Maintaining Reading Flow in E-Readers with Interactive Grammar Augmentations for Language Learning

Fiona Draxler
fiona.draxler@ifi.lmu.de
LMU Munich
Munich, Germany

Viktoriia Rakytianska
v.rakytianska@campus.lmu.de
LMU Munich
Munich, Germany

Albrecht Schmidt
albrecht.schmidt@ifi.lmu.de
LMU Munich
Munich, Germany

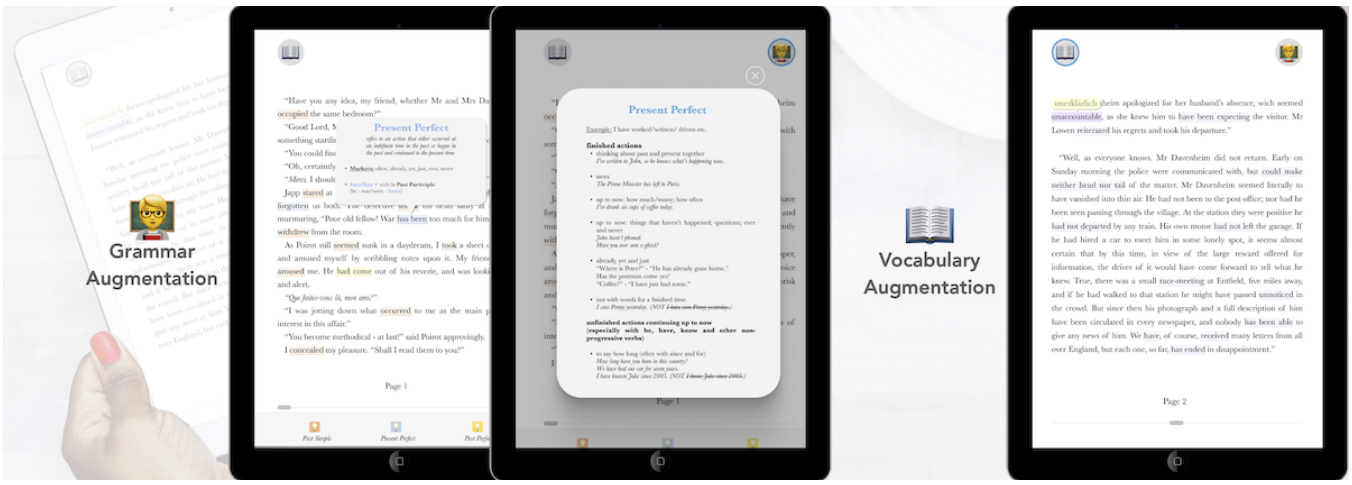


Figure 1: Interactive textual enhancement for e-readers: Grammar and Vocabulary Augmentation

ABSTRACT

Books can be a valuable resource for language learners, providing entertainment and showcasing authentic language usage. To further support learners, e-book platforms already include interactive vocabulary aids. Similarly, research has started to investigate the feasibility and usability of grammar aids. However, grammar aids can hinder reading flow. In this paper, we design an interactive e-reading interface with different levels of grammar support. We perform a within-subject user study (n = 24) where we assess the relationship between the reading flow, usability, and usefulness with these designs in a fiction reading scenario. Our findings show that more detailed designs are considered more useful, but also more disruptive and less usable than simpler designs. Hence, grammar support interfaces need to balance the focus on learning and the reading experience.

MUM 2022, November 27–30, 2022, Lisbon, Portugal

ACM ISBN 978-1-4503-9820-6/22/11…$15.00
https://doi.org/10.1145/3568444.3568459

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments**; • **Human-centered computing** → *E-book readers*; Interactive systems and tools.

KEYWORDS

reading flow, e-books, language learning

ACM Reference Format:
Fiona Draxler, Viktoriia Rakytianska, and Albrecht Schmidt. 2022. Maintaining Reading Flow in E-Readers with Interactive Grammar Augmentations for Language Learning. In *21th International Conference on Mobile and Ubiquitous Multimedia (MUM 2022), November 27–30, 2022, Lisbon, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3568444.3568459

1 INTRODUCTION

Reading foreign-language books gives learners a perfect opportunity to combine entertainment and learning. Reading comprehension is one of the core capabilities in a language [29] and unsurprisingly, it develops with increased exposure to texts [30]. To some extent, new words and grammatical structures can be grasped intuitively while reading [10, 19]. For fully understanding what a word means, or when one tense or another is used exactly, external help can be beneficial. E-readers can be a particularly useful tool here thanks to integrated dictionaries and the possibility to implement grammar augmentations. However, this support should not compromise enjoyment and reading flow. In other words, it is necessary to develop reading support systems that enable readers to balance information load and interruptions.

For vocabulary, it has already been shown that integrated dictionaries or glosses can be briefly glanced at [24, 46], minimising the time away from the text in comparison to external dictionaries. For grammar learning, research so far has focused on functionality and usability rather than the effect on the reading process. For example, Meurers et al. [27] and Zilio and Fairon [47] implemented dynamic interfaces that process the text to identify and augment grammatical structures. Draxler et al. [11] investigated the user experience achieved with different augmentation types and positions, and their perceived applicability to different types of texts. With this paper, we extend the work on interactive grammar visualisations for language learning. Specifically, we address the impact of interactive grammar visualisations on reading flow and the relationship with perceived usefulness and usability.

We developed a prototypical e-reader interface with three different grammar visualisations for English past tenses at varying levels of explicitness (see Figure 1). Each design features a translation mode, which represents the current state of e-readers. Level 0 (*Translation Only*) includes no grammar support. Level 1 (*Highlights*) highlights tense-related words. Level 2 (*Highlights & Pop-ups*) includes highlights and shows a small pop-up window when readers tap on a word, and Level 3 (*Highlights, Pop-ups & Overlays*) additionally provides large overlay screens that explains the grammar in more detail. The interface was based on the preferred designs proposed by Draxler et al. [11]. Using the prototypes, we conducted a within-subject study ($n = 24$) where participants read 12 pages of an Hercule Poirot detective story. We measured reading flow, perceived usefulness, and usability of the four designs from Levels 0 to 4. Additionally, we recorded user interaction. We hypothesised that more explicit grammar augmentations would be considered more useful, but also more harmful to the reading flow than less explicit grammar augmentations or *Translation Only*. We expected the best usability for an average level of support, i.e. for *Highlights & Pop-ups*.

The time spent reading a text gradually increased with the levels of explicitness. We also identified a tendency for lower levels of reading flow with more complex interfaces, but this only proved significant for one combination on one of the subscales. Conversely, analyses provided evidence in favour of there being no difference. Usefulness was indeed highest for designs that presented explicit information (*Highlights & Pop-ups* and *Highlights, Pop-ups & Overlays*). Finally, usability was rated best for *Translation Only* and the medium-support design *Highlights & Pop-ups*.

Overall, we extend prior concepts for grammar support with an interactive e-reader interface that realises support at different levels of explicitness. We investigate how the grammar augmentations influence reading flow and how this correlates with perceived usefulness and usability. Our findings inform the design of grammar augmentations for language learners that serve as learning scaffolds while preserving the enjoyment of reading. Specifically, we recommend a level of detail as implemented in *Highlights & Pop-ups*, with case-specific information but no disruptive elements like a full-screen dialogue. Moreover, readers should be able to activate and deactivate the augmentations to match them to their language level and current reading context.

2 RELATED WORK

Reading has long been an essential activity for language learning that makes learning engaging and entertaining, and digital technology further extends the range of possible learning support. This section summarises past research on reading for language learning, learning support strategies and interfaces, and reading flow.

2.1 Language Learning while Reading

Reading in a foreign language positively influences vocabulary acquisition, spelling skills, and grammar learning [20, 30], which makes reading an essential part of language learning. Both intensive and extensive reading are beneficial [5]. During intensive reading, readers pay close attention to constructs, words, and ideas, typically in short texts [36, p. 212], and learn to assess the structure of a text [5]. Extensive reading, on the other hand, focuses on general understanding achieved through faster reading. It is important for improved text comprehension, acquiring vocabulary, and establishing reading habits [1, 14, 18]. Especially in intensive reading, glossing and textual enhancement are common practice. Glossing refers to looking up words in a dictionary or margin notes while reading and is, thus, a tool for vocabulary acquisition and comprehension [22]. Textual enhancement puts the focus on specific (grammatical) constructs in the writing, for example, by underlining text or printing it in bold [37]. There is also some evidence that textual enhancement can improve grammar skills, although the research is not fully conclusive [21].

2.2 Language Learning Support in E-Readers

With e-readers, it has also become possible to add interactive reading support for vocabulary and grammar, which are both essential prerequisites for text comprehension [7]. Vocabulary support can be provided through *pop-up* dictionaries that appear directly in or above the text [45]. Several research projects have also proposed systems that integrate natural language processing to provide relevant information on words or phrases on demand (e.g., [3, 43, 47]). Ho et al. [16] propose utilising eye gaze to selectively show phrase-based translations and Han et al. [15] show explanatory images instead of textual information. Systems such as DysWebxia [33] support readers with dyslexia by replacing difficult words with synonyms on demand. In the commercial domain, readers can tap words on e-readers such as the Kindle to show dictionary translations or explanations[1]. Similarly, macOS and iOS include a system-wide lookup feature that also includes dictionary entries[2]. Even reading first-language texts can become an opportunity for vocabulary learning when sentences or words are replaced with foreign-language equivalents [2, 42]. Grammar visualisations are less common, but existing systems offer additional explanations [11, 47] or present auto-generated exercises [34]. The Readle[3] app displays grammar explanations for pre-selected German texts and Grammarly[4] gives brief grammar explanations in context, although not a learning system per se. Following recent research, we design and apply an interactive system for vocabulary and grammar learning.

[1] https://www.fluentu.com/blog/kindle-foreign-language-dictionary
[2] https://support.apple.com/en-gb/guide/mac-help/mchl3983326c/mac
[3] https://readle-app.com
[4] https://grammarly.com

2.3 Reading Flow, Text Comprehension, and Disruptions

The experience of flow is an important factor contributing to reading enjoyment [41] and is also beneficial for learning [9, 12]. Csíkszentmihályi [8] defined flow as an *optimal experience*. In a state of flow, humans achieve high levels of focus and performance without perceiving strong physical or mental strain [28]. They are absorbed in their activity and often do not notice the time passing. Reading is a frequently mentioned activity in the context of flow and is connected to experiences such as feeling that one is part of the story [26]. Thissen et al. [41] propose the Reading Flow Short Scale (RFSS) for measuring reading flow. It consists of ten items that make up a global score and the two subscales *(Smooth) Processing* and *Absorption* in line with the Flow Short Scale by Rheinberg et al. [35]. Smooth processing encompasses items describing the feeling of effortless competence, whereas absorption refers to the feeling of being immersed in the task. Reading flow is closely related to the concept of narrative absorption, which more broadly considers immersion in narrative media such as virtual-reality experiences or games [31]. Leroy [23] identified usability as a predictor of the flow dimension *fluency of performance* in the context of gaming. However, individual factors decide whether or not someone experiences optimal flow during reading [18], and even the best reading interface will not change this. However, it does make sense to design reading interfaces in a way that they do not cause additional disruptions. For example, glosses that are positioned in close proximity to the text reduce lookup times in comparison to physical dictionaries [39, 40] and, thus, avoid disruptions [13]. Taylor [40] further recommends that users should be able to hide glosses in digital interfaces. Therefore, we also enable users to hide and show augmentations in our designs. Moreover, we apply the RFSS to critically assess the perceived reading flow.

3 DESIGN OF THE E-READER INTERFACE

To study reading flow with augmentations for language learning, we implemented a prototypical e-reader interface with translation and grammar support with four different levels of support. For the lowest support level, we only added a translation feature. Starting from the second level, we add grammar highlights, and the third and fourth levels additionally include pop-up windows and/or overlays with more information. The text augmentation was targeted at German speakers learning English. Specifically, we prepared grammar support for past tenses, which is an aspect of the language that German speakers typically struggle with [38]. We realised the interface as a click dummy with the prototyping software Figma[5]. This enabled us to monitor interaction with the visualisations and to seamlessly embed the interface in our study setup.

Shared Design Elements. The base design was modelled after typical e-reader interfaces. Page numbers are indicated at the bottom of the page and users can turn pages with a swipe motion. Each design has three reading modes: *default*, *translation mode*, and *grammar mode*. The modes are activated using the circular buttons at the top of the page (cf. Figure 3). The default shows the plain text only. Translation mode adds lilac highlights to selected words

[5]https://www.figma.com

placing, on the table. He seidenes Taschentuch avily on the metal teapot, and polished it with a silk handkerchief. The kettle was on the

Figure 2: Tooltip in translation mode

and expressions to indicate that a translation is available. Tapping on a highlighted word shows a translation tooltip (cf. Figure 2). Finally, grammar mode adds orange, yellow, and light blue highlights for verbs in the three target tenses and the corresponding marker words. Depending on the design, the highlights are clickable or not. Only one reading mode at a time can be active. The different design concepts were based on the preferred design elements identified in [11], i.e. using highlights and pop-ups but no footnotes or text between designs. Below, we describe the final designs.

Level 0: Translation Only. This baseline design comprises the translation mode, but no grammar mode. Thus, it represents current state of the art of e-readers with integrated dictionaries, such as the Kindle.

Level 1: Highlights. This the most basic design that includes grammar support. It adds highlights to verbs and marker words as shown in Figure 3a. Highlights were also a central element for hinting at grammar structures in [11, 47]. The bottom bar serves as a legend to the highlights. The intention of this design is *focus on form*. Hence, the elements are not interactive.

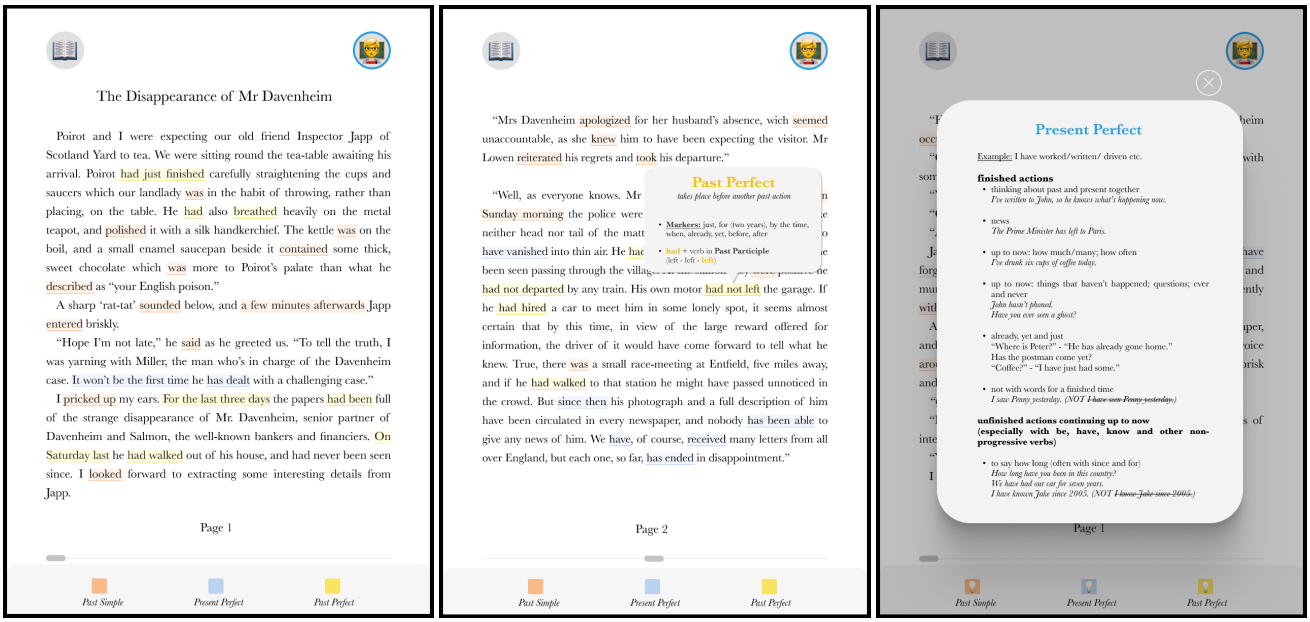
Level 2: Highlights & Pop-ups. This design extends Design 1 with contextual pop-ups that appear when a highlighted verb or marker word is tapped on (cf. Figure 3b). The pop-up shows basic information on the respective tense, possible markers, and the conjugation of the verb. This design was added because the study in [11] had shown that users are particularly interested in case-specific information.

Level 3: Highlights, Pop-ups & Overlays. As the space in the contextual pop-ups is limited, Design 2 was further extended with big overlays that show comprehensive information on the tenses (cf. Figure 3c). The overlays are opened by tapping on the light bulb icons in the bottom bar. Again, this interaction method was modelled after the explanation overlays integrated in [11].

4 USER STUDY

In a within-subject study with 24 participants, we evaluated the influence of grammar visualisations on reading flow, perceived usability, and usefulness while reading foreign-language texts in an e-reader interface. Specifically, we compared the three grammar visualisations and the baseline visualisation that only included the vocabulary mode. We chose the vocabulary mode as a baseline because integrated dictionaries are, by now, a standard feature of e-readers. With our study, we address the following research questions:

RQ1: How do users interact with translation and grammar visualisations in an e-reading interface?
RQ2: Does adding grammar visualisations have an impact on reading flow?



(a) **Highlights** (b) **Highlights + Contextual Pop-Ups** (c) **Highlights, Contextual Pop-Ups and Detailed Overlays**

Figure 3: Reading interfaces for Grammar Support

RQ3: Does adding grammar visualisations have an impact on perceived usability and usefulness of the e-reading interface?

We expected that reading flow decreases with the amount of presented information but that the additional information increases usefulness. On the other hand, we predicted usability to be best for a medium level of support.

4.1 Procedure

After signing the consent forms, participants answered a short questionnaire on demographics, English grammar skills, and experience with e-reading, and language learning. All subsequent steps were conducted as a user test on the Maze[6] platform. The first step on Maze was an introduction to basic features of the e-reading interface such as turning pages. Following the introduction, the participants tested all four designs with four different texts of 3 pages each. After each text, they answered the questions of the System Usability Scale [4], the Reading Flow Short Scale [41], two custom text comprehension questions, and three grammar exercises. While they were reading, we logged the interaction with the prototypes, e.g. taps on words and toggling of the reading modes. The texts and designs were balanced with a Latin square design to avoid sequence effects and to ensure that each design was tested with each text. As the final step, the participants rated the usefulness of all designs, stated if they had focused mostly on the text, the vocabulary, or the grammar, and commented on possible improvements. Approximately half the participants completed the study online and the other half on site, in a comfortably prepared reading area. The overall process was identical in both cases, as all steps were described in the survey and Maze project, so that no further instructions or guidance were necessary.

4.2 Participants

We recruited 24 participants (12 male, 12 female) with a mean age of 34.2 years (SD = 9.6 years, min = 23, max = 58 years) via social media and personal contacts. Of these, two self-assessed their English level as B1, ten as B2, seven as C1, and five as C2 on the European Reference Scale CEFR[7]. Twenty-one were native speakers of German, one of Russian and Ukrainian, one of Slovakian, and one was a native speaker of Telugu. Fifteen participants had experience with reading on tablets, but only three reported that they read digital books for language learning.

5 RESULTS

This section reports observed user interaction patterns, perceived reading flow, and usability with the different visualisation concepts. We complement the analysis with qualitative statements gathered in the final questionnaire. If not mentioned otherwise, we applied Repeated-Measures ANOVAs to compare measures for the four design concepts. We added a Greenhouse-Geisser correction whenever the assumption of sphericity was violated. We adjusted p-values of post-hoc tests with Holm correction and added Cohen's d for effect sizes. In addition, we include a Bayesian ANOVA for the reading flow scores because this also allows us to estimate the probability of there being no difference between conditions [44].

[6] https://maze.co/

[7] https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

We report Bayes factors indicating the likelihood ratio of the null and alternative hypothesis.

5.1 User Interaction and Reading Behaviour

From the Maze data export, we extracted the the exact timing for all interface states and user interaction such as taps on words. Table 1 lists the average values of the recorded interaction measures. There were significant differences in the total reading time per condition ($F(2.187) = 3.690$, $p < 0.05$). Specifically, the time increased from Level 0 to Level 3; the reading time for *Translation Only* was significantly lower than for *Highlights, Pop-ups & Overlays* ($t = 3.15$, $p < 0.05$, $d = 0.61$). Users frequently used both the grammar and the translation modes. The translation tended to be active for a larger share of the time than the grammar mode. The number of switches was highest in the *Highlights* condition, and lowest for *Translation Only*, where only the translation mode could be toggled on and off. On average, participants looked up the translations of 10 to 11 words or expressions across all designs and 4.5 to 6.5 grammar explanations, either as small contextual pop-ups or detailed overlays (in the two designs where this was possible). We asked participants what they focused on most while reading: vocabulary, grammar, or text comprehension (free text, multiple mentions possible). Only five participants said that grammar was their main focus, 13 listed vocabulary, and 15 text comprehension. Across all designs, the average correctness rate of the text comprehension questions ranged from 75% to 87.5% per text, which indicates that there were no major issues with the difficulty level of the text.

5.2 Reading Flow

We separately analysed the *global flow* and the subscales *absorption* and *processing* of the Reading Flow Short Scale [41] (cf. Figure 4). All average scores were better than the neutral value 4, with a tendency for better flow in the *Translation Only* condition. Table 2 shows that there were no significant differences in reading flow between the different designs. Conversely, a Bayesian ANOVA suggests that the null hypothesis (no difference) is substantially more likely than alternative hypothesis for global flow and absorption. For processing, there is an anecdotal tendency towards the alternative hypothesis. A Bayesian post-hoc test shows substantial evidence for higher processing in the *Translation Only* ($M = 4.9$, $SD = 0.92$) mode compared to the condition *Highlights, Pop-ups & Overlays* ($M = 4.4$, $SD = .96$, $BF_{10} = 6.53$[8]). For all other post-hoc tests, the null hypothesis is more likely, i.e. the Bayes factors are < 1.

5.3 Usability

There were significant differences between the System Usability Scale values of the four designs ($F(3) = 6.344$, $p < 0.001$). Post-hoc tests revealed that *Translation Only* ($M = 90.3, SD = 8.9$) was significantly more usable than *Highlights* ($M = 80.0, SD = 11,7$; $t = 3.952$, $p < 0.001$, $d = 0.951$) and *Highlights, Pop-ups & Overlays* ($M = 82.4, SD = 12.6$; $t = 3.034$, $p < 0.05$, $d = 0.730$). *Highlights & Pop-ups* ($M = 87.2, SD = 9.7$) was considered more usable than *Highlights* ($t = 2.755$, $p < 0.05$, $d = 0.663$).

After using all four designs the participants rated how useful they were on a scale from 1 to 5 (cf. Figure 6, higher scores are better). According to a Friedman test, there were significant differences between the conditions ($\chi^2(3) = 13.2$, $p < 0.01$). Holm-corrected Conover post-hoc tests showed that *Highlights* ($MD = 3$) were seen as significantly less useful than *Highlights & Pop-ups* ($MD = 4$; $t = 2.746$, $p < 0.05$) and *Highlights, Pop-ups & Overlays* ($MD = 4$; $t = 2.684$, $p < 0.05$).

5.4 Qualitative Feature Evaluation

We also asked two open-ended questions to further assess the quality of the overall concept. Here, we split participant statements by topic and clustered the partial statements[9]. For the first question, "What did you like about the prototype", the most frequent comment was that the system was easy to use (ten participants). Nine participants said that they liked the grammar feature, and another nine the translation feature. For example, P24 wrote *"Great features! Every reading app should have both translation and grammar (tenses)"*. Six participants positively commented on the design, e.g. P19: *"The colour coding of the different past tenses was very clear"*. Four participants explicitly stated that they felt (almost) no impact on their reading flow, e.g. P1: *"The coloured highlights did not affect the reading flow"*. Finally, P15 and P19 found it important that the reading and grammar support can be switched on and off.

Fifteen participants responded the second question: "What would you change?". They proposed a number of changes to the design or interaction methods. For example, four participants suggested a higher number of clickable words, e.g. P24: *"[...] all words can be translated, or entire sentences"*. Three participants would like to activate the grammar and translation modes simultaneously. The interaction could be simplified, e.g. P20: *"for the translation: that it disappears automatically after a short time"* or adjusted based on the reader's learning progress and language level, e.g. P10: *"popup of new learning contents, for example, and later only highlights"*. P7 said that switching to the grammar mode made them loose focus on the content and recommended moving the information to a sidebar. Three participants hinted that they had issues with finding the buttons for opening the overlays in *Highlights, Pop-ups & Overlays*. Finally, additional feature ideas were pronunciation support (P22), tests or exercises (P24), and the possibility to add notes (P5). There were no comments on the content of the grammar explanations.

6 DISCUSSION

Reading for language learning demands a balance between text comprehension, enjoyment, and attention to language. The analysis of user interaction, reading flow, and the participants' opinions show that interactive grammar visualisations can support learners during reading. However, they need to be usable, provide sufficient information, and should not disrupt the reading flow. We discuss these aspects on the basis of our design concepts.

6.1 Reading Flow, Reader Focus, and Interaction

The Reading Flow Short Scale and the participant statements indicate that the addition of grammar support did not significantly impact reading flow. Notably, there were no significant differences for the global reading flow, absorption, and processing. The Bayesian

[8] uncorrected for multiple comparisons

[9] Participant statements were translated to English

Table 1: Average value and standard deviation of user interaction measures

	Translation Only		*Highlights*		*Highlights & Pop-ups*		*Highlights, Pop-ups & Overlays*	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Total reading time (m:ss)	5:49	2:32	6:24	2:17	6:55	2:19	7:21	2:48
Time share translation mode	87.3%	26.1%	44.9%	33.7%	51.4%	35.2%	50.7%	31.8%
Time share grammar mode	-	-	46.9%	33.6%	44.6%	33.7%	45.3%	29.7%
Number of mode switches	1.29	1.78	12.8	9.99	7.83	4.92	7.50	4.93
Translation Tooltip	11.2	7.0	10.0	7.9	10.6	7.1	10.2	8.6
Small Grammar Popup	-	-	-	-	4.5	2.4	3.5	2.2
Grammar Overlay	-	-	-	-	-	-	3.0	1.4

Table 2: Reading Flow Short Scale

Measure	**Frequentist ANOVA**	**Bayesian ANOVA**
Global Flow	$F(2.0) = 1.57, p > 0.05$	$BF_{01} = 3.26$
Absorption	$F(2.0) = 1.02, p > 0.05$	$BF_{01} = 5.98$
Processing	$F(2.1) = 2.82, p > 0.05$	$BF_{01} = 0.81$

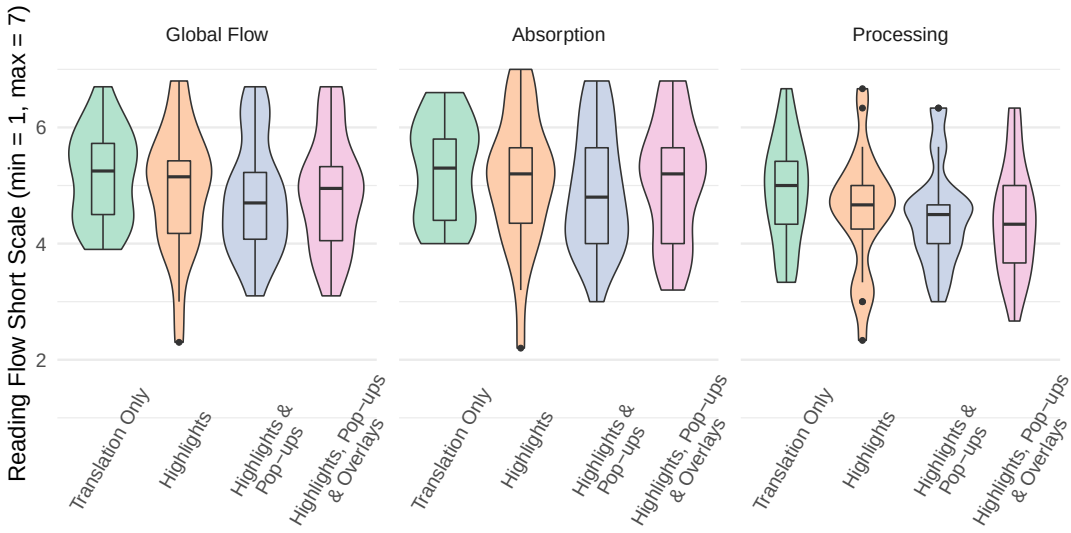


Figure 4: Reading Flow Short Scale

analysis only provided substantial evidence in favour of lower processing for *Highlights, Pop-ups & Overlays* than for *Translation Only*. This is not surprising because the big overlay in *Highlights, Pop-ups & Overlays* covered a large part of the text, which means that readers need to search for the position where they left off. Overall, the median reading flow was slightly lower than in a 20-minute reading scenario investigated by Thissen et al. [41], where participants read a novel of their choice. The qualitative statements confirm the findings from the RFSS: four participants explicitly mentioned that their reading flow was not significantly interrupted and only one participant suggested improving reading focus, specifically, with a side-by-side mode. Similar ideas have already been explored in interruption research. For example, leaving a part of the primary task window (here, the text) visible can speed up task resumption (here, continuing to read) [17]. If designs do include an overlay that hides the text, indicators that guide readers back to the last position could also decrease disruptions [25].

Overall, participants in our user study showed active engagement with the grammar and translation support: One of the two modes was almost always switched on, and the participants frequently tapped on words to reveal grammar hints and translations.

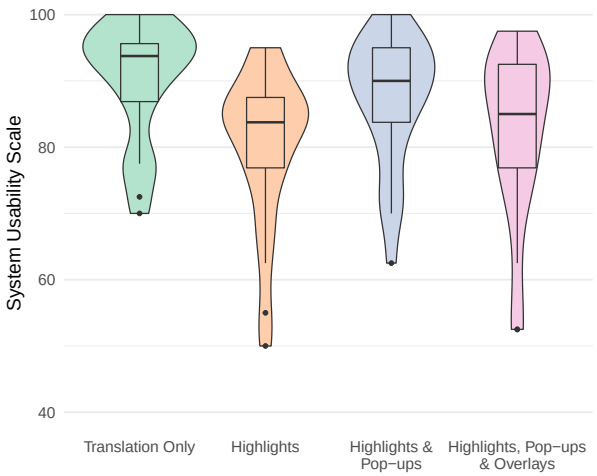


Figure 5: System Usability Scale of the four designs

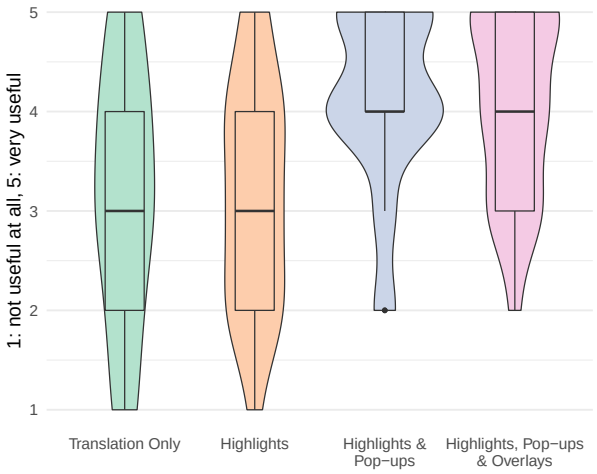


Figure 6: Perceived usefulness of the four designs

Thirteen individual participants explicitly mentioned the translation and/or grammar mode when they described what they liked about the system. Despite the fact that the main focus of the study was grammar, participants tended to focus more on translation and text comprehension than on grammar. This was apparent in their own statements but also the tendency for a larger time share spent in the translation mode than in the grammar mode and more words clicked in the translation mode. In addition, the number of consulted translations did not decrease with increasing levels of grammar support. It is possible that the participants were more familiar with translations or found them to be more immediately important for text comprehension. More importantly though, it indicates a need for additional motivation to also consider grammar, especially in an intensive reading context. However, it is not evident how the readers' attention can be actively directed towards grammar without compromising reading flow.

6.2 Balancing Information Load, Usefulness, and Usability

The perceived flow, usability, and usefulness indicate a trade-off between the reading task and the amount of information presented through the translation and grammar features. While additional features provide support for learning, they can also distract readers and decrease their focus and flow. Specifically, the order of *Highlights* and *Highlights & Pop-ups* is reversed for usability and reading flow, with higher usability and lower reading flow for *Highlights & Pop-ups*. Moreover, perceived usefulness was higher for the two conditions that included grammar explanations (*Highlights & Pop-ups* and *Highlights, Pop-ups & Overlays*) than those that provided no grammar support or static grammar highlights only (*Translation Only* and *Highlights*). Hence, it is important that readers can activate and deactivate potentially disrupting elements of the interface. For example, they could activate higher support levels for intensive reading sessions and less support for extensive reading.

6.3 The Users' Perspective for Future Applications

Combining our work with prior research on the design and computational linguistics informs the design of comprehensive grammar augmentations in digital reading. Notably, grammar classification methods as proposed by Zilio and Fairon [47], Meurers et al. [27], and Reynolds et al. [34] contribute the foundations for making grammar support scalable. Draxler et al. [11], on the other hand, introduce design recommendations that we extend and refine in our study by additionally considering the reading flow and its interplay with usefulness and usability. Thus, we provide a basis for future developments, such as adaptations based on a reader's focus and goals. Similarly, our participants suggested adaptations based on the language level. This is also a frequently and successfully applied practice in other language learning environments [32].

6.4 Limitations and Future Work

The chosen grammatical structure and type of text may have influenced the results. However, the large effect sizes observed for the differences in usability indicate that the identified trends are likely to replicate with additional longer or more frequent reading sessions. In addition, the current study did not compare learning with the different visualisation concepts. This was the case because we prioritised the comparison of reading flow effects within subjects, while a between-groups design would have been preferred for isolating learning benefits on a specific grammar topic. From a technical perspective, the implementation with Figma and Maze was helpful for composing the study setup including questionnaires. However, the possibilities for interaction tracking were very limited and required substantial preprocessing. For evaluating future prototypes, we recommend platforms that are easy to extend with custom logging commands. As a next step, the designs could be combined with natural language processing for identifying grammar constructs that can be highlighted and explained. For example, the detection methods developed by Meurers et al. [27] and Zilio

and Fairon [47] could be integrated into our design. Moreover, the reading could be complemented with text-based auto-generated exercises [6] to consolidate the acquired knowledge.

7 CONCLUSION

Vocabulary aids are already common practice in digital reading, especially for language learners. It is only a matter of time until grammar support will also be widely available, given the fast development of multi-lingual natural language processing for identifying grammatical structures in any digital text. With our paper, we contribute an essential building block from a users' perspective: our study showed the impact of grammar aids on the reading process of language learners. This is an important determiner of an enjoyable reading experience, and in the long run, for continued motivation to read. Specifically, we highlight the need for adaptive, non-disruptive interfaces that provide targeted information without overwhelming learners. We show that grammar information is considered useful, and that readers can maintain reading flow when the information is well-integrated into the reading interface.

ACKNOWLEDGMENTS

This work was partly conducted within the Amplify project which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 683008). It was also funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

REFERENCES

[1] Faisal Al-Homoud and Norbert Schmitt. 2009. Extensive reading in a challenging environment: a comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research* 13, 4 (Oct. 2009), 383–401. https://doi.org/10.1177/1362168809341508

[2] Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. 2022. VocabEncounter: NMT-powered Vocabulary Learning by Presenting Computer-Generated Usages of Foreign Words into Users' Daily Lives. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–21. https://doi.org/10.1145/3491102.3501839

[3] Mahmoud Azab, Ahmed Salama, Kemal Oflazer, Hideki Shima, Jun Araki, and Teruko Mitamura. 2013. An nlp-based reading tool for aiding non-native english readers. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*. 41–48.

[4] John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[5] Patricia L. Carrell and Joan G. Carson. 1997. Extensive and intensive reading in an EAP setting. *English for Specific Purposes* 16, 1 (Jan. 1997), 47–60. https://doi.org/10.1016/S0889-4906(96)00031-2

[6] Maria Chinkina, Simón Ruiz, and Detmar Meurers. 2017. Automatically generating questions to support the acquisition of particle verbs: evaluating via crowdsourcing. In *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017*, Kate Borthwick, Linda Bradley, and Sylvie Thouësny (Eds.). Research-publishing.net, 73–78. https://doi.org/10.14705/rpnet.2017.eurocall2017.692

[7] Yunjeong Choi and Dongbo Zhang. 2021. The relative role of vocabulary and grammatical knowledge in L2 reading comprehension: a systematic review of literature. *International Review of Applied Linguistics in Language Teaching* 59, 1 (March 2021), 1–30. https://doi.org/10.1515/iral-2017-0033

[8] Mihály Csíkszentmihályi. 1990. *Flow: The psychology of optimal experience*. Vol. 1990. Harper & Row New York.

[9] Mihály Csíkszentmihályi, Kevin Raymond Rathunde, and Samuel Whalen. 1997. *Talented teenagers: the roots of success and failure* (1st paperback ed ed.). Cambridge University Press, Cambridge [England] ; New York, N.Y.

[10] Richard R Day, Carole Omura, and Motoo Hiramatsu. 1992. Incidental EFL vocabulary learning and reading. *Reading in a foreign language* 7 (1992), 541–541.

[11] Fiona Draxler, Christina Schneegass, Nicole Lippner, and Albrecht Schmidt. 2019. Exploring visualizations for digital reading augmentation to support grammar learning. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia - MUM '19*. ACM Press, Pisa, Italy, 1–11. https://doi.org/10.1145/3365610.3365623

[12] Séverine Erhel and Eric Jamet. 2019. Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior* 91 (Feb. 2019), 106–114. https://doi.org/10.1016/j.chb.2018.09.020

[13] Cyrus K. Foroughi, Nicole E. Werner, Daniela Barragán, and Deborah A. Boehm-Davis. 2015. Interruptions disrupt reading comprehension. *Journal of Experimental Psychology: General* 144, 3 (June 2015), 704–709. https://doi.org/10.1037/xge0000074

[14] Siao-cing Guo. 2012. Using Authentic Materials for Extensive Reading to Promote English Proficiency. *English Language Teaching* 5, 8 (July 2012), p196. https://doi.org/10.5539/elt.v5n8p196

[15] Cheng-Hsien Han, Chi-Lan Yang, and Hao-Chuan Wang. 2014. Supporting second language reading with picture note-taking. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*. ACM Press, Toronto, Ontario, Canada, 2245–2250. https://doi.org/10.1145/2559206.2581204

[16] Tien-Yu Ho, Hao-Chuan Wang, and Shong-Hong Lai. 2018. Non-native Language Reading Support with Display of Machine Translation Based on Eye-Tracking and Sentence-Level Mapping. In *Proceedings of the Sixth International Symposium of Chinese CHI on - ChineseCHI '18*. ACM Press, Montreal, QC, Canada, 57–63. https://doi.org/10.1145/3202667.3202675

[17] Shamsi T. Iqbal and Eric Horvitz. 2007. Conversations Amidst Computing: A Study of Interruptions and Recovery of Task Activity. In *User Modeling 2007*, Cristina Conati, Kathleen McCoy, and Georgios Paliouras (Eds.). Vol. 4511. Springer Berlin Heidelberg, Berlin, Heidelberg, 350–354. https://doi.org/10.1007/978-3-540-73078-1_43

[18] Cheryl Kirchhoff. 2013. L2 extensive reading and flow: Clarifying the relationship. *Reading in a foreign language* 25, 2 (2013), 192–212. Publisher: ERIC.

[19] Stephen Krashen. 1989. We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *The Modern Language Journal* 73, 4 (Dec. 1989), 440–464. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

[20] Stephen D. Krashen. 2004. *The power of reading: insights from the research* (2nd ed ed.). Libraries Unlimited ; Heinemann, Westport, Conn. : Portsmouth, NH.

[21] Ryan M LaBrozzi. 2016. The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. *Language Teaching Research* 20, 1 (Jan. 2016), 75–91. https://doi.org/10.1177/1362168814561903

[22] Olaf Lenders. 2008. Electronic glossing – is it worth the effort? *Computer Assisted Language Learning* 21, 5 (Dec. 2008), 457–481. https://doi.org/10.1080/09588220802447933

[23] Raphael Leroy. 2021. Immersion, Flow and Usability in video games. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–7. https://doi.org/10.1145/3411763.3451514

[24] Andrea Lofgren. 2022. Unraveling Contradictions: Which Glosses Facilitate Reading Comprehension Among ELLs, and Why? *Journal of Language Teaching and Research* 13, 1 (Jan. 2022), 1–11. https://doi.org/10.17507/jltr.1301.01

[25] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 2953–2962. https://doi.org/10.1145/2702123.2702539

[26] Fausto Massimini, Mihály Csíkszentmihályi, and Antonella Delle Fave. 1988. Flow and biocultural evolution. (1988). Publisher: Cambridge University Press.

[27] Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, and Niels Ott. 2010. Enhancing authentic web pages for language learners. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics, 10–18. https://doi.org/10.5555/1866795.1866797

[28] Jeanne Nakamura and Mihály Csíkszentmihályi. 2014. The Concept of Flow. In *Flow and the Foundations of Positive Psychology*. Springer Netherlands, Dordrecht, 239–263. https://doi.org/10.1007/978-94-017-9088-8_16

[29] Paul Nation. 2007. The Four Strands. *Innovation in Language Learning and Teaching* 1, 1 (April 2007), 2–13. https://doi.org/10.2167/illt039.0

[30] Paul Nation. 2015. Principles guiding vocabulary learning through extensive reading. *Reading in a foreign language* 27, 1 (2015), 136–145. Publisher: ERIC.

[31] Federico Pianzola, Giuseppe Riva, Karin Kukkonen, and Fabrizia Mantovani. 2021. Presence, flow, and narrative absorption: an interdisciplinary theoretical exploration with a new spatiotemporal integrated model based on predictive processing. *Open Research Europe* 1 (July 2021), 28. https://doi.org/10.12688/openreseurope.13193.2

[32] Jan L. Plass and Shashank Pawar. 2020. Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education* 52, 3 (July 2020), 275–300. https://doi.org/10.1080/15391523.2020.1719943

[33] Luz Rello and Ricardo Baeza-Yates. 2014. Evaluation of DysWebxia: a reading app designed for people with dyslexia. In *Proceedings of the 11th Web for All Conference on - W4A '14*. ACM Press, Seoul, Korea, 1–10. https://doi.org/10.1145/2596695.2596697

[34] Robert Reynolds, Eduard Schaf, and Detmar Meurers. 2014. A VIEW of Russian: Visual Input Enhancement and adaptive feedback. In *Proceedings of the third workshop on NLP for computer-assisted language learning*. 98–112.

[35] Falko Rheinberg, Regina Vollmeyer, and Stefan Engeser. 2003. Die Erfassung des Flow-Erlebens. In *Diagnostik von Motivation und Selbstkonzept*, Joachim Stiensmeier-Pelster and Falko Rheinberg (Eds.). Number N.F., 2. Hogrefe, Göttingen Bern, Chapter 14, 261–279.

[36] Jack C. Richards and Richard W. Schmidt. 2013. *Longman Dictionary of Language Teaching and Applied Linguistics* (0 ed.). Routledge. https://doi.org/10.4324/9781315833835

[37] Daphnée Simard. 2009. Differential effects of textual enhancement formats on intake. *System* 37, 1 (March 2009), 124–135. https://doi.org/10.1016/j.system.2008.06.005

[38] Michael Swan and others. 2001. *Learner English: A teacher's guide to interference and other problems*. Vol. 1. Cambridge University Press.

[39] Alan M. Taylor. 2009. CALL-Based Versus Paper-Based Glosses: Is There a Difference in Reading Comprehension? *CALICO Journal* 27, 1 (Sept. 2009), 147–160. https://doi.org/10.11139/cj.27.1.147-160

[40] Alan M Taylor. 2014. Glossing frequency and L2 reading comprehension: the influence of CALL glossing. *Calico Journal* 31, 3 (2014), 374–389.

[41] Birte A. K. Thissen, Winfried Menninghaus, and Wolff Schlotz. 2018. Measuring Optimal Reading Experiences: The Reading Flow Short Scale. *Frontiers in Psychology* 9 (Dec. 2018), 2542. https://doi.org/10.3389/fpsyg.2018.02542

[42] Andrew Trusty and Khai N. Truong. 2011. Augmenting the web for second language vocabulary learning. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 3179. https://doi.org/10.1145/1978942.1979414

[43] Rafael Veras, Erik Paluka, Meng-Wei Chang, Vivian Tsang, Fraser Shein, and Christopher Collins. 2014. Interaction for reading comprehension in mobile devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. ACM Press, Toronto, ON, Canada, 157–161. https://doi.org/10.1145/2628363.2628387

[44] Eric-Jan Wagenmakers, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Jonathon Love, Ravi Selker, Quentin F. Gronau, Martin Šmíra, Sacha Epskamp, Dora Matzke, Jeffrey N. Rouder, and Richard D. Morey. 2018. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review* 25, 1 (Feb. 2018), 35–57. https://doi.org/10.3758/s13423-017-1343-3

[45] Jing Wang. 2014. Reading Strategies and Reading Comprehension Retention with and without a Pop-up Dictionary by Beginning Learners of Chinese. *CALICO Journal* 31, 2 (May 2014), 244–260. https://doi.org/10.11139/cj.31.2.244-260

[46] Polle T. Zellweger, Susan Harkness Regli, Jock D. Mackinlay, and Bay-Wei Chang. 2000. The impact of fluid documents on reading and browsing: an observational study. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*. ACM Press, The Hague, The Netherlands, 249–256. https://doi.org/10.1145/332040.332440

[47] Leonardo Zilio and Cédrick Fairon. 2017. Adaptive system for language learning. In *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on*. IEEE, 47–49.