Radboud University Nijmegen

Advanced Research Methods

Artificial Grammar Learning using Rapid Serial Visual Presentation

Authors:
Franziska Burger (s4500466)
Jules Kruijswijk (s4140230)
Volker Strobel (s4491491)
Inez M. Wijnands (s4149696)
Radboud University Nijmegen

Supervisor:
Dr. Makiko Sadakata
Donders Institute for Brain, Cognition and Behavior

SOW-MKI42a Advanced Research Methods
Artificial Intelligence
Faculty of Social Sciences
Radboud University Nijmegen

Abstract

1 Introduction

In our society today, a surefire way of making profit is by developing methods and tools that promise to help us manage life in the fast lane. One of the older methods with this purpose that is gaining in popularity again is the rapid intake of written information, also known as speed reading.

1.1 Speed Reading

Interest in increasing one's reading rate was sparked when Javal drew attention to eye movements during reading through a series of articles published in the 1870s [8, 9, 22]. Rather than reading words letter by letter, reading was now seen as an alternating process of fixations and saccades. Each fixation allows good readers to take in on average approximately three words. During saccades, nothing can be seen [1]. Research was fueled further after the American psychologist Renshaw employed tachistoscopes to train pilots of the U.S. Navy in the second World War [2, 6, 17]. A tachistoscope is an apparatus that allows the rapid serial visual presentation (RSVP) of stimuli. With this method, saccades could be eliminated by presenting words in the same location. The research regarding the effectiveness of RSVP to increase reading speed produced converging evidence that comprehension suffers under this approach [3, 4, 12, 19, 21, 23]. Despite such evidence, using RSVP for speed reading has recently resurfaced in popular culture in the form of smartphone and wearable applications, such as the one by Spritz [1]. Several studies have shown the importance of reading in first and second language acquisition, with reading improving not only comprehension but also inducing an intuition of the underlying syntactic structure [5, 11, 20]. Thus, we are not quite ready to write off RSVP as a method for speed reading yet. In this project, we study the possibilities of the method in artificial grammar learning (AGL).

1.2 Artificial Grammar Learning

In the 1960s, Reber studied the implicit learning of stimulus structure. By conducting a series of experiments, he defined three requirements [16, p. 190] for such learning to occur:

1. The rules governing the stimulus material must be complex.
2. Participants should pay close attention to stimuli without searching for underlying rules.
3. Participants should be unable to verbalize the knowledge they obtained during learning.

[1] http://www.spritzinc.com/

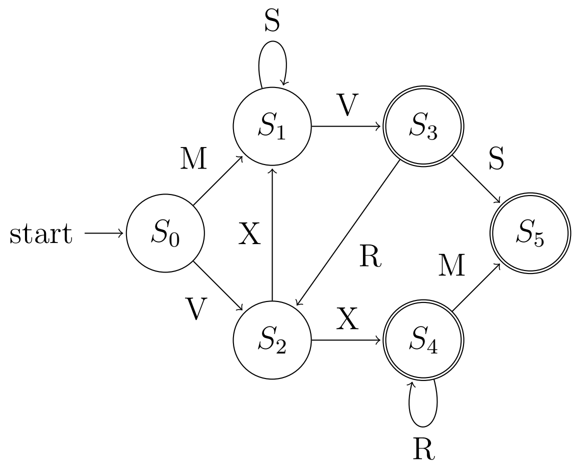

Figure 1: An artificial grammar is given by a finite state machine. States are indicated by circles labeled S_i. S_0 is the initial state. Every grammatical item has to start with a transition via an arrow from the initial state to a state directly connected to it (in this case S_1 or S_2). States encircled by two lines are final states (S_3, S_4, S_5). The machine can be exited from any final state. The arrows indicate legal transition directions (in this example, it is legal to move from S_3 to S_2 but not the other way around). Words are formed by concatenating the letters assigned to each arrow in the order of making the transition (grammatical words are for example MVS, VXRRR).

Reber employed artificial grammars to ensure meeting the first requirement. Artificial grammars are finite state machines with the transition from one state to the next corresponding to one element of the grammar. Grammatical patterns, or words, are created by making only allowed transitions from the start state to the final state (see Figure1). The AGL paradigm consists of a learning and a testing phase. In the learning phase, participants are confronted with a set of grammatical words[2]. In the testing phase, acquired knowledge is tested by presenting new words and asking participants to judge their grammaticality. This raises the question: How is it possible to develop a feeling for apparently nonsensical patterns? One theory is that implicit learning is statistical learning. This means that the grammatical patterns in the training phase are used to train a classifier within the brain, which tells us the probability for a new stimulus to be a word of the grammar or not [13, 18]. While Reber exposed participants to a total of 20 exemplars three times and each exemplar was regarded for ten seconds [16], later studies showed similar results for shorter exposure [7, 10][3].
In this project, we are interested in the possibility of implicitly learning an artificial grammar by viewing exemplars presented in an RSVP paradigm. Specifically, we examine two research questions:

1. Does AGL occur when words are presented individually at the same location at a normal reading rate?

[2]In his initial studies [14], Reber asked participants to memorize these but later went on to study learning by mere exposure [16].

[3]Knowlton and Squire exposed participants to stimuli for three seconds per item and then asked them to reproduce the item. When participants could not get it correct on this reproduction attempt, this could be repeated twice, so that theoretically participants could have seen items for a total of nine seconds. However, it is mentioned that few participants needed a third exposure to reproduce the item correctly. Gomez and Gerken exposed 1 year-old toddlers to auditory exemplars of an artificial grammar that included vowels for a total of two minutes.

2. Does it also occur at faster presentation speeds (as used by speed reading tools) when
   (a) the same amount of material is presented?
   (b) material is presented for the same amount of time?

To our knowledge, the limitations of AGL have not yet been studied with regard to the presentation speed of exemplars. We generally regard AGL as possible when exemplars are presented at normal and rapid speeds. This is due to the high attentional demand of RSVP (REFERENCE NEEDED) in combination with the importance of attention in implicit learning. However, if implicit learning is statistical learning, we further believe that longer exposure times will most likely lead to more reliable probabilistic inferences, with a repeated exposure potentially making up for a shorter exposure. Therefore, we expect an affirmative answer to research questions 1 and 2b.

2 Methods

2.1 Participants

Thirty right-handed healthy subjects participated in this experiment (nine females, twenty-one males; age: 19−2.. years (mean = years; S.E. = years)). All the participants gave informed consent.

Participants were primarily found amongst AI students at the Radboud University and other students from the Faculty of Social Sciences. Participants were excluded from the experiment when they exhibit certain symptoms or disorders, amongst epilepsy and dyslexia. The following exclusion criteria were used:

- Cognitive deficits (or unfamiliarity with the English language) that made comprehension of the information letter and instructions difficult, or motor impairment that made comprehension of the task or pushing a button impossible.
- Epilepsy. Rapid serial visual presentation uses fast presented stimuli, so this criteria was added as a precaution.
- Dyslexia. We used artificial words in the experiment and the participant should have no trouble distinguishing between letters and registering the order of letters.

2.2 Experimental paradigm

The experiment had a between subject design. To asses how well the subjects performed on artificial grammar learning using rapid serial visual presentation, we used three groups:

1. The task was performed using three times 20 words for training and the stimuli were presented at a normal reading rate of 150 words per minute.
2. The task was performed using three times 20 words for training and the stimuli were presented at a fast presentation rate of 450 words per minute.

3. The task was performed using nine times 20 words for training and the stimuli were presented at a fast presentation rate of 450 words per minute.

The participants were counter-balanced over groups. In each group were ten participants (group 1: six females, four males; age: ..–... years (mean = 22 years; S.E. = 2.4 years), group 2: one female, nine males; age ..–.. years (mean = 22.5 years; S.E. = 2.5 years), group 3: two females, eight males; age ..–... years (mean = 22.1 years; S.E. = 2.6 years)).

For each group the experiment followed the same structure (shown in Figure 3) and started with a questionnaire. The participant was asked to fill out their age, gender, the number of languages they speak and their experience with speed reading on a scale from 1 (none) to 7 (usage on a regular basis).

The participant had to perform three sessions in the experiment. The main task of the experiment is shown in Figure 4). The first session was a practice session to get acquainted with RSVP. The session began with the first three sentences of *Beauty and the Beast* CITATION, at a rate of 150 words per minute. Hereafter, the participant was prompted that the rate would increase to 200 words per minute after a button press. Three to four sentences were presented for each rate, until 500 words per minute was reached.
The second session was the training session. This session differed between groups. The participants were unaware of differences between groups, and they did not know at which rate the words would be presented, the instructions were the same for all participants and only included that we would use one of the rates presented during practice. The training consisted of three or nine (depending on the condition) takes of 20 words, all from the grammar shown in Figure 2. After every take there was time for a small break, so the participants would not get too weary to concentrate on the words.

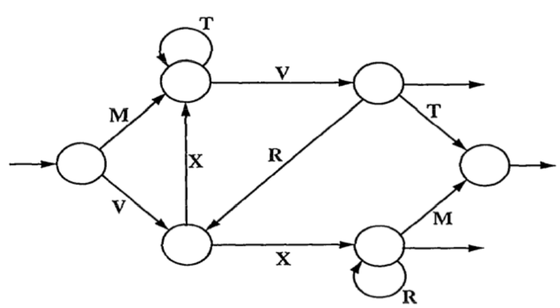

Figure 2: Artificial grammar used for the experiment CITATION.

The third session was the test session. The participants were instructed that the words from the training session followed certain rules. Even though they did not learn these rules explicitly, they were asked to answer questions following their 'gut feeling'. The participant were then presented with 100 questions, all consisting of a word - either grammatical or nongrammatical - and the question whether they judged the word as grammatical or nongrammatical. There was a total of 50 words, equally many grammatical and nongrammatical words. Each word was represented twice in the questions to check for consistency. The response time and correct response rate were measured. After the task, the participant is asked to fill out a post-questionnaire containing one question: how he decided whether a word was ruleful or unruleful. This was used to decide which rules the participant used - if any.

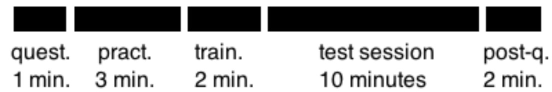

Figure 3: This shows the estimated timeline, including small breaks, of the experiment.

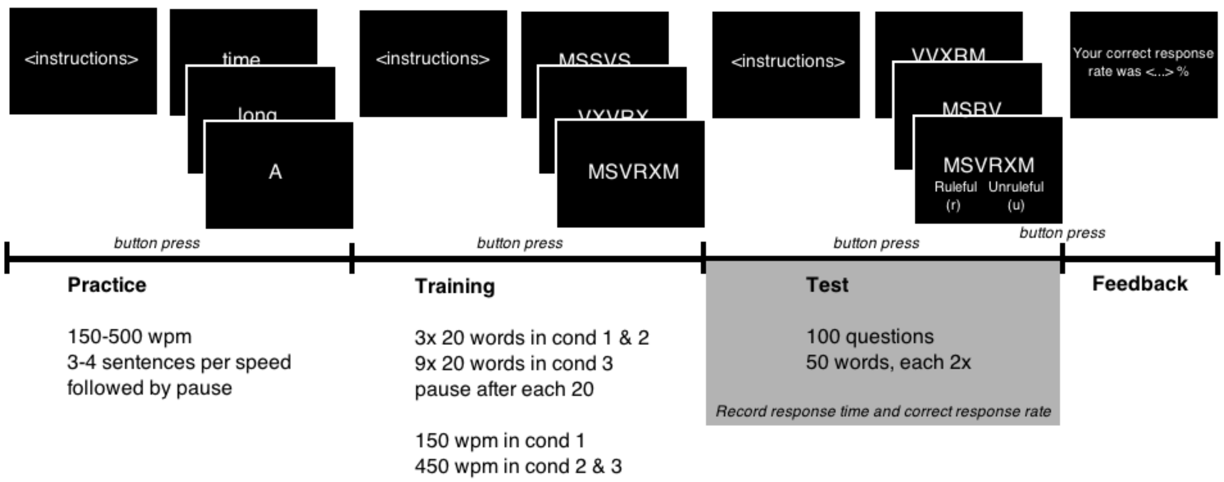

Figure 4: This figure shows the timeline of the task. The task starts with the welcome screen with instructions about the practice session. The practice session starts at 150 words per minute and after each 3-4 sentences the participant is notified and the speed increases with steps of 50. After practice, instructions for the training session are given and the training starts after a countdown of 3 seconds. During the test session, no feedback is given, only after all questions are answered is the correct response rate given. This concludes the task.

2.2.1 Feedback

In the experiment no feedback is given during the task. This is to ensure that the participants are not learning the rules during the test session and increasing performance for the remainder of the questions. Only after all questions are answered, the correct response rate is shown.

2.3 Materials

The experiment was done in a soundproof room with a desktop computer. A program to execute the test as explained earlier was developed in Python using the PsychoPy package (`http://www.psychopy.org/`). The stimuli that are based on the grammar as seen in Figure 2 and were used in the test can be found in Table 1.

2.4 Statistical analysis

3 Results

The total number of correct classifications on the 100 test items, P_C, was determined for each participant. To compare observed to chance level performance, the Wilcoxon signed-rank test was employed for the *normal* and the *fast-time* conditions, while a one-sample t-test could be used for the *fast-amount* condition. [4] The tests were performed one-tailed because only above chance level performance was of interest. As can be seen from the results of these tests, illustrated in Table X (insert Table), no group performed significantly above chance. Therefore, it would not be sensible to make any statistical between-group comparisons. However, although not significant, both the *normal* and the *fast-amount* groups show weak indications that learning may have occurred. Luckily, there are other measures that can provide additional information on this matter. The 100 test items consisted of 50 unique items which were all repeated once. Thus, participants could classify each unique item in one of four ways:

1. Correct-Correct (CC): classified correctly on both presentations
2. Correct-Erroneous (CE): classified correctly on the first presentation but incorrectly on the second
3. Erroneous-Correct (EC): classified incorrectly on the first presentation but correctly on the second
4. Erroneous-Erroneous (EE): misclassified on both presentations

The sum of CC and EE denotes overall consistency. Importantly, "when the status of the item is known, it is always classified correctly; when it is not known, a guess is made." [15, p. 227]. Using this simple model as a basis, Reber states that CE, EC, and EE should not be statistically distinguishable from each other, since they all reflect guesses. On the other hand, if EE is significantly greater than the average of CE and EC, it can be inferred that judgments were based on rules that are not representative of the grammar. Furthermore, if the participants actually implicitly learned a correct albeit partial representation of the grammar, CC should be significantly higher than each of the other three variables. Finally, if the difference between CE and EC or EC and CE is significant, it is indicative of forgetting or learning during the testing phase respectively (for an in-depth discussion see Reber [15]). For each group, Table XX shows the means for the four variables. Intuitively, the illustrated behavior of our participants seems representative of guessing behavior on all test items. Paired t-tests for the various combinations of these variables (all are distributed normally within each group) confirms this intuition. EC and CE do not show a significant difference. Thus, there is neither evidence for learning nor for forgetting during the testing phase. When comparing EE to the average of CE and EC, $t(9) \leq -4.216$ and $p < .01$ was obtained for all groups. Furthermore, CC was only significantly higher than EE in the *normal* group ($t(9) =$

[4] The Shapiro-Wilk test was used to check whether the small samples per condition were drawn from a normally distributed population [24]

1.843, $p = .049$) with the *fast-amount* group showing a trend ($t(9) = 1.758$, $p = .057$). Taken together, this implies that all participants based a considerable amount of their decisions on rules that were not part of the grammar. Only in the *normal* group (and possibly in the *fast-amount* group) were significantly more correct than incorrect rules applied. However, it must be noted that these differences are much smaller and thus much move vague than in the artificial grammar learning literature. Lastly, we correlated the number of languages participants speak as well as their prior speed reading experience with their P_C-scores across groups. Slightly negative but non-significant correlations were obtained: $r = -.30$, $p = .11$ for the number of languages and $r = -.23$, $p = .22$ for the speed reading experience.

References

[1] G. Ahuja and P. Ahuja. *How to Increase Your Reading Speed: Procedures and Practices*. Sterling Publishers Pvt. Ltd, 1995.

[2] R. Benschop. What is a tachistoscope? historical explorations of an instrument. *Science in context*, 11(01):23–50, 1998.

[3] J. R. Bormuth and C. C. Aker. Is the tachistoscope a worthwhile teaching tool? *The reading teacher*, 14(3):172–176, 1961.

[4] O. S. Causey. What the colleges are doing in reading improvement programs; yearbook of the annual meeting of the southwest reading conference for colleges and universities (3rd, texas christian university, december 11-12, 1953). 1954.

[5] C. Chomsky. Stages in language development and reading exposure. *Harvard Educational Review*, 42(1):1–33, 1972.

[6] E. C. Godnig. The tachistoscope - its history and uses. *Journal of Behavioral Optometry*, 14(2):39–42, 2003.

[7] R. L. Gomez and L. Gerken. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2):109–135, 1999.

[8] E. B. Huey. The psychology and pedagogy of reading: With a review of the history of reading and writing and of methods, text, and hygiene of reading, 1908.

[9] E. Javal. Essai sur la physiologie de la lecture. In *Annales D'Oculistique*, 1879.

[10] B. J. Knowlton and L. R. Squire. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1):169, 1996.

[11] S. D. Krashen. *The power of reading: Insights from the research*. Libraries Unlimited Englewood, CO, 1993.

[12] G. W. McConkie, K. Rayner, and S. J. Wilson. Experimental manipulation of reading strategies. *Journal of Educational Psychology*, 65(1):1, 1973.

[13] P. Perruchet and S. Pacton. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, 10(5):233–238, 2006.

[14] A. S. Reber. Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, 6(6):855–863, 1967.

[15] A. S. Reber. Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118(3):219, 1989.

[16] A. S. Reber and R. Allen. Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, 6(3):189–221, 1978.

[17] S. Renshaw. The visual perception and reproduction of forms by tachistoscopic methods. *The Journal of Psychology*, 20(2):217–232, 1945.

[18] J. R. Saffran. Constraints on statistical language learning. *Journal of Memory and Language*, 47(1):172–196, 2002.

[19] E. R. Schotter, R. Tran, and K. Rayner. Don't believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science*, page 0956797614531148, 2014.

[20] L. S. Siegel and E. B. Ryan. Development of grammatical-sensitivity, phonological, and short-term memory skills in normally achieving and learning disabled children. *Developmental Psychology*, 24(1):28, 1988.

[21] J. Swalm and M. Kling. Speed reading in the elementary school. *The Elementary School Journal*, pages 158–164, 1973.

[22] N. J. Wade and B. W. Tatler. Did javal measure eye movements during reading. *Journal of Eye Movement Research*, 2(5):1–7, 2009.

[23] P. A. Witty. Rate of reading: A crucial issue. *Journal of Reading*, 13(2):102–163, 1969.

[24] B. Yazici and S. Yolacan. A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, 77(2):175–183, 2007.

Training words (grammatical)	*Test words (grammatical)*	*Test words (nongrammatical)*
MSSSSV	MSSSV	MMVRX
MSSVS	MSSSVS	MSM
MSV	MSSV	MSRV
MSVRX	MSSVRX	MSRVRX
MSVRXM	MSV	MSSVSR
MVRX	MSVRXR	MSVRSR
MVRXRR	MSVRXV	MSVV
MVRXSV	MSVS	MXVRXM
MVRXV	MVRXM	MXVS
MVRXVS	MVRXR	MVRSR
VXM	MVRXRM	RRRXV
VXRR	MVRXVS	RVS
VXRRM	MVS	SSVS
VXRRRR	VXR	SVSSXV
VXSSVS	VXRM	SXRRM
VXSVRX	VXRRR	VRRRM
VXSVS	VXRRRM	VVXRM
VXVRX	VXSV	VXMRXV
VXVRXV	VXSSV	VXRRS
VXVS	VXSSSV	VXRS
	VXV	VXRVM
	VXVRX	VXX
	VXVRXR	XRVXV
	VXVRXV	XSSSSV
	VXVS	XVRXRR

Table 1: Words used for the experiment.

Group	**Chance Level Comparison**	**p-value**
normal	Wilcoxon Signed Rank test	.068
fast-amount	One-sample t-test	.057
fast-time	Wilcoxon Signed Rank test	.237

Table 2: TODO

Parameter	Group					
	normal		**fast-amount**		**fast-time**	
	Mean	SD	Mean	SD	Mean	SD
P_C	53.3	5.8	52.9	5.2	50.9	4.8
CC	37.6	5.8	41.6	5.7	35.4	7.4
CE	16.2	6.8	12.4	4.1	14.8	2.4
EC	15.4	3.9	10.2	5.7	16.2	1.9
EE	30.8	7.5	35.8	7.7	33.6	8.2
consistency	68.4		77.4		50.9	

Table 3: TODO