Fishing for Words: A game to learn grammatical gender in Portuguese
Pedro Bertrand Cabral
pedro.b.cabral@tecnico.ulisboa.pt
Instituto Superior Técnico, Lisboa, Portugal
June 2019

Abstract

Learning a language can be very useful and fun, but also a lot of work. Some languages include a peculiar feature that can provoke stumbles in learners whose native language lacks it: grammatical gender. Mastering gender requires the learner to memorize a (mostly) arbitrary distinction between masculine words, feminine words and sometimes more. Fortunately, modern times have brought us technologies such as videogames, an interactive medium focused around learning and training skills. If players are willing to spend hundreds of hours on a single game solely for the sake of entertainment, would they not be willing to spend them if, besides being fun, the game taught them a valuable skill? It was in the spirit of harvesting the potential of videogames as learning tools that this thesis was developed.

We developed a game for Android and Windows to help players memorize the gender of Portuguese words at the A2 level of proficiency, including two alternative word-selection systems (which balance the game’s difficulty in real-time according to the player’s performance). A full experimental procedure was developed and tested, with the purpose of validating its effectiveness, which revealed several pitfalls through which an experimenter evaluating such a game might fail to obtain results. In other words, we provide insight into how to get the participants to play the game and stick with it, which ours did not.

Keywords: Grammatical Gender, Language, Education, Games, Videogames, Learning

1. Introduction

In the face of ever-increasing globalisation and ease of communication and travel, knowing a foreign language is a valuable asset. Learning additional languages can lead to new work and business opportunities and open paths to other cultures, people, perspectives and ways of thought. Learning a language is also, on the other hand, a significant investment of time, money and effort. Fortunately, the last fifty years have introduced cheap and portable computers, the internet and videogames, technologies which can be used as novel tools to learn in a more fun, quick way. It is in the spirit of harvesting the potential of these technologies as learning tools that this thesis was developed.

As games, apps and platforms centred on learning languages rise in popularity, there is a need to analyse and compare the various approaches adopted by these existing efforts in light of older techniques and research done on the subject of language learning. Results from learning theory might help improve the methods applied by these systems to present challenges to the user through prediction or real-time analysis, identifying what topics the user is more likely to struggle with. Techniques used in game design and gamification can similarly be applied to motivate and guide the learner, grasping their interest, attention and effort away from distractions which might have defeated traditional learning exercises.

One specific problem of language education is that of learning grammatical genders. In many languages, nouns (and in some cases other words) are classified as being either masculine or feminine (and, in many languages, into other classes and genders), which is the case in Portuguese and Spanish, for example, but not English or Chinese. This masculine-feminine distinction is usually seemingly arbitrary and can vary even between closely related languages. The combination of gender variation between languages and linguistic interference (the way the languages a person learns affect each other) is one of the core issues we will attempt to address in this thesis through the development of a game for learning gender: Fishing for Words.
1.1. Problem description
People learning (or attempting to learn) any gendered language (in particular, Portuguese) must invariably memorize the gender of each noun, along with the grammar rules that govern them. Naturally, mastering this aspect of the language is a prerequisite to mastering the language itself. This can be a slow process, even for otherwise advanced learners, not only because there are many nouns to learn, but because the hints that may be used to guess the correct gender of a given word typically have exceptions, so one must learn both the regularities and the irregularities.

Any game that is proposed to accomplish a particular goal, be it artistic, educational, financial or scientific, must inexorably be evaluated to determine whether it reaches that goal. Since our goal is to create a game to help with learning the gender of Portuguese words, it makes sense to test it in the same manner. Thus, our scientific problem, which this thesis aims to answer, is this:

*How can a game to learn grammatical gender in Portuguese be tested to ensure that it fulfils its purpose?*

2. Literature Review
In this section, we will present our theoretical research on the relevant topics for the project, mainly exploring the necessary linguistic background and how games and gamification can be used to make tasks more engaging.

2.1. Language Proficiency Scale
The Common European Framework of Reference for Languages (CEFRL) defines 6 levels of proficiency for users of a language, from beginner to native-like: A1, A2, B1, B2, C1 and C2.

2.2. Grammatical Gender
2.2.1 In general
In many natural languages, nouns (and, depending on the language, other words, such as adjectives, determiners and numerals) are divided into several classes, often including gender as one of the criteria of these classes. Other possible criteria include animate versus inanimate, countable versus uncountable and human versus non-human. In languages with these features, grammatical gender allows speakers to easily clarify a being’s sex by inflecting the noun, for example, and they are useful in anaphora disambiguation (in some sentences, they allow the reader/listener to infer whom or what each pronoun is referring to). [13] [16] [5]

However, words’ genders often seem to have been assigned arbitrarily, thus someone attempting to learn these languages must memorize the correct gender for each noun, as well as the rules pertaining to them in that particular language, which can include inflection (modification) of the noun itself, choosing other words to be in agreement with the noun or picking the correct gender according to the situation (in some languages, certain words can have more than one gender, usually those having to do with people or animals).

2.2.2 In Portuguese
The Portuguese language retains two grammatical genders from its ancestral Latin: feminine and masculine, which are applied to nouns, pronouns, adjectives and determiners (not to verbs, though these may have a gendered pronoun affixed to them, written with a hyphen and not affecting the verb’s stress). Some words may be used with either gender, which may cause them to suffer inflection, change meaning entirely and/or simply reflect the biological sex of the entity they refer to. We will refer to noninflected words that may be used with either gender as double-gendered.

Typically, masculine words end with -o while feminine words end with -a, but there are exceptions to this rule. There are also many other word endings that hint at a particular gender, most of which with exceptions of their own. This is a key concept used in our Tag System, which is described in Section 3.1.2. [2]

2.3. Spaced repetition
Spaced repetition is a simple technique used for memorizing large numbers of items and retaining them in long-term memory. The basis for this method is that humans more easily retain information studied repeatedly over long periods with time intervals between moments of study. The classic system is the Leitner method published in the 1970s, in which flashcards (cards on which the items to be remembered are written) are placed in numbered boxes and moved upward when remembered correctly and downward when remembered incorrectly. The lower the box’s number, the more frequently the cards within will be practiced, therefore the least-remembered items will be remembered more often (for example, box 1 may be practiced every day and box 2 every 2 days). [10] [23] Computer implementations of spaced repetition (such as Anki, Mnemosyne, Brainscape or Duolingo) usually implement some variation of Leitner’s system or of the more complex SuperMemo system. [9] The Leitner system is described in more detail in the description of Section 3.1.1.

Spaced repetition can be applied to any field of knowledge that requires mechanical memorization, but it is particularly useful for learning languages. It cannot be used by itself to become fluent in a language because the fundamental purpose of languages is to communicate with humans, which re-
Flow is a term coined by Csikszentmihalyi to describe the psychological state in which a person is completely immersed in the task they are performing. It is a crucial concept in games, as games are typically played for enjoyment, and experiencing enjoyment is part of the state of flow. Flow can even be used a way of measuring enjoyment in games. [15] [21] [20] Persons in this state are described as being totally absorbed in the activity at hand, being concentrated on completing it while becoming unaware of the passage of time, their physical surroundings and even their own self existence and their basic physiological needs. Another component of the state of flow is that the person derives pleasure from the task (it is intrinsically enjoyable, or “autotelic”), they feel that they are using their maximum capabilities (the task is challenging to them), that they feel that they have control of the task and that they can overcome all the challenges which will be presented to them further ahead (in other words, the person must not feel helpless).

A crucial condition for flow to occur and be maintained is the balance between the person’s abilities and the task’s difficulty (or perceived difficulty). This balance is not maintained, the person cannot achieve flow and is likely to abandon the task.

As explained above, activities in which a state of flow is achieved are rewarding by themselves, which causes the individual to seek to repeat the activity. And because the state of flow implies that the individual is working at his full potential, systematically achieving flow in a task causes individuals to become more and more proficient at it, which in turn fosters the necessary conditions for flow to be maintained. [18] [19]

2.5. Gamification

Gamification is an umbrella term for applying elements from game design to non-game tasks with the goal of increasing the users’ motivation and engagement in those tasks. Though a proper definition for the term is disputed, it has grown in popularity in the last decade. [12] It is not a technique restricted to the digital medium: they have been applied to areas such as commerce, education, health, sustainable consumption, work and innovation. [6] [7]

Later in this document we will look at Duolingo, a particular case of a gamified language learning platform.

2.6. Games and apps

In this next section, we will look into games, apps and platforms for language learning. These are interesting to us as examples of what can be done and some of the implications of what can be done, providing insight and inspiration for our own solution.

2.6.1 Duolingo

Duolingo¹ (2012) is a gamified language-learning platform created by von Ahn and Hacker. In Duolingo, the user selects the language they wish to learn and progresses along a (mostly) linear chain of lessons, each of which consists of a short text introduction to the topic at hand and a series of very short (typically involving a single sentence) exercises. In each exercise, the user is asked to perform a task such as translating, reading aloud or write what they hear, which elicits immediate feedback from Duolingo (whether the answer was adequate and what mistakes were made). This give-answer-get-feedback loop is essential to Duolingo’s approach and one of its gamification mechanisms.

Duolingo employs space repetition, which is visible to the users in the lesson menu. The system estimates how well the user remembers each lesson based on how many times they practiced it and when, with lessons practiced less or less recently being considered “weaker”, prompting the system to remind the user to practice them. Besides being quite popular, it has been shown by studies that it is an effective learning tool. [25] [24]

2.6.2 Crystallize

Crystallize² is an academic project created at Cornell University. It is a 3D online multiplayer videogame focused on teaching Japanese to English speakers which has gone through several iterations, though the following description is of the version released on 2015-09-06. In Crystallize, the learners play foreign students at a Japanese school where they roam, interacting with computer-controlled Japanese students, known as NPCs (non-player-characters). Players progress in the game by adding words to their lexicon by speaking with these NPCs through menus in which they must select appropriate words from their current lexicon. Thus, Crystallize’s main gameplay loop can be described as: learn word, practice word, use word.

The study conducted by Andersen et al revealed that, although the interaction between players is very limited, simply having other players present was also reported to be enjoyable and motivating, which is in line with Bartle’s taxonomy of player

¹www.duolingo.com
²http://crystallize-online.com/
types: socializers are described as MMORPG players who are mainly interested in chatting with other players, observing and being part of the community while treating the game itself as a side activity. [1] The game’s logs showed that players sometimes communicated in Japanese, which provides extra practice. [22] [3]

Players also engaged with the NPCs. The paper explains that the game’s design was thought out in a way that would provide context, which research tells us helps people remember by tying ideas to situations.

2.6.3 Content-based platforms
Some platforms, like LingQ, Blu Blu, Yabla, Memrise and FluentU seek to motivate learners by allowing them to select aggregated third-party content (video, audio and text) in the target language that they wish to experience, so they are (ideally) exposed to content which interests them most. These platforms provide access to genuine material in the target language, contrasting with material written specifically with the purpose of language education (or recorded by professional voice actors), which language educators regard as beneficial. [11] [14]

One problem this kind of platform faces is content curation. Because they aggregate third-party content, sometimes even uploaded by users rather than experts employed by the website, content must be filtered for quality, difficulty, correctness and type. This can be hard to achieve and it is a common target of criticism. [17]

3. Implementation
Our game, Fishing for Words, mostly takes place on a main screen representing a lake in a rainforest, in the centre of which stands a fisherman on their boat. The boat is flanked by a pink barrier on the left and a blue barrier on the right, representing the feminine and masculine genders, respectively. In each round, the lake is filled with 5 fish, each with a Portuguese noun written on it. The player must then solve each word by passing each fish through either the pink or blue barrier to paint it the right colour (corresponding to the written word’s gender) and place it on the top-right corner of the screen. The game will flash green or red after each fish, indicating whether the player’s answer was correct or not, respectively.

The game includes a score system. In each round, the player is presented with 5 aquariums for the 5 fish. Each aquarium has a different score: 5, 10, 15, 20 and 25 points, and they are all presented to the player in a random order, one by one. Once the player is presented with an aquarium, they must solve one of the remaining fish/words in the lake: if they solve it correctly, they will receive the number of points the presented aquarium is worth, otherwise they receive zero points. Therefore, to maximize their score, a smart player will choose an easy word-fish when presented with a high-scoring aquarium and a difficult word-fish when presented with a low-scoring aquarium (to minimize the chance of answering incorrectly). Each game is composed of 3 rounds with 5 fish each, so in each game it’s possible to accumulate a maximum of 225 points. As the player accumulates points, they unlock new boats as a reward, which appear during the game but have no effect on game’s mechanics.

For the first two games, the player will play a tutorial, which consists of slightly modified versions of the game that make its rules easier to understand: having a single fish instead of 5, darkening non-essential parts of the screen and large, animated objects to grab the player’s attention.

3.1. Word selection systems
One critical aspect of the game is what words from the list are shown to the player at a given time. The words must be varied and provide a challenge of appropriate difficulty, otherwise the player will become either bored or frustrated and will abandon game (as explained in Section 2.4), especially since, in Fishing for Words, the player has no control whatsoever over this process, besides playing the game.

The basic idea of these systems was touched upon in Section 2.3. Ideally, the student should be reminded of a piece of knowledge just when they are about to forget it []. The basic implication of this idea is that, as you might expect, words the student has more difficulty with should be practiced more often. With this knowledge in mind, we implemented two systems for word selection in the game, one based on the classic Leitner System and one designed by us (the Tag System), which are described in the next sections.

3.1.1 Leitner System (Flashcards)
One of the systems used in the game to select the next words to present to the player is known as the Leitner System, named after Sebastian Leitner, who published it in 1972. It was devised before personal computers became common, so it was originally accomplished using physical boxes and cards (flashcards), and it was created to help learners memorize any kind of knowledge (not just language). [8]

Our version of the Leitner System is very similar to the original (though it uses a computer in place of physical components). We have 443 “cards” (in our case, the Portuguese words that the player is meant to learn) and 5 boxes numbered 1, 2, 3, 4 and 5. Initially, every word is in box 1. Box 1 will
be reviewed every session, box 2 will be reviewed every 2 sessions, box 3 will be reviewed every 3 sessions and so on. In a given session, the player processes each word (from all the boxes due to be reviewed during that session) by answering the word’s grammatical gender. If the player answers correctly, the word gets advanced to the next box (if it was in box 1, it goes into box 2 and if it was in box 2, it goes into box 3, etc). If the player answers incorrectly, the word gets placed in box 1, regardless of its previous position. In Fishing for Words, this entire process is done behind the scenes, so the player only sees the words they have to answer and whether the gender they answered was correct.

Because boxes with lower numbers are repeated at higher frequencies, the player will see words they have answered wrongly more often than those they have answered correctly.

### 3.1.2 Tag System

We call the other system for selecting the words to be presented to the player the Tag System. It’s a modified version of a dynamic game difficulty adjustment system developed by João Catarino [4]. As discussed in Section 2.2.2, the ending of Portuguese words hints at what their gender might be (the most common of these being -a for feminine nouns and -o for masculine nouns). Section 3.1.2 shows the word endings used for the game and the gender they hint at.

The idea of this system is that players only need to learn these word-ending patterns and their exceptions, rather than every individual word. This is accomplished by adding a tag (a piece of extra information) to each word, to indicate which pattern it falls into. When a word is presented to the player and they answer it, their performance for that word’s tag is recorded (1 for correct, 0 for incorrect). Note that the performance is recorded for the tags, not for individual words! When the game needs to select words to present to the player, it will use the latest 10 performance values recorded for a given tag to estimate what the player’s performance will be for words with that tag and then use that estimate to present a challenge composed of words with an appropriate level of difficulty.

### Difficulty

The difficulty in Fishing for Words is not constant, but a sine function modified to have a range between 0 and 0.7 and to have a wider wave, as the equation below shows. The function’s input is the game number, while each game is divided into 3 rounds. So the input is actually increased in third, with for the very first round $x = 0$, for the second round $x = 1/3$, etc.

$$\sin\left(\frac{1.5x - \frac{\pi}{2}}{2}\right) + 1 \cdot 0.7$$  

where:

$x$ = game number

When presenting words to the player, the game creates a challenge composed of 5 words where the player’s estimated performance for those words is as close as possible to the value indicated by the difficulty function. Recall that a word’s estimated performance is the same as the estimated performance of that word’s tag. A tag’s estimated performance is calculated using a list containing (at most) the last 10 recorded performances recorded for that tag. The desired performance is subtracted from each obtained performance from the list and the result is squared: the average of these operations is the estimated performance for that tag. If no performance values have yet been recorded ($N = 0$), then the tag’s performance value is assumed to be 0. This process is illustrated in ??

$$\text{EstimPerformance} = \begin{cases} 
0 & N = 0 \\
\sum_{i=1}^{N} (x_i - d)^2 & N \neq 0 
\end{cases}$$

where:

$x_i$ = obtained performance values
$d$ = desired performance value
$N = n^0$ of performance values, $N \in \{0, 1, ..., 10\}$

### Tags

In our game, each word has exactly one tag. This tag was chosen depending on two factors: the word’s gender and its ending (based on the endings shown in Section 3.1.2). These tags are simply a small text where the first character is “F”, “M” or “D” (to indicate that the word’s gender is feminine, masculine or double, respectively) or “S” (always and only used when the ending does not give a clue about the word’s gender). The rest of a tag’s text indicates the word’s ending and, if multiple endings are valid, the longest is chosen. Therefore, if a word ends in -ão, -e or doesn’t correspond to any ending in the table, it will begin with S, otherwise, it will begin with the letter corresponding to its true gender.

### 3.2. Word List

Because our game’s purpose is to teach the gender of Portuguese words, it was necessary to obtain a list of words to be taught and classify them according to their gender (and assign tags to them,
Table 1: Noun endings that hint at the word’s gender. “Double” refers to non-inflected words that can be used with either gender (usually occupations or professions) and “Either” refers to endings that are not found more frequently in any particular gender.

<table>
<thead>
<tr>
<th>Feminine</th>
<th>Masculine</th>
<th>Double</th>
<th>Either (no hint)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-a</td>
<td>-o</td>
<td>-ante</td>
<td>-ão</td>
</tr>
<tr>
<td>-dade</td>
<td>-r</td>
<td>-ente</td>
<td>-e</td>
</tr>
<tr>
<td>-agem</td>
<td>-i</td>
<td>-ista</td>
<td>(any other endings)</td>
</tr>
<tr>
<td>-esa</td>
<td>-l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-essa</td>
<td>-z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ina</td>
<td>-s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-inha</td>
<td>-i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-isa</td>
<td>-u</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ora</td>
<td>-ema</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-friz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ana</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-oa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-onã</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ção</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Examples of word tags. Notice that the word’s gender determines the first character of the tag unless its ending is a (no hint) ending. Furthermore, words with a gender different from the gender the ending hints at are of particular interest, because they are exceptions that must be memorized.

<table>
<thead>
<tr>
<th>Word</th>
<th>Word Gender</th>
<th>Ending</th>
<th>Gender ending hints at</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>plano</td>
<td>masculine</td>
<td>-o</td>
<td>masculine</td>
<td>Mo</td>
</tr>
<tr>
<td>elefante</td>
<td>masculine</td>
<td>-ante</td>
<td>double</td>
<td>Mante</td>
</tr>
<tr>
<td>mapa</td>
<td>masculine</td>
<td>-a</td>
<td>feminine</td>
<td>Ma</td>
</tr>
<tr>
<td>ajuda</td>
<td>feminine</td>
<td>-a</td>
<td>feminine</td>
<td>Pa</td>
</tr>
<tr>
<td>situaçã o</td>
<td>feminine</td>
<td>-ção</td>
<td>feminine</td>
<td>Fção</td>
</tr>
<tr>
<td>belga</td>
<td>double</td>
<td>-a</td>
<td>feminine</td>
<td>Da</td>
</tr>
<tr>
<td>amante</td>
<td>double</td>
<td>-ante</td>
<td>double</td>
<td>Dânte</td>
</tr>
<tr>
<td>lá</td>
<td>feminine</td>
<td>(none)</td>
<td>(no hint)</td>
<td>S_</td>
</tr>
<tr>
<td>avião</td>
<td>masculine</td>
<td>-ão</td>
<td>(no hint)</td>
<td>São</td>
</tr>
</tbody>
</table>

as discussed in Section 3.1.2). The words chosen were taken from a lexicon students of Portuguese should know at the A2 level created by Prof. Jorge Pinto for the Perfil de PLE project (a partnership between Instituto da Cultura e Língua Portuguesa and Centro de Exames de Português Língua Estrangeira).

The first step to process the word list was to remove undesired words. The two main classes of gendered words in the Portuguese language are nouns and adjectives, but we decided that the game would only include the former because almost every adjective can be used with nouns of either gender, so our game would not be as suited for them (a more interesting exercise for adjectives could be centred around learning how to correctly inflect them for gender, for example). Thus, the most obvious candidates for removal were adjectives and genderless words, such as verbs and adverbs, but the task was not as simple as it might seem, because all three of these classes can be nominalized (turned into nouns), thus becoming gendered. For example, the verb *comer* “to eat” can be nominalized to mean “food”.

Various criteria were used to solve these ambiguities over whether or not a given word should be excluded from the list, some more subjective than others. Ultimately, the decision based on how adequate the word was for the students’ curriculum.

### 3.2.1 Processing the Word List

The word list processing was partly automated, as a web scraping program was used to search each word in the Príberam online dictionary to check their gender (which were later re-checked manually). Automation was also used to annotate the words with tags, necessary for the tag system described in Section 3.1.2.

### 3.3. Data Collection

*Fishing for Words* features a system to automatically collect data about how the game is being played. A log file is simply a text file where certain game-events are recorded (and timestamped) and a log file is created for each new game played or each time the game is opened (whichever comes first). As long as the player’s device has an internet connection, the game will periodically transmit the contents of unsent log files to a server that simply stores them so that an experimenter can later retrieve and analyse them. The game also features a small indicator which shows how many files are still unsent, so the player can leave the game on if necessary.

To match log files to players, the game uses an offline login system that requires the hash values (MD5) of the participants’ email addresses to be added to the game before it is distributed (the game will ask the player for their email address and use the stored hash value to ensure that it’s correct).

### 4. Experimental Procedure

#### 4.1. Usability Tests

Informal usability tests to assess how easily players could understand the game were performed throughout the game’s development cycle with native Portuguese speakers, but the game was also tested with 5 foreign students of Portuguese at Faculdade de Letras da Universidade de Lisboa about two months before the experiment proper. These tests revealed that some players understood the game after just a few seconds, while others would take longer and tended to ask for help. These user tests led to the implementation of a basic tutorial in the game, described in Section 3.

#### 4.2. Experiment

The experiment was carried out at 9 classes of A2-level students of the Portuguese language being
taught at Faculdade de Letras da Universidade de Lisboa, with a total of 80 students agreeing to participate in the research. Sections 4.3 to 4.5 will describe the steps of the experiment.

4.3. Pre-test Visit
The first step of the experiment proper involved visiting the students’ classes to explain the research being performed and to ask for their participation, which included asking them to fill out a consent form, a form about their linguistic background and the pre-test, a written test with exercises about gender in Portuguese.

4.4. Email Interaction
After the pre-test visit, we examined the number and characteristics of our participants in order to divide them into three experimental groups: one control group (which would not receive the game), one group to receive a version of the game using the Leitner System (see Section 3.1.1) and a third group to receive a version of the game using the Tag System (see Section 3.1.2). Once we had decided the members of these groups, the students who were part of the latter two received the game by email (on the 13th of April) and were asked to play for 10 days (until the 23rd of April).

In this case, all 34 students who could be identified and claimed to regularly use Android or Windows were included in the groups due to receive the game. Preference was given to the Android version, so students with only Android or both operating systems only received the game for Android, while those who only had Windows received the game for Windows.

4.5. Post-test Visit
The 23rd of April was the day of the second visit to the classrooms, where the students were asked to:
1. Answer the post-test, a written test that is mostly identical to the pre-test, with the purpose of enabling comparison of the students’ performance before and after two weeks had past.
2. Fill out the questionnaire, a form asking their opinion on various matters which essential to our evaluation of the methodology.

One of the students, who stood out for having played the game much longer than anyone else, was informally interviewed.

NOTE: The students were on holiday from the 6th to the 15th of April.

5. Results
5.1. Overall Results
During the two weeks the students were given to play the game, the logs received indicated that only a total of 3 students had played: two for less than 15 minutes and one for around 2 hours and 30 minutes (taking about 2 hours to unlock all the boats). While this sample of gameplay data is not large enough to perform statistical analysis about how playing the game might impact the students’ knowledge (and possibly comparing the Leitner system with the Tag system), the students’ answers to the questionnaire provide us with valuable information to speculate about the reason for this low amount of participation.

Thus, the results of this paper are divided into two main sections: the analysis of the two written tests performed by the students and the qualitative evaluation of the game and methodology based on the outcome of the experiment and the students’ feedback.

5.2. Written Test Results
The students performed two written tests (the pre-test and the post-test) with a two week interval between them, described in Sections 4.3 & 4.5, respectively. Our sample size for these written tests was 50 students. Only exercises 1 and 2 were taken into account, as exercise 3 was not well explained, rendering it invalid and exercise 4 was exclusive to the post-test, which made comparison impossible.

To study the students’ change in performance over this time interval, we counted how many words each student answered correctly, yielding a grade from 0 to 10 for each exercise. In exercise 1 (in which sentences with a blank space had to be filled out with the appropriate gendered definite article, o or a), only the gender was taken into account when grading the tests, so even ungrammatical sentences were considered correct, as long as the gender was correct.

In exercise 2 (in which each word must be matched to a fixed number of gendered definitive articles o and a), if a single item in each column was left unmatched, the pair was automatically matched (it was assumed that the student simply forgot the last pair).

The Shapiro-Wilk test revealed that the grades did not follow a normal distribution and a Wilcoxon signed-rank test showed that there was no statistically significant change between the pre-test grades and the post-test grades, yielding $Z = -0.414$, $p = 0.679$ for exercise 1 and $Z = -0.625$, $p = 0.532$ for exercise 2.

5.3. Opinion Questionnaire
We have identified several factors that may have led to fewer students playing the game, some of which are supported by the students’ answers in the questionnaire (described briefly in Section 4.5).
5.3.1 Platform Support
The game only works on the Android and Windows operating systems and only 35 of our 80 participants used at least one of the two. Thus, supporting a wider range of devices (in particular, iOS, which 40 out of the 80 students reported using) would likely lead to greater acceptance. Lack of iOS support was a common complaint.

5.3.2 Installation
In the case of Android, the game was distributed as an APK file which had to be installed. Out of 17 students that said they had received the game, 6 did not try to install it and 3 were not able to due to technical problems. ?? shows that most participants agreed with the statement “The instructions by email on how to install the game were clear” and, indeed, none of them indicated lacking instructions as a reason to not install (though some of the “technical issues” they experienced could be due to lack of familiarity with this method of installation).

5.4. Video Tutorial
The game’s tutorial is rather basic and not all players immediately understand the game. As a time-effective way of mitigating this issue, a video tutorial was sent to the participants, but many students did not watch it. The questionnaire answer show that players have a slight tendency to find the game and tutorial easy to understand (rather than confusing), but there’s clearly room for improvement, with one comment reading:

“It’s difficult to understand what should we do in this game. Better to make some demo or write rules. Also I couldn’t understand how to finish the game and close this app. There is no any buttom for this.”

5.4.1 Game tone
The game may seem too childish for some students due to its cartoony graphics, which leads to them losing interest. This may be a difficult problem to avoid because language learners, as a target audience, cover a wide range of demographics and tastes, but it may be alleviated by adding diversity to the game, so players have more options on which parts of the game to explore and experience and also by carefully tuning the way it’s presented (“marketed,” if you will) to avoid alienating too many potential players. ?? shows mixed reactions to the statement “The game looks like it’s for children” as well as “You like playing smartphone games” showing that a surprisingly high number of students simply aren’t interested in games. Two of them wrote:

“I like apps like Duolingo because they are concrete. I don’t like learning games.”

“I do not like games.”

5.4.2 Time-frame
Because the time-frame during which the students were asked to play coincided with their Easter holiday, they may have had less incentive to play, whereas a game to study for an upcoming exam might have been more alluring to them. Most participants felt that the time-frame given to play the game was not enough. In the open-ended question, one student replied that “It would be nice to have longer time to try it. It came in the middle of school work, and right before Easter. Simply didn’t have time!”

5.4.3 Perceived importance of grammatical gender
Grammatical gender is not an essential part of communication in Portuguese, as a speaker can typically still be understood when they break gender agreement. This may mean that some students choose to attribute less importance to these grammar rules, especially at the A2 level of proficiency (to which our participants belonged). Furthermore, for most Portuguese words, gender can be guessed correctly knowing only the simplest hint (that an -o ending indicates masculine and an -a ending indicates feminine) and, in our word list, this was true for 63.80% of words. With this insight, we conclude that a game to teach gender may be more appealing to students at higher proficiency levels, which can already communicate effectively and are more interested in perfecting their language skills (as opposed to simply being able to communicate). In the open-ended question, a student replied:

“Sorry, I don’t even practice with Duolingo, which is a fully developed “game”. I spend too much time at my phone anyways. My priority would be to learn words before knowing more genders”, which indicates that they seem memorizing gender less important than learning vocabulary, even though they answered Neither Agree nor Disagree in relation to ‘It’s important to practice gender in Portuguese at the A2 level.” Another student stated:

“I hope to have more pronunciation training.” also implying that another part of the curriculum is more important than gender, though they replied Strongly Agree in relation to the sentence.

5.5. Gameplay Data
Though few students played the game, a single participant stood out for having played for around 2 hours and 30 minutes, unlocking all boats.

The estimated time per session of 3 minutes turned out to be remarkably accurate, at least for the student that played the most, whose average
session was 2m54s long, but the average score per session was 183, rather than our estimate of 112.

6. Conclusions

6.1. Contributions

**Fishing for Words** – The main contribution of this thesis is the game itself, as it is prepared to be used as a test-bed for different word-selection models (as we did in this thesis) and can be extended or modified for other purposes, particularly games focused on binary choice. It can also be used as-is by students of Portuguese, although its feature set is limited.

**Annotated word list for A2** – One of the components of the game that can be used separately is the word list. A baseline list of nouns that A2 students of Portuguese should know annotated with their gender may be useful for other projects of this sort.

**Word processing program** – Complementing the word list, the code used for accessing the Priberam online dictionary to retrieve words’ genders and automatically tag them (according to their ending) might be useful to reuse in other linguistic projects.

**Tag System** – Another component of the game that could be used separately is the Tag System. Although this research does not draw conclusions about its effectiveness, the simplified version developed for *Fishing for Words* may have advantages for some applications (in relation to the more general approach developed by João Catarino[4]).

**Methodology** – Our final contribution is the methodology used to test the game with the students. Anyone wishing to test the game (with or without modifications) in a context similar to our research will find it useful to implement (or at least take into consideration) our suggestions and results gathered from this experiment, including the result that no change in performance was observed between the pre-test and the post-test in students who had not played the game.

6.2. Conclusions – Written Tests

The statistical analysis of the pre-test and post-test results (Section 5.2) shows that students of Portuguese do not gain nor lose performance in a written grammatical gender test after two weeks (one of which they spent on holiday). This information may be useful for future tests of this sort: though a control group is always indispensable for any such research, this data may serve as a baseline for experiments in a similar context, as it gives researchers an idea of what to expect.

One must, of course, consider the many uncontrolled variables present. The sample size was 50 participants of A2-level proficiency, all living in Portugal. Extrapolating from these results should be done with care, especially so if many differences can be identified between these conditions and those for which we are extrapolating.

6.3. Conclusions – Methodology

On evaluating our experiment’s methodology, we conclude that some factors had more significance than others in their contribution to the low number of players. The factors we considered are as follows:

**Perceived importance of grammatical gender**

We found that while students generally consider practicing grammatical gender to be important at the A2 level, some of them commented that it is of less importance to them than other facets of the curriculum (vocabulary and pronunciation).

Linguistically, we know that it is possible to communicate effectively without following the rules of gender, though it is obviously necessary to be able to use the language correctly. Furthermore, the gender of most nouns can be guessed correctly by knowing only the most basic rule.

Though the evidence for low perceived importance of grammatical gender in students of Portuguese could be stronger, it compels us to recommend that future gender-based studies that rely on the participants’ interest be preferentially performed on more advanced students (even native-like C2 speakers are known to have trouble with gender).

**Platform support** – Having the game work on more platforms would certainly increase the number of players: a little more than half our participants could not play simply because they did not use Android nor Windows, so some students who may have been interested in playing the game did not get an opportunity to play it.

A more advanced form of multiplatform support would be to allow each player to play on more than one platform, though this raises the issue of synchronicity (the player’s progress must be synchronized across all platforms) and the research problem that players may have different gameplay experiences depending on which platform they choose, which introduces an extra variable to the experiment.

**Usability** – The game suffers from some usability problems that could be avoided, both in understanding the game itself and during installation. The players who tried the game did not play it for very long (with a single exception), therefore increasing player retention is an important factor.

This was shown by both the preliminary usabil-
ity tests and the questionnaire answers, particularly by the written feedback. However, the usability tests also indicated that players found the game easy and quick to understand when it was explained to them by the researcher.

One solution for improving usability is to improve the tutorial, which would allow anyone to learn to play the game without any added external help, as long as it was properly tested with users.

An effective solution without the need for further game development would be to have researchers guide the participants through the installation process in the classroom and explain how to play the game, which should eliminate both the need for a more elaborate tutorial and any installation issues that the participants might have otherwise encountered. This, of course, is only viable if the game is being tested in a similar context, where the research interacts directly with the participants (as opposed to, for example, publishing the game on the internet for anyone to play).

6.4. Future Work

The previous section was an evaluation of our methodology, including suggestions for how to improve it. Future research on Fishing for Words may implement these changes to achieve a more active participation by the participants.

Further research could focus on comparing the word-selection models we implemented (or novel ones), or test whether a game like this is useful to word-selection models we implemented (or novel ones), or test whether a game like this is useful to mitigate the effects of linguistic interference.

Ultimately, it should not be lost that in games, the focus is on the player’s experience being fun, and word-selection models are but one piece of that puzzle, because a player who has fun is a player who stays motivated and has a higher chance of not giving up on learning the language they wish to learn.

References