

## **Examining the use of an online version of the maze task as a pedagogical tool for second language learning**

### **Examinando o uso de uma versão on-line da tarefa do labirinto como ferramenta pedagógica para a aprendizagem de segunda língua**

Cândido Samuel Fonseca de Oliveira<sup>1</sup>, Elizabeth Duane Santos da Costa<sup>2</sup>, Karina Lúcia Ribeiro Canabrava<sup>3</sup>, Nathalie Alacoque da Silva Barros<sup>4</sup>

*CEFET-MG e POSLIN/UFMG, Brasil; CEFET-MG e POSLIN/UFMG, Brasil; CEFET-MG, Brasil; UFMG e CEFET-MG, Brasil*

#### **RESUMO**

Este artigo busca explorar a tarefa labirinto como uma ferramenta pedagógica para aprimorar o conhecimento de aprendizes sobre a L2. Evidências recentes sugerem que o treinamento linguístico pode modificar o conhecimento linguístico (Wells et al., 2009; Enkin; Forster, 2014, Fraundorf; Jager, 2016, Hopp, 2016, 2020). Analisamos o comportamento de brasileiros aprendizes do inglês como segunda língua em relação à construção de objeto duplo após um treinamento on-line com a tarefa labirinto, comparando os resultados de um pré-teste e um pós-teste, nos quais medimos julgamentos de aceitabilidade e tempos de reação. Os resultados sugerem melhora em ambas as medidas. No entanto, não houve diferença significativa em comparação com um grupo de controle que treinou outras estruturas e um terceiro grupo que assistiu a uma pequena palestra sobre estruturas do inglês.

#### **PALAVRAS-CHAVE:**

Tarefa labirinto. Aprendizagem implícita. Treinamento linguístico. Ferramenta pedagógica. Aquisição de segunda língua.

#### **ABSTRACT**

This paper aims to explore the maze task as a pedagogical tool to enhance learners' knowledge of the L2. Recent evidence suggests that linguistic training can modify linguistic knowledge (Wells et al., 2009; Enkin; Forster, 2014, Fraundorf; Jager, 2016, Hopp, 2016, 2020). We analyzed the behavior of Brazilian learners of English with respect to the double-object construction after they performed an online maze task training by comparing the results from a pretest and a posttest, in which we measured acceptability ratings and reaction times. The results suggest improvement in both measures. However, the improvement rate was not significant compared to that in a control group trained on other structures and that in a third group that watched a short lecture on English structures.

#### **KEYWORDS:**

Maze task, Implicit learning, Linguistic training, Pedagogical tool, Second language acquisition.

*Recebido em: 15/05/2020*

*Aceito em: 14/08/2020*

<sup>1</sup> E-mail: coliveira@cefetmg.br | ORCID: 0000-0001-7578-6288.

<sup>2</sup> E-mail: elizabethduane@cefetmg.br | ORCID: 0000-0002-7489-6108.

<sup>3</sup> E-mail: karinacanabrava@cefetmg.br | ORCID: 0000-0002-9117-6129.

<sup>4</sup> E-mail: nathaliealacoque@gmail.com | ORCID: 0000-0003-2568-0623.

## Introduction

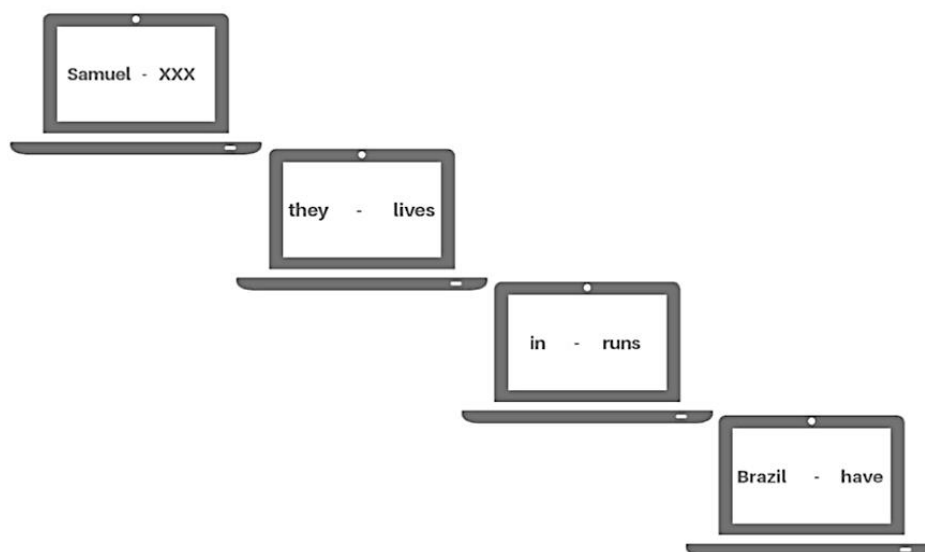
The maze task (MT) is a psycholinguistic technique that measures online processing time (Forster et al., 2009; Witzel et al., 2012; Oliveira, in press). The method has been gaining ground in the psycholinguistic field in recent years (e.g., Qiao et al., 2012; Kizach et al., 2013; O'bryan et al., 2013; Souza et al., 2014; Nyvad et al. 2015; Wang, 2015; Witzel; Witzel, 2016; Marcilese et al., 2017; Oliveira et al., 2017, Sikos et al., 2017; Souza; Oliveira, 2017; Barreto et al., 2018; Li et al., 2017; Hilpert; Saavedra, 2018; Mansbridge; Tamaoka, 2018; Suzuki; Sunada, 2018), especially because spillover and inattention effects (usually measured by comprehension questions) are mitigated during the task. During an experimental session, participants use a keyboard or a similar device to move through sentences word by word or chunk by chunk, and the software<sup>5</sup> records the time spent in each of these parts. One peculiarity of the MT is that it does not present only the words that compose the sentence. Instead, together with each part of the sentence, there is an additional option that does not complete the sentence being formed. Thus, what the software actually measures is the time participants take to make the right choice for each part of the sentence. The first piece of the sentence is usually the only part that is presented alone, since participants do not have any previous information about the sentence. After that, two options are presented for each part of the sentence, and only one of them allows the sentence to continue being formed. The incorrect option is usually a nonword (lexicity maze or L-maze) or an ungrammatical sequence for the sentence (grammaticality maze or G-maze) (Witzel et al., 2012). Figure 1 illustrates how the sentence “Samuel lives in Brazil” could be displayed in a G-maze.

---

<sup>5</sup> Different software programs were used to conduct the maze task, including DMDX and Linger.

---

Figure 1 – Visual representation of a sentence displayed in a MT.



Psycholinguistic studies that have a strong interface with education have recently received special attention (Maia, 2018). Enkin and Forster (2014), for example, proposed and investigated the use of the MT as a pedagogical tool in second language (L2) learning. The authors maintained that they offered a singular interdisciplinary approach to studying Second Language Acquisition (SLA) by bridging the gap between psycholinguistic experimentation and language learning. The rise of online learning during the COVID-19 pandemic has highlighted the importance of online educational tools for our society and the MT has a great potential to help learners developing their linguistic skills through online practices. The present study is the first part of a larger project that explores the pedagogical potential of this psycholinguistic technique. We created an online version of the MT using the PsyToolkit<sup>6</sup> software platform to determine whether practicing an L2-specific argument structure construction with this method could generate learning effects after only two training sessions. In the next section, we detail the study of Enkin and Forster (op. cit.).

### 1. The maze task as a pedagogical tool for second language learning

Enkin and Forster (2014) suggested that the MT is a promising technique for language learning since it enforces production processes as well as comprehension processes. They

---

<sup>6</sup> [www.psytoolkit.org](http://www.psytoolkit.org)

contended that MT training could reinforce classroom instruction, as learners must correctly form sentences as quickly as possible and are informed when they make a mistake. Additionally, the authors claimed that learners could develop both their implicit and explicit knowledge of the L2 using the MT as a pedagogical tool. The implicit-explicit learning dichotomy has been widely discussed in SLA and other fields (Reber, 1976; Krashen, 1982; Hayes; Broadbent, 1988; Dekeyser, 2003; Ellis, 2005; Rohrmeier; Rebuschat, 2012; Ellis, 2015). Implicit learning is the intuitive, subconscious, unselective process by which learners absorb, without conscious operations, the linguistic distribution of the input. Explicit learning, on the other hand, is the conscious, strategic and selective process by which learners intentionally absorb information.

The idea that the MT can result in changes in learners' linguistic knowledge is in line with recent evidence that suggests that different linguistic experiences – naturalistic or controlled – influence language processing and representation (Squires, 2014; Fraundorf; Jager, 2016; Hopp, 2016, 2020; Oliveira et al., 2017; Arnold et al., 2018). By way of illustration, Wells et al. (2009) analyzed how linguistic experience in a laboratory setting could influence how native speakers of English processed relative clauses. The authors exposed half of the participants to a reading task in which 2/3 of the sentences consisted of subject and object relative clauses. The other half read sentences with different types of structures. After reading each sentence, participants were exposed to two statements and were instructed to select the one that was true according to the information in the sentence. Participants performed this reading task once after the pretest and four other times in two different sessions. The results from the self-paced reading pretests and posttests indicated that participants who were exposed to the relative clauses exhibited shorter reaction times (RTs) to this structure than those in the control group. The authors concluded that these findings support the idea of statistical learning in sentence comprehension processes.

To test the potential utility of the MT as a pedagogical tool that gives learners linguistic experience capable of enhancing their L2 implicit and explicit knowledge, Enkin and Forster (2014) analyzed the performance of English-Spanish bilinguals with low levels of proficiency after they underwent training sessions with the method. Approximately half of the participants practiced L2-specific structures, and the other half practiced L1-similar structures. The authors hypothesized that the former group would show more learning benefits because previous evidence has indicated that training on syntactically complex sentences can benefit performance on syntactically less complex sentences, but not the reverse. To confirm this potential learning effect, after the training sessions, the authors conducted a MT test to examine whether or not learners exhibited

shorter RTs to both types of structures, which was considered a measure of implicit knowledge. In addition, they conducted a grammaticality judgment task to analyze whether the training also had a carryover effect benefiting participants' explicit knowledge. Furthermore, they carried out a production-based paper-and-pencil pretest-posttest and administered an attitude questionnaire to analyze the participants' perceptions of the task.

The experiment was conducted in a language training program for beginners. Twenty-one native speakers of English who were undergraduate students enrolled in a second semester beginner-level Spanish class participated in the experiment. To take this course, the students had to pass a proficiency test, which, according to the authors, ensured that the participants had comparable skill levels in the L2. The target items comprised three sentence structures – object relative clauses, direct object pronouns, and the copulative verbs *ser* and *estar*. According to the authors, object relative clauses and direct object pronouns can both be structurally L1-similar or L2-specific. For the copulative verb, *ser* was used in the L1-similar versions of the test, and *estar* was used in the L2-specific versions. The authors argued that the sentence types used were level-appropriate and familiar to the participants.

Participants took a paper-and-pencil pretest and, one week later, they started the MT training, which was carried out in three sessions, one per week. Participants practiced 20 sentences in each training session, 15 sentences containing the target structures (either L1-similar or L2-specific, depending on the training group) and 5 grammatical fillers that were the same for both groups. During these sessions, the participants were asked to try the sentences again if they made a mistake, and the location of the mistake was immediately pointed out. In the fourth week, participants took a paper-and-pencil posttest and the maze task test, which had a total of 32 new sentences, 28 of which were the target items. In the fifth week, they performed the grammaticality judgment task, which had 24 experimental items balanced in terms of grammaticality: 12 grammatical and 12 ungrammatical sentences. The target items were three grammatical L1-similar sentences and three grammatical L2-specific sentences. All the remaining sentences were experimental fillers. Finally, at the conclusion of the study, the participants completed a questionnaire in class for feedback on the acceptability and usefulness of the task.

The results suggested that the training produced a learning effect in all the postintervention tests. In the MT test, participants who trained on the L2-specific sentences exhibited significantly shorter RTs for the L2-specific sentences than participants who trained on L1-similar sentences, but the two groups did not differ significantly from each other in terms of

their RTs for the L1-similar sentences. Regarding the grammaticality judgment task, the results showed that participants who trained on L2-specific sentences exhibited significantly fewer errors on both L2-specific sentences and L1-similar sentences. Finally, the results from the production-based paper-and-pencil pretest-posttest also indicated a performance improvement, and the results from the attitude questionnaire indicated that the participants were receptive to the task. The authors interpreted the results as evidence that this type of linguistic training could assist learners in developing procedural representations and could also help them develop explicit knowledge.

As contended by DeKeyser (2003), there are no perfect tests for distinguishing the results of implicit and explicit learning; hence, tests should elicit knowledge under conditions that are more likely to make speakers retrieve implicit or explicit knowledge, and researchers should then infer to what extent the learning itself may have been implicit or explicit. Based on Enkin and Forster (2014), we assumed in our study that the MT was more likely to make speakers retrieve implicit knowledge than lectures on L2 structures, which were more likely to make speakers retrieve explicit knowledge. In addition, consistent with the same authors, implicit knowledge was measured by RTs in a MT test, and explicit knowledge was measured by an acceptability judgment task (AJT). In the next section, we present the linguistic structure that was used to investigate student's learning process in the present study.

## **2. The Double Object Construction**

To test the learning effect of the MT as a pedagogical tool, we observed the behavior of learners of English whose first language was Brazilian Portuguese (BP) with respect to an argument structure construction. Argument structure refers to the property of predicators to subcategorize one or more phrases to fully realize their semantic properties (Salamoura; Williams, 2007). We selected an argument structure construction that is licensed in the L2 but not in the L1. The participants in our study had lower levels of proficiency, and L2-specific constructions tend to be learned only at higher levels of proficiency. Recent evidence, for example, shows that the double object construction (DOC) (Zara et al., 2013), the resultative construction (Oliveira, 2014; 2016) and the induced movement construction (Souza, 2012), which are argument structure constructions licensed in English but not in BP, are only acquired at higher levels of proficiency by BP-English bilinguals. In our study, we investigate the effects of MT training on the learning of the

DOC by Brazilian learners of English with lower levels of proficiency.

The DOC is characterized by the structure NP<sub>1</sub>-VP-NP<sub>2</sub>-NP<sub>3</sub>. The VP denotes a transfer of a theme argument NP<sub>2</sub> (book/story) from an agent argument NP<sub>1</sub> (John) to a recipient-like argument NP<sub>3</sub> (Mary). The VP in the DOC typically denotes a physical or possessive transfer, as in (1), but it may also indicate a mental transfer, as in (2). Examples of verbs that form DOCs are ‘give’, ‘hand’, ‘lend’, ‘offer’, ‘promise’, ‘return’, ‘sell’, ‘show’ and ‘tell’.

(1) John gave Mary the book

(2) John told Mary the story

This construction seems to be the most productive structure for conveying the idea of transfer in English, but in most cases, it can also be realized by a different syntactic structure: the prepositional dative construction (PDC), illustrated in (3) and (4). It is important to highlight that both constructions are still productive, as attested by new verbs, such as “text”, that appear in both of them (Yang; Montrul, 2016). The apparent affinity among these two syntactic forms is termed by many authors in the linguistics literature the dative alternation or shift (Pinker, 1989; Hovav; Levin, 2008). The two constructions seem to be strongly related, and the conditions underlying speakers’ preferences are still disputed (Pinker, 1989; Krifka, 2003; Hovav; Levin, 2008; Agirre, 2015). Researchers who claim that the two constructions generate different meanings, in general terms, argue that the DOC is usually used to refer to a change in possession (‘cause to have’), whereas the PDC is usually used to convey the idea of movement to a goal (‘cause to go to’). Other studies indicate that other factors seem to play a role in speakers’ preference for the DOC or the PDC, such as information structure, linguistic typology (especially latinate verbs), referential prominence and argument length (Levin, 1993; Malchukov et al., 2010).

(3) John gave the book to Mary

(4) John told the story to Mary

BP and English contrast sharply with regard to the productivity of the DOC. In English, not only is the DOC more frequent than the PDC, but it is also acquired first by native speakers (Campbell; Tomasello, 2001). Although there are authors who claim that the DOC is an emergent structure in certain dialects of BP (Scher, 1996; Gomes, 2003), recent findings show that it is

considerably less frequent than the PDC. Zara (2014) investigated the behavior of BP-English bilinguals with respect to the dative shift in BP by conducting a corpora study. The author analyzed the Humanas corpus, which is an online BP corpus with 1,786,289 words from a variety of texts related to the human sciences collected from websites, newspapers, journals and magazines. Moreover, the author analyzed the C-ORAL-Brazil corpus, which is a spontaneous speech corpus of BP collected in one of the regions said to have the DOC as an emergent construction, the metropolitan region of Belo Horizonte, Brazil. The corpus has 208,130 words from 139 texts that represent a variety of communicative situations in both public and private settings. The results indicated that the PDC (5) was the preferred dative construction in both the written (58%) and spoken (46%) corpus. The DOC (6) was the fourth preferred construction in both the written (7%) and spoken corpus (3%), behind the pronoun ditransitive (7) and the inverted prepositional ditransitive (8). Therefore, the dative shift has a different status in BP than it does in English, since in the former, the PDC is by far the preferred structure and the DOC has only a marginal status.

(5) Maria deu um livro para João.

Mary give.PST DET book PREP John.

'Mary gave a book to John.'

(6) Maria deu João um livro.

Mary give.PST John DET book.

'Mary gave John a book.'

(7) Maria lhe deu um livro.

Mary [to him] give.PST DET book.

'Mary gave him a book.'

(8) Maria deu para João um livro.

Mary give.PST PREP John DET book.

'Mary gave a book to John.'

Zara (op. cit.) also presented findings that shed light on BP-English bilinguals' acquisition of the DOC in English. The author compared two subcorpora of native speakers of English, Written



(written) and Spoken (oral), which are part of a larger British corpus called the International Corpus of English (ICE-GB), to two distinct subcorpora of BP-English bilinguals, Br-ICLE and LINDSEI-BR. Whereas Written is composed of 423,702 words and Spoken of 637,562 words, Br-ICLE is composed of 159,364 words and LINDSEI-BR of 30,952 words. The results indicated that participants with high and intermediate levels of proficiency were sensitive to the appropriate contexts for the use of the DOC. Zara et al. (2013) conducted an AJT to examine how BP-English bilinguals perceived the DOC in contrast to other dative sentences, including the PDC. The authors found that only bilinguals with higher levels of proficiency exhibited acceptability ratings that indicated that they had acquired the DOC. Together, these results suggest that our target population, BP-English bilinguals with lower levels of proficiency, is unlikely to have acquired the DOC in English, indicating the appropriateness of using this structure to evaluate the learning effects of the MT training.

In the next section, we present the methodology of the study we conducted to investigate the learning effects of the MT training on the acquisition of the DOC.

### **3. Materials and Methods**

Our experiment sought to examine whether BP-English bilinguals with lower levels of proficiency could improve their knowledge of a construction specific to the English language through training sessions involving the MT in a laboratory setting. Based on Enkin and Forster (2014), our hypothesis was that this training would have a learning effect on bilinguals' implicit and explicit knowledge. The repetition of this training could move them forward on the path towards the entrenchment of the construction.

To confirm this potential learning effect, we compared participants' RTs and acceptability ratings for the DOC construction in a pretest and a posttest. Our experimental setup was very similar to that used by Enkin and Forster (*op. cit.*), as we took the ratings in an AJT as a measure of explicit knowledge and the RTs in a MT (a different task than the one used in the training session) as a measure of implicit knowledge. We hypothesized that the participants who were exposed to the target construction during the training session would exhibit shorter RTs and higher acceptability ratings on the posttest than on the pretest. To better understand the MT learning effects, we included a group of participants who performed the task without the target construction. Finally, we also included a group of participants who, instead of performing the

---

training session with the MT, were exposed to a lecture about English structures, which included the target construction. By comparing the results of participants who were exposed to the lecture with those of participants who were exposed to the MT training session, we expected to shed light on the differences and similarities between implicit and explicit learning.

Notably, we used the term *acceptability* instead of *grammaticality* for the judgment task. Grammaticality is an inherent feature of grammar, i.e., it is a theoretical construct in linguistic theory, whereas *acceptability* refers to the speaker's perception of a sentence's well-formedness (Bard et al., 1996). Thus, as we intended to obtain granular information on participants' perception of the well-formedness of the DOC construction and not simply their error rate, the term *acceptability* seemed more appropriate.

The predictor variables for this experiment were the (i) intervention type: (a) MT training with the DOC (MT w/ DOC), (b) MT training without the DOC (MT w/ no DOC) and (c) lecture on English structures including the DOC (lecture); and (ii) participants' level of proficiency: (a) L-0, (b) L-1 and (c) L-2. The outcome variables were (i) the RTs for the DOC in the MT tests (pretest and posttest), (ii) the acceptability ratings given to the DOC in the acceptability tests (pretest and posttest), (iii) the improvement rate in the MT and (iv) the improvement rate in the AJT. These variables will be further detailed later in this section.

To conduct the experiments with larger groups in a computer lab, we utilized online versions of both the AJT and the MT. Whereas the former was conducted with an online survey form, the latter had to be designed by adapting a script from a different method, since we were not familiar with any online software that could run the MT. In the next subsection, we detail how we programmed the MTs that were used in the pretest, posttest and training sessions in this study.

### **3.1. An online version of the maze task**

To implement the online version of the MT, we used the PsyToolkit<sup>4</sup> software platform, which is a free web-based service designed for setting up, running, and analyzing online questionnaires and RT experiments (Stoet, 2016). The PsyToolkit website provides an online environment to write and run scripts. Furthermore, there are many tutorials and a library with a number of ready-to-use demo experiments. We used the Simon task script<sup>5</sup> to design our MT.

PsyToolkit uses a particular script language, which is a list of command codes with a specific

---

syntax. There are three main sections in the PsyToolkit script: tasks, tables and blocks. As reported in Stoet (2010), a task describes the sequence of exactly one trial of an experiment, a table describes the experimental conditions, and the block sets up the way that trials are selected and combined. We modeled each sentence of the experiment as a table with each chunk pair in one row, as shown in (9). The first and second columns are the chunk pairs shown at each time. The correct option was set up in the first column. Although we fixed the first column as the correct option in the table, the chunks were displayed randomly on the left or on the right when we implemented the task section.

(9) Table script example:

1. table sentence1
2. "I lend" "x-x- x-x"
3. "Bob pencils" "blue sky"
4. "every Saturday" "East India"
5. "in class" "tall trees"

The task section performs only one row of the table per trial. Each time, one of the table rows is chosen to be used in the current task. Each table column is identified by the symbol @ plus the number. Then, the symbol @1 refers to the first column of a selected row in the table. The script is illustrated in (10):

(10) Task script example:

1. #Everything following hash mark is a comment.
- 2.
3. task sample1 #task section name
4. table sentence1 #table/sentence required
5. keys a l #define a = 1(left), l = 2(right)
6. set \$x random from 1 2 #set \$x with 1 or 2, randomly
7. if \$x == 1 #check if \$x is 1
8. set &left @1 #set &left with column @1 value

---

9.	set &right @2	#set &right with column @2 value
10.	set &correct 1	#key a (left) should be pressed
11.	fi	#end of the conditional if block
12.	if \$x == 2	#check if \$x is 2
13.	set &left @2	#set &left with column @2 value
14.	set &right @1	#set &right with column @1 value
15.	set &correct 2	#key l(right) should be pressed
16.	fi	#end of the conditional if block
17.	draw off	
18.	show text &left -250 25 255 255 255	#display the text at left
19.	show text &right 250 25 255 255 255	#display the text at right
20.	draw on	
21.	readkey &correct 7000	#wait for keyboard response
22.	save TASKNAME TABLEROW &left &right &correct RT STATUS	
23.	if STATUS == CORRECT	#test the STATUS
24.	if TABLEROW == 4	#if it is the last chunk
25.	show text "CORRECT!" 0 0 35 142 35	
26.	delay 1500	
27.	clear 1	#clear last text
28.	fi	
29.	fi	
30.	if STATUS == WRONG	
31.	show text "INCORRECT!" 0 0 255 0 0	
32.	delay 1500	
33.	clear 1	#clear last text
34.	end tasklist	#finish the current task
35.	fi	
36.	if STATUS == TIMEOUT	
37.	end tasklist	#finish the current task
38.	fi	

The participants used the 'A' and 'L' keys to make their choices, which corresponded to the

---

---

numbers 1 and 2, respectively (10, line 5). For each row selected, the experiment randomly sets the chunk position on the screen (one on the right and the other on the left) and sets the correct key answer for that pair (lines 6-20). The "readkey" command (line 21) waits for a key, which represents the user's choice, and determines whether the pressed key is equal to the variable "&correct". The STATUS variable saves the comparison result. That value is used to determine the next action. If the STATUS is CORRECT (1), the task selects the next chunk pair. If the incorrect chunk is selected, the message "Incorrect!" is shown, and the experiment continues to the next sentence. If the participant completes a sentence, the message "Correct!" appears, and the experiment continues with the next sentence. If the STATUS is WRONG (2) or TIMEOUT (3), the current task ends, and the experiment continues with the next sentence (lines 23-38).

Some of the experimental data were stored in a file on the website for later analysis (line 22): the task name (TASKNAME), the current table row (TABLEROW), the response time (RT) and the result of the user's choice (STATUS). In APPENDIX 1, there are three examples of recorded output data.

In the block section, the programmer can specify how the table lines are selected. The default configuration is random, but there are alternatives (e.g., fixed order, repeat on trial). For this experiment, the table rows were sequentially selected from 1 to 4 without repetition. The complete script is available at <https://github.com/elizabethduane/mazetask>.

Once the experiment was set up, we embedded it in an online questionnaire (Stoet, 2016). This questionnaire was sent to the participants, who could complete each one in a computer browser. All results were saved on the PsyToolkit web server and could be downloaded later. An example of an experiment and questionnaire is available at [https://github.com/elizabethduane/mazetask\\_example](https://github.com/elizabethduane/mazetask_example).

### **3.2. Participants**

The experiment participants were 182 Brazilian vocational high school students from CEFET-MG (Federal Center for Technological Education of Minas Gerais) Contagem Campus. They were students from the 1<sup>st</sup> and the 3<sup>rd</sup> (final) year of the three different courses offered by the aforementioned campus: electro-electronics, environmental control and informatics. These participants constitute a so-called convenience sample since their English teacher was the first author of the present paper. Their ages ranged from 14 to 19 years. They took part in this study

---

voluntarily and filled out the Terms of Informed Consent form approved by CEP/CONEP (process 11637518.5.0000.8507). The English language course is part of the school curriculum and is taken by all students. The participants were residents of the Belo Horizonte metropolitan area, where BP is the predominant language used in social interaction. Participants were tested in groups of about fifteen people (all from the same class) and each of these groups were assigned to a different intervention (MT w/ DOC, MT w/ no DOC and lecture) in such a way that all the intervention groups included students from all courses and all course years we tested.

The students had different levels of low proficiency, which we measured via the Vocabulary Levels Test (VLT) (Nation, 1990). We selected this test because, in addition to being quick and practical, it has also proven to be efficient for classifying bilinguals according to their proficiency level (Souza; Silva, 2015). The VLT is a word-meaning association test. The assumption of the test is that proficiency can be measured in terms of lexical access competence. The test is divided into five levels that reflect access to lexical items whose frequency decreases progressively in English corpora. The test assumes that access to words with progressively lower frequency reflects the size of the L2 mental lexicon. To enhance the test's discriminatory power, the VLT was administered with a 10-minute time limit.

All the participants had levels of proficiency that varied among the three most basic levels on the VLT, which we named level 0 (L-0), level 1 (L-1) and level 2 (L-2). The VLT estimates that participants classified as L-0 have a lexicon that includes fewer than the 2,000 most frequent English lexemes. At L-1, this knowledge includes more than the 2,000 most frequent lexemes but less than the 3,000 most frequent lexemes. At L-2, participants' knowledge includes more than the 3,000 most frequent lexemes and less than the 5,000 most frequent lexemes. A total of 102 participants were classified as L-0, 23 as L-1 and 57 as L-2.

### **3.3 Materials**

We employed an AJT and a MT test representing measures of explicit and implicit knowledge, respectively, as suggested by Enkin and Forster (2014). The AJT had 64 experimental items, 8 of which were instances of the DOC. The distractor items were 24 grammatical sentences unrelated to the target structure and 32 ungrammatical sentences. The target sentences are shown below:

---

- 
- (11) We give Ann flowers for her birthday every year.
  - (12) I sell Mark books about food very often.
  - (13) I tell John stories about monsters at school.
  - (14) They bring Michael presents from Brazil every year.
  - (15) I teach Mary math every night at home.
  - (16) They send Lucy e-mails about work at night.
  - (17) We lend Joan money every year on Christmas.
  - (18) We write Judy letters every year in July.

The MT test had 32 experimental items. Eight sentences were instances of the DOC, and the remaining sentences were grammatical distractor sentences unrelated to the target structure. The sentences are presented below, and the slashes in the sentences illustrate how they were divided into chunks.

- (19) I lend/ Bob pencils/ every Saturday/ in class.
- (20) You give/Jack hugs/ every time/ he's sad.
- (21) I sell/ Jane rings/ every month/ on payday.
- (22) I bring/ Lana breakfast/ every morning/ in bed.
- (23) You get/ Matt books/ every week/ since kindergarten.
- (24) We send/ Dave gifts/ every year/ after Christmas.
- (25) They teach/ Kim music/ every Monday/ at school.
- (26) You write/ Jack poems/ every weekend/ for fun.

The MT employed as a training tool for practicing the structure also consisted of 32 sentences. One group of participants had the task with 8 sentences containing the target structure and 24 grammatical fillers. Only 1/4 of the sentences had the target structure to make learning more implicit. The second group had the task with 32 grammatical sentences unrelated to the target structure. The third group was introduced to the structure through a 15-minute lecture on English structures (the DOC, questions with the verb to be, third person present tense, and the adjective-noun structure). We avoided using low-frequency words in all the experimental items, and we pseudorandomized the items so that they were always exhibited in a different order.

---

### 3.4. Procedures

The experiment was conducted in groups in computer labs at the CEFET-MG Contagem Campus with Dell OptiPlex 790MT Intel Core I3 PCs with 1TB HD, 8 GB RAM, 18,5" monitors and the Windows 7 operating system. Each experimental session involved 20 participants, and the tasks were performed individually. The entire experimental session lasted approximately one hour and thirty minutes. The AJT, the MT test, and the MT training took place online. In each task, participants were introduced to a set of instructions, which were read by one of the experimenters and displayed on the desktops. The instructions were all in Portuguese so that they could be clearly understood. All the MTs included a break halfway through the task, and short breaks could also be taken during the AJT.

The pretests consisted of an AJT and a MT test, performed in that order. The AJT was conducted with Google Forms. In the task, the participants were instructed to assess the acceptability of each sentence with a 5-point Likert scale, assigning one of the following values to each sentence: 1 (totally unacceptable), 2 (unacceptable), 3 (neutral), 4 (acceptable), and 5 (perfectly acceptable). This scale has been argued to be the most suitable for this type of task (Souza; Oliveira, 2014). In the MT test, the participants were informed that they should form sentences by selecting, from each pair of words, the ones that best suited the sentences being formed. They had to use the "A" (left) and "L" (right) keys to make their choice. After the presentation of the instructions, the participants performed a practice session with 8 sentences, which were not used in any other part of the study. After that, participants could ask the experimenter questions. The sentences were presented continuously and in a random order. Each pair of words was visible to the participants for up to four seconds, and if a choice was not made by then or if they picked the wrong option, the next sentence was automatically displayed. We measured the RTs for the second segment of the DOC sentences, since this is the part of the sentence that participants are likely to be unfamiliar with due to its ungrammaticality in the L1.

We administered the first MT training session immediately after the pretest. Since the activity was the same as the previous activity, the instructions were not repeated. Sixty-five participants again performed the MT training containing the target structure but with different sentences; 57 also performed the MT training but containing only grammatical fillers unrelated to the target structure; and 60 participants watched the lecture on English structure, which included



the DOC. The lecture, however, was administered only after the VLT test. We administered the VLT in a paper-and-pencil format. Then, participants performed another MT training session. To determine the effect of the training, a posttest consisting of the MT test and the AJT was administered at the end of the procedure in that order.

#### 4. Results

The main hypothesis that motivated our experiment was that MT training would generate an effect on learners' implicit and explicit knowledge of the English DOC. Our hypothesis was based on the findings from Enkin and Forster (2014), described in the second section. The existence of this effect would be indicated by improvement on a MT test and an AJT. More specifically, we expected participants to show shorter RTs and higher acceptability ratings on the posttest than on the pretest. Since this improvement in RTs could also result from familiarity with the task, we compared the results of participants who practiced the DOC in the training session with those of participants who performed the training without the DOC. To illuminate possible differences between implicit and explicit learning, we also included in the analysis a group that, instead of undergoing these training sessions, watched a short lecture on English structures. Finally, we divided each group into three proficiency levels to observe whether proficiency played a major role in the capacity to learn through these interventions.

To verify our hypothesis, we compared participants' improvement rate on both the MT test and the AJT in relation to the DOC sentences. The fixed value of  $\alpha$  to reject the null hypothesis was .05. All trials where an error occurred or in which participants violated the four-second time limit for the target segment were excluded. RTs were trimmed to discard observations under 500 ms. We also excluded from our analysis participants who did not make it to the target part of the eight DOC sentences at least four times. Participants who were not included in our analysis probably had difficulty performing the MT with English structures not licensed in the L1. In total, we excluded 23 participants, all of whom were classified as L-0. The fact that all the excluded participants were at the lowest VLT level corroborates that they had difficulty with the task due to proficiency issues. Thus, we analyzed 159 participants: 79 L-0, 23 L-1 and 57 L-2 students. Regarding the interventions, 57 participants were in the MT w/ DOC group, 51 were in the MT w/ no DOC group and 51 were in the lecture group.

Figure 2 illustrates the mean and standard deviation RTs to the double object of the DOC

---

during the MT test for all participants. Figure 3 illustrates these results according to participants' proficiency level. Figure 4 illustrates the acceptability rating means and standard deviations for the DOC sentences for all participants. Figure 5 illustrates these results according to the participants' proficiency level.

Figure 2 – Mean  $\pm$  SD maze task RTs before and after the interventions (n = 159).

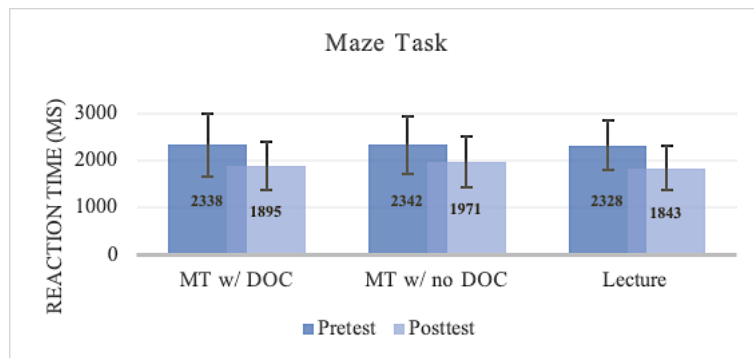


Figure 3 – Mean  $\pm$  SD maze task RTs before and after the interventions by proficiency level (n = 159).

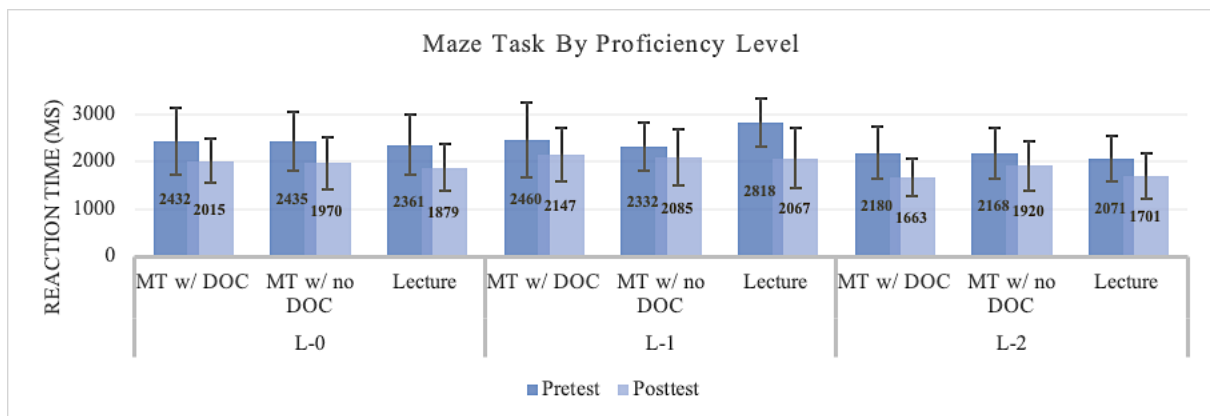


Figure 4 – Mean  $\pm$  SD acceptability ratings before and after the interventions (n = 159).

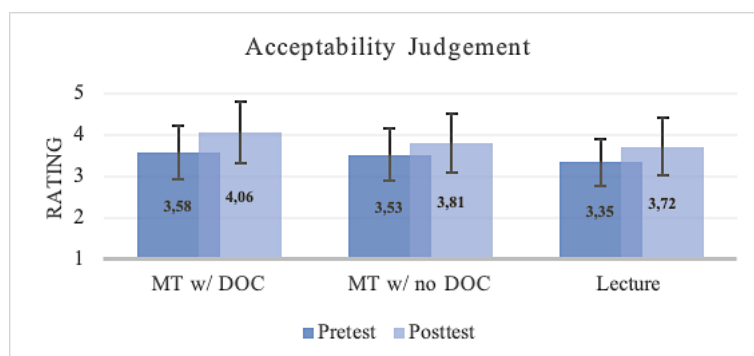
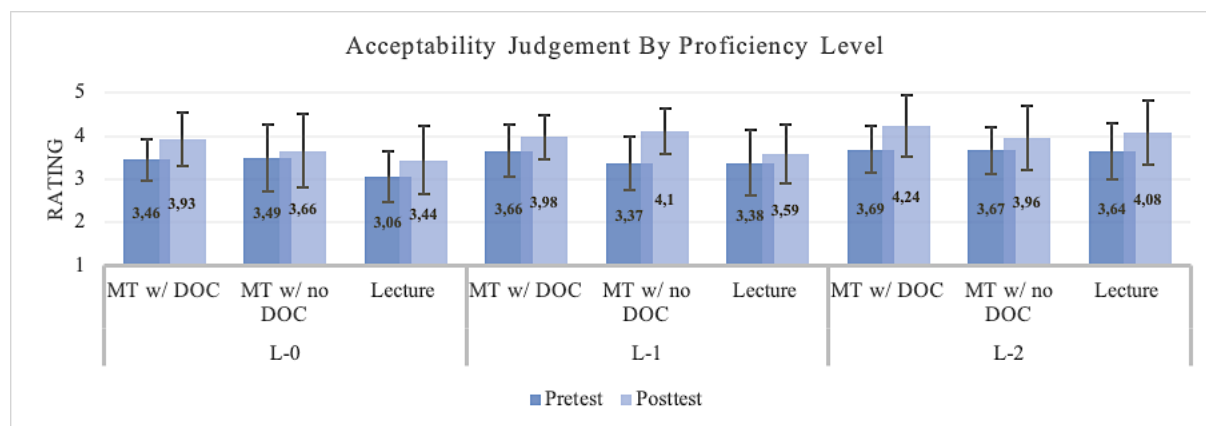


Figure 5 – Mean  $\pm$  SD acceptability ratings before and after the interventions by proficiency level (n = 159)

Both the RTs and the acceptability ratings were tested for normality with the Kolmogorov-Smirnov test. Because the data had a normal distribution, within-group analysis was conducted with an appropriate paired T-test that compared the results from the pretests and posttests. Table 1 shows the mean and standard deviation of the RTs to the target structures before and after each intervention. The p-value is shown for each test.

The MT training with the DOC resulted in a significant difference between the pretest and posttest both in RTs ( $p < .001$ ) and in acceptability ratings ( $p < .001$ ). Similarly, the MT training without the DOC yielded a significant difference between the pretest and posttest in both the MT ( $p < .001$ ) and in the AJT ( $p < .04$ ). Finally, the lecture also resulted in significant differences in RTs ( $p < .001$ ) and in acceptability ratings ( $p < .001$ ). Therefore, all the groups exhibited improvement in the form of significantly shorter RTs and higher acceptability ratings after the intervention.

Table 1 – Mean  $\pm$  SD maze task RTs and acceptability ratings before and after the interventions (n = 159).

	Maze Task			Acceptability		
	RTs			Ratings		
	RT-Pre	RT-Post	p-value	AJ-Pre	AJ-Post	p-value
<b>Maze training with the DOC (n = 57)</b>	2337.89 $\pm$ 663.36	1895.47 $\pm$ 479.03	<.001*	3.58 $\pm$ 0.52	4.06 $\pm$ 0.65	<.001*
<b>Maze training without the DOC (n = 51)</b>	2342.47 $\pm$ 585.04	1970.63 $\pm$ 546.44	<.001*	3.53 $\pm$ 0.69	3.81 $\pm$ 0.79	.036*
<b>Lecture (n = 51)</b>	2327.92 $\pm$ 602.87	1842.51 $\pm$ 520.54	<.001*	3.35 $\pm$ 0.68	3.72 $\pm$ 0.79	<.001* <sup>†</sup>

\*Statistically significant difference ( $p < 0.05$ ).

Between-group analysis was conducted with analysis of variance (ANOVA) to compare the

improvement rate observed in each intervention group. The improvement data were obtained by calculating the difference between the posttest and the pretest for all the participants for both the AJT and the MT test. The comparison of the three interventions in terms of improvement rate did not yield a significant difference for the MT ( $p=.583$ ) or the AJT ( $p=.366$ ). The results are presented in table 2.

Table 2 – Mean  $\pm$  SD of the improvement rate (IR) in the maze task RTs and acceptability ratings ( $n = 159$ ).

	Maze training with the DOC (n = 57)	Maze training without the DOC (n = 51)	Lecture (n = 51)	p-value
Maze Task IR	-442.42 $\pm$ 624.12	-371.84 $\pm$ 529.42	-485.41 $\pm$ 502.52	.583
Acceptability Judgment IR	0.48 $\pm$ 0.54	0.28 $\pm$ 0.93	0.38 $\pm$ 0.64	.366

Finally, we analyzed the interaction between intervention and proficiency level. As illustrated in table 3, the interaction did not reach significance for either the MT ( $p=.287$ ) or the AJT ( $p=.408$ ). Thus, all the groups performed similarly regardless of proficiency level or the intervention. These results are discussed in the next section.

Table 3 – Mean  $\pm$  SD of the improvement ratio (IR) in the maze task RTs and acceptability ratings by proficiency level ( $n = 159$ ).

	Maze training with the DOC (n = 57)			Maze training without the DOC (n = 51)			Lecture (n = 51)			p-value
	0	1	2	0	1	2	0	1	2	
Proficiency	n=28	n=7	n=22	n=29	n=7	n=15	n=22	n=9	n=20	
Maze Task RTs	-416.6 $\pm$ 722	-312.7 $\pm$ 357	-516.5 $\pm$ 565	-465.9 $\pm$ 585	-246.8 $\pm$ 477	-248.3 $\pm$ 418	-482.1 $\pm$ 607	-750.1 $\pm$ 385	-369.8 $\pm$ 383	.287
Acceptability Ratings	0.46 $\pm$ 0.60	0.32 $\pm$ 0.52	0.54 $\pm$ 0.49	0.17 $\pm$ 1.07	0.73 $\pm$ 0.52	0.28 $\pm$ 0.74	0.38 $\pm$ 0.79	0.21 $\pm$ 0.37	0.44 $\pm$ 0.56	.408

Statistical significance was set at  $\alpha < 0.05$ .

## 5. Discussion

Our findings indicated that all the participants in the three proficiency groups and in the three intervention groups exhibited the same behavior towards the DOC in both the MT and in the AJT. Their RTs to the double objects in the MT were significantly shorter on the posttest than on the pretest, and their acceptability ratings of the DOC sentences were higher on the pretest than on the posttest. We expected this result due to participants becoming more familiar with the task.

However, we also expected participants who practiced the DOC in the maze training and those who watched a short lecture on English structures to perform better than those in the control group (maze training without the DOC). The results showed that the learning effect was the same across proficiency levels and intervention types. In sum, our results suggest that the improvement we observed occurred regardless of the MT training with the target structure.

There are several possible reasons why our training did not result in a better improvement rate than that in the control group, which was not exposed to the target structure during the intervention. First, the pretests, the posttests and the different interventions all occurred on the same day in a laboratory context. Moreover, participants were exposed to eight DOC sentences in each of the two training sessions, in the pretest and in the posttest; thus, they were exposed to DOC sentences 24 times before the MT posttest and 32 times before the acceptability judgment posttest. Participants in Enkin and Forster (2014) practiced the target structure in 3 different sessions (one training per week), and RTs were measured in a fourth test during the fourth week; the AJT was performed in the fifth week. In each session, participants were exposed to 15 target structures; thus, they were exposed to the target structure 45 times before the final MT and 60 times before the AJT. Furthermore, participants in Enkin and Forster (op. cit.) were exposed to only 5 distractor sentences in each session, whereas our participants were exposed to 24 distractors in each MT session. These factors may have hindered the learning effects on our target group; consequently, these participants performed similarly to participants who performed the maze training without the DOC sentences. In sum, the smaller number of trainings, smaller number of exposures to the target structure and larger number of distractor items may have mitigated the possible learning effects of the MT.

Another possible interpretation is that the DOC is difficult to acquire for learners with low levels of proficiency whose first language does not license this construction, which is the case for BP. We presented a set of studies in the Double Object Construction section showing that argument structure constructions that are licensed in English but not in BP, including the DOC, tend to be fully learned only at higher levels of proficiency. In Enkin and Forster (2014), participants also practiced structures that were not licensed in their L1, but the sentences did not instantiate argument structure constructions unlicensed in the L1. Instead, they were sentences with nonovert pronouns (direct object and object relative clauses) and sentences with the copulative verbs *ser* and *estar*. In addition, the author pointed out that participants were familiar with all these structures before they started the training sessions. Thus, it is possible that the DOC

---

imposes a level of difficulty that, unlike the structures tested by Enkin and Forster (op. cit.), hinders the learning effects of both the MT training and the lecture.

A different interpretation is that the MT training does not generate a learning effect. In this case, we would interpret the learning effects encountered by Enkin and Forster (2014) as being due to differences between the tested groups. Their study used 21 participants who were enrolled in a second semester beginner-level Spanish class and, according to the authors, the students had to pass a proficiency test to take the course, which ensured that they all had comparable skill in the L2. However, the authors analyzed participants' behavior only after the training; hence, there might have been unnoticed differences between the tested groups that could have been detected with a pretest before the training session.

An alternative view is related to the fact that our participants were exposed to the target structure during the pretest, whereas the participants in Enkin and Forster (op. cit.) never practiced half of the structures that were on the MT test. Thus, the MT pretest itself might have been enough to generate a learning effect on the MT posttest, and both of them might have generated an effect on the AJT posttest. If this is true, what our results show is that the learning effect of one MT session (pretest) was similar to the effect of three MT sessions (pretest + two training sessions) on the MT posttest, and the two MT sessions (pretest + posttest) had the same effect as four MT sessions (pretest + two training sessions + posttest) on the AJT. Considering that the sentences on the MT pretest and the MT posttest were the same, it is possible that the pretest had a greater effect on the posttest than the training sessions. According to this interpretation, the MT training did generate a learning effect on both RTs to and acceptability ratings of the DOC. However, we cannot distinguish this effect from the learning effects resulting from familiarity with the task.

Due to the novelty of the topic we investigated, we will not argue in favor of any of these interpretations. We would rather keep them as a starting point for future investigations because, as we mentioned previously, this study is the first part of a larger project focused on the use of the MT as a pedagogical tool. It is our understanding that comparing results obtained using different protocols of data collection is of paramount importance to the cumulative advancement of knowledge through scientific research. To the best of our knowledge, the present paper and the study by Enkin and Forster (2014) are the first to test the possible benefits of the MT for L2 syntax learning. The fact that they yielded different results highlights the need for more research on this topic. Though the findings appear to be in striking contrast, they may prove to be complementary

---

---

as this investigation moves forward. In the next section, we present our final remarks.

## 6. Conclusions and suggestions for future research

This study sought to shed new light on the possible learning effects of L2 syntactic training with the MT. Enkin and Forster (2014) proposed that this psycholinguistic tool could benefit L2 learners' implicit and explicit knowledge. We tested this hypothesis with a study that, unlike the aforementioned one, included both pretests and posttests in all tasks and a comparison with the learning effects of a lecture on syntactic structures. This study also had a higher number of distractor items and a higher number of participants. Despite not finding strong evidence of learning effects, our study has contributed to the investigation of the utility of this psycholinguistic method for SLA pedagogy.

Our results suggest several important methodological strategies for future studies. First, both pretests and posttests should be administered to ensure that the differences found on the posttests did not exist before the intervention. Second, a longitudinal study in which results are gathered in different stages of the learning process would be useful to shed light on how much practice with the MT is necessary for learning to take place with respect to both implicit and explicit knowledge. Third, testing different linguistic structures may be useful in order to determine if there are different learning effects according to the structure type and according to the presence of the structure in the L1. Finally, both our study and Enkin and Forster (op. cit.) tested low-proficiency participants and used the MT task itself as a measure of implicit learning. Future studies should use a method other than the MT task to measure RTs so that the effects related to familiarity with this tool can be mitigated. Additionally, L2 learners with a higher level of proficiency may exhibit a different behavior, which could offer further insight into the learning effects of this method.

Another important outcome of this study was the creation of an online version of the MT task. The adaptation of the script from the PsyToolkit website allowed us to carry out the research in a computer lab and, consequently, recruit a considerable number of participants. The participants did not have any problems accessing or completing this online version of the task, and, moreover, we could easily access and explore the data. Conducting the experiment in a computer lab allowed us to test participants in similar contexts and administer the tasks following a similar procedure, which mitigated the possible effects of external factors.

---

It is plausible that a few limitations may have influenced the results. We did not use participants with a high level of proficiency, and we did not balance the number of participants across each level. Thus, we had many participants classified as L-0 or L-2, but few classified as L-1. The fact that, instead of conducting a longitudinal study, we carried out a long laboratory session is likely the foremost cause of the discrepancy between our results and those obtained by Enkin and Forster (2014). Finally, the use of the MT task in both the training and test sessions, as in Enkin and Forster (op. cit.), may have muddied the results, since it is difficult to distinguish the linguistic learning effects from the method learning effects. We hope that, notwithstanding these limitations, our study can contribute to a better understand of the MT task's potential as a pedagogical tool.

### Acknowledgements

The authors thank the CEFET-MG (Federal Center for Technological Education), CNPq (National Council for Scientific and Technological Development) and FAPEMIG (Minas Gerais State Research Support Foundation) for their financial support.

### References

- AGIRRE, A. The acquisition of dative alternation in English by Spanish learners. *Vigo International Journal of Applied Linguistics*, v. 12, p. 63-90, 2015.
- ARNOLD, J.; STRANGMANN, I.; HWANG, H.; ZERKLE, S.; NAPPA, R. Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, v. 102, p. 41–54, Oct. 2018.
- BARD, E.; ROBERTSON, D.; SORACE, A. Magnitude estimation of linguistic acceptability. *Language*, v. 72, n.1, p. 32-68, Mar. 1996.
- BARRETO, S.; MARCILESE, M.; OLIVEIRA, A. Idiomaticidade, familiaridade e informação prévia no processamento de expressões idiomáticas do PB. *Letras de Hoje*, Porto Alegre, v. 53, n. 1, p. 119-129, Mar. 2018.
- CAMPBELL, A.; TOMASELLO, M. The acquisition of English dative constructions. *Applied Psycholinguistics*, v. 22, n. 2, p. 253-267, 2001.
- DEKEYSER, R. Implicit and explicit learning. In.: DOUGHTY, C; LONG, M. H. (Eds.). *The handbook of second language acquisition*. Oxford, UK: Blackwell, 2003. p. 131-348.
- ELLIS, N. Implicit and explicit learning: Their dynamic interface and complexity. In.: REBUSCHT, P. (Ed.). *Implicit and explicit learning of languages*. Amsterdam: John Benjamins, 2015. p. 2-23.
- ELLIS, R. Measuring Implicit and Explicit Knowledge of a Second Language: A Psychometric Study. *Studies in Second Language Acquisition*, v. 27, n. 2, p. 141-172, Jun. 2005.



ENKIN, E.; FORSTER, K. The Maze Task: Examining the Training Effect of Using a Psycholinguistic Experimental Technique for Second Language Learning. *Journal of Linguistics and Language Teaching*, v. 5, p. 161-180, 2014.

FORSTER, K.; GUERRERA, C.; ELLIOT, L. The maze task: measuring forced incremental sentence processing time. *Behavior Research Methods*, v. 41, n. 1, p. 163-17, Feb. 2009.

FRAUNDORF, S. JAEGER, T. Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, v. 91, p. 28-58, 2016.

GOMES, C. Dative alternation in Brazilian Portuguese: typology and constraints. *Language Design Journal of Theoretical and Experimental Linguistics*, n. 5, p. 67-78, 2003.

HAYES, N.; BROADBENT, D. Two modes of learning for interactive tasks. *Cognition*, v. 28, n. 3, p. 249-76, Apr. 1988.

HILPERT, M.; SAAVEDRA, D. The unidirectionality of semantic changes in grammaticalization: An experimental approach to the asymmetric priming hypothesis. *English Language and Linguistics*, v. 22, n. 3, p. 357-380, Nov. 2018.

HOPP, H. Learning (not) to predict grammatical gender processing in second language acquisition. *Second Language Research*, v. 32, n.2, p. 277-307, Feb. 2016.

HOPP, H. Morphosyntactic adaptation in adult L2 processing: Exposure and the processing of case and tense violations. *Applied Psycholinguistics*, v. 41, n. 33, p. 627-656, May 2020.

HOVAV, M.; LEVIN, B. The English dative alternation: The case for verb sensitivity. *J. Linguistics*, v. 44, p. 129-167, Mar. 2008.

KIZACH, J.; NYVAD, A.; CHRISTENSEN, K. Structure before meaning: Sentence processing, plausibility, and subcategorization. *PLoS ONE*, v. 8, n. 10, Oct. 2013.

KRIFKA, M. Semantic and pragmatic conditions for the dative alternation. In *KASELL International Conference on English Language and Linguistics*, Seoul, 25-26 Jun. 2003 (p. 1-32), 2003.

KRASHEN, S. *Principles and Practice in Second Language Acquisition*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

LEVIN, B. *English verb classes and alternations*. A preliminary investigation. Chicago: University of Chicago Press, Sep. 1993.

LI, R.; ZHANG, Z.; NI, C. The impact of world knowledge on the processing of mandarin possessive reflexive *ziji*. *Journal of psycholinguistic research*, v. 46, n. 3, p. 597-615, 2017.

MAIA, M. *Psicolinguística e Educação*. Campinas, SP: Mercado de Letras, Jun. 2018.

MALCHUKOV, A.; HASPELMATH, M.; COMRIE, B. Ditransitive constructions: a typological overview. In: \_\_\_\_\_. *Studies in ditransitive constructions: A comparative handbook*. Berlin/New York: Walter de Gruyter. 2010. p. 1-64.

MANSBRIDGE, M.; TAMAOKA, K. The (in)sensitivity of plural -s by Japanese learners of English.

---

*Open Journal of Modern Linguistics*, v. 8, p. 176-198, 2018.

MARCILESE, M.; RODRIGUES, E.; AUGUSTO, M.; HENRIQUE, K. Efeitos de distância linear e marcação no processamento da concordância verbal variável no PB. *Revista de Estudos da Linguagem*, v. 25, n. 3, p. 1291-1325, Jun. 2017.

NATION, P. *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle, 1990.

O'BRYAN, E.; FOLLI, R.; HARLEY, H.; BEVER, T. Evidence for the Use of Verb Telicity in Sentence Comprehension. In.: FOLLI, R.; SEVDALI, C.; TRUSWELL, R. (Eds.). *Syntax and its Limits*. Oxford: Oxford University Press, 2013.

OLIVEIRA, C. Uma investigação sobre a aquisição das regras de formação da construção resultativa por bilíngues de alta proficiência do par linguístico português do Brasil e inglês. *Letrônica*, Porto Alegre, v. 7, n. 1, p. 66-85, 2014.

\_\_\_\_\_. *Processing, representation and learnability of the resultative construction by Brazilian Portuguese-English bilinguals*. 2016. 202f. Dissertation (Ph.D. in Linguistics) College of Letters, Federal University of Minas Gerais, Belo Horizonte, Brazil, 2016. Retrieved from <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/RMSA-AM2UMP>

\_\_\_\_\_. Métodos on-line em psicolinguística: a tarefa labirinto (maze-task). *Cadernos de Tradução*. (in press).

OLIVEIRA, C.; SOUZA, R.; OLIVEIRA, F. Bilingualism effects on L1 representation and processing of argument structure. *Journal of the European Second Language Association*, v.1, n. 1, p. 23-37, 2017.

NYVAD, A.; KIZACH, J.; CHRISTENSEN, K. (non-) arguments in long-distance extractions. *Journal of Psycholinguistic Research*, v. 44, n. 5, p. 519–531, 2015.

PINKER, S. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press, 1989.

QIAO, X.; SHEN, L.; FORSTER, K. Relative clause processing in mandarin: Evidence from the maze task. *Language and Cognitive Processes*, v. 27, n. 4, p. 611-630, May 2012.

REBER, A. Implicit learning of synthetic languages: the role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, v. 2, p. 88–94, 1976.

ROHRMEIER, M.; REBUSCHAT, P. Implicit Learning and Acquisition of Music. *Topics in Cognitive Science*, v. 4, n. 4, p. 525-553, Oct. 2012.

SALAMOURA, A.; WILLIAMS, J. Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics*, v. 28, n. 4, p. 485-501, Oct. 2007.

SCHER, A. P. *Construções com dois complementos no inglês e no português do Brasil: um estudo sintático comparativo*. 1996. 128f. Dissertation (Masters in Linguistics) - Institute of Studies of Language, UNICAMP, Campinas, Brazil, 1996. Retrieved from

---

<http://repositorio.unicamp.br/jspui/handle/REPOSIP/271031>

SIKOS, L.; GREENBERG, C.; DRENHAUS, H.; CROCKER, M. Information density of encodings: The role of syntactic variation in comprehension. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society (CogSci 2017), London, UK, 26-29 July 2017 (p. 3168-3173), 2017.

SOUZA, R. Two languages in one mind and the online processing of causatives with manner-of-motion verbs. *Revista Virtual de Estudos da Linguagem*, v. 10, n. 6, p. 220-239, 2012.

SOUZA, R.; OLIVEIRA, C. The learnability of the resultative construction in English L2: a comparative study of two forms of the acceptability judgment task. *Revista da Abralin*, v. 8, n. 2, p. 375-410, 2014.

\_\_\_\_\_. Are bilingualism effects on the L1 byproducts of implicit knowledge? Evidence from two experimental tasks. *Revista de Estudos da Linguagem*, v. 25, n. 3, p. 1685-1716, 2017.

SOUZA, R.; COELHO, S.; SANTOS, A.; NASCIMENTO, T. Processamento de sujeitos de orações com verbos inacusativos no português brasileiro. *Revista Fórum Linguístico*, v. 11, n. 3, p. 328-339, 2014.

SOUZA, R.; SILVA, J. Exploring the measurement of vocabulary size to differentiate Brazilian Portuguese-English bilinguals' access to grammatical knowledge in the L2. *Revista Linguística*, v. 11, p. 187-204, 2015.

SQUIRES, L. Social differences in the processing of grammatical variation. *University of Pennsylvania Working Papers in Linguistics*, v. 20, n. 2, p. 178-188, 2014.

STOET, G. PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, v. 44, n. 1, p. 24-31, Nov. 2016.

STOET, G. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, v. 42, p. 1096-1104, Nov. 2010.

SUZUKI, Y.; SUNADA, M. Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. *Bilingualism: Language and Cognition*, v. 21, n. 1, p. 32-46, Jan. 2018.

WANG, X. Language control in bilingual language comprehension: evidence from the maze task. *Frontiers in psychology*, v. 6, Sep. 2015.

WELLS, J.; CHRISTIANSEN, M.; RACE, D.; ACHESON, D.; MACDONALD, M. Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, v. 58, p. 250-271, 2009.

WITZEL, J.; WITZEL, N. Incremental sentence processing in Japanese: A maze investigation into scrambled and control sentences. *Journal of psycholinguistic research*, v. 45, n. 3, p. 475-505, Jun. 2016.

WITZEL, J.; WITZEL, N.; FORSTER, K. Comparisons of Online Reading Paradigms: Eye Tracking, Moving-Window, and Maze. *Journal of Psycholinguistic Research*, v. 41, n. 2, p. 105-128, Apr. 2012.

YANG, C.; MONTRUL, S. Learning datives: The Tolerance Principle in monolingual and bilingual acquisition. *Second Language Research*, v. 33, n. 1, p. 119-144, Oct. 2016.

ZARA, J. V. *To give you a book or to give a book to you: um estudo sobre a variação sintática na expressão linguística de eventos de transferência de posse na interlíngua português/inglês*. 2014. Thesis (PhD in Linguistic Studies) – Faculty of Languages, Literature, and Linguistics, Federal University of Minas Gerais, Belo Horizonte, Brazil, 2014. Retrieved from <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/MGSS-9HSMZ2>

ZARA, J.; OLIVEIRA, F.; SOUZA, R. Selective transfer in the acquisition of double object constructions by Brazilian learners. *Alfa: Revista de Linguística (São José Rio Preto)*, São Paulo, v. 57, n. 2, p. 519-544, 2013.

## Appendix 1

Table 1 – Example of data output when the sentence was correct (Status = 1).

Taskname	Tablerow	&left	&right	&correct	RT	Status
sample1	1	I lend	x-x- x-x	1	1421	1
sample1	2	blue sky	Bob pencils	2	1452	1
sample1	3	east India	every Saturday	2	923	1
sample1	4	tall trees	at class	2	989	1

Table 2. Example of data output with an incorrect choice for the third chunk pair (Status = 2).

Taskname	Tablerow	&left	&right	&correct	RT	Status
sample1	1	I lend	x-x- x-x	1	1421	1
sample1	2	blue sky	Bob pencils	2	1452	1
sample1	3	east India	every Saturday	2	923	2

Table 3. Example of data output data when there was a timeout (Status = 3).

Taskname	Tablerow	&left	&right	&correct	RT	Status
sample1	1	x-x- x-x	I lend	2	1284	1
sample1	2	Bob pencils	blue sky	1	7000	3