trans-kom ISSN 1867-4844 https://www.trans-kom.eu
trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom 18 [2] (2025): 625–655

Raphael Berthele & Isabelle Udry & Didier Yourassoff

The effects of digital tool use on complexity and accuracy of written texts in English as a foreign language

Abstract

Foreign language writing in European languages is increasingly done with the help of digital online tools (AI powered translators and/or dictionaries). We tested the effect of such tool use on the process and the product of foreign language writing in 274 adolescent learners of English as a foreign language. We compare to what extent writing with such tools changes indicators of (lexical, syntactic) complexity and of accuracy. Both dimensions were positively associated with task achievement as assessed by human raters. We also investigate the impact of a short instruction on the characteristics and features of online tools and their use in writing. The results show a substantial increase in accuracy and some increase in lexical complexity when digital tools are used. The instruction led to an increase in the proportion of dictionary use in relation to the use of the online translators but had little impact on the actual text output.

1 Introduction

In this paper we present evidence from a writing task across different conditions. We address the question how written output produced by learners of English as a foreign language changes depending on the authors' use of digital online resources such as automatic translation machines and online dictionaries. The topic is timely and relevant as these tools are omnipresent, and second and foreign language users will use them (if allowed). Our interest covers several research questions:

(1) What are the digital tools intermediate learners of English as a foreign language (EFL) use when they are writing texts for formal purposes?

(2) How much time do they spend online when writing texts, and how much do they switch between applications during the writing process?

(3) What is the effect of the use of digital tools on the outcome of the writing task? More specifically:
   (a) How do features of the texts change if tools are used?
   (b) Is the product of the writing task 'better' with respect to task achievement if tools are used?

The first and the second questions relate to the writing process. They are exploratory in nature, as we do not know much about EFL learners' behaviour in this domain and thus we do not have testable hypotheses. There are some questionnaire-based studies that elicit the tools learners claim to use (see e. g. Udry/Berthele 2023 for such a study from the Swiss context), but to our knowledge there are no studies that observe which tools exactly learners use in what way and for how long when doing writing tasks. The third question relates to the product and will be addressed via hypothesis testing, based on theories and findings we will discuss in the next section.

We draw on two influential frameworks that investigate the effects of task characteristics on language learning and acquisition, namely Peter Skehan's (Skehan/Foster 1999, 2001; Skehan 2014) Trade-off or Limited Capacity Hypothesis model, and Peter Robinson's (Robinson 2001) Cognition Hypothesis or Triadic Componential Framework. As we discuss below, scholars have argued for different patterns of associations (trade-offs or mutual boosts) between dimensions of the product, e. g. regarding accuracy on the one hand and complexity of the texts on the other. Thus, in the case of the third question, we will investigate the association of the tool use behaviour (as a task characteristic) and text characteristics (within the framework of CAF, Complexity, Accuracy, Fluency, cf. Housen 2021).

2 Literature review

2.1 Models of writing

Influential models of first language (L1) writing theorize interrelated and recursive mechanisms that involve planning of content, formulating ideas, and monitoring output to ensure it is coherent with the intended message (Hayes/Flower 1980; Kellogg 1996). Executing these processes requires the allocation of attentional resources and is therefore linked to the construct of working memory (WM; Kellogg 1996). In Baddeley's (Baddeley/Hitch 1974; Baddeley 2000) still influential multicomponent model, WM comprises two slave systems with limited capacity that are responsible for the short-term maintenance of different types of information: the phonological loop stores phonological information, and the visuo-spatial sketchpad stores visual and spatial information. A third component, the central executive, coordinates the two slave systems and assists in the integration of the various stimuli. The episodic buffer combines information into meaningful and lasting representations, thus providing the link between WM and long-term memory. The cognitive burden placed on each of the WM components is thought to vary along each stage of the writing process (Kellogg et al. 2013).

The basic sub-processes of planning, formulating, and monitoring are thought to be roughly the same in L1 and in the second/foreign language (L2/FL) writing. Nevertheless, text production in a foreign language is associated with greater cognitive effort. In addition to coping with conceptual requirements set by the task (i. e. planning and producing coherent text), L2 writers must come to terms with the cognitive demands of

expressing their ideas in a language in which they have to get by with more limited resources and less automatized processes compared to their L1. Among those more limited resources are generally smaller vocabularies per language and grammatical knowledge that is limited by the constraints of the respective proficiency in the target language. Processing limitations, among other things, involve comparatively less efficient lexical retrieval and less automatic syntactic encoding, and less ease in creating text cohesion and revision (Kormos 2011). Unsurprisingly therefore, L2 writing processes have been found to be less automated, usually resulting in shorter, less cohesive, and less accurate output that tends to be lexically and syntactically poorer than texts written in the L1 (Silva 1993).

2.2 Task effects

How allocation of attention affects language performance and development has long been a question in L2 or FL acquisition research, also in the domain of writing (see e. g. Kuiken/Vedder 2008; Kormos 2011; Frear/Bitchener 2015). In current task-based approaches to FL learning and teaching (Willis 1996), tasks are the main element that can be adapted in complexity to alter the cognitive demands placed on learners (Skehan 1996).

Robinson (2001: 294) distinguishes between two dimensions along which task complexity can be manipulated: resource-direction (i. e. more or fewer elements, higher or lower cognitive demands in terms of reasoning and planning) and resource-depletion (e. g. a task involving little planning time or not providing much background knowledge). Robinson posits that higher *resource-directing* demands raise the functional or communicative demands of a task.

To match these high functional demands, learners arguably generate more elaborate language structures and produce more complex L2 speech and text. In contrast, increasing the *resource-depleting* dimension of a task is thought to decrease the memory resources available for attending to language form, lowering the quality of language output in more complex tasks. Notwithstanding these predictions, studies that have addressed the issue have come up with inconclusive results (for a discussion see Johnson 2017). As discussed below, translation tools can be expected to affect both resource direction (increase or decrease in reasoning demands) and resource-depletion (increase or decrease of planning time).

2.3 Complexity and accuracy

Language performance is commonly described along the dimensions of complexity, accuracy, and fluency (CAF; Housen/Kuiken 2009; Pallotti 2009). In task-based learning settings, greater (lexical and syntactic) complexity in language production is associated with higher language competence; greater accuracy is the result of successful error avoidance; and fluency is the ability to produce language continuously and at adequate speed (Skehan 2009: 510). Performance in these dimensions not only depends on task complexity but is underpinned by a range of individual learner characteristics, including

attentional resources (Housen/Simoens 2016). Based on the uncontroversial insight outlined previously that cognitive processing capacity is limited, contrasting conceptions have emerged to explain the effects of task-induced cognitive involvement on language performance.

The trade-off hypothesis (or Limited Capacity Hypothesis model) by Skehan and Foster emphasizes the point that cognitive processing ability is individually constrained. As learners reach their attentional limits, they are forced to focus on either meaning or form. When faced with difficult tasks that require more attention, learners are likely to prioritize content over language form (Kuiken/Vedder 2008). The model assumes that complexity, accuracy, and fluency are in competition, and higher performance in one of these dimensions can occur at the expense of the others (Skehan/Foster 2001; Skehan 2009). In contrast, Robinson's (2001) Cognition Hypothesis posits the existence of multiple attentional pools that are non-competing and can be accessed differentially as a function of task demand.

Each model makes different predictions in terms of language output. On the one hand, the trade-off perspective expects a negative association of accuracy and complexity (Skehan/Foster 1997, 2001). On the other hand, the Cognition Hypothesis claims that greater task complexity may promote attentional activation as it facilitates noticing and integration of new linguistic features, thus pushing learner performance towards greater accuracy and complexity (Robinson 2001: 34). At the same time, the Cognition Hypothesis assumes difficult tasks to be associated with lower fluency. In other words, while fluency contrasts with complexity and accuracy, the latter two are expected to correlate positively (Skehan 2003: 5).

2.4 Tool use and writing task: Expected effects

Based on the literature outlined in the previous section, different effects of tool use on the writing product can be hypothesized depending on

(a) how the influence of tool use is conceptualized in terms of its contribution to task complexity (resource-directing and resource-depleting dimensions)
(b) how tool use affects the different writing sub-processes (planning, formulating, monitoring).

Tool use may affect the resource-directing dimension of task complexity both positively and negatively. Consulting external resources can raise reasoning demands as it requires learners to interrupt their writing, manage the look-up process, and transfer the result back to their texts. This is especially relevant for participants who use online dictionaries in a targeted way to search selectively for L2 vocabulary as needed during writing. That said, if such processes are familiar and automated, the time and effort invested may be minimal, reducing reasoning demands.

In contrast, participants who compose their text in their L1 first and then feed it into a translator to produce the L2 text may experience reduced demands on the resource-directing dimension, since they avoid constructing meaning directly in the L2. However,

this approach may also reduce opportunities for deeper engagement with the target language. These different behaviours (targeted search vs. full-text composition in L1 – machine translation in L2) can be investigated by tracking the time spent online and the types of tools used (dictionary or translator).

In terms of the writing process, tool use is expected to ease formulation (translating ideas into the L2) and monitoring (revising the output). Online translators, in particular, may free attentional resources that can be directed to planning and text coherence. With reference to Robinson's framework, this shift may reduce the resource-depleting impact of complex task features.

To sum up, depending on the intensity and modality, tool use may have opposing effects: On the one hand, it may reduce task complexity by supporting formulation, in particular by allowing learners to write in a well-mastered language (L1) as opposed to a less well mastered target language (L2). This can speed up the process, especially if large chunks or entire texts are first written in the learner's L1 and then translated. But more targeted tool use also allows learners to quickly fill language gaps, potentially more efficiently than relying solely on their (by definition) limited L2 knowledge. Overall, if tool use is efficient, it likely results in more accurate and more elaborate output. On the other hand, tool use can also increase task complexity by requiring learners to switch between programs and interfaces, which may be time consuming for less experienced users. Such task switching can divert attention from composition and negatively affect text coherence and complexity.

2.5 Operationalizing complexity and accuracy

In this contribution, we analyse various complexity metrics, and we also include the number of tokens in the final product per time spent on the task among these metrics, as is done by other scholars in the field (e. g. Struc/Wood 2011: 50). Another approach would consider this last metric as a measure of fluency. Fluency metrics typically examine overall quantity of language produced per temporal unit, pausing behaviour or the number and quantitative features of productive bursts in the writing process. Although analysing fluency and pausing behaviours is possible, there are many open questions as to what a pause (in the first or in the target language) means and how it should be defined and operationalized (see Michel et al. 2020 for discussion). Due to space restrictions, we plan to address these questions in a separate contribution and will focus on the effects on the textual product in this article.

As outlined previously, tool use – if efficient – is expected to help with formulation (spelling and grammar) and lead to more accurate texts. Also, if participants choose to use a translator, typing their ideas in the L1 as they go, they are able to formulate more complex utterances which arguably will also lead to more complexity in the target language. In other words, these participants would dispense less effort on formulating, leaving more attentional resources to be dedicated to planning and possibly revising.

The trade-off or cognition hypotheses were developed without taking into account additional language-support tools, such as those used by our participants. The predictions of the two models thus need to be used cautiously in our context. The competition between accuracy and complexity predicted by the trade-off hypothesis need not apply when digital tools are used, as attention to formal aspects in the L2 may be 'outsourced' partly or entirely to the tools. Thus, the tools take away cognitive load related to form and free up resources for meaning. This may apply in particular if entire sentences are fed into an online translator. Texts could turn out to be both more accurate and more complex, which admittedly would be compatible with the cognition hypothesis. However, competition between the different dimensions would remain if the process of meaning consultation required more cognitive effort (i. e. resulting in increased reasoning demands).

To sum up, there is little doubt that digital tool use should yield more accurate output, while the expectations regarding the textual complexity are unclear: First, we do not know to what extent tool use changes task complexity and second, there are competing hypotheses regarding the existence of trade-offs and regarding the way tool use might have an impact on the different requirements imposed by the task.

3 Participants, material, and procedure

3.1 Participants

We tested a total of 274 participants who were between 16 and 21 years old. All participants were enrolled in a vocational training program with parallel schooling (leading to the Swiss professional baccalaureate "Berufsmatura/maturité professionnelle"). Their domains of vocational training involved IT, commerce, health care, and design and arts. The aim of the study program is to get to a level of B2 in English as a foreign language after four years of classes. The participants were enrolled either in a French-speaking (N=55) or a German-speaking (N=178) or a bilingual French-German (N=41) program. The sample consists of classes of pupils whose teachers responded positively to a mass email call for participation in a "study on digital tools and English foreign language learning".

3.2 Task

The task (see the supplementary material[1] for details) was to write an email to a potential employer located in an English-speaking area in order to apply for an internship or open position. Slightly different variants of the task were used in the repeated-measures design to avoid boredom and self-copying. The instructions were given in the school language (German or French) to ensure comprehension. Specific content elements were required in each writing task. These elements pertained to information about the applicants'

[1] The effects... (2025).

education/curriculum, motivation, and future career plans. The students were asked to be as informationally rich and specific as possible. They had 30 minutes at their disposal to write their texts which turned out to be largely sufficient for task completion.

The texts were produced in the respective classrooms on laptop computers provided by the researchers. They were written in Microsoft Word with spell-checking and grammar-checking options disabled. The writing was logged via InputLog (Leijten/Van Waes 2013). Log files were available for all texts, listing keyboard and mouse events with timestamps including resources consulted online in table form that allows the researcher to reconstruct the students' activities.

For the sake of maximizing ecological validity, in the condition in which tools were allowed, no constraints were imposed on the participants as to the type of tool (online dictionary or translator) or how they wished to employ them. Thus, the learners were free to use the tools just for looking up specific words or problems or to write the whole text in their L1 and then have it translated in one go.

3.3 Procedure and text corpus

The data were collected in three steps: First (T1), the baseline texts were written by all participants without access to any digital tools. Second (T2), the pupils wrote a second text in a randomly assigned condition: either again without any online tools (identical to T1, control condition), or with online tools. A subset of the students who wrote in the online condition received an introduction to the functioning and use of online translation machines and dictionaries. More specifically, a one-lesson classroom intervention was designed to raise students' awareness of the strengths and limitations of these tools.

The session began with a guided discussion in English to activate prior knowledge and explore learner attitudes towards translators and dictionaries (10 minutes). This was followed by an individual translation task in which students compared outputs from DeepL and Pons to highlight differences in contextual accuracy and appropriateness between translators and dictionaries (10 minutes). In the main phase, students engaged in three scaffolded tasks to develop strategies for using these tools effectively, including identifying appropriate use cases, detecting idiomatic mismatches, and testing translation reliability through re-translation (e. g. from L1 – L2 – L1; 20 minutes). The lesson concluded with a plenary feedback session and a brief instructional video summarizing key takeaways (5 minutes). The activities were designed to equip learners with practical techniques for informed tool use in foreign language writing. All materials can be downloaded from Online translation tools (s. a.).

Table 1 gives an overview of the number of texts per condition in the data.

condition	**texts**
T1 baseline	238
T2 control (offline)	78
T2 no instruction (online)	83
T2 instruction (online)	73
Total	472

Table 1: Number of texts per time and condition

The students were clustered in classes, the number of students per class who participated in the different writing sessions (text 1 and text 2) are given in the supplementary material (Table S3).

After the two text writing sessions, we had a total of 472 texts at our disposal for analysis. The InputLog files were processed with a custom-tailored script that extracted information on the time spent working in Word, the type of online resources used (e. g. DeepL, Pons, etc.), the number of switches between Word and those resources, and the respective time spent interacting with the online tools.

To test the relevance of the different metrics analyzed for the writing task, we also had a panel of 10 human raters assess the texts along different dimensions, one of which was the extent to which the texts are deemed to achieve the communicative goals of the writing task. The associations of the average rating of task achievement with all other variables analyzed below will be discussed in the results section.

3.4 Complexity metrics

The complexity metrics were generated via a workflow that uses and adapts freely available software.

The first analysis involved parsing all texts with the Stanford parser (Socher et al. 2013). Based on the output of the parser, we calculated different metrics of complexity as proposed by Lu (2010). Using a modified R Script based on the R implementation of Lu's L2SCA (https://sites.psu.edu/xxl13/l2sca/) we produced 14 different metrics of complexity, both on the level of T-units and of clauses.

A second analysis sent the texts through the UDPIPE annotation software (Wijffels 2018) in order to lemmatize the texts. Based on the lemmatized texts, different lexical diversity metrics could be generated via functions provided by the koRpus package (Michalke 2020). Moreover, an R script unifies the spelling to UK and looks up the corpus frequencies in a film subtitle corpus (van Heuven et al. 2014). Based on these frequencies, metrics of lexical rarity or sophistication can be calculated.

This workflow produces an overwhelming number of potentially usable metrics. Based on previous research (McCarthy/Jarvis 2010; Biber/Gray/Kornwipa 2011; Vanhove et al. 2019), we selected a subset of metrics in order to address the research questions spelled out above.

For lexical diversity, we use two type-token based metrics, a simple one and a complex one. The simple metric is Guiraud's index.

$$Guiraud's\ index: \frac{numberofTypes(lemmas)}{\sqrt[2]{numberofTokens}}$$

We chose this index as it is considered a valid metric for lexical richness (diversity) when texts are assessed within the framework of the levels of the CEFR (Treffers-Daller/ Parslow/Williams 2016). Moreover, as shown in Vanhove et al. (2019), Guiraud's index turns out as equally valid a predictor of human ratings of lexical diversity of text production as more complex metrics. Guiraud's index corrects to some extent for text length but is not entirely independent of it. As our texts are rather variable in length (min 71, max 560 tokens, sd 86.0), a metric that better corrects for text length was also included. We chose MTLD72MA, which is an updated version of the MTLD72 (discussed in McCarthy/Jarvis 2010). Broadly speaking, the MTLD is the ratio of tokens and so-called 'factors', that is subsequent sequences of tokens until a certain type-token-ratio threshold is reached (here: 0.72). MTLD72MA (MA stands for "moving-average") calculates these factors with moving windows, starting each time at the subsequent token (see Michalke 2020 for more details). Lexical sophistication (also sometimes called rarity) is measured using the mean Zipf-transformed frequency of the unique lemmas of each text (following Vanhove et al. 2019: 515).

There are many possible metrics that operationalize syntactic complexity of written texts. Certain metrics, such as T-unit complexity ratio or clausal coordination, seem less appropriate for written texts: As shown by Biber, Gray and Kornwipa (2011), an increase in this type of complexity is characteristic of conversational patterns, not writing. In writing, phrase complexity metrics (such as complex nominals and complex phrases) should be taken into account when investigating development. From the metrics generated by our pipeline, we thus chose the number of complex nominals per clause and the mean length of clause, as recommended by Biber, Gray and Kornwipa (2011) for the assessment of complexity in academic written prose. We use Lu's (2010: 483) operationalization and count as complex nominals all nouns with a modifying adjective, possessive, prepositional phrase, or with a relative clause, participle, or appositive, but also nominal clauses and gerunds and infinitives (the latter two only when in subject position).

The students had 30 minutes at their disposal, but there was some classroom-life related variability in how much time they really dedicated to the writing task. To compare how much text the learners produce within a given time unit, we divided the text length (number of tokens produced) by the total time the students worked on their laptops. This

metric ranges considerably from 2.83 tokens to 25 tokens per minute (mean=9.89, sd=3.33).

3.5 Accuracy metric

The text files produced by the students were processed both automatically and manually. The manual processing involved error correction of the texts, for two reasons: First, error correction is needed to assess the accuracy of the texts, and second spelling errors needed to be corrected so that the automatic analysis could identify the lemmas correctly.

Accuracy was calculated following Foster and Wigglesworth (2016: 104) by segmenting the texts into clauses. In a second step, three raters independently evaluated the clauses for spelling errors, syntactic/grammatical errors, and semantic/lexical errors. Punctuation was disregarded. All error types were coded individually, but no differential weighting was applied in the scoring system. Interrater reliability was assessed using the intra-class correlation coefficient (ICC) appropriate for interval-scaled data and multiple raters (Hallgren 2012). The ICC (two-way, consistency, average measures) was 0.98 (95 % CI: 0.969–0.987), indicating excellent agreement between raters. The accuracy metric used here is the ratio of correct clauses per total number of clauses produced in a given text. We generated two variants of the metric, one taking into account capitalization and the other not. The two are so highly correlated (0.96) that for the remainder of this contribution we only use the metric that does not take capitalization into account.

Table 2 gives an overview of the metrics used in our analyses.

abbreviation	**metric**	**construct**	**level**	**reference**
MLC	mean length of clause	complexity	clause	Lu (2010)
CN.C	complex nominals per clause	complexity	clause	Lu (2010)
Guiraud	Guiraud's index	complexity	vocabulary	Treffers-Daller/ Parslow/Williams (2016)
MTLD72MA	moving-average measure of textual lexical diversity	complexity	vocabulary	McCarthy/Jarvis (2010)
LemmameanZipfUnique	mean Zipf-transformed corpus frequency of the unique lemmas of the text	complexity	vocabulary	Vanhove et al. (2019)
accuracy	proportion of error-free clauses	accuracy	clause	Foster/Wiggles-worth (2016)
nTokensPerMinute	number of tokens produced per minute	complexity (fluency)	vocabulary	Struc/Wood (2011)

Table 2: The metrics used in the analyses

4 Results

4.1 Tool use

The first analysis is based on the InputLog data. To answer research questions (1) and (2) above, we identify the tools used by the students and calculate the number of switches to these tools on average for the two online conditions.

Figure 1 shows the relative switches to dictionaries and translation machines. The preferred choice rather clearly is the DeepL translator. However, this does not mean that the students predominantly had larger text chunks translated by this tool: As survey data show (Udry/Berthele 2023), learners often use translators as dictionaries, in that they simply type in a single target word rather than larger chunks. A qualitative examination of randomly picked InputLog files confirms this use of DeepL, GoogleTranslate or Microsoft Translator for single lexical items also in our sample.

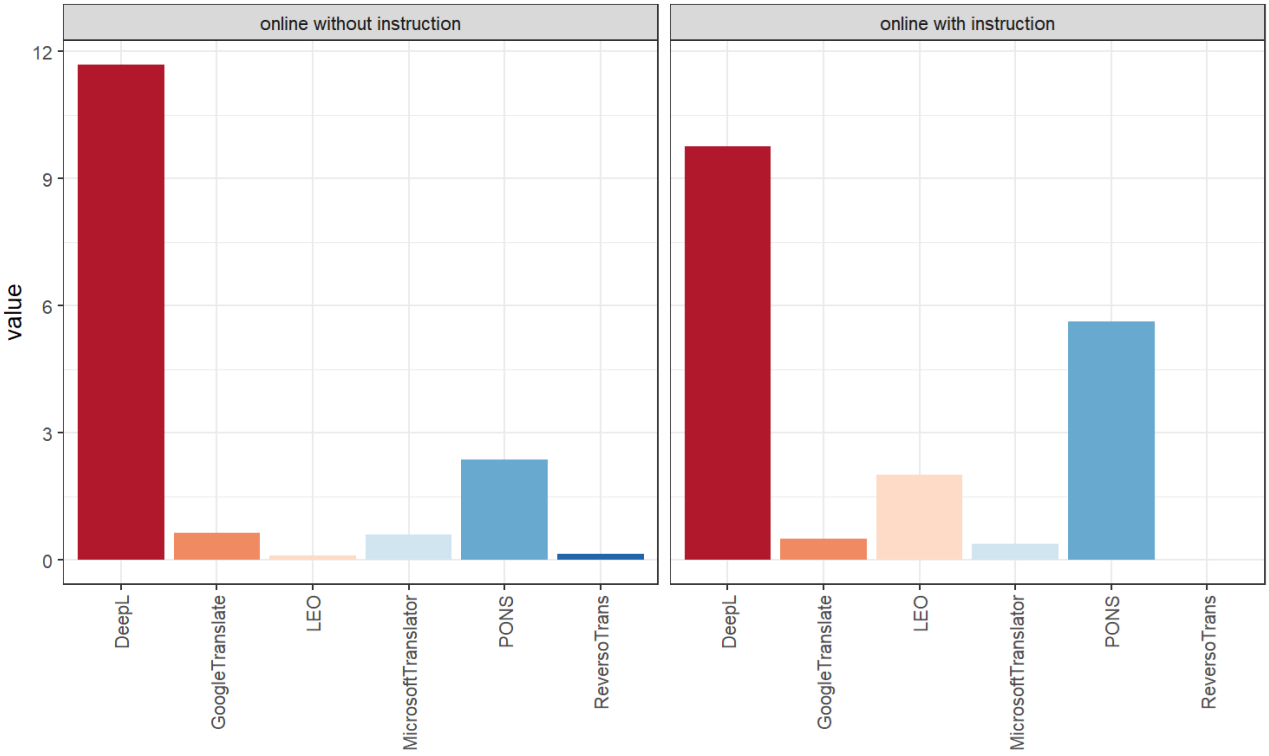

Fig. 1: Mean number of switches to specific tools in the writing task.

Table 3 shows the combination of tools across the two online conditions. The counts do not represent the number of switches, but only if a given tool was used or not.

Tools used	**instruction**	**no instruction**
DeepL	21 (30 %)	48 (59 %)
PONS	9 (13 %)	8 (10 %)
None	3 (4 %)	6 (7 %)
GoogleTranslate	0 (0 %)	4 (5 %)
PONS + DeepL	17 (25 %)	4 (5 %)
MicrosoftTranslator + DeepL	0 (0 %)	2 (2 %)
MicrosoftTranslator + ReversoTrans	0 (0 %)	2 (2 %)
DeepL + GoogleTranslate	0 (0 %)	1 (1 %)
DeepL + LEO	6 (9 %)	1 (1 %)
MicrosoftTranslator	0 (0 %)	1 (1 %)
MicrosoftTranslator + DeepL + GoogleTranslate	1 (1 %)	1 (1 %)
MicrosoftTranslator + LEO	0 (0 %)	1 (1 %)
PONS + GoogleTranslate	2 (3 %)	1 (1 %)
PONS + LEO	2 (3 %)	1 (1 %)
LEO	3 (4 %)	0 (0 %)
MicrosoftTranslator + PONS	3 (4 %)	0 (0 %)
PONS + DeepL + LEO	2 (3 %)	0 (0 %)

Table 3: Tools and tool combinations used across the two conditions. The percentages add up to 100 % per condition. The number of cases is slightly lower than in Table 1 because in some cases the InputLog software failed to record usable log files.

Our task instructions did neither constrain the learners regarding the choice of digital online tools nor regarding the intensity of their use. Therefore, the texts were produced with considerably variable tool use behaviour. Figure 2 sheds some light on these patterns by showing the number of switches to online tools and the total time spent online.

The figure shows that if a learner switches very frequently, they typically spend a bit less than half of the total time allocated for the task online. If switches are relatively infrequent, this represents a continuum of behaviour between two opposed poles: Some participants spend most of the time online (upper region on the y-axis), which means that they write their text in an online translation machine and copy and paste the translated text into Word (either once towards the end of the task or in a piecemeal way). Another subgroup spends relatively little time online and switches rarely or occasionally to online tools. At first sight this seems to characterize the behaviour of learners who genuinely make an effort to write in the target language and who use the tools in a focused way to solve specific language problems. This is the case for the points in Figure 2 that represent users who predominantly use dictionaries (dark blue). Qualitative exploration of the data shows that there is an alternative pattern: Some learners write their text in L1 in Word and copy and paste it later into a translation machine (and then come back and replace it with the translated text). Beyond these two extreme subgroups, many learners show an intermediate number of switches and spend an intermediate time span online.

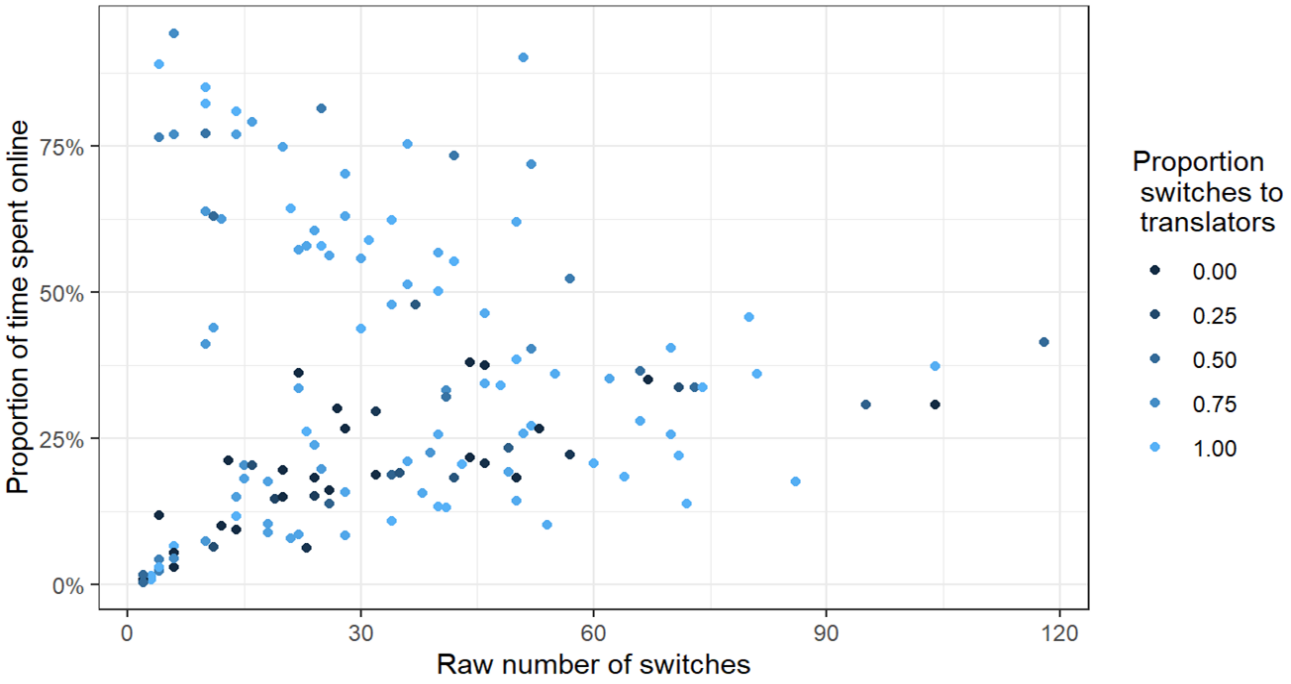

Fig. 2: Number of switches from Word to an online tool and the proportion of time spent online (online with and without instruction). The darker the points, the more dictionaries were used, the lighter the points, the more translators were used.

In sum, the analyses show that students prefer online translators (Figure 1) and that their switching behaviour varies greatly (Figure 2).

4.2 Text characteristics

The third research question above concerns the effects of digital tool use on the writing outcome. As discussed in the literature review above, there are different effects imaginable regarding the effects of the tool use on the written texts. If tools are allowed and used, the task changes in nature, which arguably influences the two components complexity and accuracy of the output. We investigate first the impact of the tool use on complexity and accuracy separately.

4.2.1 Complexity

The expectation regarding (lexical and structural) complexity is unclear. On the one hand, we hypothesize that using the tools adds complexity to the task, which might reduce complexity in the output due to increased cognitive load. On the other hand, the tools, if used productively, help planning and revising the text, which might decrease task complexity and lead to increased complexity in the output. Given the inconclusive previous evidence on the matter, our hypothesis on the effect of tool use on complexity is thus undirected:

Hypothesis on complexity and tool use (“condition tool use” in the models described below):

> If participants use online tools when writing the second texts, these texts’ complexity will change more compared to the first texts written (without tools) than if the second texts are again written without tools. Given the competing ideas and predictions discussed above, the complexity change induced by tool use can be positive or negative, and not all metrics of complexity may be affected in the same direction.
>
> Differences in complexity are operationalized within subjects as the difference between the second and the first texts.

Hypothesis on complexity and instruction on tools (“condition tool instruction” in the models described below):

> Complexity relative to the first texts written is higher if the online tool use is preceded by instruction compared to the condition without instruction. The directionality of this effect is grounded in the assumption that even brief instruction may foster the use of more complex means of expression.

As shown in Figure 3, the structural complexity (here we show the proportion of complex noun phrases per clause, see supplementary material for figures of all variables) does not change if students are using online tools, and there is no difference between the instructional conditions.

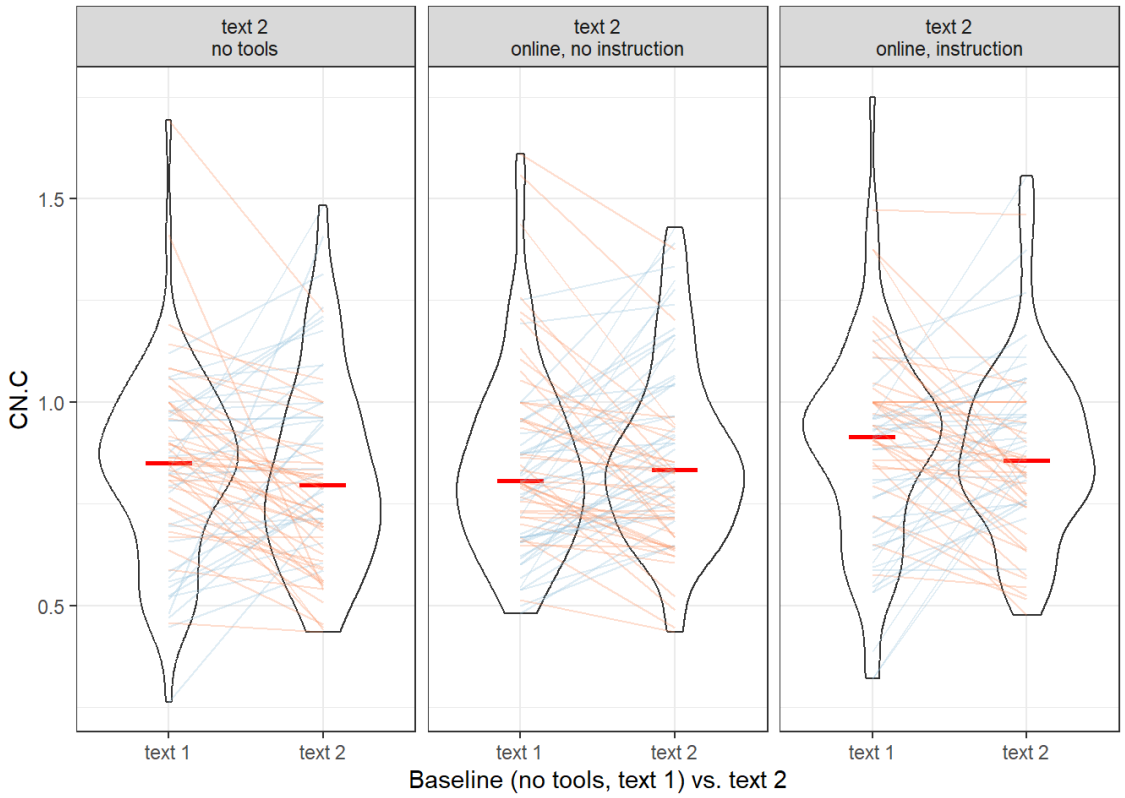

Fig. 3: Complex noun phrases per clause across the different conditions. Text 1 was always written without any tools, the second was written either again without tools (left panel), with tools but no instruction (middle panel), or with tools with instruction (right panel). The figure shows the density of the distributions and the median (horizontal line) and individual changes (red lines decrease between the two measurement points, blue lines increase).

The lexical complexity, as far as lexical diversity is concerned, increases slightly, in particular if we consider the MTLD72MA metric (Figure 4).

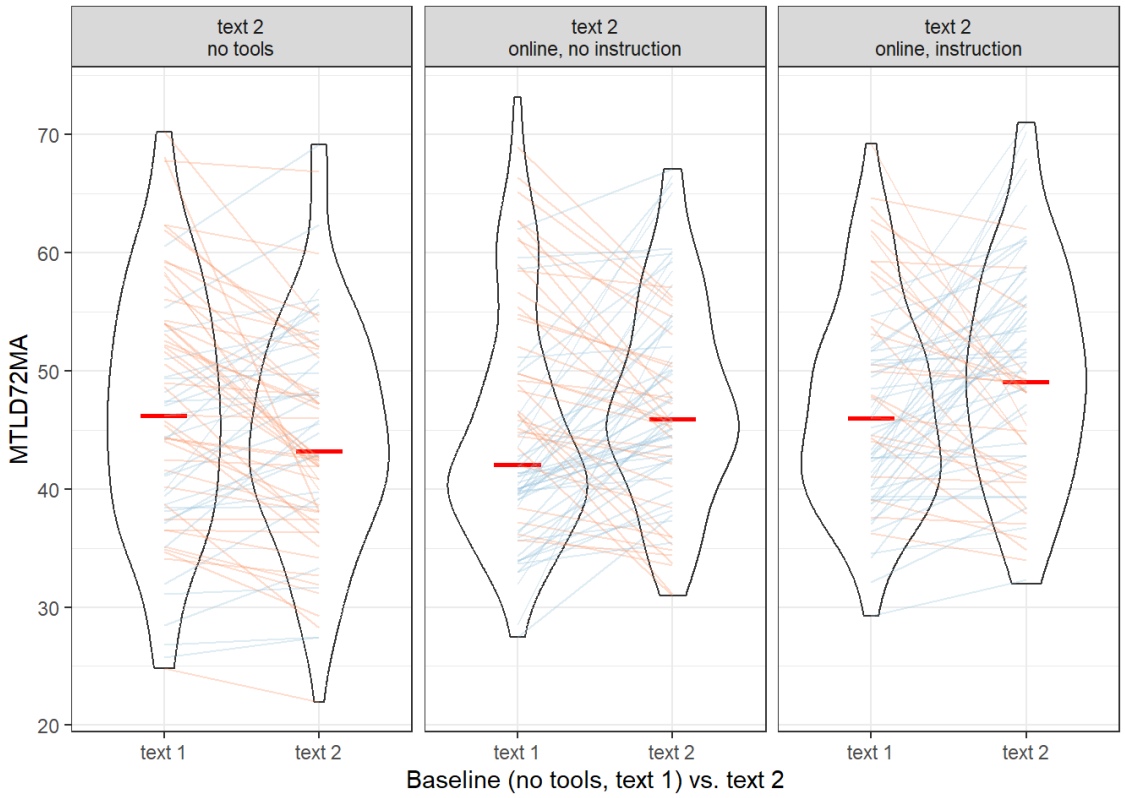

Fig. 4: Changes in lexical diversity (MTLD72MA) across the three conditions.

There is also a slight increase of the group average in the Guiraud metric, but only in the condition without instruction (see supplementary material, Figure S7).

Moreover, an alternative metric for lexical complexity, the relative rarity of the words used, increases visibly for the two online conditions (see Figure 5), but there is also a slight increase in the offline condition for the second text written by the students.

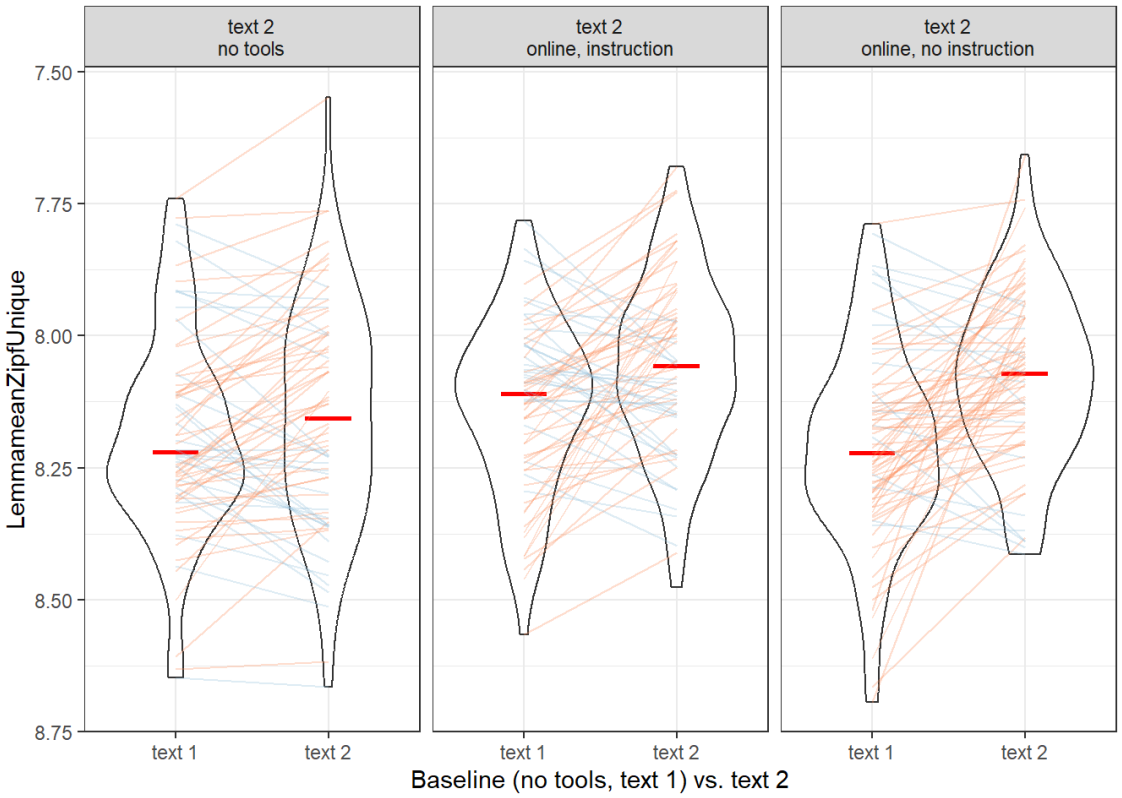

Fig. 5: Changes in lexical rarity as measured by the Zipf-transformed mean subtitle corpus frequency of the unique lemmas.

The number of tokens per time unit in the final product might be affected in different ways by tool use. Texts could be longer if large parts are written directly in the translator in the L1, as multilinguals are generally more fluent in their L1. This effect might even be larger if learners who work in the condition without tools spend much time thinking as they cannot get any help from online tools. Conversely, texts with the tools could be shorter, for at least two reasons: First, tool use takes up time and may thus compete with writing time. Second, texts may be shorter as the translators arguably reduce the typical translate-and-paraphrase phenomenon that leads to longer translations (Danielsson 2007). Thus, in this view, using automatic translators would lead to more condensed structures than non-assisted word-by-word translation.

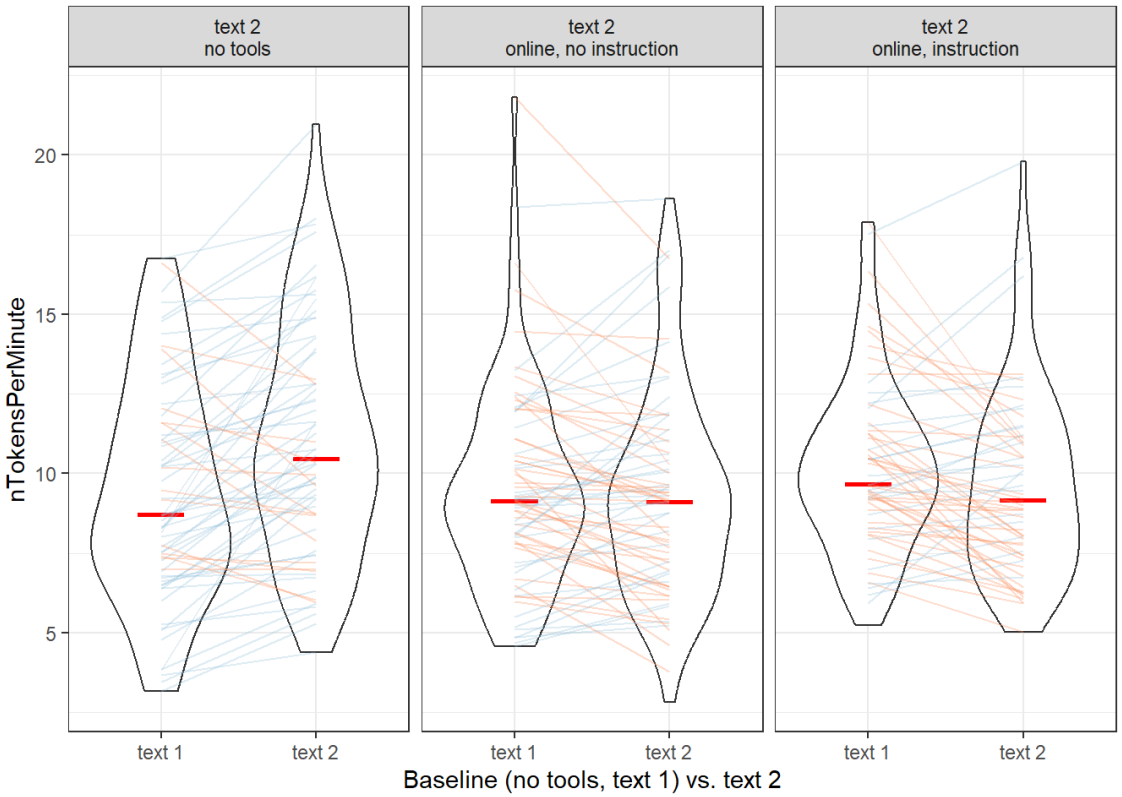

Fig. 6: Changes in number of tokens in final text per minute across the three conditions.

There seems to be a weak negative effect of tool use on the number of tokens written per minute (Figure 6). This is particularly the case if the offline condition for text 2 is taken into account: Students who work offline in the second task as well write more tokens per minute, whereas students who go online for text 2 write slightly fewer tokens. However, as the individual lines between text 1 and text 2 show, there is considerable inter-individual variability regarding this variable. As Figure 7 shows, the time spent online is weakly negatively associated with the number of tokens produced per minute for all types of switching (low, intermediate or frequent switches).

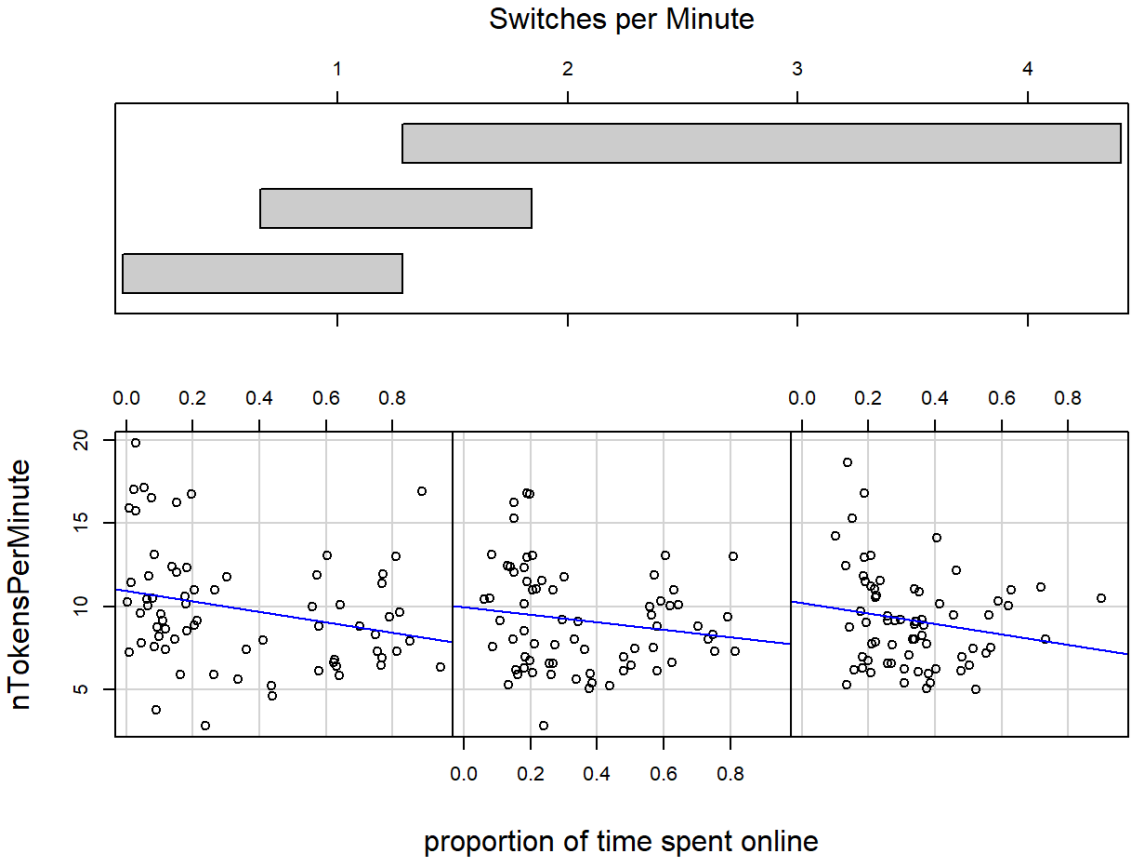

Fig. 7: Proportion of time spent online and the number of tokens written per minute; the three panels represent learners that show low, average, and high switching rates between the word processor and the online tools. The upper panel shows the range of average number of switches per group shown in the three panels.

As regards complexity, the data show that lexical complexity slightly increases with tool use, but that the two syntactic structural metrics do not change (see Figures 4 and 5 and S4 – S8 in the supplementary material). The two different online conditions do not show a differentiating impact on the texts produced.

In addition to visual inspection, we fitted regression models to further examine the data. The nominal predictors (the conditions in which the texts were written) were contrast coded following Vanhove (2020) so that the regression models directly test our hypotheses related to research question 3 above. All following models use the formula, with the different metrics as dependent variable (DV).

DV ~ studentconditions + (1|class)

Since our learners are clustered in classes, we included random effects per class (1|class) in the models. A first version with random intercepts per class and random slopes per class and condition (1+condition|class) yielded singular fits so that we only

included the random intercepts per class in the final models (see supplementary material for details). The variance explained by the class level (intra-class correlation) is very low (0.01, 0.02, and 0.06 for the three metrics respectively, cf. supplementary material for full detail). This means that the learners' responses are only very little influenced by their respective classes.

The results for the three lexical metrics discussed above are given in Table 4. Tables S6 and S7 in the supplementary material show the null results for the two syntactic complexity metrics for which the visual inspection already suggested that there is no effect across the conditions.

MTLD 72 MA	unstandardized		standardized			
Predictors	estimate (SE)	confidence interval	estimate	confidence interval	t value	p value
condition tool use (vs. no tool use)	5.01 (1.4)	2.27 – 7.76	0.51 (0.14)	0.23 – 0.79	3.58	0.001
condition instruction (vs. no instruction)	0.98 (1.64)	-2.24 – 4.19	0.1 (0.17)	-0.23 – 0.42	0.6	0.555
Guiraud						
condition tool use (vs. no tool use)	0.13 (0.09)	-0.05 – 0.32	0.21 (0.15)	-0.08 – 0.49	1.41	0.161
condition instruction (vs. no instruction)	-0.22 (0.11)	-0.44 – 0.00	-0.34 (0.17)	-0.68 – 0	-1.96	0.055
Lemma mean Zipf unique						
condition tool use (vs. no tool use)	0.07 (0.03)	0.02 – 0.13	0.38 (0.15)	0.08 – 0.68	2.52	0.013
condition instruction (vs. no instruction)	-0.05 (0.04)	-0.12 – 0.03	-0.24 (0.19)	-0.61 – 0.13	-1.28	0.209

Table 4: The impact of going online or not (condition tool use) and the difference of being online with or without instruction (condition tool instruction) on three lexical complexity metrics (raw and z-standardized dependent variables). The t and p values show that the tool use condition leads to significant changes for the MTLD 72 MA and the Lemma mean Zipf metrics.

Applying the conventional threshold of 5 % for the p value, the models listed in Table 4 show a significant effect of the online conditions compared to the offline condition for two out of the three metrics of lexical complexity. The effect has a negative sign for the lexical sophistication measure. Since lower values on this measure correspond to rarer, more sophisticated vocabulary, this negative effect should be interpreted as an increase in lexical sophistication. For lexical diversity, there is a positive effect on the MTLD72MA metric, but no significant effect for the Guiraud Index. A possible explanation for these non-aligned metrics is that the latter is less independent of text length than the former, and the texts in the online conditions tend to be somewhat shorter (see below, Table 6). The effect of condition only explains a small proportion of the overall variance in the data.

This is expressed by the low R^2 scores in all models fitted (Conditional R2 of 0.064, 0.047, and 0.102 respectively for the three lexical metrics, see supplementary material, Table S17). In addition to these lexical measures, text length also seems to be affected by tool use: As shown in Figure 7 and in the regression model (Table 5), the number of tokens produced per minute significantly decreases when online tools are being used.

4.2.2 Accuracy

The expectation regarding accuracy is straightforward and admittedly a bit trivial: We expect the texts written with the tools to be more accurate, as in particular the translation tools not only produce correct spelling, but also (more) correct syntax/grammar.

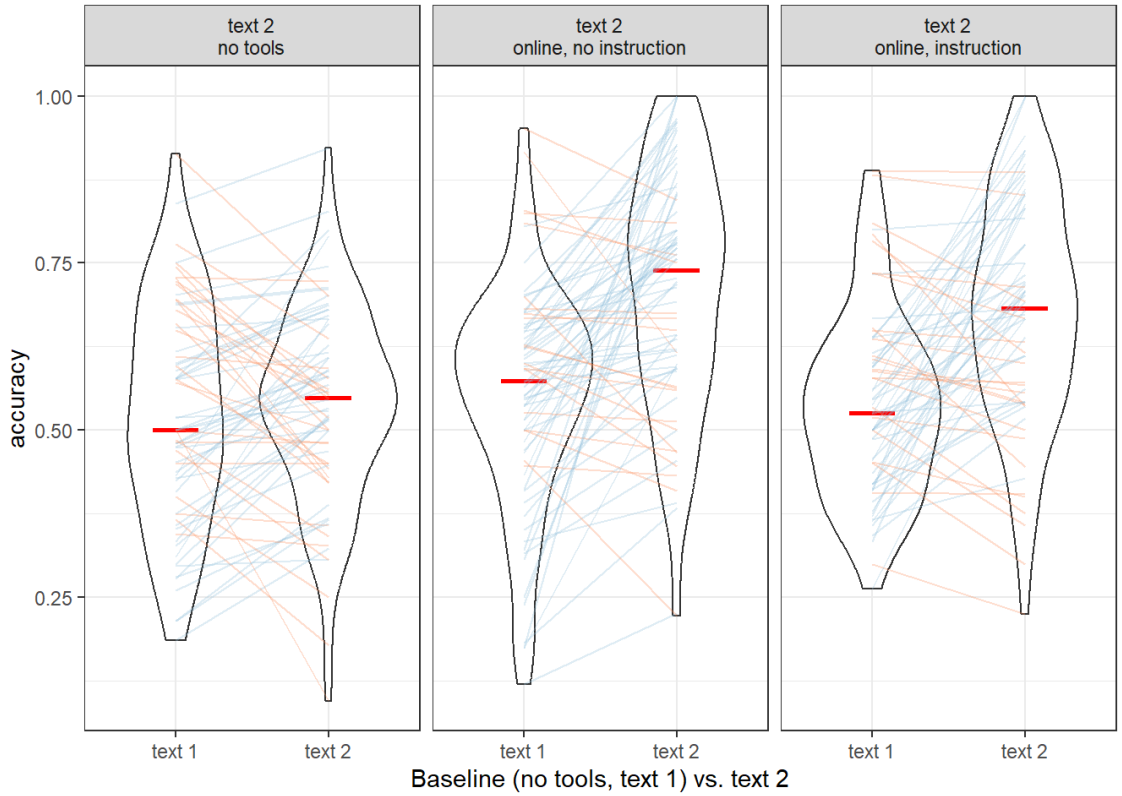

Fig. 8: Changes in accuracy across the three conditions.

There is a considerable positive impact of tool use on accuracy, for both online conditions compared to the offline condition. Accuracy in the two online conditions is positively associated with the time spent online (Figure 9), and there is no interaction with the average number of switches between the online tool and the word processor. Thus, regardless of whether a learner switches very frequently or infrequently, the more time they spend online, the more accurate the text will be. This difference, as shown by the regression analysis in Table 5, is statistically significant. Again, no impact is found of the instruction on online tools.

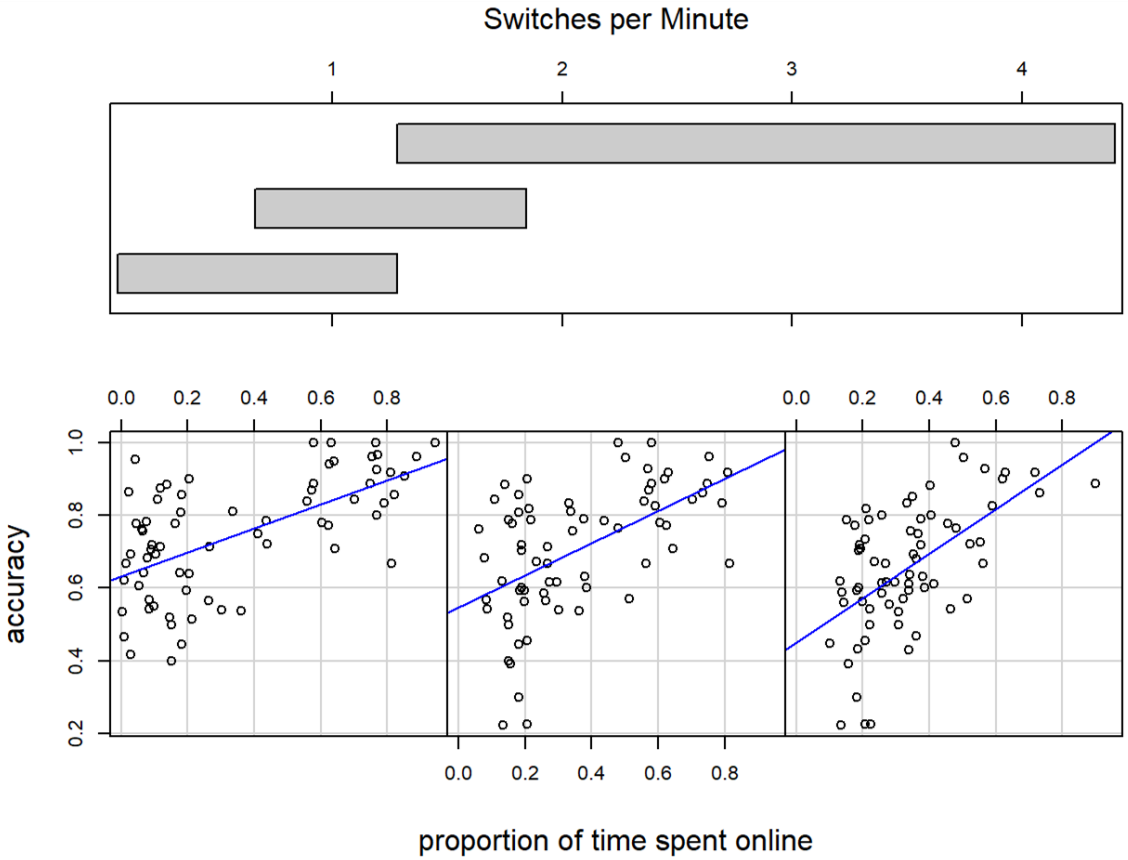

Fig. 9: Proportion of time spent online and accuracy; the three panels represent learners that show low, average, and high switching rates between the word processor and the online tools.

n tokens per minute	unstandardized		standardized			
Predictors	estimate (SE)	confidence interval	estimate	confidence interval	t value	p value
condition tool use (vs. no tool use)	-1.98 (0.36)	-2.7 – -1.27	-0.80 (0.15)	-1.09 – -0.51	-5.44	<0.001
condition instruction (vs. no instruction)	-0.4 (0.47)	-1.32 – 0.52	-0.16 (0.19)	-0.53 – 0.21	-0.85	0.403
accuracy						
condition tool use (vs. no tool use)	0.13 (0.03)	0.07 – 0.19	0.65 (0.15)	0.36 – 0.93	4.46	<0.001
condition instruction (vs. no instruction)	-0.01 (0.04)	-0.08 – 0.06	-0.06 (0.18)	-0.41 – 0.28	-0.36	0.722

Table 5: The effect of going online or not (condition tool use) and the difference of going online with or without instruction (condition tool instruction) on the text length and accuracy metrics (raw and z-standardized dependent variables).

The model summaries in Table 5 show a statistically significant impact of going online on both variables, a negative one for the change in number of tokens produced per minute, and a positive effect for change in accuracy. There is no difference between the two online conditions. The effect of condition again explains only a relatively small proportion of the variance in the data, which is expressed by the low R^2 scores in all models fitted (Conditional R2 of 0.214 for n tokens per minute and 0.126 for accuracy, see supplementary material for full detail). As in the previous analysis, the intra-class correlation is low (conditional ICC 0.069 and 0.033).

4.2.3 Task achievement and trade-offs

Metrics of complexity and accuracy do not necessarily measure features that are directly related to the degree to which the task (here: write an application letter/email) can be considered successfully achieved by the text. It seems nevertheless interesting to explore how metrics of complexity and accuracy are associated with task achievement. Table 6 shows that more complexity and more accuracy are consistently positively associated with the attainment of the objectives formulated in the task instruction as assessed by human raters (task achievement). The correlations of the three lexical complexity metrics and task achievement are all moderate. Lexical sophistication shows a relatively high linear association, in particular in the offline condition, suggesting that the use of more sophisticated vocabulary coincides with better fulfilment of the communicative task. The relatively low association of the number of tokens per minute is not surprising, as this variable is not directly visible in the output and the positive association is thus an effect of this variable's positive association with surface features of the texts.

We have discussed different expectations regarding trade-offs among the three dimensions in the literature review section above. There is disagreement in the literature as to whether there is competition between some of the components or, on the contrary, mutual positive influence. If accuracy and complexity go hand in hand, as Robinson (2001: 306) claims, no trade-off between the two dimensions and thus no negative association between metrics operationalizing the two constructs is expected. On the contrary, an increase in one would be associated with an increase in the other. Following this rationale, we can hypothesize that if the tools help with accuracy, they should thus also help with complexity.

To get an overview of the direction and strength of the linear associations, we can take a look at correlation matrices. Figures S14 and S15 in the supplementary material show heat-maps of all associations of our metrics in the online and offline conditions. Since we are most interested in the trade-offs between accuracy and all other variables, Table 6 shows the correlation coefficients for each variable with accuracy in both conditions. The table also shows the linear association between the complexity metrics and task achievement as rated by a panel of human raters.

Variable	correlations with accuracy (offline)	correlations with accuracy (online)	correlations with task achievement (offline)	correlations with task achievement (online)
MLC	-0.16	0.05	0.30	0.12
CN.C	-0.10	0.02	0.30	0.06
Guiraud	0.35	0.08	0.52	0.35
MTLD72MA	0.28	0.14	0.40	0.17
LemmameanZipfUnique	0.12	0.21	0.45	0.27
nTokensPerMinute	0.09	-0.11	0.24	0.22
task achievement	0.50	0.36	NA	NA

Table 6: Correlations of all variables with accuracy and with task achievement (offline and online conditions). The metric for lexical sophistication (LemmameanZipfUnique) was recoded here so that higher values represent more sophistication.

If there were a clear trade-off between accuracy and the other metrics, we would expect there to be negative correlations. There are two weak negative correlations in the offline condition for the two syntactic metrics; all other coefficients are positive. In the online condition, all associations except one are positive. There is a weak negative association with the number of tokens produced per minute, a variable that is also negatively associated with other complexity variables.

Overall, no strong relationships between accuracy and complexity emerge from the data. There is weak evidence for the aforementioned trade-off between complexity and accuracy in the offline condition. In the online condition, there is no evidence for such a trade-off. The metrics we discussed all show the expected positive association with task achievement, which confirms that they are related to what human raters perceive as relevant aspects of the texts with respect to the communicative goal they are supposed to reach. In the online condition, these associations are weaker, which could be an effect of less variability in these metrics than in the offline condition.

5 Discussion

The real-world question underlying our research project is in what respect exactly texts written with online tools, both translation machines and/or dictionaries, are different compared to texts the same learners write without any such help. The dimensions along which we analysed the texts were complexity (lexical and syntactic) and accuracy (proportion of error-free clauses).

5.1 Dimension-specific impact of digital tool use

The analyses above show that the texts are indeed different, but not regarding all analytical dimensions. The most important difference is the least surprising one: Online

tools clearly help learners to produce more accurate texts in the target language. Not only are texts produced with these tools overall more accurate, there is also a weak positive linear relationship between the time spent online and the accuracy of the texts.

The second difference pertains to the lexical complexity: Texts written with online tools tend to include more sophisticated (i. e. less frequently occurring) lemmas than texts written without such tools. Moreover, there is a tendency towards a more diverse use of lemmas compared to the condition without online tool use. These differences are not enormous, but statistically significant. Syntactic complexity was analysed using two metrics where higher complexity is associated with longer clauses and more complex noun phrases. Neither of these two indicators is associated in a significant way with the conditions investigated here. Thus, it seems that the main impact of the use of digital tools on complexity lies on the level of the lemmas selected in writing.

The impact on the amount of text produced is again measurable: Writing in the online conditions tends to lead to slightly fewer tokens produced per time unit. Different explanations are possible for this finding. On the one hand, the effect might be that doing the writing task while using online tools divides the time between writing in the word processor and interacting with the online tools, thus leading to shorter texts simply because less time is dedicated to the actual writing. This is corroborated by the insight that learners who spend more of the task time online write less than those who spend little time online (Figure 7). On the other hand, using online tools, in particular translation machines, produces more idiomatic and therefore often less periphrastic text, which could also account for overall shorter productions. The two explanations are not mutually exclusive, and the effect overall is not dramatic.

5.2 Associations and trade-offs

Based on the literature on the topic and on the unclear expected effects of the online condition on task complexity, no clear predictions were formulated regarding potential trade-offs between accuracy and complexity. As discussed above, the online variant of the writing task is arguably more cognitively demanding due to the switches between resources, which may entail lower complexity of the written text, but on the other hand, it may also foster more complexity in the texts due to the scaffolding impact of dictionaries and translators.

The exploration of these associations shows no evidence for consistent and strong negative associations of the two dimensions of complexity and accuracy. In the offline condition, while the lexical metrics all point towards a positive association of richness (more different and more sophisticated words are used when texts are more accurate), the two syntactic metrics show weak negative associations. In the online condition, these weak negative estimates disappear (and converge to zero), while the lexical metrics stay positive but weak (diversity) or become stronger (sophistication). Using language-support tools thus does not seem to increase the competition between the two attentional foci (one on accuracy, the other on complexity). Thus, while the offline condition yields inconsistent results with a possible weak trade-off regarding syntactic complexity and

accuracy, the data gathered in the online condition seem to point more towards the absence of a trade-off and to a weakly positive relationship between the two components.

Overall, no strong trade-off effects can be found in either condition, and if anything, the data seem slightly more compatible with Robinson's idea that complexity and accuracy are not in competition but go hand in hand.

5.3 Limitations and outlook

The analysis presented here pertains to the quality of the writing product in terms of complexity and accuracy. Several limitations can be identified that could be considered in future research addressing the effects of digital tool use on various L2 writing processes as well as on target language learning. First, more process-oriented information could be taken into account. One hypothesis outlined above was that access to online dictionaries and translators would allow participants to dedicate more attentional resources to the planning and revision of their texts. Examining the log files in more detail with respect to time dedicated to each stage of writing could address this question.

Further analyses also could involve a less theory-driven and more data-driven approach to text classification. Given the ease with which the corpus tools used can produce a multitude of metrics, many more indicators, in particular of complexity on the lexical, phrasal, and clausal level, could be produced and the associations of single or combinations of indicators with the online/offline conditions could be further investigated. Moreover, as we have texts written by learners both enrolled in French and German curricula, the specific impact of the first and dominant languages, French and German, could be investigated.

Finally, this study focuses on digital translation and dictionary tools as they are commonly used in foreign language writing today. However, as one reviewer rightly points out, a broader comparison including offline tools such as printed dictionaries would further contextualize our findings. We have addressed this question in a related study (Udry/Berthele 2025), which compared the effects of using online and paper dictionaries on vocabulary retention following a meaning-focused reading task. That study found significantly better vocabulary learning gains for the online dictionary condition, suggesting that the medium and format of support tools can indeed influence learning processes. While our present study did not include paper-based tools, these earlier findings reinforce the importance of differentiating between tool types, both in terms of their design (digital vs. print) and their mode of use (whole-text translation vs. selective word lookup). Future research should further explore how these factors interact across different language skills and learner populations.

6 Conclusion

Our study investigates two domains: First, we shed light on the question what intermediate learners use in terms of online tools, and how they use them when carrying out a writing task. Second, we investigate the impact of these (varying) uses of online language tools on the product of the writing task, with a focus on the two dimensions complexity and accuracy.

Our results show that the dimensions under investigation change in a differentiated way when such tools are used, and that the main impact of tool use concerns the accuracy and the lexical complexity of the texts.

Writing tasks in foreign/second languages that involve different degrees of digital tool usage are common in real life. We believe it is important to acknowledge this also in foreign/second language teaching and think of ways to include digital translation aids in everyday classroom practice (Berthele/Udry 2023; Udry/Berthele 2023). For example, the tools may be used to encourage noticing (Schmidt 1993) of salient text features by having students compare and/or correct juxtaposed learner-generated and automatically generated (translated) texts.

The effect of our intervention, aiming at making tool use more adequate and efficient, was negligible in terms of measurable changes in learners' written output. One likely reason is that the intervention was short, spanning only one lesson. While this may have been sufficient to raise awareness, it was not enough to consolidate or automatize new behaviours. Further studies should take this into account and vary the intensity and duration of the instructional component to allow for more extensive practice and reflection. Moreover, the question arises under what circumstances learners actually acquire language-related knowledge through tool use. Thus, investigations of varying degrees of involvement load while looking up words or translating text chunks would be a welcome extension of the work presented here.

Lastly, including tools that not only translate but also generate texts, such as Microsoft Bing Chat or ChatGPT, should be included in the empirical investigation of digital tool use in foreign/second language learning and teaching.

References

Baddeley, Alan D. (2000): "The episodic buffer: A new component of working memory?" *Trends in Cognitive Sciences* 4 [11]: 417–423 – https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, Alan D.; Graham Hitch (1974): "Working memory." Gordon H. Bower (ed.): *Psychology of learning and motivation*. New York etc.: Academic Press, 52–62

Berthele, Raphael; Isabelle Udry (2023): "Welche fremdsprachlichen Lernziele der Schulen können mit digitalen Tools aber ohne Lernen erreicht werden? Simply playing the ostrich will not help in the long run." *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 28 [1]: 443–461

Biber, Douglas; Bethany Gray, Poonpon Kornwipa (2011): "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?" *TESOL Quarterly* 45 [1]: 5–35 – https://doi.org/10.5054/tq.2011.244483

Danielsson, Pernilla (2007): "In other words: Using paraphrases in translation." *Proceedings of Translating and the Computer* 29: 29–30– https://aclanthology.org/2007.tc-1.5

"The effects of digital tool use on complexity and accuracy of written texts in EFL." (2025). OSFhome – https://osf.io/6f4xm/?view_only=02419aa420ed476a8e828f0bc2cce2e7 (31 August 2025)

Foster, Pauline; Gillian Wigglesworth (2016): "Capturing accuracy in second language performance: The case for a weighted clause ratio." *Annual Review of Applied Linguistics* 36: 98–116 – https://doi.org/10.1017/S0267190515000082

Frear, Mark Wain; John Bitchener (2015): "The effects of cognitive task complexity on writing complexity." *Journal of Second Language Writing* 30: 45–57 – https://doi.org/10.1016/j.jslw.2015.08.009

Hallgren, Kevin A. (2012): "Computing inter-rater reliability for observational data: An overview and tutorial." *Tutorials in Quantitative Methods for Psychology* 8 [1]: 23–34

Hayes, John R.; Linda S. Flower (1980): "Identifying the organization of writing processes." Lee W. Gregg, Erwin R. Steinberg (eds): *Cognitive processes in writing*. London/Oxon/New York: Routledge, 3–30

trans-kom **ISSN 1867-4844**

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: https://www.trans-kom.eu

Housen, Alex (2021): “Difficulty and complexity of language features and second language instruction.” Carol A. Chapelle (ed.): *The encyclopedia of applied linguistics.* Hoboken N. J.: Wiley, 1–7 – https://doi.org/10.1002/9781405198431.wbeal1443.pub2

Housen, Alex; Folkert Kuiken (2009): “Complexity, accuracy, and fluency in second language acquisition.” *Applied Linguistics* 30 [4]: 461–473 – https://doi.org/10.1093/applin/amp048

Housen, Alex; Hannelore Simoens (2016): “Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition.” *Studies in Second Language Acquisition* 38 [2]: 163–175

Johnson, Mark D. (2017): “Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis.” *Journal of Second Language Writing* 37: 13–38 – https://doi.org/10.1016/j.jslw.2017.06.001

Kellogg, Ronald T. (1996): “A model of working memory in writing.” Michael Levy, Sarah Ransdell (eds): *The science of writing: Theories, methods, individual differences, and applications*. Mahwah (NJ): Lawrence Erlbaum Associates, 57–71

Kellogg, Ronald T.; Alison P. Whiteford, Casey E. Turner, Michael Cahil, Andrew Mertens (2013): “Working memory in written composition: An evaluation of the 1996 model.” *Journal of Writing Research* 5 [2]: 159–190

Kormos, Judit (2011): “Task complexity and linguistic and discourse features of narrative writing performance.” *Journal of Second Language Writing* 20 [2]: 148–161 – https://doi.org/10.1016/j.jslw.2011.02.001

Kuiken, Folkert; Ineke Vedder (2008): “Cognitive task complexity and written output in Italian and French as a foreign language.” *Journal of Second Language Writing* 17 [1]: 48–60 – https://doi.org/10.1016/j.jslw.2007.08.003

Leijten, Mariëlle; Luuk Van Waes (2013): “Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes.” *Written Communication* 30 [3]: 358–392

Lu, Xiaofei (2010): “Automatic analysis of syntactic complexity in second language writing.” *International Journal of Corpus Linguistics* 15 [4]: 474–496 – https://doi.org/10.1075/ijcl.15.4.02lu

McCarthy, Philip M.; Scott Jarvis (2010): “MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment.” *Behavior Research Methods* 42 [2]: 381–392 – https://doi.org/10.3758/BRM.42.2.381

Michalke, Meik (2020): koRpus: An r package for text analysis. [Manual] – https://reaktanz.de/?c=hacking&s=koRpus

Michel, Marije; Andrea Révész, Xiaojun Lu, Nektaria-Efstathia Kourtali, Minjin Lee, Laís Borges (2020): “Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study.” *Second Language Research* 36 [3]: 307–334

“Online translation tools.” (s. a.). Université de Fribourg / Universität Freiburg – https://digitalvocabulary.wordpress.com/ (31 August 2025)

Pallotti, Gabriele (2009): “CAF: Defining, refining and differentiating constructs.” *Applied Linguistics* 30 [4]: 590–601 – https://doi.org/10.1093/applin/amp045

Robinson, Peter (2001): *Cognition and second language instruction*. Cambridge: Cambridge University Press

Schmidt, Richard (1993): “Awareness and second language acquisition.” *Annual Review of Applied Linguistics* 13: 206–226

Silva, Tony (1993): “Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications.” *TESOL Quarterly* 27 [4]: 657–677 – https://doi.org/10.2307/3587400

Skehan, Peter (1996): “A framework for the implementation of task-based instruction.” *Applied Linguistics* 17: 38–62

Skehan, Peter (2003): *A cognitive approach to language learning*. 2nd edn. Oxford etc.: Oxford University Press

Skehan, Peter (2009): "Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis." *Applied Linguistics* 30 [4]: 510–532 – https://doi.org/10.1093/applin/amp047

Skehan, Peter (2014): "Limited attentional capacity, second language performance and task-based pedagogy." Peter Skehan (ed.): *Processing perspectives on task performance*. Amsterdam etc.: Benjamins, 211–260

Skehan, Peter; Pauline Foster (1997): "Task type and task processing conditions as influences on foreign language performance." *Language Teaching Research* 1 [3]: 185–211

Skehan, Peter; Pauline Foster (1999): "The influence of task structure and processing conditions on narrative retellings." *Language Learning* 49 [1]: 93–120 – https://doi.org/10.1111/1467-9922.00071

Skehan, Peter; Pauline Foster (2001): "Cognition and tasks." Peter Robinson (ed.): *Cognition and second language instruction*. Cambridge: Cambridge University Press, 183–205

Socher, Richard; John Bauer, Christopher D. Manning, Andrew Y. Ng (2013): "Parsing with compositional vector grammars." *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria.* Vol. 1: *Long Papers*, 455–465 – https://aclanthology.org/P13-1045/

Struc, Nicolai; Nicholas Wood (2011): "A corpus-based investigation of syntactic complexity, fluency, sentence variety, and sentence development in L2 genre writing." 麗澤大学紀要 / *Reitaku University Journal* [93]: 45–79 – https://api.semanticscholar.org/CorpusID:60385775

Treffers-Daller, Jeanine; Patrick Parslow, Shirley Williams (2016): "Back to basics: How measures of lexical diversity can help discriminate between CEFR levels." *Applied Linguistics* 39 [3]: 302–327 – http://applij.oxfordjournals.org/content/early/2016/04/07/applin.amw009.abstract

Udry, Isabelle; Raphael Berthele (2023): "Digitale Übersetzungsprogramme und Online-Wörterbücher im Fremdsprachenunterricht: Erkenntnisse aus einer Umfrage bei Lehrpersonen und Lernenden der obligatorischen, post-obligatorischen und tertiären Bildung." *Linguistik Online* 120 [2]: 45–167

Udry, Isabelle; Raphael Berthele (2025): "A case study of online and paper bilingual dictionaries and their impact on vocabulary learning through meaning-focused reading." *International Journal of Lexicography* [advance access publication June 2025] – https://doi.org/10.1093/ijl/ecaf004

van Heuven, Walter J. B.; Pawel Mandera, Emmanuel Keuleers, Marc Brysbaert (2014): "Subtlex-UK: A new and improved word frequency database for British English." *Quarterly Journal of Experimental Psychology* 67 [6]: 1176–1190 – https://doi.org/10.1080/17470218.2013.850521

Vanhove, Jan (2020): "Tutorial: Obtaining directly interpretable regression coefficients by recoding categorical predictors." *Blog Posts Jan Vanhove* – https://janhove.github.io/analysis/2020/05/05/contrast-coding (20 January 2025)

Vanhove, Jan; Audrey Bonvin, Amelia Lambelet, Raphael Berthele (2019): "Predicting human lexical richness ratings of short French, German, and Portuguese texts using text-based indices." *Journal of Writing Research* 10 [3]: 499–525 – https://doi.org/10.17239/jowr-2019.10.03.04

Wijffels, Jan (2018): *Udpipe tokenization, parts of speech tagging, lemmatization and dependency parsing with the "UDPipe" "NLP" toolkit* – https://CRAN.R-project.org/package=udpipe (20 January 2025)

Willis, Jane (1996): "A flexible framework for task-based learning." David Willis, Jane Willis (eds): *Challenge and change in language teaching*. Oxford: Macmillan Education, 52–62

trans-kom 18 [2] (2025): 625–655

Authors

Raphael Berthele is Professor in multilingualism at the University of Fribourg/Freiburg (Switzerland). He studied and worked at the Universities of Fribourg, Tübingen, Berkeley, and Berne. His research interests cover different areas from cognitive to social aspects of multilingualism.
E-mail: raphael.berthele@unifr.ch
Website: https://institut-plurilinguisme.ch/fr/personnel/berthele-raphael
ORCID: 0000-0003-1403-4116

Isabelle Udry is a researcher and lecturer at the Institute of Multilingualism at the University of Fribourg and the Zurich University of Teacher Education. She is interested in the cognitive processes that underlie foreign language learning and their contribution to developing teaching methodology. Her research is focused on individual differences, language aptitude, and the use of digital technology in the language classroom.
E-mail: isabelle.udry@unifr.ch
Website: https://institut-plurilinguisme.ch/de/staff/isabelle-udry
ORCID: 0000-0003-3879-7257

Didier Yourassoff recently gratuated with a master's degree in Cognitive Science from the University of Neuchâtel. During his studies, he also worked at the Institute of Multilingualism at the University of Fribourg. His master's thesis examined the description of sounds through typically visual properties.
E-mail: didier.yourassoff@unine.ch

Empfehlungen

TransÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Herausgegeben von Hartwig Kalverkämper und Sylvia Reinart

Hervé Delplanque / Stephanie Schwerter (dir.): **L'incompréhension culturelle à l'épreuve du droit.** ISBN 978-3-7329-1071-7

Easy – Plain – Accessible

Herausgegeben von Silvia Hansen-Schirra und Christiane Maaß

Laura Marie Maaß: **Erwartungen, Einstellungen, Erfahrungen. Zur Interaktion zwischen hörenden Gebärdensprachdolmetschenden und ihrer tauben Kundschaft.** ISBN 978-3-7329-1161-5

Sarah Ahrens: **Einfache Sprache in der Gesundheitskommunikation. Patientinnenaufklärung für Frauen mit Deutsch als Zweitsprache.** ISBN 978-3-7329-1132-5

Anke Radinger: **Researching Subtitling Processes. Methodological considerations for the investigation of AI-assisted subtitling workflows.** ISBN 978-3-7329-1029-8

Forum für Fachsprachen-Forschung

Herausgegeben von Hartwig Kalverkämper

Claudio Di Meola / Joachim Gerdes / Livia Tonelli (Hg.): **Sprachvariation im Deutschen zwischen Theorie und Praxis. Fachsprachlichkeit, Inklusion, Didaktik, Übersetzung, Kontrastivität.** ISBN 978-3-7329-1042-7

Anna Malena Pichler / Barbara von der Lühe / Felicitas Margarete Tesch (Hg.): **Deutsch als Fremd- und Fachsprache im internationalen Kontext.** ISBN 978-3-7329-1047-2

Daniel Green (Hg.): **The Future of Teaching Law and Language.** ISBN 978-3-7329-1090-8

Audiovisual Translation Studies (AVTS)

Herausgegeben von Marco Agnetta und Alexander Künzli

Marco Agnetta / Astrid Schmidhofer / Alena Petrova (Hg.): **Bild – Ton – Sprachtransfer. Neue Perspektiven auf Audiovisuelle Translation und Media Accessibility.** ISBN 978-3-7329-0921-6

Alexander Künzli / Klaus Kaindl (Hg.): **Handbuch Audiovisuelle Translation. Arbeitsmittel für Wissenschaft, Studium, Praxis.** ISBN 978-3-7329-0981-0

Transkulturalität – Translation – Transfer

Herausgegeben von Martina Behr, Larisa Schippel und Julia Richter

Michael Schreiber: **Die Übersetzungspolitik der Französischen Revolution und der napoleonischen Zeit. Studien zu Übersetzungen ins Deutsche, Italienische, Niederländische und Kreolische.** ISBN 978-3-7329-1141-7

Leipzig Middle East Studies

Herausgegeben von Sebastian Maisel

Ahmed Ibrahim Abdallah Mohmmed: **Übersetzung als interkultureller Dialog. Sudanesische Romane in deutscher Übersetzung.** ISBN 978-3-7329-1087-8

Malek Al Refaai: **Challenges in Interpretation Services for Syrian Refugees.** ISBN 978-3-7329-1193-6

Magdalena Zehetgruber / Bernadette Hofer-Bonfim / Elisabeth Peters / Johannes Schnitzer (Hg.): **Linguistic Diversity in Professional Settings.** ISBN 978-3-7329-1089-2

Lola Debüser / Jekatherina Lebedewa: **Ein deutsch-russisches Leben.** ISBN 978-3-7329-1181-3

Alle Bücher sind auch als E-Books erhältlich.

Frank & Timme

Wittelsbacherstraße 27a, 10707 Berlin • Telefon +49 (0)30 88 66 79 11 • info@frank-timme.de • www.frank-timme.de