54.00	1.491	9.563	9.563
1st qu.	1,875.1	1513.2	76.56	2.114	36.937	36.828
Median	2,333.2	1935.1	82.50	2.757	85.805	85.125
Mean	2,426.7	2014.3	81.46	2.700	124.225	114.973
3rd qu.	2,954.2	2486.5	88.75	2.932	162.969	159.899
Max.	5,037.6	4192.0	96.25	4.079	1521.297	624.734
SD	826.0031	737.1506	10.76115	0.6718379	153.8698	103.6124

A regression model in R (R Core Team 2024) with all above predictor variables retrieved no significant effects of L2 proficiency and typing speed in the copy task on pause behaviour (see Appendix B). We therefore ran a second model without the variables "LexTale_EN_c" and "keystroke_time_copy_sec_c", which allowed us to include data from

[3] As an alternative perspective on typing speed, we also calculated the time required to type a key, for ease of understanding given in ms (after removing the NAs for participants without a copy task): Min. = 245.1, 1st qu. = 341.1, Median = 362.7, Mean = 395.6, 3rd qu. = 473.1, Max. = 670.6, SD = 106.9926

[4] The survey filled in by participants after completion of the four translation tasks also included questions about interruptions, but the participant's respective answers don't provide definite information.

[5] For comparison, word-internal pauses per second: Min. = 134.5, 1st qu. = 220.6, Median = 281.6, Mean = 288.1, 3rd qu. = 340.1, Max. = 541.4, SD = 78.3733

[6] After removing the NAs for participants without a copy task

[7] After removing one influential data point (textID "BACHMCBDAL07_A")

all 38 participants. The regression model yielded the outcome summarised in Table 5. Recall that data for one text was removed because of an extreme value for "no_activity_sec_c".

Table 5: Summary of the mixed linear regression model

	Estimate Std.	**Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	18.7922	63.7631	37.9403	0.295	0.76982	
direction_sum1	95.5683	32.6844	113.1476	2.924	0.00418	**
register_sum1	360.4499	33.7880	117.5891	10.668	< 2e-16	***
no_activity_sec_c	2.0692	0.4546	143.5747	4.552	1.13e-05	***
direction_sum1:register_sum1	-183.2756	33.3175	117.1153	-5.501	2.25e-07	***
Random effect for participant (151 observations, 38 groups): Variance = 113667, SD 337.1						

Model diagnostics reveal that the residuals are somewhat right-skewed, meaning that the error is not entirely normally distributed. Moreover, there is a weak tendency for residuals to fan out when moving along the fitted values, suggesting that variance of the residuals is not entirely homogeneous. There is no indication of collinearity, as the variance inflation factors for all predictors are all below 2 and thus well below the conservative threshold suggested by Winter (2019: 114). Model comparison was carried out with the *afex* package (Singmann et al. 2018) for nested mixed models using likelihood ratio tests. It yielded significant effects for all three predictors and the interaction, showing that the model with all factors captures the variance best (see Appendix B for the complete model output).

All predictors are significantly associated with the sum of inter-word pauses per text, and this effect is complex, characterised by a strong interaction effect between translation direction and register. In both registers, when translating into the L1, overall inter-word pause length is predicted to decrease by 191.1 seconds[8] compared to the opposite translation direction. In both directions, when the register is popular-scientific, the difference in inter-word pause length to reviews is 720.9 seconds. However, the interaction effect reveals that the directionality effect is actually register-specific: When participants translate reviews into the L1, the overall inter-word pause length is predicted to be 912.0 seconds shorter than when translating a popular-scientific text into the L2 and even 1,087.4 seconds shorter than popular-scientific translation into the L1. Consequently, L2>L1 translation is characterised by shorter inter-word pause lengths only in the register of reviews. In the popular-scientific register, translating into the L2 actually appears to have a facilitating effect: A participant is predicted to require roughly 175 seconds less in inter-word pauses than in the other translation direction. Last but not least, each additional second without a computer activity before or after text production turns out to increase the overall inter-word pause length by 2.07 seconds.

Including a random intercept for participants reveals a great deal of variation between participants accounted for by the model: Variation for 95% of participants (+/- two standard deviations of the random effect) around the intercept (i.e. sum of inter-word pauses per text) ranges roughly from -655 to +693 seconds.

[8] Recall that the regression model uses sum coding so that the estimates in Table 5 for both binary predictors only specify half of the actual difference between the two levels.

The conditional R^2 of 0.70 suggests a very good model fit, where the conditional R^2 captures the variance described by both fixed and random effects. The marginal R^2 for the fixed effects only is 0.49, suggesting that accounting for the effect of participants substantially adds to the described variance. As R^2 is actually a measure of effect size measuring the strength of the relationship between variables (Winter 2019: 77), this suggests a medium and, when accounting for the participants' effect, even strong effect.

4.2. Editing procedures

The annotation of the above-mentioned production segments for editing procedures yielded the characterisation of the four texts summarised in Table 6. It should be noted that the overall number of translation units involving the cohesive items identified in the STs diverges strikingly across texts. Translations into English of the German popular-scientific text yield a total of 936 relevant macro units, whereas translations of the English popular-scientific text only yield less than half that amount (435 macro units). This latter result is not surprising as this is also the source text with the lowest number of cohesive devices (see Table 2), which will naturally reflect in the distributions in the translations. The German original review results in 852 relevant macro units in the translations and the English review yields slightly more with 873 relevant macro units. This skewed distribution means that only the proportions of the different editing procedures per text can be compared at all and even these need to be interpreted very cautiously. A comprehensive analysis and discussion of editing procedures would require annotation of *all* macro units in the INCEpTION interface, which is outside the scope of this paper.

Table 6: Proportions of editing procedures in segments of interest across translation directions (by register) in per cent

	L1 > L2			**L2 > L1**		
	Popsci	**Review**	**Combined**	**Popsci**	**Review**	**Combined**
P0	56.52	68.66	62.30	43.45	62.31	56.04
P1	40.92	29.69	35.57	53.79	33.33	40.14
P2	1.39	1.17	1.29	0.46	2.63	1.91
P3	1.18	0.47	0.84	2.30	1.72	1.91
Macro units involving edits (= P1+P2+P3)	43.48	31.34	37.70	56.55	37.69	43.96

Table 6 reveals a difference between the translation directions with translation into the L1 requiring overall more edits per translation unit, with overall 43.96% of the translations into the L1 requiring edits and only 37.70% of translations into the L2. The table also shows that this difference is carried mainly by the review register. Around two thirds of the translation units are unchallenged productions (P0) in both directions, whereas the popular-scientific texts force the participants to edit more. In L2>L1 translation, less than half of the translation units examined here remain unchallenged. Whether the difference between the two types of texts can be interpreted as a register effect or is actually simply a reflection of the specific linguistic characteristics of the four source texts cannot be reliably decided.

5. Discussion

This study set out to investigate two related research questions, namely (1) whether there is a difference in translation task execution depending on the translation direction and, more specifically, (2) whether it differs with respect to editing and temporal patterns as well as by register. The above results show that, indeed, there is an effect of directionality in terms of both editing and required time as measured by inter-word pauses, but this effect is complex and driven by the source text and its register. Overall pause length is affected by the translation direction, but even more so by the variable of register. Inter-word pause lengths, which we linked to issues of lexical retrieval above, are indeed shorter for the review text in L2>L1 translation than in the opposite direction. While we cannot entirely rule out other explanations, individual factors such as keyboard-related issues are not highly likely, as participants worked in their own working environment, and are additionally accounted for by the random effect in the mixed model. The non-specialised wording of the register means that participants do not have to contemplate the wording in the potentially more familiar translation direction. When translating into the L2, more pauses are required, and this effect does not depend on L2 proficiency. By contrast, when the register is popular-scientific writing, the overall task appears more effortful, requiring more pauses between words in both directions, and there is no advantage of translating into the L1 anymore. Apparently, making sense of the source text now requires so much more time that text production takes much longer into the L1. Following Whyatt's (2019) reasoning, it may actually be more important for the participants to understand the source text well. So, the source text being in the L1 actually has a facilitating effect in this register. As a result, slightly less time is required for contemplating translation in the L2 than in the opposite direction. It should be kept in mind, though, that this study is based on only one source text per register and translation direction. The effect observed here could be entirely due to the specific challenges of the four texts. As Penha-Marion et al. (2024) report, a small set of source texts can be quite influential. This means that future work should involve a much-increased data set with a sufficiently high number of source texts that mitigates the influence of individual texts on the overall results.

We can tentatively interpret the significant effect of time without computer activity on inter-word pause length, both increasing together, as suggesting that those participants who take more time familiarising themselves with the source text and re-reading their target text may also take longer in general in decision-making (compare Lehka-Paul (2020) for a different observation and discussion of various behavioural preferences of translators). At the same time, the high variance between our participants as revealed by the random intercepts in the regression model indicates larger individual differences in their translation behaviour. This concerns not only their overall translation strategy, that is, how much time they spend outside of actual text production and whether they revise at all, but also individual differences in the time spent on translation and production speed. The participant metadata also indicates a wide range in language proficiency based on their LexTALE results for both German and English.

Concerning the editing strategies, P0 is overall the most frequent category. This observation based on the descriptive statistics (see Table 6) is even more pronounced in the direction L1>L2 with straightforward, continuous production of the micro unit. The overall prevalence of P0 corresponds to the findings reported by Alves & Gonçalves (2013: 115–16), who link P0 to routinised translation behaviour. While this result may suggest competent decision-making, participants may on the other hand have found fewer options in the less familiar language. P1 and P2 strategies, which represent more complex macro units involving editing of the produced item either in the drafting or in the end revision phase, are comparably more frequent in the translation direction L2>L1. The participants may have been more critical of their target texts and may draw on a wider range of possible translation solutions in their L1. At the same time, this result may reflect that the antecedents in the English source texts allow more variation in how they are subsequently referred to in the text. These are *pans* and *bacteria*. The former, for example, is ambiguous, in that it could mean both or either *Töpfe* or *Pfannen* in German. This ambiguity is frequently not recognised from the start but only throughout the text, which led participants to change their first translation solution, sometimes multiple times. The antecedent in the English popular-scientific source text, *bacteria*, also grants access to various synonyms in German. The antecedents and references to them in the German source texts seem to be more specific in comparison, though this might also reflect a lower range of vocabulary or less familiarity with diversifying lexis in the participants' L2.

The reviews seem to require less editing or production of cohesive items spread across several micro units in translation, as reflected by the higher proportion of the P0 type of text production pattern in this register. In contrast, an interrupted production frequently occurs in the case of unfamiliar and highly specific terminology in the popular-scientific texts, where participants may more frequently pause during the production of an item, resulting in a higher number of P1 patterns in that register. This kind of P1 pattern can be distinguished from P1 macro units that are characterised by text operations other than production, i.e., those that involve editing the produced item at a later stage in the drafting phase. It will be worthwhile to extend the categorisation and make this distinction in future analyses, because the two patterns reflect different behaviours under the same P1 label. The unfamiliar subject matter and lexis of the popular-scientific register seem to slow down the participants as well as trigger more editing. Some participants also seemed to perceive the more scientific or technical texts as more difficult and remarked upon this after the experiment. Some noted that they would have found background information or reference texts helpful or would have usually done research to produce a target text they are more confident in. On the other hand, at least one participant remarked that they worked with both of these registers, i.e., product reviews and popular-scientific texts, in their translation programme.

P2 and P3 editing strategies were overall very infrequent. Both of these patterns are defined by including editing in the end revision phase. This last phase during the translation process was frequently very short or even completely missing in many text productions, which might reflect the student participants' limited translation experience or their lack of a strategy involving editing of a rough draft (cf. Ehrensberger-Dow 2013).

That said, the proportions of the more complex editing procedures P1–P3 are higher in translation into the L1 than into the L2. However, it must be kept in mind that this study only uses one text each for each register and translation direction combination. This means that the results are affected by source-text specific features. Translating from the English popular-scientific source text, for example, more uncertainty with certain items regarding their spelling could be observed in the translation process, such as recursive variations of *Kolibakterien*. The German source text seems to allow less variation in its translation solutions in comparison, with a higher frequency of P0 overall. It can also be argued that, despite careful selection of the source text and pilot-testing (both texts describe certain chemical and physical processes in a somewhat narrative order), the English source text is more technical in nature. Though we do not consider the translation quality in this study, one sentence in that text (see Example (1) below) seemed to pose a particular challenge syntactically and was frequently misunderstood by the translation students.

(1) Then, to transform the muconic acid into adipic acid, they used a second type of E. coli, which produced hydrogen gas, and a palladium catalyst.

The majority of participants translated this sentence in such a way that the last nominal group *a palladium catalyst* was understood to be part of the non-defining relative clause. Their translations thus suggest that the E. coli bacteria also produced the palladium catalyst instead of that catalyst also being used to transform the acid. Expressing complex methods such as described in this sentence poses a particular challenge when translators are unfamiliar with the subject matter.

6. Conclusion

The results presented above, and the regression analysis in particular, throw new light on the question of directionality in translation by revealing the complex influence of register on translational behaviour across translation directions (bearing in mind that this influence is currently at least partly due to specific characteristics of the source texts). The results suggest that time-related measures shed some light on processing effort in the performance of translators working from L1 into L2 or from L2 into L1. Nevertheless, they also show that factors like register (cf. Whyatt 2019) and idiosyncratic behaviour (cf. Penha-Marion et al. 2024) have a significant effect on task execution. Considering the specific measures such as pause length and editing strategies reveals that participants produce translations of reviews in less time as reflected by shorter overall sums of inter-word pauses and require less revision with fewer complex macro units, when looking at specific cohesive items. This is especially the case when translating in the more familiar direction L2>L1. In the popular-scientific register, on the other hand, more time is taken overall and specifically in the translation direction L2>L1. In that register, the source text seems to pose more of a challenge, and participants may also spend more time on their target text with higher metalinguistic critical awareness of their target text in this more specialised register. Future studies must include a larger number of source texts to mitigate potential source text-specific effects. Collecting more translations from more participants will also help reduce the influence of individual participants' preferences and translation strategies, though a wide range of variation with regard to their behavioural measures can still be expected. More generally, extending the classification discussed here for only the editing procedures of elements in the identified cohesive chains in the source texts to *all* translation units in the translations will permit statistically analysing associations of editing procedures with the various predictor variables as demonstrated for inter-word pause length.

In addition to the mentioned limitations concerning the textual material, the experimental set-up also comes with certain constraints. The remote solution and lack of eye-tracking means that phases of inactivity on the participants' side cannot reliably be attributed to source or target text reading or understood to indicate cognitive effort. Though the processing of the translation process data was handled individually, technical issues such as delays during remote access or keyboard incompatibilities will nevertheless affect the results. The set-up allows some control, for example over the use of resources during the experiment, though the consultation of additional sources cannot be completely eliminated, as the participants worked 'behind closed cameras' for reasons of privacy and creating a more natural situation. Participants were mostly not accustomed to the working environment of Translog.

One participant remarked that they found it unfamiliar not to be able to edit the source text, which made orienting themself in the text more difficult. Two participants also commented on the order of the texts: This order was randomised but one participant perceived the two popular-scientific as particularly difficult as they followed the two easier reviews while another said the vocabulary from the English source texts was helpful when translating the German ones into English.

Since the participants in the experiment had little or no professional experience in translation, the next step will be to replicate the study using the same methodology with professional translators. Furthermore, the annotated data also allows an investigation of how the different translation units and cohesive relations differ in terms of their complexity as captured by the number of editing procedures performed as well as their complexity in terms of distance measures. A longer distance between the micro units within a macro unit potentially means that processing effort is higher for items that need to be recalled in comparison to items revised immediately or in close proximity to their initial production (Alves & Gonçalves 2013). Which editing strategy is triggered by the cohesive relation in the source text is also of interest: Depending on whether, for example, the item is a repetition or stands in a hyponymy relation to the antecedent, the translator may have more options to revise their solution and therefore edit the item more or less frequently. This also has to do with different stylistic preferences in what the translation students were taught in their programme. For example, some translator trainers – at least in a German context – might maintain the stylistic instruction to avoid repetition (arguably depending on the context). The translators may deviate from the source text choice in this case, whereas a hyponym, for example, is already a more specific choice made from the possible types of the antecedent, which is why a deviation might be less likely in this case unless there is linguistic ambiguity in the source or target language.

In sum, we hope to have shown that our experimental design involving remote data collection results in larger data sets. The use of mixed regression modelling of inter-word pause length involving multiple predictor variables as well as an interaction term sheds new light on the complex interplay of various factors affecting linguistic choice in the two translation directions. Future work that addresses the above points while using larger numbers of participants and considering a more complex set of variables demonstrated here can lead to a more comprehensive understanding of the intricacies of directionality in translation.

Acknowledgements

This study was supported by the German Research Foundation (DFG) project BiTPro, project no. 542780701. We would like to thank the CAPES Foundation for grant no. 88887.803531/2023-00 (Capes/PrInt), which funded Fabio Alves' sabbatical stay at RWTH Aachen University.

We are furthermore grateful to our source text proofreaders, pilot participants, student annotators and our student assistant Leonie Krah for their support and to the reviewers for their helpful remarks.

Appendix A: Source texts

EN_POP (from the online research news source *ScienceDaily*; https://www.sciencedaily.com/releases/2023/11/231101134747.htm)

Plastic-eating bacteria turn waste into useful starting materials for other products
Mountains of used plastic bottles get thrown away every day, but microbes could potentially tackle this problem. Researchers in ACS Central Science report that they have recently developed a plastic-eating E. coli that can efficiently turn polyethylene terephthalate (PET) waste into adipic acid, which is used to make nylon materials, drugs and fragrances. Previously, a team of researchers had engineered a strain of E. coli to transform the main component in old PET bottles, terephthalic acid, into something tastier and more valuable: the vanilla flavor compound vanillin. At the same time, other researchers had engineered microbes to metabolize terephthalic acid into a variety of small molecules, including short acids. A new team from the University of Edinburgh wanted to further expand E. coli's biosynthetic pathways to include the metabolism of terephthalic acid into adipic acid. The team developed a new E. coli strain that produced enzymes that could transform terephthalic acid into compounds such as muconic acid and adipic acid. Then, to transform the muconic acid into adipic acid, they used a second type of E. coli, which produced hydrogen gas, and a palladium catalyst. The team found that attaching the engineered microbial cells to alginate hydrogel beads improved their efficiency, and up to 79% of the terephthalic acid was converted into adipic acid. Using real-world samples of terephthalic acid from a discarded bottle and a coating taken from waste packaging labels, the newly engineered E. coli system efficiently produced adipic acid.

EN_REV (from the SFU Review Corpus; Taboada & Grieve 2004)

Pans that are worth buying
I have finally found a set of pans that are worth the money I paid for them. I thought I needed a new stove; instead a new set of pans was a better deal. I researched all of the different sets available and decided that these were the best fit for the money I had available. I absolutely love how fast the copper core bottom pans heat up. I can now boil water in half the time and at a much lower setting than before. They are sturdy and I don't feel like the handle is going to come loose when I carry something to the table. I also enjoy the fact that the handles are welded to the outside of the frying pan; this ensures that there are no rough surfaces on the interior for food to stick to. I also enjoy being able to cook at high temps without the sides of the skillet heating up. The heat source is concentrated at the base of the pan instead of in the handle and sides. This is safer and more efficient. The only disadvantage I have found so far is the clean-up. The instructions say the pans are dishwasher safe but it is not recommended due to possible cosmetic changes. My wife once used a metal spoon to stir and this left scratches on the surface. It did not affect the use, only the appearance.

DE_POP (from the online science news site *Wissenschaft-aktuell*; https://www.wissenschaft-aktuell.de/artikel/Mikroplastik_in_Wolken1771015591004.html)

Mikroplastik in Wolken
Mikroplastik hat viele Quellen von Fleece-Jacken über Kunststoffverpackungen und Reifenabrieb bis zu Kosmetika. Diese winzigen Partikel belasten die Umwelt auf dem gesamten Globus und konnten sogar schon in den Schneeproben der Arktis nachgewiesen werden. Nun

analysierte eine chinesische Forschergruppe ihren Anteil in Wolken. Tatsächlich fanden sie so beachtliche Mengen, dass ein Einfluss der Partikel auf die Wolkenbildung selbst nicht ausgeschlossen werden kann. Die Arbeitsgruppe an der Shandong University in Qingdao fing die Feuchtigkeit der Wolken über dem Berg Tài Shān mit einer speziellen Apparatur ein. Insgesamt enthielten 24 der 28 gesammelten Proben Mikroplastik. Jeder Liter des Kondensats enthielt im Durchschnitt 463 Plastikpartikel. Das entspricht einem Teilchen Mikroplastik auf etwa fünf Kubikmeter feuchter Luft. Die meisten Mikroplastikteilchen waren nicht größer als 100 Mikrometer und konnten dadurch leicht durch die Atmosphäre über weite Strecken transportiert werden. Sie bestanden aus einer Vielzahl verschiedener Kunststoffe wie Polypropylen, Polyethylen und Polyamid. Mit Wettermodellen rekonstruierten die Forschenden den Ursprung dieser Luftverschmutzung. So gelangte das meiste Mikroplastik aus dicht besiedelten Regionen im Süden Chinas bis zum Tài Shān. In weiteren Analysen entdeckten sie, dass viele Plastikpartikel durch Verwitterung eine raue Oberfläche hatten. An dieser Oberfläche konnten sich leichter weitere Umweltgifte wie Blei oder Quecksilber anlagern. Zudem sei es wahrscheinlich, dass Mikroplastik in der Atmosphäre als Kondensationskeime für Tropfen dienen, also direkt die Bildung von Wolken und Niederschlag unterstützen. Diese Funktion des Mikroplastiks könnte einen Einfluss auf Wettervorhersagen und sogar Klimamodelle haben.

DE_REV (from the USAGE review corpus; Klinger 2014)

Schönes Design mit durchdachter Funktionalität
Nachdem man den Toaster ausgepackt hat, fällt einem als Erstes sein Gehäusedesign angenehm ins Auge. Die schlanke Bauform und der verwendete Edelstahl verleihen ihm ein schönes und hochwertiges Aussehen. Was mir am Gehäuse am besten gefällt, ist, dass dieses wärmeisoliert ist. Wie ich finde ein wichtiges Kriterium für Haushalte mit Kindern. Der Toaster beherrscht neben dem Rösten auch das Aufwärmen und Auftauen von Toast beziehungsweise Brotscheiben. Einen gewissen Seltenheitscharakter hingegen hat die Möglichkeit, Toast einseitig zu rösten. Wozu ich das gebrauchen kann, weiß ich zwar noch nicht, aber die Funktion funktionierte beim Toastbrot sehr gut. Die Auswahl seiner Röstgrade erfolgt über einen Drehschalter an der Seite, und mit 7 Stufen ist die Auswahl zwar nicht außergewöhnlich, aber allemal ausreichend. Klasse finde ich, dass der Brötchenaufsatz im Toaster integriert ist und man somit keine weiteren Teile in der Küche verstauen muss, bis sie gebraucht werden. Bei Bedarf wird dieser einfach nach oben geklappt und schon kann es losgehen. Durch die längliche Form des Toasters kann man locker 3 Brötchen auftauen und der Aufsatz hatte damit auch keine Probleme. Was mich als Einziges an dem Gerät stört, ist der Piep-Ton nach Beendigung des Toastens. Wobei der Ton mich an sich nicht stört, sondern nur die Lautstärke. Für meinen Geschmack hätte man die Lautstärke verringern können. Fazit: Alles in allem summieren sich die einzelnen Ausstattungsmerkmale des Philips zu einem runden Gesamtprodukt, das Funktionalität und Design wunderbar unter einen Hut bringt.

Appendix B: R script

Descriptive stats

Key participant information

```
# age
summary(bachMeta$age)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   19.00   22.25   24.00   24.39   25.00   30.00

# age, standard deviation
sd(bachMeta$age)

## [1] 2.736534

# gender
table(bachMeta$gender)

## 
##   Divers Männlich Weiblich 
##        1        8       29

# education
table(bachMeta$gender) # 18 BA students, 20 MA students

## 
##   Divers Männlich Weiblich 
##        1        8       29

# L1
table(bachMeta$B005_01) # 34 German, 4 list an additional L1

## 
##                       deutsch                       Deutsch 
##                             2                            31 
## Deutsch, Finnisch, Ungarisch            Deutsch, Polnisch 
##                             1                             1 
##                        Deutsh Hochdeutsch und Plautdietsch 
##                             1                             1 
##             Türkisch & Deutsch 
##                             1

#LexTale test results for German
summary(bachMeta$LexTale_DE)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   66.25   79.06   87.50   85.62   92.50   98.75

# Lextale German, standard deviation
sd(bachMeta$LexTale_DE)

## [1] 8.808537

# Lextale test results for English
summary(bachMeta$LexTale_EN)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   54.00   76.56   82.50   81.46   88.75   96.25

# Lextale English, standard deviation
sd(bachMeta$LexTale_EN)

## [1] 10.76115
```

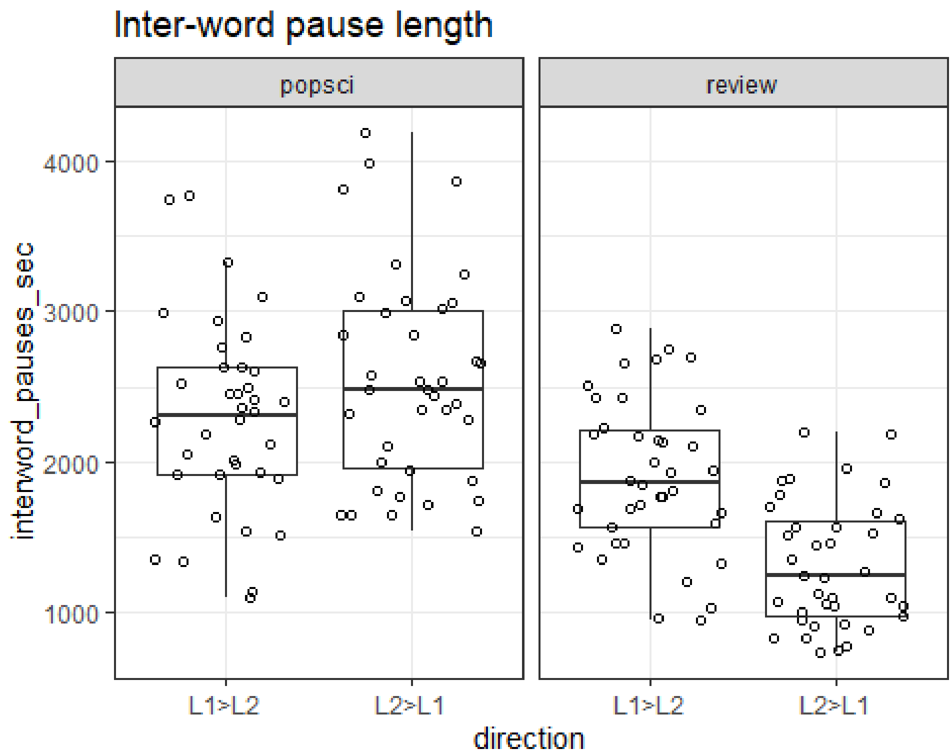
Pause measures

```
# Inter-word pauses
summary(bachMain$interword_pauses_sec)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   733.6  1513.2  1935.1  2014.3  2486.5  4192.0

# Inter-word pauses, standard deviation
sd(bachMain$interword_pauses_sec)

## [1] 737.1506

# Word-internal pauses
summary(bachMain$pauses_wordinternal_sec)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   134.5   220.6   281.6   288.1   340.1   541.4

# Word-internal pauses, standard deviation
sd(bachMain$pauses_wordinternal_sec)

## [1] 78.3733

# no. of keystrokes per second in copy task
summary(bachMain$keystroke_time_copy_sec, na.rm = TRUE)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.491   2.114   2.757   2.700   2.932   4.079      16

# no. of keystrokes per second in copy task, standard deviation
sd(bachMain$keystroke_time_copy_sec, na.rm = TRUE)

## [1] 0.6718379

# time required per keystroke in ms in copy task
summary(bachMain$time_keystroke_copy_ms, na.rm = TRUE)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   245.1   341.1   362.7   395.6   473.1   670.6      16

# time required per keystroke in ms in copy task, standard deviation
sd(bachMain$time_keystroke_copy_ms, na.rm = TRUE)

## [1] 106.9926

# time without an activity before and after text production
summary(bachMain$no_activity_sec)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
##    9.563   36.937   85.805  124.225  162.969 1521.297
```

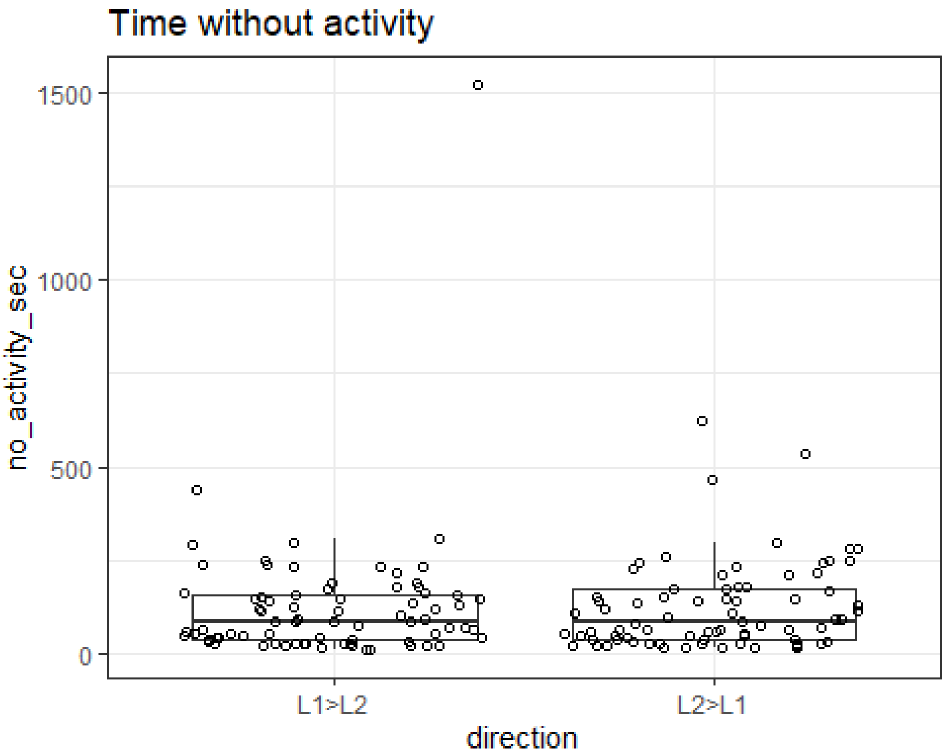
```
# time without an activity before and after text production, SD
sd(bachMain$no_activity_sec)

## [1] 153.8698
```

Visual summaries main experiment

Sums of inter-word pause lengths per text



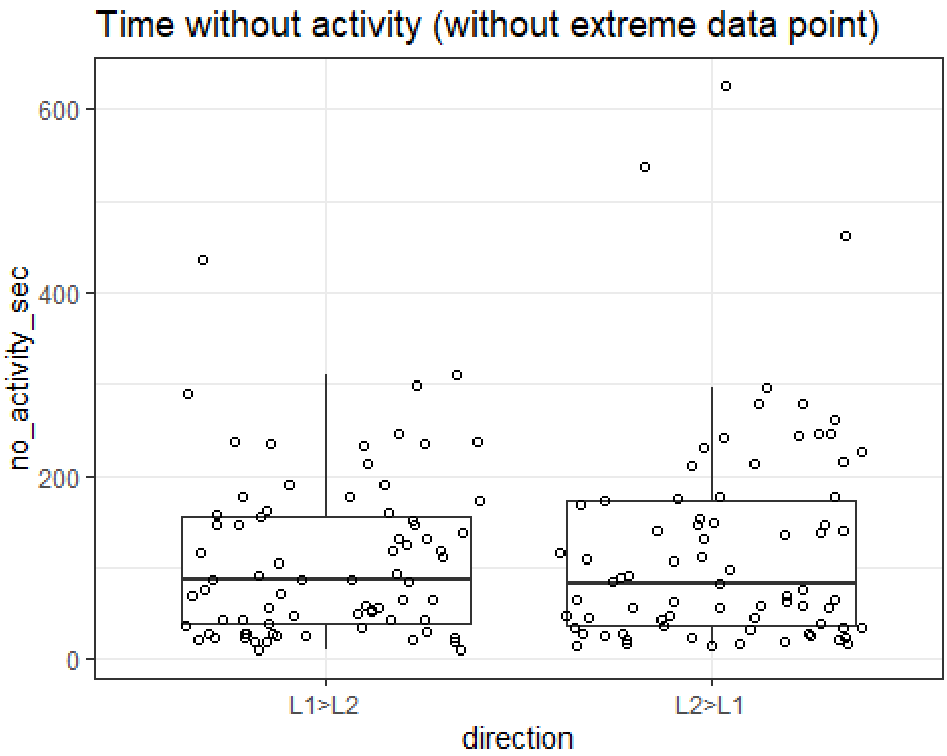
Times without any activity before and after text production



The boxplot clearly shows an extreme data point. Let's therefore inspect the data set without the extreme value. First, descriptive stats without the data point:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   9.563  36.828  85.125 114.973 159.899 624.734

## [1] 103.6124
```

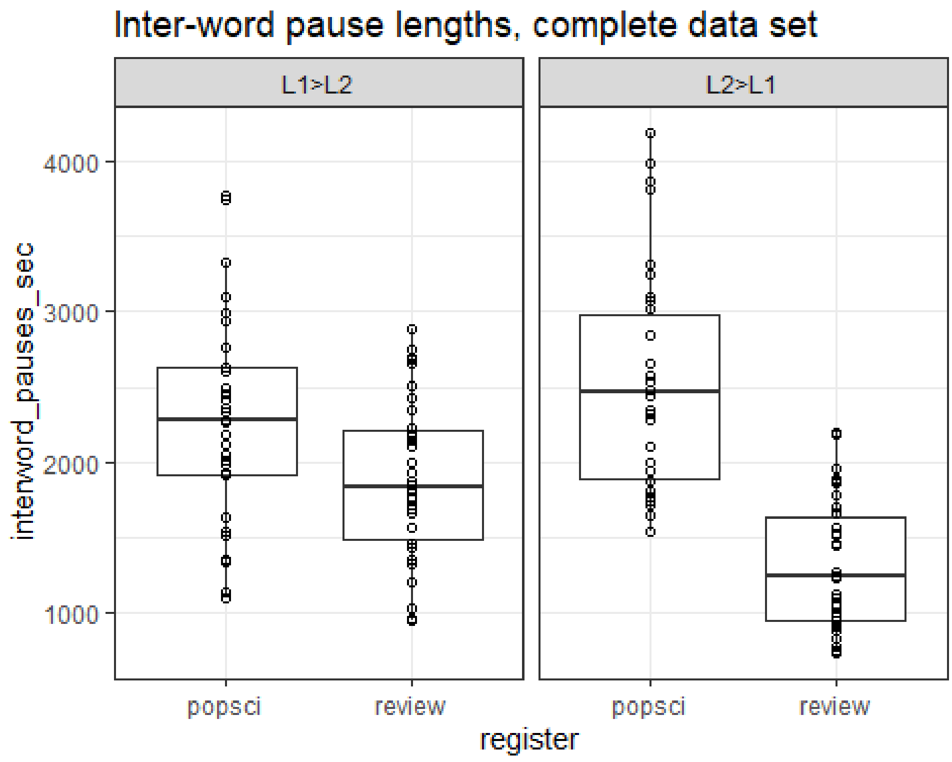
Now, the boxplot without the data point:

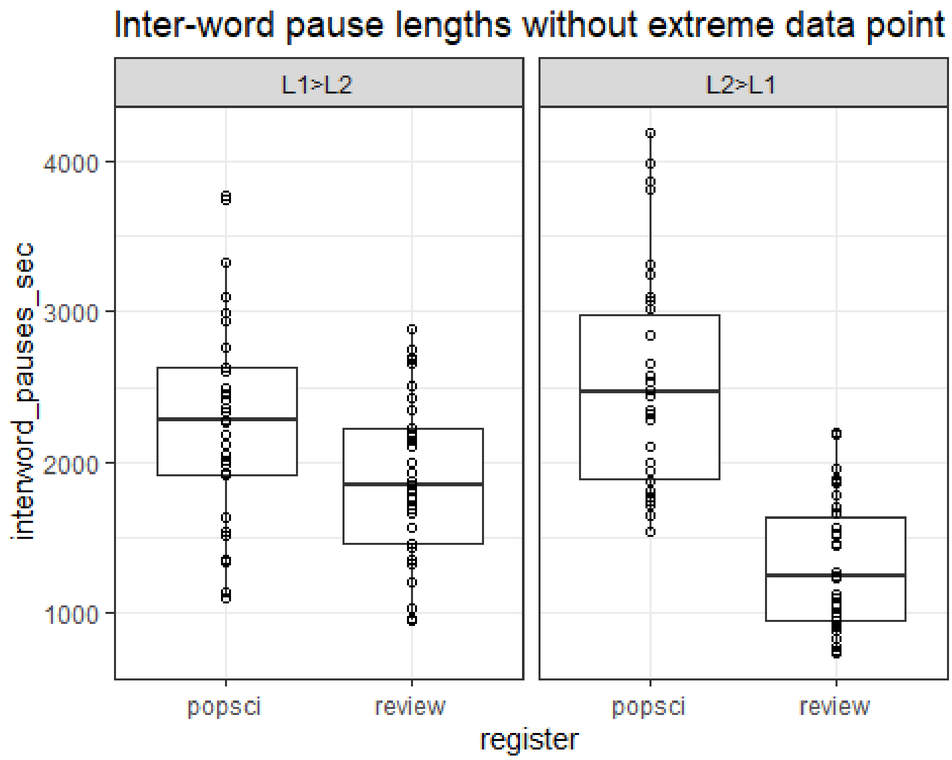


Regression analysis for inter-word pause length

We saw above that there is one extreme value for no activity. All data for this particular text looks ok except for its value for no activity. It is possible that there was simply a gap at the end of the task.

Let's visually check how the distribution of `interword_pauses_sec` looks with and without the extreme data point.





The interquartile range for L1>L2 reviews becomes slightly wider. Otherwise the data looks ok. We will therefore run the model without the extreme data point.

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## interword_pauses_sec_c ~ direction_sum * register_sum + LexTale_EN_c +
##     keystroke_time_copy_sec_c + no_activity_sec_c + (1 | participantID)
##    Data: bachMain135_s
##
##      AIC      BIC   logLik deviance df.resid
##   2074.4   2100.5  -1028.2   2056.4      126
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3634 -0.5780 -0.0503  0.5301  3.6245
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 122935   350.6
##  Residual                  172178   414.9
## Number of obs: 135, groups:  participantID, 34
##
## Fixed effects:
##                              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)                    12.651     70.163   34.163   0.180  0.85798
## direction_sum1                 98.438     35.941  101.475   2.739  0.00728 **
## register_sum1                 347.147     37.289  106.005   9.310 2.01e-15 ***
## LexTale_EN_c                   -3.839      6.628   34.847  -0.579  0.56613
## keystroke_time_copy_sec_c      13.788    104.942   34.138   0.131  0.89624
## no_activity_sec_c               2.174      0.500  127.593   4.349 2.77e-05 ***
## direction_sum1:register_sum1 -177.019     36.440  104.300  -4.858 4.19e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) drct_1 rgst_1 LT_EN_ ky____ n_ct__
## directn_sm1  0.013
## registr_sm1 -0.026 -0.038
## LexTal_EN_c -0.018 -0.008  0.035
## kystrk_t___ -0.012 -0.002  0.016  0.054
## n_ctvty_sc_  0.077  0.104 -0.285 -0.145 -0.075
## drctn_s1:_1  0.010  0.011 -0.046 -0.035 -0.020  0.194
```

As neither L2 proficiency nor typing speed in the copy task turn out to have an effect on pause behaviour, we fit another model with all 38 participants (but without the extreme data point) without these variables.

Final model

```
# Complete data set minus the extreme case
bachMainCompl <- bachMain_s |>
  filter(textID !="BACHMCBDAL07_A")

pause_mdl_compl <- lmer(interword_pauses_sec_c ~ direction_sum * register_sum + no
_activity_sec_c + (1|participantID), data = bachMainCompl, REML = FALSE)
summary(pause_mdl_compl)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## interword_pauses_sec_c ~ direction_sum * register_sum + no_activity_sec_c +
##     (1 | participantID)
##    Data: bachMainCompl
##
##      AIC      BIC   logLik deviance df.resid
##   2302.8   2324.0  -1144.4   2288.8      144
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3840 -0.5618 -0.0401  0.5598  3.6973
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 113667   337.1
##  Residual                  159956   399.9
## Number of obs: 151, groups:  participantID, 38
##
## Fixed effects:
##                              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)                   18.7922    63.7631  37.9403   0.295  0.76982
## direction_sum1                95.5683    32.6844 113.1476   2.924  0.00418
## register_sum1                360.4499    33.7880 117.5891  10.668  < 2e-16
## no_activity_sec_c              2.0692     0.4546 143.5747   4.552 1.13e-05
## direction_sum1:register_sum1 -183.2756    33.3175 117.1153  -5.501 2.25e-07
##
## (Intercept)
## direction_sum1               **
## register_sum1                ***
## no_activity_sec_c            ***
## direction_sum1:register_sum1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) drct_1 rgst_1 n_ct__
## directn_sm1  0.009
## registr_sm1 -0.019 -0.030
## n_ctvty_sc_  0.058  0.082 -0.266
## drctn_s1:_1  0.008  0.009 -0.048  0.210
```

Coefficients

```
# random effect coefficients for each participant
# coef(pause_mdl_compl)$participantID
summary(coef(pause_mdl_compl)$participantID[1])
```

```
##   (Intercept)
##  Min.   :-652.06
##  1st Qu.:-174.06
##  Median :  44.28
##  Mean   :  18.79
##  3rd Qu.: 186.23
##  Max.   : 721.37
```

Check for collinearity

```
vif(pause_mdl_compl)

##              direction_sum               register_sum 
##                   1.006929                   1.076071 
##          no_activity_sec_c direction_sum:register_sum 
##                   1.129798                   1.046316
```

Model comparison

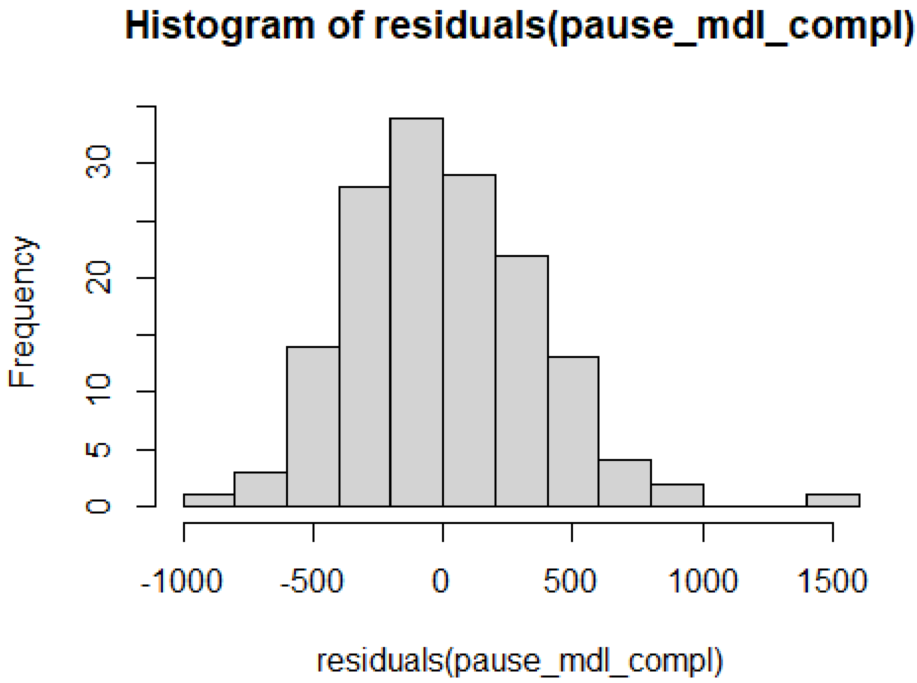
```
pause_mdl_compl_afex <- mixed(interword_pauses_sec_c ~ direction * register + no_a
ctivity_sec_c + (1|participantID), data = bachMainCompl, method = "LRT")

## Contrasts set to contr.sum for the following variables: direction, register, pa
rticipantID

## Numerical variables NOT centered on 0: no_activity_sec_c
## If in interactions, interpretation of lower order (e.g., main) effects difficul
t.

## REML argument to lmer() set to FALSE for method = 'PB' or 'LRT'

pause_mdl_compl_afex

## Mixed Model Anova Table (Type 3 tests, LRT-method)
## 
## Model: interword_pauses_sec_c ~ direction * register + no_activity_sec_c + 
## Model:     (1 | participantID)
## Data: bachMainCompl
## Df full model: 7
##               Effect df     Chisq p.value
## 1          direction  1   8.26 **    .004
## 2           register  1 78.54 ***   <.001
## 3  no_activity_sec_c  1 19.08 ***   <.001
## 4 direction:register  1 26.83 ***   <.001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

R2

```
r.squaredGLMM(pause_mdl_compl_afex$full_model)

##            R2m       R2c
## [1,] 0.4877996 0.7005753
```

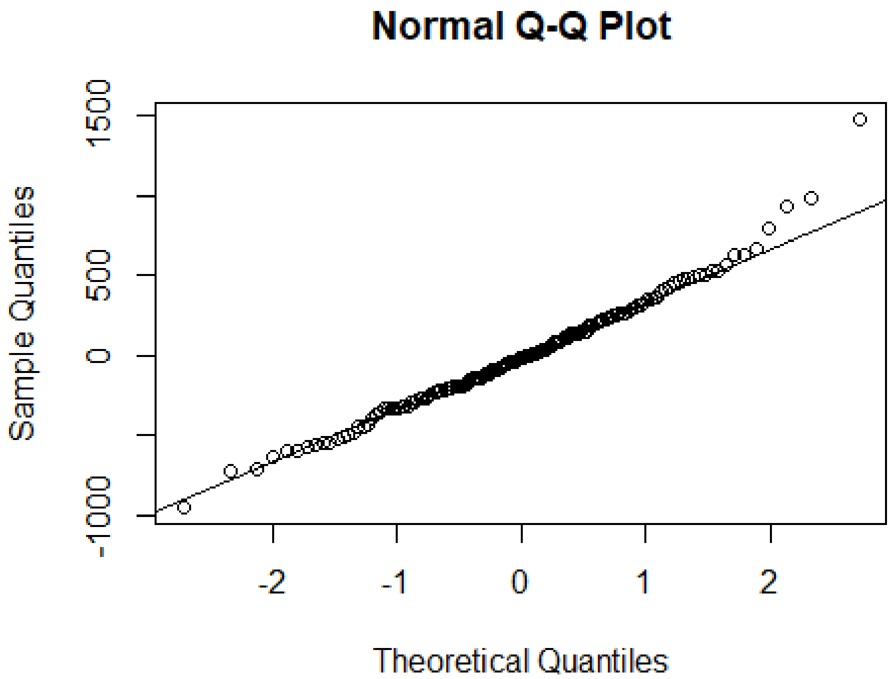
Is the error normally distributed?

```
hist(residuals(pause_mdl_compl))
```



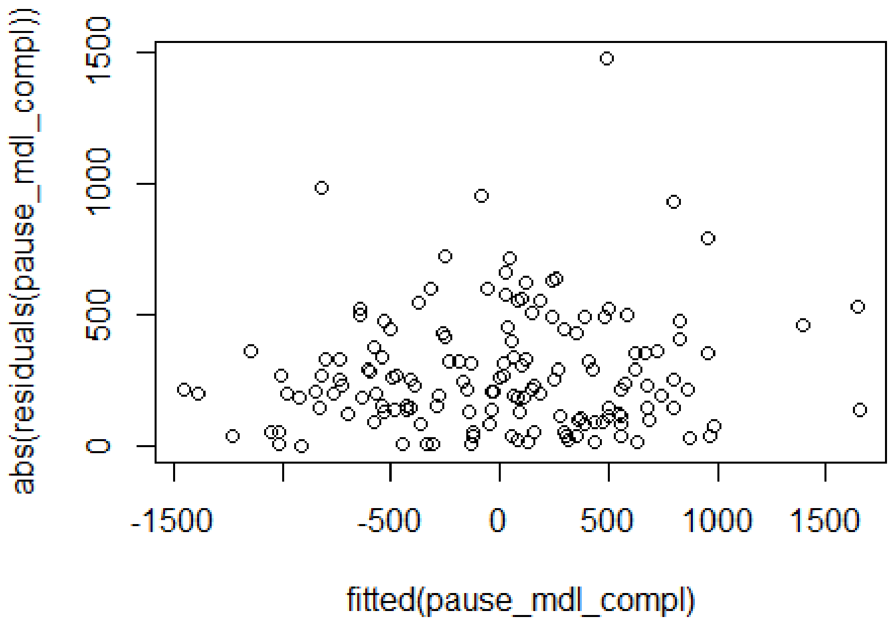
There is somewhat of a right-skew in the model, which is also visible in the q-q plot:

```
qqnorm(residuals(pause_mdl_compl))
qqline(residuals(pause_mdl_compl))
```



Check for homoskedasticity

```
plot(fitted(pause_mdl_compl),abs(residuals(pause_mdl_compl)))
```



References

Alves, Fabio & Pagano, Adriana & Da Silva, Igor. 2009. A new window on translators' cognitive activity. Methodological issues in the combined use of eye tracking, key logging and retrospective protocols. In Mees, Inger M. & Alves, Fabio & Göpferich, Susanne (eds.), *Methodology, Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur Press. 267–92.

Alves, Fabio & Vale, Daniel. 2009. Probing the unit of translation in time: Aspects of the design and development of a web application for storing, annotating, and querying translation process data. Across Languages and Cultures 10(2): 251–73. doi: 10.1556/Acr.10.2009.2.5.

Alves, Fabio & Vale, Daniel. 2011. On drafting and revision in translation. A corpus linguistics-oriented analysis of translation process data. *TC3: Translation: Corpora, Computation and Cognition* 1(1): 105–122. Reprinted in Hansen-Schirra, Silvia & Neumann, Stella & Čulo, Oliver (eds.), *Annotation, exploitation and evaluation of parallel corpora: TC3 I*. Berlin: Language Science Press. 89–110. doi: 10.5281/zenodo.283500.

Alves, Fabio & Gonçalves, José L. 2013. Investigating the conceptual-procedural distinction in the translation process. A relevance-theoretic analysis of micro and macro translation units. *Target* 25(1): 107–24.

Cao, Lu & Doherty, Stephen & Lee, James F. 2023. The process and product of translation revision. empirical data from student translators using eye tracking and screen recording. *The Interpreter and Translator Trainer* 17(4): 548–65.

Carl, Michael. 2012. Translog-II: A program for recording user activity data for empirical reading and writing research. In *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'12). (http://www.lrec-conf.org/proceedings/lrec2012/pdf/614_Paper.pdf) (Accessed 2024-12-02)

Carl, Michael. 2021. Micro units and the first translational response universal. In Michael Carl (ed.), *Explorations in Empirical Translation Process Research*. Cham: Springer. 233–57. doi: 10.1007/978-3-030-69777-8_9.

Carl, Michael & Schaeffer, Moritz & Bangalore, Srinivas. 2016. The CRITT translation process research database. In Carl, Michael & Bangalore, Srinivas & Schaeffer, Moritz (eds.), *New Directions in Empirical Translation Process Research Exploring the CRITT TPR-DB*. Cham: Springer. 13–54.

Chang, Vincent C.-Y. & Cheng, I-Fei. 2023. Translation directionality and the Inhibitory Control Model. A machine learning approach to an eye-tracking study. *Frontiers in Psychology* 14. doi: 10.3389/fpsyg.2023.1196910.

Ehrensberger-Dow, Maureen. 2013. *Capturing translation processes final report.* School of Applied Linguistics IUED Institute of Translation and Interpreting. Zurich University of Applied Sciences. (https://www.zhaw.ch/storage/linguistik/forschung/uebersetzungswissenschaft/publikation-2013-capturing-translation-processes-final-eport.pdf) (Accessed 2024-12-02)

Evert, Stefan & Neumann, Stella. 2017. The impact of translation direction on characteristics of translated texts. A multivariate analysis for English and German. In De Sutter, Gert & Lefer, Marie-Aude & Delaere, Isabelle (eds.), *Empirical Translation Studies. New Theoretical and Methodological Traditions*. Berlin: de Gruyter. 47–80.

Elston-Güttler, Kerrie E. & Paulmann, Silke & Kotz, Sonja A. 2005. Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience* 17(10): 1593–1610. doi: 10.1162/089892905774597245.

Ferreira, Alina & Schwieter, John. W. & Gottardo, Alexandra & Jones, J. 2016. Cognitive effort in direct and inverse translation performance. Insight from eye-tracking technology. *Cadernos de Tradução* 36(3): 60–80.

Ferreira, Aline & Gottardo, Alexandra & Schwieter, John W. 2018. Decision-making processes in direct and inverse translation through retrospective protocols. *Translation, Cognition & Behavior* 1(1): 98–118.

Ferreira, Aline & Gries, Stefan & Schwieter, John W. 2021. Assessing indicators of cognitive effort in professional translators. A study on language dominance and directionality. In Tra&Co Group (ed.), *Translation, Interpreting, cognition: The way out of the box*. Berlin: Language Science Press. 115–43.

Fonseca, Norma. 2015. Directionality in translation. Investigating prototypical patterns in editing procedures. *Translation & Interpreting: The International Journal of Translation and Interpreting Research* 7(1): 111–25.

Göpferich, Susanne & Bayer-Hohenwarter, Gerrit & Stigler, Hubert. 2008. *TransComp–The Development of Translation Competence. [Corpus and Asset Management System for the Longitudinal Study TransComp]*. Graz: University of Graz.

Halliday, M.A.K. & Hasan, Ruqaiya. 1976. *Cohesion in English* (1st edition). London: Routledge. doi: 10.4324/9781315836010.

Jakobsen, Arnt Lykke. 2002. Translation drafting by professional translators and by translation students. In Sánchez Trigo, Elena & Díaz Fouces, Óscar (eds.), *Traducción & Comunicación: Ciclos de conferencias organizados en la Universidad de Vigo* (Vol. 3). Collección Congresos 36. Vigo: Servicio de Publicacións da Universidade de Vigo. 89–103.

Jensen, Kristian Tangsgaard Hvelplund & Sjørup, Annette C. & Balling, Laura Winther. 2009. Effects of L1 syntax on L2 translation. *Copenhagen Studies in Language* 38: 319–36.

Jia, Juan & Wei, Ziyu & Cheng, Heben & Wang, Xiaolu. 2023. Translation directionality and translator anxiety. Evidence from eye movements in L1-L2 translation. *Frontiers in Psychology* 14. doi: 10.3389/fpsyg.2023.1120140.

Keller, Frank & Gunasekharan, Subahshini & Mayo, Neil & Corley, Martin. 2009. Timing accuracy of web experiments. A case study using the WebExp software package. *Behavior Research Methods* 41 (1): 1–12. doi: 10.3758/BRM.41.1.12.

Klie, Jan-Christoph & Bugert, Michael & Boullosa, Beto & De Castilho, Richard E. & Gurevych, Iryna. 2018. The INCEpTION platform. Machine-Assisted and knowledge-oriented interactive annotation. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. 5–9. (https://aclanthology.org/C18-2002.pdf) (Accessed 2025-09-03).

Klinger, Roman. 2014. *The USAGE review corpus for fine-grained, multi-lingual opinion analysis*. Bielefeld University. doi: 10.4119/unibi/2664575.

Lehka-Paul, Olha. 2020. Behavioural indicators of translators' decisional styles in a translation task. A longitudinal study. *Yearbook of the Poznań Linguistic Meeting* 6 (1): 81–111. https://pressto.amu.edu.pl/index.php/yplm/article/view/26593/24206. (Accessed 2025-09-07).

Lemhöfer, Kristin & Broersma, Mirjam. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2): 325–43. doi: 10.3758/s13428-011-0146-0.

Martín, Ricardo Muñoz & Apfelthaler, Matthias. 2022. A task segment framework to study keylogged translation processes. *Translation & Interpreting* 14 (2): 8–31. doi: 10.12807/ti.114202.2022.a02.

Pavlović, Nataša. 2007. Directionality in Collaborative Translation Processes. Tarragona: Universitat Rovira i Virgili. (Doctoral dissertation.)

Pavlović, Nataša. 2013. Exploring directionality in translation studies. *Explorations in English Language and Linguistics* 1(2): 149–65.

Pavlović, Nataša & Hvelplund Jensen, Kristian. 2009. Eye tracking translation directionality. In Pym, Anthony & Perekrestenko, Alexander (eds.), *Translation Research Projects 2*. Tarragona: Intercultural Studies Group. 93–109.

Penha-Marion, Laura & Gilquin, Gaëtanelle & Lefer, Marie-Aude. 2024. The effect of directionality on lexico-syntactic simplification in French><English student translation. In Kotze, Haidee & Van Rooy, Bertus (eds.), *Constraints on language variation and change in complex multilingual contact settings*. Amsterdam: John Benjamins Publishing Company. 153–90.

Pietryga, Marcelina. 2022. Perception of directionality in translation among students. *UWM Olsztyn Acta Neophilologica XXIV* 1: 87–109. doi: 10.31648/an.6807.

Pokorn, Nike K. 2005. *Challenging the traditional axioms. Translation into a non-mother tongue*. Amsterdam: Benjamins.

Pokorn, Nike K. & Blake, Jason & Reindl, Donald & Peterlin, Agnes P. 2019. The influence of directionality on the quality of translation output in educational settings. *The Interpreter and Translator Trainer* 14(1): 58–78. doi: 10.1080/1750399X.2019.1594563.

R Core Team. 2024. R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. (https://www.R-project.org) (Accessed 2024-12-02).

Rodríguez-Inés, Patricia. 2013. ¿Cómo traducen traductores y profesores de idiomas? Estudio de corpus [How do translators and language teachers translate? A corpus study]. *Meta* 58(1): 165–90. doi: 10.7202/1023815ar.

Rodriguez-Inés, Patricia. 2017. Analysis of the Translation Competence corpus from PACTE's experiment. In Hurtado Albir, Amparo (ed.), *Researching Translation Competence by PACTE Group*. Amsterdam: Benjamins. 243–66.

Schaeffer, Moritz & Huepe, David & Hansen-Schirra, Silvia & Hofmann, Sascha & Muñoz, Edinson & Kogan, Boris & Herrera, Eduar & Ibáñez, Agustin & García, Adolfo M. 2020. The Translation and Interpreting Competence Questionnaire. An online tool for research on translators and interpreters. *Perspectives* 28(1): 90–108. doi: 10.1080/0907676X.2019.1629468

Schmaltz, Marcia & Leal, Ana & Wong, Derek & Chao, L. Sam & Da Silva, Igor A.L. & Alves, Fabio & Pagano, Adriana & Quaresma, Paulo. 2016. Cohesive relations in text comprehension and production. An exploratory study comparing translation and post-editing. In Carl, Michael & Bangalore, Srinivas & Schaeffer, Moritz (eds.), *New directions in empirical translation process research*. Springer: Singapore. 239–63.

Singmann, Henrik & Bolker, Ben & Westfall, Jake & Aust, Frederik. 2018. *Afex. Analysis of Factorial Experiments.* (https://CRAN.R-project.org/package=afex) (Accessed 2024-12-02)

Stewart, Dominic. 1999. Translators into the foreign language. Charlatans or professionals? *International Journal of Translation* 4: 41–67. (http://hdl.handle.net/10077/2912) (Accessed 2025-09-03)

Taboada, Maite & Grieve, Jack. 2004. Analyzing appraisal automatically. American association for artificial intelligence spring symposium on exploring attitude and affect in text. Stanford. March 2004. *AAAI Technical Report SS-04-07*. 158–61.

Whyatt, Bogusława. 2019. In search of directionality effects in the translation process and in the end product. *Translation, Cognition & Behavior* 2(1): 79–100. doi: 10.1075/tcb.00020.why.

Winter, Bodo. 2019. *Statistics for linguists: An introduction using R.* Routledge. doi: 10.4324/9781315165547.

Zoë Miljanović
RWTH Aachen University
Institut für Anglistik
Kármánstr. 17/19
52062 Aachen
Germany
e-mail: zoe.miljanovic@ifaar.rwth-aachen.de

Fabio Alves
Universidade Federal de Minas Gerais (UFMG)
Faculdade de Letras
Av. Presidente Antonio Carlos 6.627
Campus Pampulha, Belo Horizonte/MG
e-mail: fabio-alves@ufmg.br

Celina Brost
RWTH Aachen University
Institut für Anglistik
Kármánstr. 17/19
52062 Aachen
Germany
e-mail: celina.brost@ifaar.rwth-aachen.de

Stella Neumann
RWTH Aachen University
Institut für Anglistik
Kármánstr. 17/19
52062 Aachen
Germany
e-mail: stella.neumann@ifaar.rwth-aachen.de