Translating a Typing-Based Adaptive Learning Model to Speech-Based L2 Vocabulary Learning

Wilschut, Thomas; van der Velde, Maarten; Sense, Florian; Fountas, Zafeirios; van Rijn, Hedderik

DOI:
10.1145/3450613.3456825

Translating a Typing-Based Adaptive Learning Model to Speech-Based L2 Vocabulary Learning

Thomas Wilschut
t.j.wilschut@rug.nl
University of Groningen
Groningen, Netherlands

Maarten van der Velde
m.a.van.der.velde@rug.nl
University of Groningen
Groningen, Netherlands

Florian Sense
f.sense@rug.nl
University of Groningen
Groningen, Netherlands

Zafeiros Fountas
f@emotech.co
Emotech Ltd.
London, United Kingdom

Hedderik van Rijn
d.h.van.rijn@rug.nl
University of Groningen
Groningen, Netherlands

ABSTRACT

Memorising vocabulary is an important aspect of formal foreign language learning. Advances in cognitive psychology have led to the development of adaptive learning systems that make vocabulary learning more efficient. These computer-based systems measure learning performance in real time to create optimal study strategies for individual learners. While such adaptive learning systems have been successfully applied to written word learning, they have thus far seen little application in spoken word learning. Here we present a system for adaptive, speech-based word learning. We show that it is possible to improve the efficiency of speech-based learning systems by applying a modified adaptive model that was originally developed for typing-based word learning. This finding contributes to a better understanding of the memory processes involved in speech-based word learning. Furthermore, our work provides a basis for the development of language learning applications that use real-time pronunciation assessment software to score the accuracy of the learner's pronunciations. Speech-based learning applications are educationally relevant because they focus on what may be the most important aspect of language learning: to practice speech.

KEYWORDS

memory, adaptive learning, vocabulary learning, pronunciation, speech

ACM Reference Format:
Thomas Wilschut, Maarten van der Velde, Florian Sense, Zafeiros Fountas, and Hedderik van Rijn. 2021. Translating a Typing-Based Adaptive Learning Model to Speech-Based L2 Vocabulary Learning. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21), June 21–25, 2021, Utrecht, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3450613.3456825

UMAP '21, June 21–25, 2021, Utrecht, Netherlands
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8366-0/21/06.
https://doi.org/10.1145/3450613.3456825

1 INTRODUCTION

Storing word representations in the mental lexicon is one of the most important aspects of learning a new language. Because the process of memorising words is tedious and effortful, methods that can improve the efficiency of this process are valuable for anyone who is learning a new language [10]. In recent years, advances in cognitive psychology have led to the development of adaptive learning systems that aim to improve the process of word learning by determining optimal learning strategies for individual learners in real time. These digital systems typically focus on teaching orthography (i.e., the letters that spell the word) and require the learner to respond by typing or selecting the correct words (e.g., [14],[22],[30],[33]). Several variables, such as accuracy and reaction times, are measured during the learning process and are used to determine optimal repetition schedules for individual learners. In practice, using such adaptive learning systems results in higher learning efficiency than learning with traditional, non-adaptive methods, which translates into better retention at the end of the study sessions [30].

Learning systems can employ various degrees of adaptivity. Almost no systems are completely static, as most flashcard-based learning systems often register the accuracy of the learners responses to determine the repetition schedule, even though the level of adaptivity is fairly coarse. Fully adaptive systems, as mentioned above, measure various learning characteristics and use these to create a repetition schedule that is continually adapted to each individual learner.

Although such adaptive learning methods have made *written* word learning more efficient, the possibilities for adaptive speech-based learning have not yet received considerable scientific attention. Some language learning systems currently employ speech recognition software to automatically assess the accuracy of pronunciations (for example, see Duolingo (duolingo.com), Graphogame (graphogame.com), Rosetta Stone (rosettastone.com) or ProTutor [9]). Other systems use text-to-speech technology to provide feedback to the learners (for example, see Alex [19], [20]). However, these systems typically only use speech technology for learner-feedback, and not for more refined adaptation. While some studies have found promising results concerning the effectiveness of these systems for pronunciation learning (see [4]), the possibilities for improving speech-based learning using adaptive algorithms has not yet been examined.

Speech-based learning systems have numerous potential advantages compared to typing-based systems. First, speech-based systems allow the learner to learn the correct pronunciation of words, which is an important part of language acquisition that is completely omitted in typing-based learning. Second, since speaking a word is usually faster than typing it, speech-based learning systems could allow for a more efficient use of the available study time. Third, speech-based learning systems could be used by people who lack the opportunity to type (e.g., while driving a car or walking) or the ability to type (e.g., young children, elderly people or people with a physical disability), making them applicable in a wide range of settings. Hence, combining the advantages of adaptivity and speech-based vocabulary learning seems particularly promising.

In this study, we applied the *Rugged Learning* system [30] to speech-based learning. Originally developed for typing-based learning, Rugged Learning aims to create maximally efficient repetition schedules for individual learners by combining the beneficial effects of retrieval practice and spacing [30], [24]. Active retrieval practice, rather than passively rehearsing the study material, greatly contributes to learning efficiency (e.g.,[23]; see [17] for a review). Spacing learning sessions over time consistently results in better long-term memory consolidation [7], [12]. The Rugged Learning system balances the two above-mentioned mechanisms by rehearsing items just before they are estimated to be forgotten. The system uses the ACT-R architecture of human declarative memory to model the activation of each word in the learner's memory [2]. Individual learning differences are captured by a single parameter called the rate of forgetting (RoF), which is computed for each item and which is continuously updated throughout the learning session using reaction times and accuracy scores. The RoF is used to determine optimal repetition schedules for each learner (see [30] and [24] for details). The system has proven itself in both lab studies [24][25] [29] and real-world applications [30][26], yet it is currently limited to orthographic inputs. Here, we build upon the existing framework which we extend to work with speech input.

When learning a language, the learner has to store an association between the meaning of words (their semantic representation) and their form, which consists of phonology (sound) and orthography. These associations are stored in a mental lexicon, which is a long-term memory store for words. The lexicon has three interacting parts that contain the semantic, orthographic and phonological representations of words, see [1]. One of the core assumptions of the Rugged Learning algorithm is that reaction times can be used as a proxy of the memory activation of a word: the faster a correct response is produced, the stronger the memory representation [3]. The assumption that reaction time can reflect memory activation is further substantiated by a long tradition of research in word acquisition and retrieval ([11], see [13] for a review). Crucially, a similar relationship should hold between reaction times and memory activation for phonology or spoken words, because the two relationships are based on functionally similar encoding and retrieval mechanisms [3], [11]. If this assumption holds, the beneficial effects of using adaptive, reaction time-based algorithms that were found for typing-based learning should also apply to speech-based learning.

In the current study, we applied an adaptive learning (AL) algorithm designed for typing-based learning to speech-based learning. Because of the assumed functional similarity between speaking- and typing-based acquisition and retrieval, we hypothesise that (1) typing- and speech-based AL will lead to similar behavioural learning outcomes, and (2) the AL benefits found in typing-based setups will generalise to speech-based learning. We tested these hypotheses by comparing a speech-based learning session using the Rugged Learning model to (A) a typing-based learning session that employed the same adaptive learning algorithm and (B) a speech-based session using a flashcard algorithm that repeated incorrectly answered questions sooner than correctly answered questions. This comparison mirrors the experiment that was conducted by van Rijn and colleagues in 2009 [30], in which the fully adaptive Rugged Learning algorithm proved to be a more effective study method compared to a less adaptive flashcard system for typing-based learning.

2 METHODS

2.1 Participants

In total, 21 people completed all parts of this experiment, of whom 7 participants were native German speakers, and 14 participants were native Dutch speakers. Participants were first-year psychology students who were between 19 and 24 years old at the moment of participation. Participants received course credit for participation. All participants gave informed consent and the study was approved by the ethics committee of the department of psychology at the University of Groningen (study code: PSY-1920-S-0323).

2.2 Design and Procedure

The study had three parts that each participant completed in the same sequence, see Fig. 1. Each part had the same structure: a 12-minute study session, in which native Dutch participants studied a set of Dutch-English word pairs and in which native German participants studied a set of German-English word pairs (see Materials), was followed by a 3-minute filler task in which participants were asked to complete simple integer sequences (see Materials). Each part ended in a test. All items that the participant encountered during the learning session were asked on the test, in the order in which they were introduced during the learning session. The parts differed in how participants were asked to respond (typing or speaking) and in the way in which the items were scheduled (using the Rugged Learning algorithm or using a flashcard algorithm).

The first part was adaptive and typing-based. At the first presentation of a word, the Dutch/German written word was presented on a computer screen together with the written English translation of this word. In subsequent presentations of the word pair, only the written Dutch/German word was presented to the participants, and they were asked to type the correct English translation of the word and received corrective feedback. The Rugged Learning adaptive algorithm determined when each item was repeated and when new items were introduced, based on learners' reaction times and accuracy scores. See [24] for a detailed description of the algorithm used.

The second part was adaptive and speech-based. As in the adaptive-typing part, the written Dutch/German word was presented to the

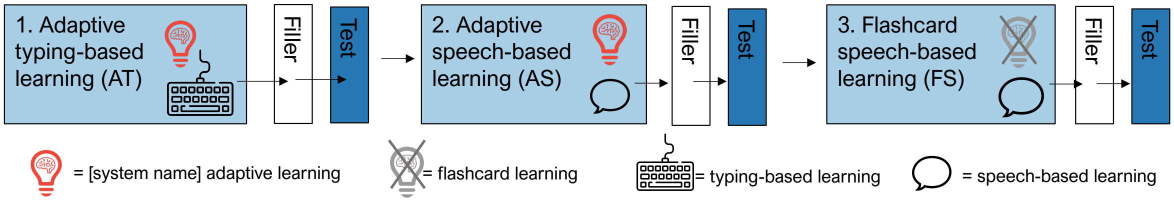
1. Adaptive typing-based learning (AT)
Filler
Test
2. Adaptive speech-based learning (AS)
Filler
Test
3. Flashcard speech-based learning (FS)
Filler
Test
= [system name] adaptive learning
= flashcard learning
= typing-based learning
= speech-based learning

Figure 1: Experimental design.

participants. Simultaneously, the participants heard the correct pronunciation of the English translation (see Materials) through headphones. Next, the written Dutch/German word was presented to the participants, and they were asked to pronounce its English translation. As in the adaptive-typing part, the Rugged Learning algorithm determined the order and time of the presentation of the words. Reaction times were measured using the voice onset. The accuracy of the answers was manually scored by the experimenter. If the answer was correct, the written prompt 'correct' was shown on the screen. If it was incorrect, the participants saw the prompt 'incorrect, the correct answer was . . . ' and again heard the correct pronunciation.

The third part was speech-based and a less adaptive flashcard algorithm was used (see below). To allow for a direct comparison to the adaptive speaking part, the number of studied words was equal to that in the adaptive speaking part (which varied between participants, depending on performance). Words were repeated based on the accuracy of earlier responses using a Leitner flashcard system [18], which groups words into three virtual boxes: All words start in Box 1 and move to Box 2 if answered correctly. If a word is answered incorrectly, it moves to the previous box. This flashcard system allows for difficult items to be rehearsed more often than easy items and has been shown to be a relatively effective study strategy [6]. The answer scoring and feedback were the same as in the adaptive speaking part.

2.3 Materials

The experiment was built with JavaScript and HTML5 using the jsPsych experiment library [8]. Since COVID-19 restrictions prevented any lab experiments, the experiment was conducted remotely. Participants were asked to be located in a quiet room and wear headphones. The experimenter's screen, which hosted the experiment, was shared with the participant using Skype (skype.com). Participants recorded audio and video that was sent back to the experimenter in real time. Voice onset times were measured by the experimenter using a physical delayed key trigger box, that registered the onset of all sounds that lasted longer than 98ms. Audio was looped using Loopback (rogueamoeba.com/loopback/), such that the voice trigger box only received the participants' audio recordings and did not receive audio from the experimenter or the example pronunciations in the experiment. The accuracy of the responses was manually scored by the experimenter using a USB gamepad during both speaking parts of the experiment.

Study materials were prepared in three lists of 30 word pairs. Lists were randomly assigned to each part of the experiment (counterbalanced across participants). Each of the three lists appeared each block the same number of times, in order to control for word difficulty. Words were selected on the basis of (1) being difficult to pronounce for native Dutch/German speakers, such as the *th*-sound in *thersitical*, (2) having an irregular orthography-phonology mapping, such as *hierarchy* or *awry*, (3) having difficult stress, such as *analysis*, or (4) being long and contain many consonants, such as *omphaloskepsis*. The proportional distribution of words from each category was equal for all three lists of words. The correct exemplar pronunciations that were provided to the participants were generated by Google's WaveNet text-to-speech algorithm (cloud.google.com/text-to-speech) in British English.

In the three-minute filler task, participants completed integer sequences in an open-question format (e.g., '3-6-12-24-?' requires response $2 \times 24 = 48$).

Words, exemplar voice materials, and filler items can be found in the online supplement at https://osf.io/cm72k/.

2.4 Analysis

The data was pre-processed and analysed using Python 3.0.3 [31], using the pandas [15] and numpy [21] packages. Video and audio data were processed in Python using the ffmpeg package [28]. Statistical analyses were conducted in R 3.4.1 [27], with the linear mixed-effects package lme4 [5]. The data was visualised using ggplot2 [32].

3 RESULTS

The main aims of this study were (1) to compare typing- and speaking based AL and (2) to examine the beneficial effects of employing an AL algorithm in speech-based learning. In order to address differences in learning efficiency between typing-based AL, speech-based AL and speech-based flashcard learning, we compared descriptive statistics for the average number of studied items, number of trials and trial durations. Furthermore, we fitted mixed effects regression models for the accuracy of responses, reaction times and RoF, see Table 1 and Fig. 2. In all regression models, we controlled for variance between items and between participants by adding these factors as random intercepts. Session (study or test) was added as a fixed effect to the models. We will first address the differences for the above-mentioned variables between the three study conditions in turn, and then discuss the results in light of the research questions in the discussion section below.

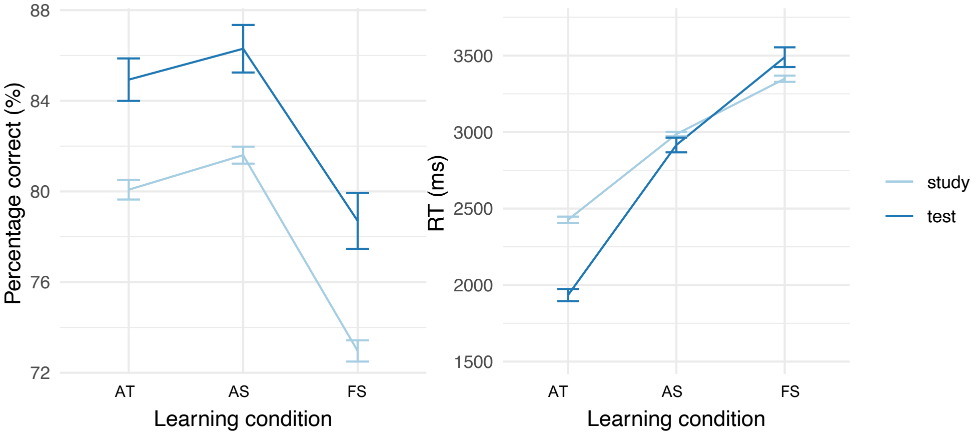

Figure 2: Average percentage of correct responses and reaction times for the three learning conditions, separate for study and test session. Error bars represent one standard error of the mean. AT = Adaptive typing; AS = Adaptive speaking; FS = Flashcard speaking.

Table 1: Mixed effects model results. Model 1 is a logistic regression, Models 2 and 3 use linear regression.

Model 1: Accuracy	β	SE		z	p
Intercept (adaptive speaking)	1.957	0.149		13.099	<0.001 ***
Adaptive typing	-0.187	0.077		-2.444	0.015 *
Flashcard speaking	-0.713	0.073		-9.741	<0.001 ***
Session (study vs test)	0.214	0.099		2.164	0.031*
Model 2: Reaction time (ms)	β	SE	df	t	p
Intercept (adaptive speaking)	2833	87.93	69	32.215	<0.001 ***
Adaptive typing	-521	47.70	7779	-10.912	<0.001 ***
Flashcard speaking	554	47.29	7778	11.796	<0.001 ***
Session (study vs test)	-58	59.45	7701	-0.977	0.328
Model 3: RoF	β	SE	df	t	p
Intercept (adaptive speaking)	0.414	0.005	0.673	81.232	<0.001 ***
Adaptive typing	-0.039	0.002	0.005	-23.331	<0.001 ***

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

The number of distinct items that were presented to the participants by the adaptive learning algorithm depended on their reaction times and accuracy (see Introduction). On average, participants studied approximately 17 items in the adaptive typing part and 13 items in the adaptive speaking part. Participants completed on average 97 trials in the adaptive typing part, 118 trials in the adaptive speaking part and 101 trials in the flashcard speaking part, which had an mean trial duration of 7400ms, 6080ms, and 7130ms, respectively. We fitted a logistic mixed effects regression model to predict binary accuracy from study condition and session using dummy coding (study = 0; test = 1). According to this model, the probability of giving a correct answer was 2.2 percentage points higher in the adaptive speaking condition than in the adaptive typing condition during the study session, and 1.9 percentage points higher during the test session[1]. There was a larger difference between adaptive speaking and flashcard speaking: the probability of giving a correct answer was 10.0 percentage points higher for adaptive speaking than for flashcard speaking during the study part, and 8.6 percentage points higher during the test part. There was a small effect of *session* on accuracy, indicating that the accuracy during test was, on average, slightly higher than accuracy during the study session. The interaction effects of session and condition were not significant, indicating that the above mentioned effects of learning condition were present both during test and study.

[1] The logistic regression coefficients in Table 1 can be converted to probabilities using an inverse logit transform. For example, adaptive speaking during the study session = $exp(1.957)/(1+exp(1.957)) = 0.876$, compared to adaptive typing = $exp(1.957-0.187)/(1 + exp(1.957 - 0.187)) = 0.854$.

Second, we fitted a linear mixed effects model to examine the differences in reaction times between the three learning conditions (interaction effects are not shown in table). Reaction times were on average 521ms shorter in the adaptive typing condition than in the adaptive speaking condition, and participants responded on average 554 ms faster in the adaptive speaking condition than in the flashcard speaking condition. In addition, there was a significant interaction effect of session and learning condition, indicating that reaction times were shorter during test in the typing condition, but not in the speaking conditions, t (7683) = -3.55, p < 0.001 (not shown in Table 1).

Third and finally, we examined the differences between RoF between the adaptive typing and adaptive speaking part (there was no estimated RoF in the flashcard speaking part, since we did not apply the adaptive learning algorithm in this condition). The RoF was, on average, approximately 0.04 points lower in the typing than in the adaptive speaking condition.

4 DISCUSSION AND CONCLUSION

Both of our initial hypotheses were confirmed: typing- and speech-based adaptive learning (AL) led to relatively similar behavioural learning outcomes, and the benefits of typing-based AL (relative to flashcards) generalised to speech-based AL. More specifically, learners typically responded faster and studied more words when typing rather than speaking, but at the cost of slightly lower accuracy. Notwithstanding these differences between the two types of learning, the results show that overall accuracy was similar for typed and spoken responses. Using either variant of adaptive learning resulted in higher accuracy and faster reaction times than using the flashcard algorithm.

Regarding the possibility of using an existing AL algorithm—that was designed for typing-based learning—to do spoken word learning, we hypothesised that the functional mechanisms of typing- and speech-based learning would be similar enough for the typing-based AL system to be applicable to speech-based adaptive learning. The results of this study strongly support this hypothesis: both typing- and speaking-based AL were superior to the flashcard algorithm in terms of average accuracy and reaction time (see Fig. 2). This suggests that voice onset times can stand in for keystroke-based reaction times to infer latent memory strength.

Despite the overall similarity in performance, some differences between typing and speech-based learning were found. Most prominently, typing-based AL was associated with shorter average reaction times, but longer trial durations, than speech-based AL. There is a plausible explanation for these differences: in the typing condition, participants may have started typing an answer *before* they completely retrieved the correct answer, and paused during their response. In other words, the retrieval process may have partially taken place during the typing of the answer. In line with this explanation, the data shows longer average durations between the first reaction (either the first key-press or the voice onset time) and the feedback (which appeared right after the completion of the answer) in the typing condition than in the adaptive speaking condition. There was a 10% difference in RoF between adaptive typing and adaptive speaking based learning. Although this difference is significant, it is relatively small compared to the range of values that has been found in previous studies (e.g. [24]). The shorter reaction times and lower RoF caused the algorithm to select more words to be studied in the adaptive typing part than in the adaptive speaking part. The average accuracy of these studied words was higher in the adaptive speaking part then in the typing part. In short, adaptive typing resulted in more items studied with lower accuracy, whereas speaking resulted in fewer items studied with higher accuracy. Hence, overall accuracy was similar in the adaptive typing adaptive speech conditions. Taken together, these results point towards a strong functional similarity between typing-based and speech-based word learning and retrieval.

Our findings lead to several suggestions for future work. In this study, spoken responses were manually scored by the experimenter. Recent technological advances allow for the automatic, real-time assessment of pronunciation accuracy. Using automatically assessed pronunciation accuracy does not only lead to more objective accuracy measures, but could also be used to provide detailed feedback to the learner, which may further enhance the effectiveness of speech-based word learning. This approach showed promising results in a pilot study conducted in our lab. In addition, pronunciation quality—expressed as the degree of overlap between the learner's pronunciation and a reference exemplar—would provide a continuous score, which might prove to be a more sensitive measure of memory strength than binary accuracy. Adaptive systems that use both continuous reaction times and accuracy have been shown to outperform systems that use binary accuracy only (e.g., [30], [16]). Future work should explore whether combing two continuous scores (voice onset time and pronunciation quality) could further improve such systems.

In conclusion, in this study we successfully applied an adaptive learning algorithm that was developed for typing-based learning to speech based learning. Despite differences in study pace between typing- and speech-based learning, it seems to be possible to use the same behavioural measures to estimate memory parameters in both learning systems. As a consequence, we were able to successfully improve the efficiency of speaking based learning using an adaptive system: learners who studied using the AL system were able to produce faster responses with 9-10 percentage points higher accuracy compared to learners who used a less adaptive, flashcard speaking based learning system. These results are important in two ways. First, they contribute to understanding the memory mechanisms involved in speech-based language learning, which have received too little attention so far. Second, this study contributes to the development of language learning systems that can be applied in a wide range of settings. Such applications have practical importance, because they incorporate one of the most important parts of language learning: to practise speech.

REFERENCES

[1] Jean Aitchison. 2012. *Words in the mind: An introduction to the mental lexicon.* John Wiley & Sons.

[2] John R Anderson, Dan Bothell, Christian Lebiere, and Michael Matessa. 1998. An integrated theory of list memory. *Journal of Memory and Language* 38, 4 (1998), 341–380.

[3] John R Anderson and Lael J Schooler. 1991. Reflections of the environment in memory. *Psychological science* 2, 6 (1991), 396–408.

[4] Joan Palmiter Bajorek. 2017. L2 Pronunciation in CALL: The Unrealized Potential of Rosetta Stone, Duolingo, Babbel, and Mango Languages. *Issues and Trends in Educational Technology* 5, 1 (2017), 24–51.

[5] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, and Fabian Scheipl. 2012. Package 'lme4'. *CRAN. R Foundation for Statistical Computing, Vienna, Austria* (2012).

[6] David Bryson. 2012. Using flashcards to support your learning. *Journal of visual communication in medicine* 35, 1 (2012), 25–29.

[7] Nicholas J Cepeda, Edward Vul, Doug Rohrer, John T Wixted, and Harold Pashler. 2008. Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological science* 19, 11 (2008), 1095–1102.

[8] Joshua R De Leeuw. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods* 47, 1 (2015), 1–12.

[9] Carrie Demmans Epp and Gordon McCalla. 2011. ProTutor: Historic open learner models for pronunciation tutoring. In *International Conference on Artificial Intelligence in Education*. Springer, 441–443.

[10] Joshua K Hartshorne, Joshua B Tenenbaum, and Steven Pinker. 2018. A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* 177 (2018), 263–277.

[11] Jörg D Jescheniak and Willem JM Levelt. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 4 (1994), 824.

[12] Jeffrey D Karpicke and Althea Bauernschmidt. 2011. Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37, 5 (2011), 1250.

[13] Willem JM Levelt. 1999. Models of word production. *Trends in cognitive sciences* 3, 6 (1999), 223–232.

[14] Robert V Lindsey, Jeffery D Shroyer, Harold Pashler, and Michael C Mozer. 2014. Improving students' long-term knowledge retention through personalized review. *Psychological science* 25, 3 (2014), 639–647.

[15] Wes McKinney et al. 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, Vol. 445. Austin, TX, 51–56.

[16] Everett Mettler, Christine M Massey, and Philip J Kellman. 2011. Improving Adaptive Learning Technology through the Use of Response Times. *Grantee Submission* (2011).

[17] Bruna Fernanda Tolentino Moreira, Tatiana Salazar Silva Pinto, Daniela Siqueira Veloso Starling, and Antônio Jaeger. 2019. Retrieval practice in classroom settings: a review of applied research. In *Frontiers in Education*, Vol. 4. Frontiers, 5.

[18] Rehana Mubarak and Daniela C Smith. 2008. Spacing Effect And Mnemonic Strategies: A Theory-Based Approach To E-Learning.. In *e-Learning*. 269–272.

[19] Cosmin Munteanu, Joanna Lumsden, Hélène Fournier, Rock Leung, Danny D'Amours, Daniel McDonald, and Julie Maitland. 2010. ALEX: mobile language assistant for low-literacy adults. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 427–430.

[20] Cosmin Munteanu, Heather Molyneaux, Julie Maitland, Daniel McDonald, Rock Leung, Hélène Fournier, and Joanna Lumsden. 2014. Hidden in plain sight: low-literacy adults in a developed country overcoming social and educational challenges through mobile learning support tools. *Personal and ubiquitous computing* 18, 6 (2014), 1455–1469.

[21] Travis E Oliphant. 2006. *A guide to NumPy*. Vol. 1. Trelgol Publishing USA.

[22] Jan Papousek, Radek Pelánek, and Vít Stanislav. 2014. Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining 2014*.

[23] Henry L Roediger III and Jeffrey D Karpicke. 2006. The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science* 1, 3 (2006), 181–210.

[24] Florian Sense, Friederike Behrens, Rob R Meijer, and Hedderik van Rijn. 2016. An individual's rate of forgetting is stable over time but differs across materials. *Topics in cognitive science* 8, 1 (2016), 305–321.

[25] Florian Sense, Rob R Meijer, and Hedderik van Rijn. 2018. Exploration of the Rate of Forgetting as a Domain-Specific Individual Differences Measure. *Frontiers in Education* 3, 112 (2018).

[26] Florian Sense, Maarten van der Velde, and Hedderik van Rijn. 2018. Deploying a Model-based Adaptive Fact-Learning System in a University Course. In *Proceedings of the 16th International Conference on Cognitive Modeling*. 138.

[27] R Core Team. 2017. R: A language and environment for statistical com-puting.

[28] Suramya Tomar. 2006. Converting video formats with FFmpeg. *Linux Journal* 2006, 146 (2006), 10.

[29] Maarten van der Velde, Florian Sense, Jelmer P Borst, and Hedderik van Rijn. [n.d.]. Alleviating the Cold Start Problem in Adaptive Learning using Data-Driven Difficulty Estimates. ([n. d.]).

[30] Hedderik Van Rijn, Leendert van Maanen, and Marnix van Woudenberg. 2009. Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th International Conference of Cognitive Modeling*, Vol. 2. 7–6.

[31] Guido Van Rossum and Fred L Drake. 2009. *Introduction To Python 3: Python Documentation Manual Part 1*. CreateSpace.

[32] Hadley Wickham. 2016. *ggplot2: elegant graphics for data analysis*. springer.

[33] Piotr A Wozniak and Edward J Gorzelanczyk. 1994. Optimization of repetition spacing in the practice of learning. *Acta neurobiologiae experimentalis* 54 (1994), 59–59.