
Learning New Words and Spelling with Autocorrections

Ahmed Sabbir Arif

Ryerson University
Toronto, Ontario, Canada
a.s.arif@gmail.com

Cristina Sylla

University of Minho
Braga, Portugal
sylla@engagelab.org

Ali Mazalek

Ryerson University
Toronto, Ontario, Canada
mazalek@ryerson.ca

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
ISS '16, November 06-09, 2016, Niagara Falls, ON, Canada
ACM 978-1-4503-4248-3/16/11.
<http://dx.doi.org/10.1145/2992154.2996790>

Abstract

We present a novel color-coded feedback method that highlights the types of corrections made by autocorrections. We then present results of a longitudinal user study with 7-8-year-old children that compared the new method with the conventional autocorrection feedback method, in terms of learning new words and spelling. Results suggested that the new method better accommodates learning new words. Interestingly, learning was observed with the conventional feedback method as well, demanding further investigation into whether predictive methods are truly a barrier to learning new words and spelling.

Author Keywords

Child; children; prediction; predictive keyboard; spelling; autocorrection; color-coded autocorrect.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Graphical user interfaces (GUI).

Introduction

Mobile devices are becoming increasingly popular among children. A survey revealed that about 72% of children aged eight and under in the U.S. have access to mobile devices, including tablets, which they are using for various purposes, such as to play games, watch videos,

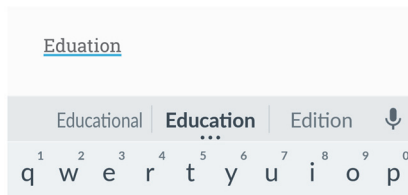


Figure 1. Current autocorrection feedback.

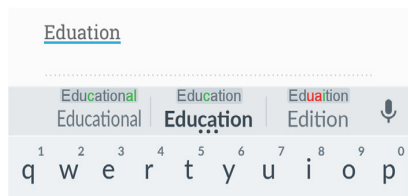


Figure 2. Color-coded autocorrection feedback.

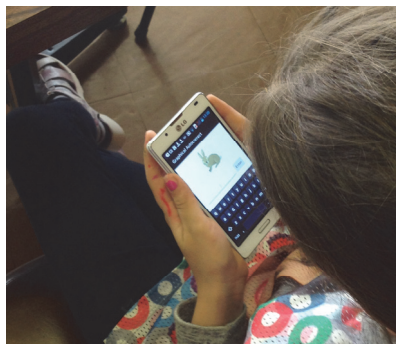


Figure 3. A child participating in the user study.

take pictures, and gain access to social networks and mobile applications [19]. Many of these activities require text entry. Yet most current mobile devices do not include children-friendly keyboards. There are some third-party virtual keyboards available for children. However, the effort and technical expertise required to find and install these keyboards are often discouraging to mobile owners. As a result, most children are forced to use keyboards that are primarily designed for adults. Research suggests that composing text with a conventional keyboard can play a role in the language learning process of young children. Conventional keyboards require children to revise, rework, and discuss their work, encouraging critical thinking. This assists children to improve their writing [11], narrative [21] and spelling [15] skills. Most current mobile keyboards, however, are augmented with predictive systems that suggest the most probable next word/s based on prefix and context, and automatically correct all probable misspelled words. Studies showed that these techniques have made mobile text entry much easier for both children [14] and adults [4]. Yet, predictive systems, particularly autocorrect, have long been considered a barrier to learning spelling by popular media, teachers, and parents [16, 22]. The argument is that because of autocorrections, children no longer have to review their input to identify and correct misspelled words [22], thus cannot learn from their mistakes. As a first step towards investigating this popular belief, here we explore whether or not learning of new words and spelling occurs with autocorrections.

Related Work

Most current text entry techniques for children are full-length physical keyboards that use the standard QWERTY layout with larger key-prints and brightly colored vowel

and numeric keys to enable learning [20]. Some use alphabetic layouts to facilitate a faster transition from novice to expert [9]. Several virtual keyboards are also available that are practically virtual versions of the existing physical keyboards [6]. Unfortunately, most of these keyboards do not account for the variable spelling abilities of children [17].

Several studies have found relationships between text entry and language skills. Papert [15] suggested that text entry accommodates critical thinking and learning by allowing children to revise and rework their ideas. Kurth [11] argued that the high visibility of text on the screen can passively improve children's spelling skills by igniting more conversation about the writing. Sturm [21] showed that text entry can improve users' narrative skills. Newell et al. [14] showed that predictive text can improve the quantity and quality of their written work. Wood et al. [23] confirmed this in a separate study, where they showed that the use of word prediction can improve spelling skills. Yet none of these works explored the impact of autocorrection on language skills.

Many have highlighted the need for providing children with meaningful visual feedback to facilitate learning. Anthony et al. [1] argued that visual feedback can help children learn a new system, especially when they are still developing the skills required to interact with it. Druin [7] showed that feedback facilitates learning by providing children with the means to evaluate themselves.

Current Visual Feedback on Autocorrections

Most current virtual keyboards provide visual feedback on autocorrections. They display the most probable word/s in a prediction bar or above the text as the user types, and highlight an autocorrected word by changing

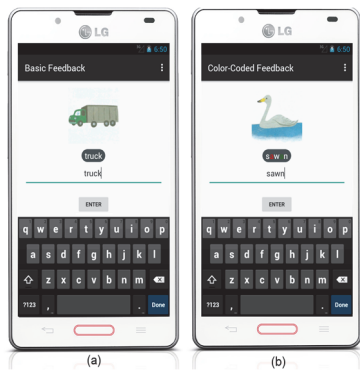


Figure 4. The device and the (a) current and (b) color-coded feedback methods.

Words per Minute (WPM)

measures how many words can be produced in one minute [2].

Success rate (%) represents the rate at which correctly spelled words were inputted.

This was calculated as the ratio of the total number of correctly spelled words to the total number of words entered.

Recall rate (%) measures the proportion of new words that were correctly recalled by the users. This was calculated as the ratio of the total number of correctly recalled words to the total number of words entered.

its font or background to a different color for about one second.

Color-Coded Visual Feedback

We designed a new feedback method for autocorrection to inform users of how an incorrect input was fixed.

When an autocorrection occurs, it highlights the corrections made in the prediction bar and in a bubble above the autocorrected word. It characterizes all edits as insertions or deletions, because substitutions are practically combinations of these two operations. It highlights all deletions with a red font and all insertions with a green, however different colors can be used to address colorblindness. The method does not highlight the types of errors, but the types of edits made by the autocorrection method. Figure 1 and 2 illustrates the current and the proposed methods, respectively.

An Experiment

The study explored whether current autocorrection feedback facilitates learning new words and spelling, and if the proposed method improves these learning rates.

Apparatus

We used two LG Optimus L7 II P710 smartphones, 121.5× 66.6×9.7 mm and 118 g. They ran on Android 4.1.2 at 480×800 (Figure 4). A custom application, developed with the default Android SDK was used. It displayed drawings of people, animals, and various objects on the screen and asked children to enter the name of the entity in the picture using the default Google Android keyboard. For this, we collected 25 nouns, and the drawings associated with them, from a workbook for young children [10]. We included a narrator feature to address unknown words, implemented using the Google Translate API [8]. When children could not

recognize the subject of a picture, they could tap on the picture to hear its name. We used words instead of short phrases for better control of the variables and to reduce the possibility of fatigue. We used nouns, as children's first learned words are usually names of people, animals, and objects [13]. We used a custom autocorrection method [3], since the default SDK does not allow access to its predictive system. We implemented the color-coded feedback using the Levenshtein distance algorithm [12]. The prediction bar was disabled to eliminate a confound.

Participants and Design

We recruited 26 grade two students from a school where English is taught as a second language, reducing the possibility of learning new words outside the classroom. We only recruited students with comparable academic standings. We used a between-subjects design with two groups: *basic* (current feedback) and *color-coded*. Each group had 13 children, 7 males and 6 females, aged 7-8, on average 7.5. There were 3 sessions, with about a day in between. Children entered the same 25 words in all sessions in random order. In summary, the design was: 2 groups × 13 children × 3 sessions × 25 words = 1,950 words, in total. Figure 3 shows the study setup.

Procedure

During the study, children entered words using the custom application. It displayed one drawing at a time and asked them to enter the name of the entity in the picture. When done, they had to press the "ENTER" button to see the next drawing. They were instructed to tap on the drawing if they could not identify/remember the entity in the picture to hear its name. They were asked to input as fast and accurately as possible, but informed that it is alright to forget/misspell words, so that they would not feel under pressure. Error correction

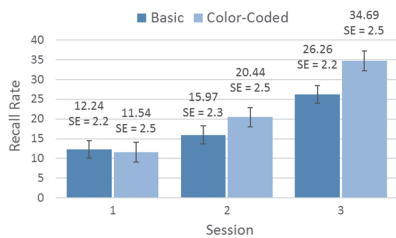


Figure 5. Average recall rate by session. Error bars represent ± 1 standard error ($\sigma_{\bar{x}}$).

was recommended, but not enforced, thus they could submit incorrect/misspelled words. We took permission from the school to run the study during class-hours in an empty room at the school. Children also had to have consent from their parents to participate. When they arrived for the first session, we demonstrated the application and the feedback method assigned to them. We also allowed them to practice with the application using a different set of words. We started the study when they felt confident. The following sessions used the same structure, excluding the demonstration and practice. The sessions were scheduled with about a day in between. After the last session, children were asked to fill out a questionnaire, where they could rate the feedback methods on a 5-point Smileyometer scale [18].

Results

We used a Mixed-Design ANOVA on the study data and a Mann-Whitney U Test on the questionnaire data.

Recall Rate

An ANOVA identified a significant effect of condition on recall rate ($F_{1,24} = 4.51, p < .05$). On average recall rate was 18.2% ($\sigma_{\bar{x}} = 1.4$) and 22.4% ($\sigma_{\bar{x}} = 1.4$) in basic and color-coded, respectively. There was also a significant effect of session in both basic ($F_{2,12} = 10.7, p < .0001$) and color-coded ($F_{2,12} = 21.9, p < .0001$). See Figure 5. To study learning, we fitted power functions to sessional success rates to model the power law of practice [5]. In Figure 6, one can see that the data correlates well to the power functions for both basic ($R^2 = 0.89$) and color-coded ($R^2 = 0.98$).

Success Rate

An ANOVA failed to identify a significant effect of condition on success rate ($F_{1,24} = 0.84, ns$). The basic

and color-coded yielded on average 63.8% ($\sigma_{\bar{x}} = 1.6$) and 65.9% ($\sigma_{\bar{x}} = 1.7$) success rate, respectively. There was also no significant effect of session in basic ($F_{2,12} = 1.6, ns$) or color-coded ($F_{2,12} = 1.2, ns$). See Figure 7. Results also revealed that both groups yielded comparable success rates in the first session: 61% ($\sigma_{\bar{x}} = 2.8$) and 63% ($\sigma_{\bar{x}} = 2.9$) in basic and color-coded, respectively ($F_{1,24} = 0.11, ns$). This suggests that one group was not substantially better at spelling than the other group when they started the study.

Investigation also revealed that 28% of all errors were caused by missing or incorrectly entered characters, 16% of all errors were caused by entering extra characters, while the remaining 56% were caused by the combination of the both. A Kruskal-Wallis test failed to find a significant effect of condition on this ($H_1 = 0.6, ns$). To explore learning, we fitted power functions to sessional success rates to model the power law of practice [5]. In Figure 8, one can see that the data correlates well to the power functions for both basic ($R^2 = 0.92$) and color-coded ($R^2 = 0.99$) conditions.

Entry Speed

An ANOVA failed to identify a significant effect of condition on entry speed ($F_{1,24} = 0.03, ns$). The basic and color-coded groups yielded on average 52.9 ($\sigma_{\bar{x}} = 0.8$) and 52.7 ($\sigma_{\bar{x}} = 0.8$) WPM, respectively. There was also no significant effect of session in basic ($F_{2,12} = 1.3, ns$) and color-coded ($F_{2,12} = 0.6, ns$) groups. See Figure 9.

User Feedback

A Mann-Whitney U test failed to identify a significance with respect to children's overall rating of the feedback systems ($U = 75.5, Z = -0.523, p > .05$). About 92% and 85% children from basic and color-coded liked the

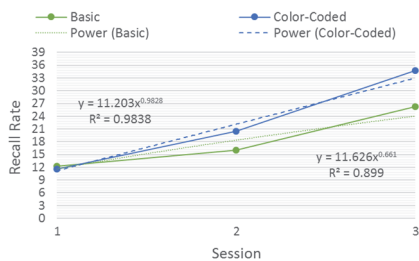


Figure 6. Average recall rate by session fitted to the power law of practice [5].

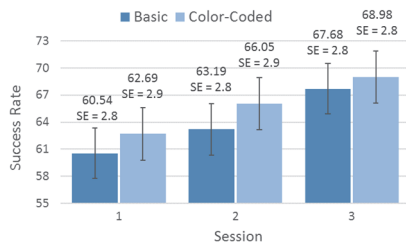


Figure 7. Average success rate by session. Error bars represent ± 1 standard error (σ_x).

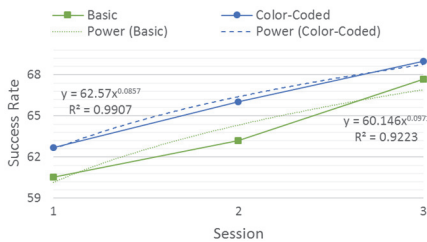


Figure 8. Average success rate by session fitted to the power law of practice [5].

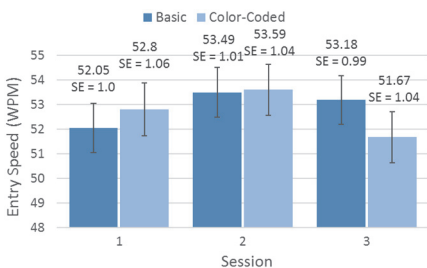


Figure 9. Average entry speed by session. Error bars represent ± 1 standard error (σ_x).

corresponding feedback methods. The remaining 8% and 15% children were mostly neutral (Figure 10).

A Mann-Whitney U test failed to identify a significance with respect to children’s perception of if the feedback methods helped them learn new words and spelling ($U = 82.0$, $Z = -0.150$, $p > .05$). 100% and 92% children from basic and color-coded felt that the corresponding feedback methods facilitated learning of new words and spelling. The remaining 8% were impartial (Figure 11). A Mann-Whitney U test failed to identify a significance with respect to willingness to use the feedback methods ($U = 75.0$, $Z = -0.552$, $p > .05$). 100% and 92% children from basic and color-coded wanted to keep using the corresponding feedback methods. The remaining 8% were neutral (Figure 12).

Discussion

There was no significant effect of condition on success rate, although on average about 3% more words were correctly spelled with color-coded than basic. Further, there was no significant effect of session in either of the groups. Figure 8 fits average success rate of both groups in all sessions to the power law of practice that shows that the data correlates well to the power functions for both groups. Hence, it is possible that learning of spelling occurs with autocorrections augmented with both feedback methods, but at too slow a pace to explore in the study. Interestingly, a significant effect of feedback was identified on recall rate. On average recall rate was about 23% higher with color-coded. There was also an effect of session. Recall rate improved substantially with time. Figure 6 fits average recall rate for both groups in all sessions to the power law of practice that shows that the data correlates well to the power functions for both groups. Both groups

started with a comparable recall rate ($\sim 12\%$), but in the last two sessions color-coded outperformed basic. In the final session, recall rate for color-coded was about 32% higher than basic. Yet, the fact that learning occurred in both groups with new words and spelling, although likely not at as fast a pace as conventional keyboards, demands further investigation into whether autocorrection can improve children’s language skills. There was no significant effect of condition on entry speed. Both groups yielded on average 53 WPM. There was also no significant effect of session in either of the groups, suggesting that children did not take extra time to process the information on the color-coded feedback. User feedback was also positive for both feedback methods. Most children liked the examined methods, felt that the methods helped them to learn new words and spelling, and thus wanted to keep using them.

Conclusion and Future Work

We presented a novel color-coded feedback method that highlights the corrections made by autocorrections. We then presented results of a user study with 7-8-year-old children that compared the proposed method with the conventional autocorrection feedback method, in terms of learning new words and spelling. Results suggested that the new method significantly improves learning new words. Interestingly, learning was also observed with the conventional method, demanding further investigation into if predictive methods are truly a barrier to learning new words and spelling. In the future, we will investigate this in another study that will compare learning rates with predictive and non-predictive input techniques.

Acknowledgments

We thank all students, teachers, and staffs of Colégio Teresiano Braga for their support. This work has been

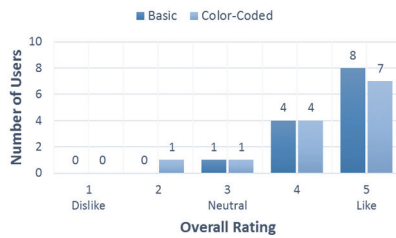


Figure 10. User feedback on if they liked the feedback systems.

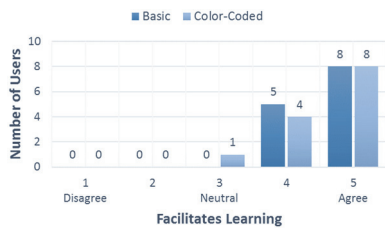


Figure 11. User feedback on whether the feedback facilitated learning new words and spelling.

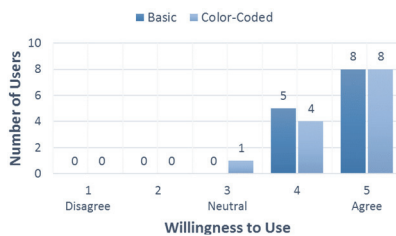


Figure 12. User feedback on if they are willing to use the examined feedback systems on their mobile devices.

supported in part by the Canada Research Chairs (CRC) Program, the Human Potential Operating Programme (HPOP) of the European Social Fund, and the Portuguese Ministry for Science, Technology and Higher Education (MCTES) Postdoctoral Grant SFRH/BPD/111891/2015.

References

- [1] Anthony, L., Brown, Q., Nias, J., and Tate, B. Examining the need for visual feedback during gesture interaction on mobile touchscreen devices for kids. *Proc. IDC 2013*, ACM (2013), 157-164.
- [2] Arif, A. S. and Stuerzlinger, W. Analysis of text entry performance metrics. *Proc. TIC-STH '09*, IEEE (2009), 100-105.
- [3] Arif, A. S. and Stuerzlinger, W. Pseudo-pressure detection and its use in predictive text entry on touchscreens. *Proc. OzCHI 2013*, ACM (2013), 383-392.
- [4] Arif, A. S. *Predicting and Reducing the Impact of Errors in Character-Based Text Entry*. York University, 2013.
- [5] Card, S. K., Newell, A., and Moran, T. P. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, 1983.
- [6] Cintero Systems. Kids Keyboard. <https://play.google.com/store/apps/details?id=com.cintero.kidskeyboard>
- [7] Druin, A. *Mobile Technology for Children: Designing for Interaction and Learning*. Morgan Kaufmann, 2009.
- [8] Google Translate API. <https://cloud.google.com/translate>
- [9] KidTech, Inc. The PC Computer Keyboard Designed Just for Kids (2015). <http://www.kidtech.com>
- [10] Kumon Publishing. *My Book of Sample Sentences: Learning about Nouns and Verbs*. Kumon Publishing North America, Inc., 2005.
- [11] Kurth, R. J. Using word processing to enhance revision strategies during student writing activities. *Educational Technology* 27, 1 (1987), 13-19.

[12] Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10, 8 (1966), 707-710.

[13] Nelson, K. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* 38, 1/2 (1973), 1-135.

[14] Newell, A. F., Booth, L., Arnott, J., and Beattie, W. Increasing literacy levels by the use of linguistic prediction. *Child Language Teaching and Therapy* 8, 2 (1992), 138-187.

[15] Papert, S. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc, 1980.

[16] Paton, G. Over-reliance on technology is undermining spelling skills. *The Telegraph* (May 22, 2012). <http://goo.gl/2MJz8c>

[17] Read, J. C. Text input for the elderly and the young. In *Text Entry Systems: Mobility, Accessibility, Universality*. Morgan Kaufmann, 2007, 271-288.

[18] Read, J. C. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology a Work* 10, 2 (2008), 119-128.

[19] Rideout, V. Zero to eight: children's media use in America. Common Sense Research Study (2013). <https://www.common sense media.org/research/zero-to-eight-childrens-media-use-in-america-2013>

[20] Solutions for Individuals with Disabilities (1998). <https://www.ablenetinc.com>

[21] Sturm, J. Using computer software tools to facilitate narrative skills. *The Clinical Connection* 11, 4 (1988), 6-9.

[22] University of Alabama at Birmingham. Could texting and autocorrect affect kids' writing skills? *ScienceDaily* (May 19, 2014). <https://goo.gl/WoDFSz>

[23] Wood, L. A., Rankin, J. L., and Beukelman, D. R. Word prompt programs: current uses and future possibilities. *American Journal of Speech-Language Pathology* 6, 3 (1997), 57-65.