Extending VITHEA in order to improve children's linguistic skills

Vânia Patrícia Padrão Mendonça
vania.mendonca@tecnico.ulisboa.pt

Instituto Superior Técnico, Porto Salvo, Portugal

October 2015

Abstract

An Autism Spectrum Disorder (ASD) is a developmental disability characterized by impairments in social communication and interaction that often comprises difficulties in the acquisition of verbal language skills. In order to improve the quality of life for individuals with ASD, and given their documented interest in computers, several researchers and companies have developed applications that aim to provide an alternative to verbal communication or a way for children with ASD to develop linguistic skills, necessary for verbal communication. However, there is a lack of applications in Portuguese that are tailored to the individual needs of each child. In this context, we present VITHEA-Kids, a platform that accounts for the needs of different children, as well as their caregivers, by providing an interface where caregivers can create exercises and customize various aspects of the interaction with the platform. We also developed a module that allows the automatic generation of multiple choice exercises, meant to be integrated in VITHEA-Kids. Our platform is under evaluation on multiple fronts: a preliminary evaluation with caregivers (which provided promising indicators), an evaluation with a child (ongoing) and an evaluation of the generation of incorrect answers (distractors) in multiple choice exercises (which resulted in acceptance rates between 61.11% and 92.22%).

Keywords: Autism Spectrum Disorder, children learning, language skills development, automatic generation of content

1. Introduction

Autism Spectrum Disorder (ASD) comprises a set of developmental disabilities characterized by persistent deficits in social communication and interaction, as well as restricted, repetitive patterns of behaviour or interests since an early developmental period [1]. The most recent worldwide estimations (2012) point to a proportion of 17 in 10000 children with autism and 62 in 10000 children with other pervasive developmental disorders in the autism spectrum [6]. Furthermore, the number of reported cases has been increasing [3].

Since individuals with ASD often face challenges regarding social and communication tasks, and some of them do not develop the ability to verbally communicate at all, or develop it at a much slower pace than their typically developing peers [13, 1], there are therapies that aim to minimize these difficulties. However, since therapy might not be affordable, investment in education for children with special needs might be insufficient [4, 5] or therapy centres might not benefit from enough conditions [8], the possibility of taking therapeutic interventions to other settings or having affordable means to perform such interventions could be desirable.

Given the interest that individuals with ASD display towards computers [16, 23], some authors researched about the effect of using software for the purpose of teaching academic skills to children with ASD [10, 19, 17, 11, 29]. Additionally, in the past few years, with the increasing popularity of mobile devices such as smartphones and tablets [9, 12], both companies and researchers have invested in the development of educational mobile applications targeting children with impairments, thus easing the practice of verbal skills in diversified settings, with or without assistance. Despite the current large offer of such applications, there are still issues left to be addressed: most applications are paid; there is a lack of applications available in Portuguese, as well as of applications that take into account each user's progress, characteristics and/or needs, specially among the ones for educational purposes. Considering this situation, the main goal of our work is to develop a software platform where children diagnosed with ASD can solve exercises in order to develop or improve a set of linguistic skills regarding the Portuguese language, having in mind the characteristics of each user.

The platform we developed, VITHEA-Kids, comprises two modules: one for the children and another for their caregivers. The child's module's features include: a set of user-created multiple choice exercises for the child to solve; a talking animated character to utter the exercises and feedback regarding the child's answers; the possibility of using prompting when the child pick an incorrect answer; the possibility of displaying user-created reinforcement images when the child hits a correct answer. As for the caregiver's module, it allows to: create and manage multiple choice exercises of two types; upload and manage multimedia resources to be featured in the exercises; create and manage the users for the child's module; customize the reinforcement images and the animated character's utterances.

In addition to this platform, we also developed a module that generates multiple choice exercises (composed

by a question, a correct answer, and a set of incorrect answers, also known as distractors) given a certain topic, a question template and a set of constraints regarding the distractors. This module makes use of a second module that extracts words and images based on hierarchies of synonym sets.

2. Background

The first reference to autism belongs to Leo Kanner, in 1943. Kanner observed 11 children who revealed symptoms of a condition that had never been reported until then [13]. Despite the peculiarities of each individual case, Kanner was able to identify a set of common characteristics, namely: inability to relate themselves to people and situations and to develop social awareness; delayed development of speech and language skills, in contrast to an exceptional rote memory; high sensitivity to loud noises, moving objects; display of an obsessive desire for the maintenance of sameness and completeness; good cognitive skills, despite an potential delay in the development of language skills. In 1944, Hans Asperger described a group of children with a similar pattern of behaviour; however, his observations differed in the sense that the children he described displayed a typical (and sometimes exceptionally good) development in what concerns to cognitive and language skills [2]. Currently, the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) places both Kanner's and Asperger's definition of autism under the diagnostic of "Autism Spectrum Disorder".

Often individuals diagnosed with ASD go through therapy in order to improve their social and communication skills. Most therapies are based in a psychology approach called Applied Behaviour Analysis (ABA), where the focus is to modify a certain behaviour in an individual. This approach follows the **operant conditioning principle**, which consists of analysing an individual's behaviour considering not only the behaviour itself, but also the events that precede such behaviour (**antecedents**) and the events that follow the behaviour (**consequences**) [28, 20]. The general procedure for an ABA based therapy consists of the following steps [20]: 1) Functional analysis in order to identify the individual's behaviours and the respective antecedents and consequences; 2) Selection of a single behaviour as the focus of the treatment (**target behaviour**); 3) Measurement of the current level of the individual's target behaviour; this level is then defined as the treatment's **baseline**; 4) Implementation of an **intervention** in order to improve the target behaviour and measurement of this behaviour during the intervention process in order to assess its effectiveness; 5) Assessment of whether the acquired skills were generalized across different settings, people and materials, or not.

The most commonly used interventions are the **reinforcement** based interventions (if the manipulated stimulus is something that pleases the individual, i.e., the individual's behaviour is rewarded). These includes several techniques [24], such as **prompting**, which consists in using an antecedent auxiliary stimulus that aims to elicit the desired response. Once such response becomes more frequent, this stimuli can be removed. An example of prompting is when the child is told to pick a certain object and the adult points at that object in order to help the child performing such task. In a least-to-most approach, the adult would just tell the child to pick the object and, if they did not pick it after some time, the adult would then point at the object; in a most-to-least approach, the adult would start by always pointing at the object immediately after asking the child to pick it, and would gradually stop doing so in the following iterations.

3. Related Work

In this work, we reviewed works concerning two distinct topics that are related to our goals: the use of technology for children with Autism Spectrum Disorder (ASD) and the use of Natural Language Generation in the context of education.

Regarding technology for children with ASD, some authors performed surveys regarding aspects such as software features preferred by children with ASD and their caregivers, as well as feedback from users' relationship with technology, namely: previous experiences, main difficulties that should be addressed, and reasons to abandon previously adopted technology. The outcome of these surveys indicates that communication and social skills are the ones caregivers prioritize. Also, both children and caregivers display interest towards the use of technology; however, caregivers have reported many difficulties in their past experiences with technology, mainly regarding aspects such as device portability, content tailoring based on the child's characteristics, ease of use and maintenance, among others.

Many researchers and companies have also developed software for different purposes, spanning from entertainment and planning to the development of the communication, academic, social or emotional skills. In terms of research, some authors performed studies where they compared the outcome of using software *versus* non-computational solutions or human intervention, regarding children's language development process, while others focused their work on evaluating the impact of a specific set of features. Most of these studies reported progress regarding the children's ability to learn new skills.

As for the use of automatic features such as generation and processing of content in the context of education, these can help teachers and educators by allowing them to save them time while creating content [26], adapting to each student's characteristics [18], contributing to a fairer grading process [25, 14, 15], etc.

For the purpose of automatically generating content, we analysed a set of resources and tools that we could use in order to extract words and images, as well as to assure the correctness of the resulting content. For the task of retrieving words in Portuguese and images to illustrate such words, we looked for lexical and/or image resources. However, the only resource which facilitates image retrieval is **ImageNet**[1], which allows to browse through nested hierarchies of synsets, in which the outer synsets contain the hyperonyms of the inner

[1] `http://image-net.org/` (last visited on 05/01/2015)

synsets. Each synset is identified with a number, which is associated with a text file including a set of image URLs.

As for guaranteeing the correction of the generated content, a tool such as TreeTagger [27] can be considered. This is a tool for the Part of Speech (PoS) and lemma annotation of sequences of words, and it can be used for any language as long as a lexicon and a manually tagged training corpus are given, which is the case of many languages, including Portuguese.

4. VITHEA-Kids: A platform for children with ASD and their caregivers

Recalling our goal of delivering a platform in which children can develop language and generalization skills, having in mind their individual needs/characteristics, we present VITHEA-Kids: a platform where children can solve exercises created by their caregivers.

From the child's point of view, our solution comprises the following features: Solving multiple choice exercises; Using prompting cues to help the child to choose the correct answer for a given exercise; Possibility of showing content tailored to the child's preferences as a way to reinforce the production of correct answers; Use of an animated character to present the exercises and provide feedback according to the child's answers. As for the caregiver's point of view, the delivered features include: Possibility of creating and managing different kinds of exercises, with full control over their content; Possibility of uploading and managing multimedia resources (images) to be featured in the exercises; Possibility to create and manage the application's users; Possibility of customizing the content to use as reinforcement for each child and the animated character's utterances in each situation (e.g., when the child logs into the application, when a correct answer is given, when a wrong answer is given, etc.).

4.1. VITHEA - Virtual Therapist for Aphasia Treatment

In order to implement our solution, we decided to extend an existing in-house platform, since that would allow us to save time on features that might have already been implemented and to direct our effort towards new challenges. Out of the applications considered, Virtual Therapist for Aphasia Treatment (VITHEA) [21, 22] stood out as the most appropriate for our needs. VITHEA is a platform originally targeted to people diagnosed with aphasia (a language disorder that causes difficulties in several tasks, such as word naming). This platform allows patients with aphasia to solve exercises in which they can practice oral naming skills. All the types of exercise follow the same structure, being composed of: (a) a stimulus in the form of text, image, video or audio; (b) a question about that stimulus; (c) a set of up to four correct answers.

When solving an exercise, the patient is provided with the stimulus and the question about it, to which they should orally reply. For an answer to be considered correct, it does not have to be an exact match to one of the correct alternatives provided by the therapist. A list with the most frequent synonyms and diminutives is also considered, so, if the patient's reply includes a word in this list, it is considered correct. Replies that include a correct keyword, even if they also include unrelated terms, will also be considered correct. Finally, if the reply is incorrect, the patient can try again for a user-defined number of attempts.

VITHEA comprises two modules: the therapist's module and the patient's module. The therapist's module is intended to be used by therapists in order to create and manage the exercises described above, as well as to manage the multimedia resources (images, audio or video) available to be used as stimuli in the exercises, create and manage users (both therapists and patients), and monitor statistical data about the patient. As for the patient's module, it provides the exercise interface to the patients and stores statistical data about each session (e.g, the patient's utterances when replying to a question). This module also features an animated talking character, called "Catarina", which is responsible to present the exercises and provide feedback regarding the patient's answers.

The process behind the resolution of an exercise comprises the following steps: for each stimulus, the animated character asks the patient a question related to a given stimulus. The patient's reply is recorded into an audio file that is sent to an Automatic Speech Recognition (ASR) module, which converts the audio into a textual representation to be compared with a set of possible correct answers. Finally, the patient receives both textual and audio feedback based on the comparison result. The audio feedback is presented by the virtual character through text-to-speech synthesis).

4.2. From VITHEA to VITHEA-Kids

Although we took advantage of VITHEA's infrastructure, as well as of some of its features, our target users differ from VITHEA's, which means VITHEA-Kids also differs significantly from VITHEA concerning its purpose and functionalities.

In this work, we use multiple choice exercises. Each exercise is composed of a question, an optional stimulus (picture or text) complementing the question, and a set of possible answers (textual or pictures, respectively), in which only one is correct. The number of distractors per exercise can vary between zero and three, easing the task of creating exercises with variations in the number of distractors. For instance, a caregiver can create exercises with zero distractors in order for the child to strengthen a certain concept, and then, when appropriate, create exercises with one or more distractors to increase the difficulty. We have implemented multiple choice exercises of two subtypes: 1) identification of an image; 2) choice of the correct picture to match a given word/expression. An example of each subtype of exercise is presented in Tables 1 and 2.

The reason behind the choice of using multiple choice exercises is that this kind of exercises has been used for children with ASD and might allow to work on skills such as vocabulary acquisition, word-picture association, and generalization.

Similarly to VITHEA, VITHEA-Kids is composed of two modules: the caregiver's module and the child's module. The caregiver's module was based on

Field	Value
Question	What is the name of this object?
Stimulus	
Correct answer	Fork
Distractors	Spoon
	Cup
	Bowl

Table 1: Image stimulus and a set of possible textual answers

Field	Value		
Question	What is the image corresponding to this word?		
Stimulus	Fork		
Correct answer			
Distractors			

Table 2: Textual stimulus and a set of possible answers (images)

VITHEA's therapist's module, however, it offers different functionality. Therefore, we had to modify this module in all its layers, since the information to be recorded and displayed regarding the exercises, multimedia resources and users strongly differs from the information recorded in VITHEA.

After logging into this module, the caregiver is presented with the main screen, which features four sections: *Exercícios* (Exercises), where the caregiver can manage both the exercises and multimedia resources; *Utilizadores* (Users), where the caregiver can manage child users' accounts; *Preferências* (Preferences), where the caregiver can customize several aspects of the interaction between each child user and the child module; *Estatísticas* (Statistics), where the caregiver should be able to access the logs of each child user's performance during a session of exercises. The latter is still being developed.

Regarding the exercises, the caregiver is now able to create the two types of multiple choice exercise presented above. When creating an exercise, the caregiver should specify the topic of the exercise (e.g.,"Animals") from a predefined set, its level of difficulty (*Introductory*, *Intermediate* or *Advanced*), the instruction, the stimulus, the correct answer and the set of distractors, as depicted in Figure 1. When the caregiver submits the exercise, the input is validated in order to check if the mandatory fields (all of them except the stimulus and the distractors) are filled. The caregiver can also list and edit the exercises available.

Concerning the multimedia resources, the caregiver is able to upload images, also specifying the title of the resource and the topic it relates to. The image formats currently supported include jpe, jpeg, jpg, png, gif, bmp, tif and tiff. As with the exercises, submission triggers input validation to check if all the fields are filled. It is also possible for the caregiver to list the available resources and well as to edit them.

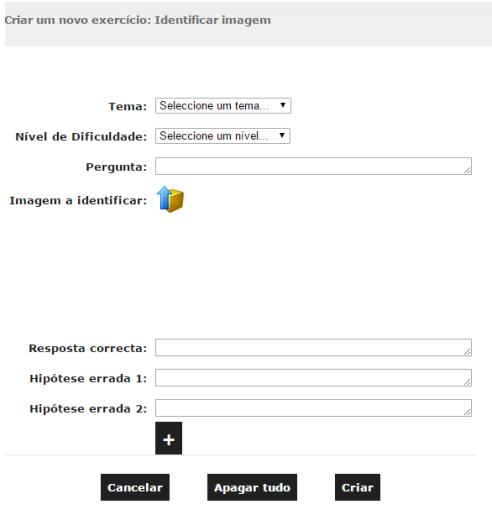

Figure 1: Creation of a new exercise.

It is also possible to create, list and manage child users. When creating a child user, the caregiver should specify a unique user name, a password, as well as the child's first and last name, birth date and sex. Submitting a child creation form triggers input validation to check if all mandatory fields are filled and whether both password fields match. It is also verified whether the specified user name is already in use by another child.

Besides the basic features described so far, we added the possibility for the caregiver to customize certain aspects of the child's module. One of those aspects is the animated character's utterances in specific situations (see Figure 2), namely: when the child logs in for the first time in the child's module; after each subsequent login in the child's module; when presenting the list of exercise types; when the child answers correctly to an exercise; when the child finishes a session of exercises.

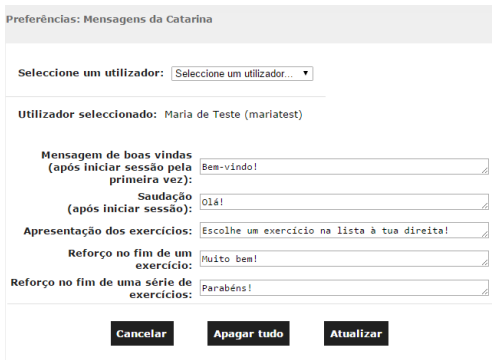

Figure 2: Animated character's utterances customization.

If these utterances are never edited for a certain child, the following default utterances will be used: "Bem-vindo" ("Welcome") when the child logs in; "Escolhe um exercício na lista à tua direita!" ("Pick an exercise in the list on the right!") when presenting the list of exercise types, "Muito bem!" ("Very good!") when the child answers correctly to an exercise and "Parabéns!" ("Congratulations!") when the child finishes a session of exercises.

Another customizable parameter is the set of reinforcement images to be used for each child user when they choose a correct answer. The caregiver can then upload images, list all available images and select which images should be used for each child (see Figure 3). When a new image is uploaded, it does not become immediately active for the existing child users, since the content might not be reinforcing for all of them.

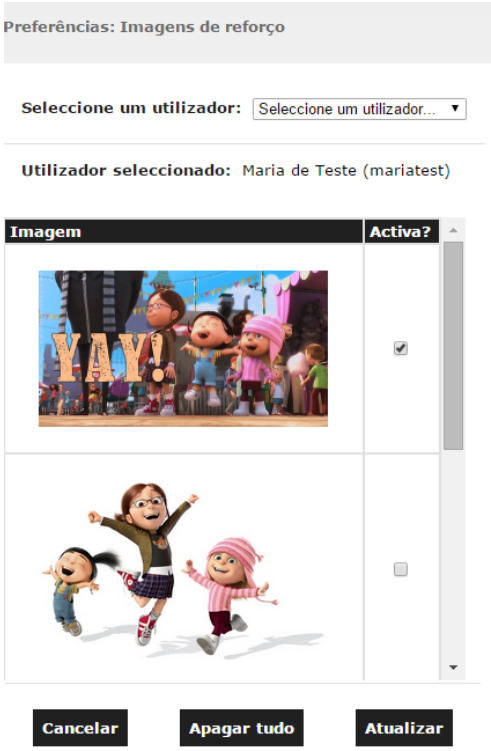

Figure 3: Reinforcement images customization.

The child's module is based on the patient's module. After trying both the web and the mobile versions, we decided to focus our efforts in adapting the mobile since the plug-in required to run the animated character (Unity Web Player) is about to be discontinued in most popular browsers (also, users might not be allowed to install it on certain computers). Besides, the caregivers we consulted displayed preference towards using a mobile device for the child's application.

In what regards to interaction, after logging in, the child is presented with the main screen, which shows the list of types of exercises. When choosing one of them, the first exercise shows up. The animated character utters the question, and the exercise area is filled with the stimulus and the possible answers in a random order (Figure 4). Picking the correct answer on the first try will lead to a reinforcement image. A click over any other answer will lead to prompting in order to help the child picking the correct answer: the clicked distractor disappears, the correct answer is highlighted and the remaining answers are uttered by the animated character. If, after that, the child picks the correct answer, a weaker reinforcement screen, composed of the question stimulus and the correct answer, is shown. The child can also click on the arrow located in the bottom right corner to skip the current exercise or to proceed to the next exercise after solving the current one. In the application's settings menu, there are also options to end the exercise session (saving all the child's performance data regarding that session) and to go back to the main screen (without saving any information). When the exercise session ends (either because the user chose so or because all the exercises under the chosen type were solved), some information about child's performance data is shown, namely: number of solved exercises; number of correctly answered exercises; number of skipped exercises; average number of times the child picked a distractor per exercise. We also added the possibility of having the animated character repeating her most recent utterance by clicking a repeat button near the character.

Figure 4: Child's module exercise.

5. Automatic Generation of Exercises

Besides VITHEA-Kids, we also developed a system for automatic generation of multiple choice exercises that aims to ease the task of creating exercises, either for VITHEA-Kids users or in other contexts.

The main idea of our system is to allow the generation of content for multiple choice exercises of two types: one where the answers are in the form of text (to identify an image) and another where the answers are in the form of images (to illustrate a word). Each exercise is generated given as input the exercise topic (e.g, "Animals"), a template to guide the question generation (composed by an immutable part and a variable whose value can change depending on the exercise's topic) and information regarding the number of distractors and their proximity to the given topic. The output is an exercised composed by a question/instruction, the correct answer to the question and a set of distractors.

From this main idea, we identified two distinct responsibilities: 1) generating content that constitutes a valid exercise, and 2) retrieving words and/or images according to the relationships among them. Therefore, we decided to delegate these responsibilities into two different modules: *ExerciseGeneration* and *WordImageLookup*, respectively. Figure 5 provides a black-box view of the main components of our system, in which the components with a thick line (*ExerciseGeneration* and *WordImageLookup*) constitute our contributions (TreeTagger is an external module).

An example of input could be the topic "Emoções" (Emotions), the question template "Que `<%variavel>` é este?"(Which `<%variable>` is this?), "3" as the number of distractors and "true" as whether the the distractors should be related to the given topic. For this input,

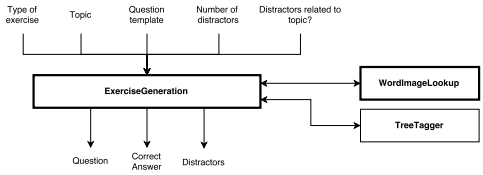

Figure 5: Overview of the automatic exercise generation architecture.

a possible output could be the question "Que emoção é esta?" (Which emotion is this?), the correct answer "alegria" (happiness) (together with an image of a happy person) and the distractors "tristeza" (sadness), "raiva" (anger) and "medo" (fear).

The *ExerciseGeneration* module is responsible for generating the content for multiple choice exercises. The *WordImageLookup* module is responsible for retrieving words and/or images from a hierarchy of synsets (sets of synonyms) organized according the relationships of hyperonymy/hyponymy among them.

Regarding the generation process, *ExerciseGeneration* module provides the methods for generating two variations of multiple choice exercises. Both follow the same approach; the only difference lies in the set of distractors returned: one of the variations returns a set of textual distractors, while the other returns a set of image distractors. As we can see in Figure 6, the exercise generation process starts with the generation of the the correct answer and the corresponding stimulus (being one of them a word and the other an image, or vice-versa, depending on the type of exercise), given a certain topic. Both the processes of generating a question and generating the set of distractors depend on having the correct answer: if the question template contains a variable, the variable will be replaced by a hyperonym of the correct answer; as for the distractors, they must be combined with the correct answer in such a way that it is possible to solve the exercise (meaning the set of distractors cannot contain the correct answer nor a synonym of it). The question and the set of distractors can be generated in any order. In both cases, the exercise's topic is given as input, but in the case of the question generation, a question template is also provided, while in the case of the set of distractors, the input includes the number of distractors to be returned and a boolean indicating whether the distractors should belong to the the exercise's topic. The question, the correct answer and the set of distractors are then assembled in a structure to be returned by either one of the two methods provided.

Our method to generate a correct answer consists in picking an hyponym of the topic provided. The process to achieve this goes as follows:

1. The ExerciseGenerator component asks *WordLookupImage* for all the leaf hyponyms of the topic;
2. If no hyponyms were found, an exception is thrown, and the process of generating a correct answer ends. Otherwise, a random synset is picked out of the returned leaf hyponyms;

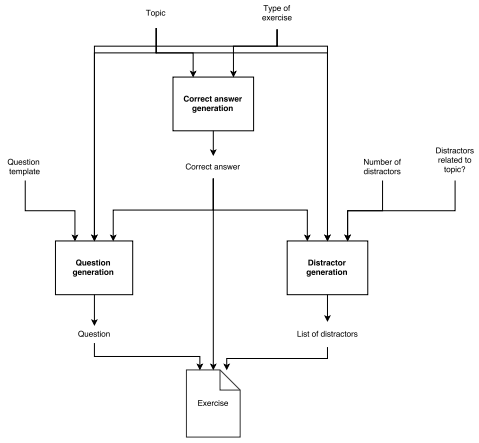

Figure 6: Overview of the automatic exercise generation process

3. The ExerciseGenerator component asks *WordLookupImage* for all the images associated with the id of the synset picked in Step 2;
4. If the set of images associated with that id is empty, we go back to Step 2. In order to avoid an endless loop, once the number of iterations exceeds the size of the hyponyms list, we pick the first synset that has a non-empty set of images associated;
5. A random word is picked out the last picked synset;
6. A random image is picked out of the image set associated with the picked synset.

Taking the example input given above, in Step 1, *WordImageLookup* would retrieve the leaf hyponyms of the word "emoção" (emotion) ("Emoções" after being normalized). If any hyponyms were found, a random synset would be selected (e.g., the one composed by the words "alegria, contentamento" – hapiness, joy). If this synset is associated with one or more images, a random word (e.g, "alegria") and a random image associated with the synset would be selected and returned.

The process of generating a question involves several steps (unless the template provided does not include a variable, in which case the template remains unaltered). This process goes as follows:

1. The ExerciseGenerator component asks *WordLookupImage* for all the root hyperonyms of the correct answer;
2. The first synset to contain a word that matches the topic is picked;
3. If no synsets contain the topic, a synset is randomly picked from the root hyperonyms list. Otherwise, a random word is picked from the synset that contains the topic;
4. The template is parsed in order to find the variable (a word surrounded by <% and >);
5. If a variable is found, it is replaced by the word picked in Step 3. Otherwise, the process ends here and the template is returned as the final question;

6. The resulting sentence is PoS tagged using TreeTagger;

7. The resulting tagged sentence ischecked for inconsistencies regarding number agreement between the variable's replacement and the rest of the sentence. If any inconsistency is found, it is fixed;

8. The same as Step 7, but regarding gender agreement.

Again, taking the previous example input, in Step 1, *WordImageLookup* would retrieve the root hyperonyms of the word "alegria". If any hyperonyms are found, the normalized topic ("emoção") is searched within the obtained hyperonyms. The first synset that contains the topic will be selected, and a random word will be selected from that synset (e.g. "emoção"). The variable in the template would then be replaced by "emoção", resulting in the sentence "Que emoção é este?". This sentence is PoS tagged in order to check the number and gender agreement. Since all the words are singular, no changes regarding number would be performed. However, there is a gender inconsistency: the noun "emoção" is feminine and the determinant "este" (this), which refers to "emoção", is masculine. Therefore, the gender of "este" will be fixed by replacing the final "e" with an "a".

Out of the three main tasks that compose the exercise generation process, the question generation is the most challenging, since it involves making sure a sentence is syntactically correct (which implies PoS tagging, inconsistency detection regarding gender and number agreement, and performing substitutions where needed). To detect issues regarding gender and number disagreement, we focused on a set of combinations likely to occur in multiple choice questions in Portuguese, namely: the variable is a noun preceded by a demonstrative determinant; the variable is a noun preceded by an article; the variable is a noun preceded by a demonstrative determinant and a verb; the variable is a noun preceded by an article and a verb; the variable is a noun preceded by a demonstrative determinant, a verb and an interrogative pronoun; the variable is a noun preceded by an article, a verb and an interrogative pronoun; the variable is a noun followed by a demonstrative determinant; the variable is a noun followed by a verb demonstrative determinant. When an inconsistency is detected, all the words that should agree in gender and number with the variable but fail this criterion suffer substitutions in order to match the variable's gender and number.

Regarding distractors, for each textual distractor to be generated, the process goes as follows:

1. If the distractor should belong to a given topic, ExerciseGenerator asks *WordImageLookup* for the list of all the leaf hyponyms of the topic. Otherwise, it asks for a list of all the leaf synsets;

2. A random synset is picked from the list obtained in Step 1;

3. If the synset contains the correct answer or any previously generated distractor for the current exercise, we go back to Step 2. In order to avoid an endless loop, once the number of iterations exceeds the size of the list, we pick the first synset that does not contain the correct answer nor any previously generated distractor. If no synsets match these criteria, an exception is thrown, and the process of generating the current distractor ends;

4. A random word is picked out the last picked synset.

This process repeats itself as many times as the number of distractors specified.

Considering the previous example input, since the distractors should belong to the topic "emoções", *WordLookupImage* is asked for the leaf hyponyms of the normalized topic ("emoção"). A random synset is selected out of the hyponyms list (e.g, "tristeza, angústia" – sadness, anguish). Since the selected synset does not contain the correct answer and no distractors have been generated so far, a random word (e.g., "tristeza") is picked. The remaining distractors are obtained through a similar process (which could return, for instance, "raiva" and "medo").

As for the process of generating image distractors, it is be similar to the one for word distractors, but here the selected synsets would have to include images. The data returned would then comprise, for example, an image of a sad person, of an angry person, and of a scared person.

6. Evaluation

In order to evaluate VITHEA-Kids, we performed two distinct procedures: one to assess the quality of the caregivers' experience while using their module, and another to assess the impact of using the child's module in teaching a new word to a child with ASD. For the evaluation of the caregiver's module, we designed an experiment consisting of five tasks and a questionnaire to assess the participants' satisfaction regarding the ease, effectiveness and efficiency of performing each task in our platform. We contacted several caregivers from diverse backgrounds (therapists, parents, etc.) and sent the application's URL, the task guide and the questionnaire URL to those who displayed interest in participating. The task guide comprised the following tasks to be performed in the caregiver's module: 1) Creation of a new multiple choice exercise; 2) Modification of a multimedia resource; 3) Creation of a new child user; 4) Customization of one of the animated character's utterances for a specific child; 5) Visualization of the reinforcement preferences for a specific child. The questionnaire featured the following sections: an introductory section to characterize the participants, by asking them what is their role as caregivers and whether their children have any impairment; a section where each task is evaluated in terms of how easy and fast it was to complete the task and of whether there were errors or not; a section where the overall user experience is evaluated through the classification of a set of affirmations; a section where the participant is asked about their interest in a potential module for the automatic generation of exercises, as well as about other suggestions that they might have.

Out of the contacts established, we obtained seven valid answers (i.e, seven complete participations) from

three therapists, three parents, zero teachers and educators and one participant who did not match any of these roles. In their role as caregivers, three participants deal with children with ASD, two participants deal with children with other impairments and three do not deal with children with special needs.

For each task, participants classified their agreement with three statements: "I was able to easily complete the task", "I was able to complete the task quickly" and "The task was completed without any error from the application" in a scale of 1 (*Totally disagree*) to 5 (*Totally agree*). Most participants agreed that the tasks were easy and fast to perform and were able to execute them without errors. However, the results differ according to the task. While Tasks 2 and 5 (Modification of an existing resource and Visualization of the reinforcement preferences, respectively) achieved full agreement, in Tasks 1 and 3 (Creation of a new exercise and Creation of a new child user, respectively) one participant reported errors. Also, in Task 4 (Customization of the character's utterances) there were two participants who did not fully agree with all the statements. In order to better understand the agreement classifications, we also asked the participants to describe their experience regarding each task. Aside from the errors reported by one of the participants, which were immediately fixed, Tasks 1 and 3 did not receive any comment, nor did Task 2. However, in Tasks 4 and 5, some participants found it hard to understand how to select a child, given the instruction in the guide. Participants were also confused with the use of the name of the animated character ("Catarina") in the application's interface.

Regarding overall user experience, participants classified the following set of statements in a scale of 1 (*Totally disagree*) to 5 (*Totally agree*): "I found the application was easy to browse"; "The discourse used in the application was easy to understand"; "The feedback given by the application (in error or confirmation messages) was clear"; "The kind of exercises allowed to be created are useful for the children I deal with"; "Using this application would allow me to save time in my daily routine"; "I would like to use this application again". All participants fully agree that the platform was easy to browse and makes use of a discourse easy to understand. However, some participants disagree concerning the clarity of the feedback given in error or confirmation messages. Furthermore, one participant considers that the kind of exercises supported is not useful for their children and there is also one participant that would not like to use this application again.

We also asked the participants about their interest in the possibility of having exercises automatically created by the platform (while keeping the possibility of manually creating them), using a scale of 1 (*Not interested at all*) to 5 (*Very interested*). Participants are either interested or very interested in the possibility of having exercises automatically generated by the platform).

Finally, we asked the participants an open question in order to get general feedback and further suggestions. These included the possibility of uploading other kinds of multimedia content, such as videos, and new exercises/activities, in order to train different skills (e.g, maths).

Given that the amount of participations (seven) is not enough to have statistical relevance, we do not provide further statistical treatment of these results.

As for the evaluation of the child's module is currently taking place. For this evaluation, we chose to follow a research design called **single-subject design**. In this design, the subject serves as their own comparison term, in opposition to group designs, in which there are two or more groups, and one of them serves as term of comparison (control group). Single-subject design is used in the majority of the studies featuring subjects diagnosed with ASD or other impairments [17, 11], since each individual with ASD has a unique set of symptoms, characteristics and needs, making it difficult to generalize results.

Single-subject design is based on the principle that, if an intervention is effective, a change in status from the period prior to intervention to the period during and post-intervention should be noticeable [7]. The three main components of this design are: **Repeated measurements** of the subject's progress regarding the target behaviour; **Baseline phase** (the period prior to the treatment. The measurements collected in this phase will serve as comparison term to those collected during and after the execution of the treatment. Baseline is represented by the letter **A**); **Intervention phase** (the period during which the treatment is applied. It should last as long as the baseline phase, in order to facilitate the comparison against baseline measured values. This phase is represented by the letter **B**).

We chose to follow an **A-B-A** design, which comprises three phases – baseline, intervention and follow-up. In our evaluation, we aim to assess whether the prompting and custom reinforcement features of our application, as well as the possibility of creating exercises with a variable number of distractors, are effective in teaching a new word to a child with ASD or not. The application is being tested with a male seven year old child diagnosed with ASD who is in the process of learning how to read and write, with the help of a therapist. The target word is "Javali". The sessions are taking place in the child's classroom, and are expected to last about fifteen days.

The first phase of the evaluation, the **baseline**, focused on accessing whether the child was able to select the word "Javali" (boar) when an image of a boar shows up, without prompting or reinforcement. As expected, the child never selected the correct answer ("Javali"), so this phase ended after four sessions, which is the minimum number of sessions required to consider that a stable pattern was observed. In the **intervention** phase, the child will have to solve exercises that will start without distractors and progressively increment the number of distractors. In this phase, prompting will be activated in case the child picks a distractor instead of the correct answer. Selecting the correct answer on the first try of a give exercise will lead to visual and audio reinforcement, while selecting it after having selected any distractor will only lead to the same image stimulus, followed by the correct answer. This phase will only end when the child selects the correct answer during two consecutive ses-

sions, without prompting. Finally, in the **follow-up** phase, prompting and reinforcement will be deactivated again in order to assess whether the child has learned the target word or not.

Finally, the evaluation of our module for automatic generation of exercises consisted solely in the assessment of the quality of the generated distractors considering a question about a certain topic and the correct answer to that question, given that, in the context of VITHEA-Kids, the distractors are a key component in learning how to generalize a concept. For this purpose, we created an interface for the exercise generation module and set up the following resources: an Extensible Mark-up Language (XML) file comprising three root synsets:"animal" (animal), "alimento" (food) and "transporte" (transportation), where each synset had twelve, twelve and seven hyponym synsets each, respectively; a set of twenty-three text files, each containing one or more image URLs related to a certain hyponym synset (which means six synsets did not have an URL file associated).

We asked thirty people to participate in the assessment process. Participants were mostly between 20 and 30 years old, resided in Portugal, their academic qualifications range from high school to master's degree and were currently university students and/or researchers. All of them have a background in computer science and engineering, although there were participants who also had a background in electro-technical engineering, law, and arts/multimedia. The generated questions, answers and distractors, as well as the responses given by the participants were saved into Comma-separated Values (CSV) files, to facilitate their processing. The evaluation comprised four steps: 1) An initial screen with the context and the instructions regarding the evaluation process; 2) A screen to evaluate the quality of the word distractors in the context of an exercise where the stimulus would be an image and the answers would be textual; 3) a screen to evaluate the quality of the image distractors in the context of an exercise where the stimulus would be a word and the answers would be images; 4) An open question for any suggestions and observations from the participants.

In Steps 2 and 3, the input is a topic (e.g.,"Animals"), a question about that topic (e.g., "What is the name of this animal?" or "Which image refers to this animal", respectively) and the correct answer (e.g., the word "cat" or an image of a cat, respectively). In each step, once the participant clicks on a button to generate the distractors, two sets of three distractors each are displayed: the first set should only be composed of distractors related to the given topic, while the second set might include distractors about any topic (including the topic taken as input). For each distractor, the participants should indicate whether they think the distractor makes sense in the context of the given input or not (i.e., if it is possible to solve an exercise composed of the given question and a set of answers containing the given correct answer and the distractor being evaluated). Participants should also rate each distractor in terms of how adequate they find it, in a scale of 1 (Poor) to 5 (Very good).

Regarding the generation of word distractors, out of the 180 generated distractors (6 per participant), 143 (79.44%) were marked as making sense. In what respects to quality, the most frequent rating was 5 (very good), given to 83 distractors (46.11%). The average rating was 3.75 (with a std dev of 1.41). As for the generation of image distractors, 138 out of the 180 generated distractors (76.67%) were marked as making sense. In what respects to quality, the majority of the distractors (91, i.e., 50.56%) was rated with 5 (very good). The average rating was 3.82 (with a std dev of 1.44).

Splitting the results for distractors that should be related to the topic and distractors that did not have a topic constraint, we notice a significant difference in the results for each of these categories, regarding both word and image distractors: 92.22% of the word and image distractors generated under a topic constraint were marked as making sense, and the majority of the word and image distractors were rated as very good (57.78% and 61.11%, respectively). The average ratings were 4.26 (with a std dev of 1.03) for word distractors and 4.24 (with a std dev of 1.11) for image distractors. As for the distractors with no topic contraint, the percentage of distractors marked as making sense drops to 66.67% (words) and 61.11% (images), and there is a greater balance among each quality rating's percentage. The average ratings were 3.24 (std dev = 1.55) for word distractors and 3.39 (std dev = 1.59) for image distractors.

Regarding suggestions and final observations, two participants suggested that, in the sets of distractors that should be related to the topic, there should be a stronger relation between the correct answer and each distractor (e.g, if "parrot" is the correct answer, the distractors should also be birds, for instance). Three participants suggested that, in the sets of distractors that did not have a topic constraint, there still should be some kind of semantic connection between the correct answer and each distractor (e.g., if the correct answer is "horse", a possible distractor could be "stable"). One participant reported that it was difficult to rate the quality of the distractors without information about the context in which such exercises would be used.

During the evaluation sessions, we were able to notice some difficulties in interpreting the aim of the study. For instance, some participants reported that they felt confused about the aim of the question about the adequacy of the distractor, which might have led to different interpretations of what it means for a distractor to make sense. In fact, given that we evaluated our generation module with a very restricted set, it is expectable that most generated distractors correctly satisfy the topic constraint and do not collide with the correct answer (we can confirm this expectation by inspecting the generated data). Another aspect that we noticed was that different participants had a completely different notion of what it could make an interesting distractor without topic contraints: some participants would penalize the distractors that belonged to the given topic, while others would penalize the distractors that were too unrelated to the topic or the correct answer. This might be explain the differences observed between the results for distractors under

a topic constraint and for the distractors without a topic contraint.

7. Conclusions and Future Work

Motivated by the difficulties that individuals diagnosed with Autism Spectrum Disorder (ASD) face in social and communication contexts, several authors have performed studies and developed applications that either aim to provide an alternative mean of communication or a way to develop language skills. With the increasing popularity of mobile devices, many applications of such kind have been made available in mobile application markets, some of which for free. Nonetheless, there are still difficulties and needs regarding the use of software for individuals with ASD that are not yet addressed. In particular, there is a desire for an individualized user experience, as well as for the accounting of each user's characteristics, that is not fulfilled by currently available options yet. Additionally, there is a lack of applications that allow to develop language skills regarding the Portuguese language.

Given this context, we developed a platform that aims to meet the needs of children with ASD and their caregivers, specially those concerning customization. Our platform, VITHEA-Kids, is composed of two modules: a mobile application for children to solve multiple choice exercises that allow them to acquire new vocabulary and to develop skills such as the ability to generalize concepts, and a web application for the children's caregivers to prepare those exercises and customize several interaction aspects in the child's application.

Additionally, in order to ease the task of creating new exercises, we developed a module to automatically generate two variants of multiple choice exercises given a set of parameters, such as the topic of the exercise and a question template. We developed this module independently so it can be integrated either in VITHEA-Kids or any other project. The words and images used in the generated exercises are obtained through a second module, which we also implemented independently so that it can be used in other contexts.

Finally, we evaluated each part of our contribution: regarding VITHEA-Kids, we asked a set of caregivers to perform some tasks on the caregiver's module and reply to a questionnaire in which they could rate their experience and add further suggestions; the majority of them agreed that the application was easy and fast to use, and would like to use the application again. We are also currently evaluating the child's module with a child with ASD. As for the exercise generation module, we set up a questionnaire in which participants would have to rate the quality of the generated distractors given a question about a certain topic and the correct answer to that question. In this process, a total of 180 word distractors and 180 image distractors were generated. Out of the word distractors, 79.44% were marked as making sense, and the average quality rating was 3.75 (std dev = 1.41). As for the image distractors, 76.67% were marked as making sense and the average quality rating was 3.82 (std dev = 1.44).

In what concerns to future work, on short term, one of the goals should be the integration of the exercise generation module in VITHEA-Kids. Yet, both modules have the potential to be extended with new features.

Regarding VITHEA-Kids, there are some features that could be added that were emphasized by therapists as being relevant, namely: the possibility of uploading other kinds multimedia resources; the possibility of using animations as reinforcement; the possibility of sorting the existing exercises in order to control the order in which they appear in the child's application; the customization of the way prompting is done. Before moving to more ambitious features, it would also be desirable to run a new evaluation process for both applications, featuring more participants. In a medium term, VITHEA-Kids could be extended with new kinds of exercises and activities, such as the ones suggested by the caregivers that evaluated the platform, as well as exercises that make use of other kinds of input (e.g, speech or text introduction).

In what regards the automatic generation of exercises, it would be desirable to explore different approaches regarding the generation of questions in addition to the one we currently follow. It would also be important to run a new evaluation featuring the generation of correct answers and questions. Concerning the word and image extraction module, in a short term it should provide a larger set of methods, namely methods to get hyperonyms and hyponyms from different levels of the resource tree, or methods to extract words given other relations between words. In a medium/long term, it would be interesting to add new features to the exercise generation module, namely the generation of paraphrases given a question. This feature could have a positive impact in applications such as VITHEA-Kids, as it could contribute to improve the child's ability to generalize different formulations of the same question.

References

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Association Arlington, VA, 5th ed. edition, 2013.

[2] H. Asperger. 'Autistic psychopathy' in childhood. In U. Frith, editor, *Autism and Asperger Syndrome*. Cambridge University Press, 1991.

[3] S. Chakrabarti and E. Fombonne. Pervasive Developmental Disorders in Preschool Children: Confirmation of High Prevalence. *American Journal of Psychiatry*, 162(6):1133–1141, 2005.

[4] Conselho Nacional de Educação. Estado da Educação 2013. Technical report, 2013.

[5] Conselho Nacional de Educação. Políticas Públicas de Educação Especial. Technical report, 2014.

[6] M. Elsabbagh, G. Divan, Y.-J. Koh, Y. S. Kim, S. Kauchali, C. Marcín, C. Montiel-Nava, V. Patel, C. S. Paula, C. Wang, M. T. Yasamy, and E. Fombonne. Global prevalence of autism and other pervasive developmental disorders. *Autism research : official journal of the International Society for Autism Research*, 5(3):160–79, June 2012.

[7] R. J. Engel and R. K. Schutt. Single-subject design. In *The Practice of Research in Social Work*, chapter 7, pages 206–246. SAGE, 2009.

[8] R. Gaudêncio. Colégios de ensino especial reabrem mas ainda sem dinheiro. in *Público*, Jan. 2015.

[9] I. Ghiciuc. Smartphone and tablet usage patters: The how's, the why's and the where's. in *Thin Slices*, Dec. 2014.

[10] M. Heimann, K. E. Nelson, T. Tjus, and C. Gillberg. Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of autism and developmental disorders*, 25(5):459–80, Oct. 1995.

[11] O. E. Hetzroni and J. Tannous. Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of autism and developmental disorders*, 34(2):95–113, Apr. 2004.

[12] IDC Portugal. Tablets vão continuar a impulsionar a procura de equipamentos inteligentes conectados em 2014 em Portugal. in *IDC Notícias*, Mar. 2014.

[13] L. Kanner. Autistic Disturbances of Affective Contact. *Acta Paedopsychiatr*, 35(4):100–136, 1943.

[14] S. Klai, T. Kolokolnikov, and N. Van Den Bergh. Using Maple and the Web to grade mathematics tests. *Proceedings - International Workshop on Advanced Learning Technologies: Advanced Learning Technology: Design and Development Issues, IWALT 2000*, pages 89–92, 2000.

[15] J. P. Leal and N. Moreira. Using matching for automatic assessment in computer science learning environments.

[16] J. F. Lehman. A Feature-based Comparison of Software Preferences in Typically-developing Children versus Children with Autism Spectrum Disorders, 1998.

[17] D. W. Massaro and A. Bosseler. Read my lips: The importance of the face in a computer-animated tutor for vocabulary learning by children with autism. *Autism : the international journal of research and practice*, 10(5):495–510, Sept. 2006.

[18] E. Melis, E. Andres, J. Büdenbender, A. Frischauf, G. Goguadze, P. Libbrecht, M. Pollet, and C. Ullrich. ActiveMath: A generic and adaptive web-based learning environment. *International Journal of Artificial Intelligence in Education*, 12(4):385–407, 2001.

[19] M. Moore and S. Calvert. Brief report: vocabulary acquisition for children with autism: teacher or computer instruction. *Journal of autism and developmental disorders*, 30(4):359–62, Aug. 2000.

[20] N. Naoi. Intervention and Treatment Methods for Children with Autism Spectrum Disorders. In *Applied Behavior Analysis for Children with Autism Spectrum Disorders*, chapter 4, pages 67–81. 2009.

[21] A. M. Pompili. Virtual therapist for aphasia treatment. Master's thesis, Instituto Superior Técnico, 2011.

[22] A. M. Pompili. New features for on-line aphasia therapy. Master's thesis, Instituto Superior Técnico, 2013.

[23] C. Putnam and L. Chong. Software and technologies designed for people with autism: What do users want? In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, pages 3–10, New York, NY, USA, 2008. ACM.

[24] J. E. Ringdahl, T. Kopelman, and T. S. Falcomata. Applied Behavior Analysis and Its Application to Autism and Autism Related Disorders. In *Applied Behavior Analysis for Children with Autism Spectrum Disorders*, chapter 2, pages 15–32. 2009.

[25] C. J. Sangwin. New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. *International Journal of Mathematical Education in Science and Technology*, 34(6):813–829, 2003.

[26] A. M. Santos, P. Santos, F. M. Dionísio, and P. Duarte. An experiment using the system CAL for generating multiple choice questions. *Matrix*.

[27] H. Schmid. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49, 1994.

[28] B. F. Skinner. *Science and Human Behavior*. Free Press paperback. Macmillan, 1953.

[29] C. Whalen, D. Moss, and J. Symon. TeachTown and Los Angeles Unified School District: NCTI Technology in the Works Final Report. Technical report, 2009.