# Automatic Question Generation for the Portuguese Language

Bernardo José Coelho Leite

**U.**PORTO

FEUP **FACULDADE DE ENGENHARIA**
UNIVERSIDADE DO PORTO

# Automatic Question Generation for the Portuguese Language

**Bernardo José Coelho Leite**

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: Prof. João Pascoal Faria
External Examiner: Prof. Luísa Coheur
Supervisor: Prof. Henrique Lopes Cardoso
Co-supervisor: Prof. Luís Paulo Reis

July 29, 2020

# Abstract

Artificial Intelligence (AI) is changing the world, either the way we live, work, or learn. Education is not left out. There is a growth in the use of AI in teaching and learning from primary to higher education. Educational tools provided by AI demonstrate their potential to improve the quality of teaching and traditional learning methods.

The act of questioning is one of the most used traditional methods by teachers to assess student's knowledge or understanding. However, it proves to be a very time-consuming task. In addition, taking into account the difficulty in creating new and rich questions, these are often repeated, already used in previous contexts.

Hence, our research focuses on means to reduce the excessive time spent in elaborating questions and optimize this process. Thus, our goal is to automatically generate questions from Portuguese educational texts and to enrich/refine ways to control the difficulty of the generated questions. We make use of AI and Natural Language Processing (NLP) techniques in order to achieve our goals.

Our methodology is divided into three distinct approaches: (A) Grammar, (B) Reading Comprehension and (C) Pronoun Reference questions. The first approach (A) is capable of generating grammar questions. For that, we follow a rule-based technique, for which we have established well-defined rules, according to the Portuguese grammar. The second approach (B) aims to generate reading comprehension (factual) questions and, for that, we tested five different methods. The first one performs a syntax-based analysis by using the information obtained from Part-of-speech tagging and Named Entity Recognition. The second carries out a semantic analysis of the sentences, through Semantic Role Labeling. The third method extracts the inherent dependencies within sentences using Dependency Parsing. The fourth takes advantage of the relative pronouns and adverbs found in the sentences. The fifth explores the usefulness and practicality of discourse connectors. Finally, for the last approach (C), we create pronoun reference questions, in which we do not only generate our questions but also the text excerpts they are generated from. We define heuristic functions that assign difficulty values for each question.

To evaluate the developed system, we have elaborated a survey that was answered by Portuguese teachers. For that, we use concrete metrics regarding question formulation, objectivity, relevance and difficulty. Then, we applied a questionnaire in a teaching environment, with students. Here, we draw conclusions about the practical utility that our questions have in a real context, when used with pupils. Overall, teachers considered the questions to be well formulated, objective and confirmed their relevance. In turn, students scored lower on the most difficult grammar exercises, which demonstrates the ability of our application to generate questions with different difficulty levels. In conclusion, the results verify the potential of our approaches, opening up the possibility to use them in a teaching environment.

**Keywords**: Artificial Intelligence, Natural Language Processing, Automatic Question Generation, Item Difficulty Control

ii

# Resumo

A Inteligência Artificial (IA) está a mudar o mundo, seja da forma como vivemos, trabalhamos ou até aprendemos. A área da Educação não fica de parte. Os últimos anos revelam uma crescente emergência na utilização da IA para a aprendizagem, desde o ensino básico ao ensino superior. As ferramentas educacionais fornecidas pela IA demonstram potencial em melhorar a qualidade do ensino e aprimorar os métodos tradicionais de aprendizagem.

A elaboração de questões é um dos métodos tradicionais mais utilizados pelos professores para avaliar o conhecimento e/ou compreensão de um aluno. No entanto, revela ser uma tarefa bastante demorada. Para além disso, tendo em conta a dificuldade em criar novas questões, estas são muitas vezes repetidas, já usadas em contextos anteriores.

Consequentemente, a nossa pesquisa foca-se em encontrar meios que permitam reduzir o tempo dispendido na elaboração de questões e na otimização desse processo. Desta forma, o nosso objetivo é a geração automática de questões a partir de textos educativos Portugueses e enriquecer/refinar formas de controlar a dificuldade dessas mesmas questões. A metodologia seguida utiliza técnicas de IA e de Processamento de Linguagem Natural para alcançar estes objetivos.

A nossa metodologia divide-se em três abordagens distintas: (A) Gramática, (B) Compreensão de Leitura e (C) Questões de Referenciação de Pronomes. A primeira abordagem (A) permite gerar questões gramaticais. Para isso, seguimos uma técnica baseada em regras, para a qual foram estabelecidas regras bem definidas, de acordo com a gramática portuguesa. A segunda abordagem (B) visa gerar questões de compreensão de leitura (factuais) e, para tal, foram testados cinco métodos diferentes. O primeiro faz uma análise baseada em sintaxe, utilizando as informações obtidas através do analisador morfológico e do reconhecimento de entidades mencionadas. O segundo elabora uma análise semântica das frases, através da rotulagem dos papéis semânticos. O terceiro método extrai as relações de dependências inerentes às frases, através da Análise de Dependências. O quarto tira proveito dos pronomes e advérbios relativos encontrados nas frases. O quinto explora a utilidade dos conectores de discurso. Finalmente, para a última abordagem (C), foram criadas perguntas sobre referenciação de pronomes, nas quais, para além das perguntas geradas, são também geradas as passagens de texto a partir das quais surgem as questões. Foram estabelecidas funções heurísticas que atribuem valores de dificuldade para cada uma das perguntas.

De forma a avaliar o sistema desenvolvido, elaboramos um inquérito que foi respondido por professores da disciplina de Português. Para tal, foram utilizadas métricas em relação à formulação, objetividade, relevância e dificuldade das perguntas. De seguida, aplicamos um questionário em ambiente escolar, com os alunos. Aqui tiramos conclusões sobre a utilidade prática que as nossas perguntas têm num contexto real, quando aplicadas aos alunos. De uma forma geral, os professores consideraram as perguntas bem formuladas, objetivas e confirmaram a sua relevância. Por sua vez, os alunos obtiveram uma menor pontuação nos exercícios gramaticais mais difíceis, o que demonstra a capacidade do sistema para gerar perguntas com diferentes níveis de dificuldade. Concluindo, os resultados confirmam o potencial das nossas abordagens, abrindo a possibilidade de as utilizar num contexto educativo.

iv

# Acknowledgements

*"Education is the most powerful weapon which you can use to change the world."*


Nelson Mandela

# Contents

# List of Figures

# List of Tables

# Abbreviations

AI         Artificial Intelligence
AQG     Automatic Question Generation
CFG      Context Free Grammar
CTI       Content Technology Inc.
CVIT     Captivating Virtual Information for Training
FLTL     Foreign Language Teaching and Learning
IELTS    International English Language Testing System
IRT       Item Response Theory
IT         Information Technology
ITS       Intelligent Tutoring Systems
MOOC    Massive open online course
NCES    National Center for Education Statistics
NLG      Natural Language Generation
NLP      Natural Language Processing
NLU      Natural Language Understanding
NQG     Neural Question Generation
POS      Part of Speech Tagging
QG        Question Generation
RC        Reading Comprehension
SAT       Scholastic Aptitude Test
SRL       Semantic Role Labeling
TOEFL   Test of English as a Foreign Language
TOEIC    Test for English International Communication

# Chapter 1

# Introduction

Education has an impact on all areas of our lives. It is a fundamental right that helps not only in the development of a country, but also helps each individual. It is through Education that we learn to prepare for life. Moreover, through education we guarantee our social, economic and cultural development.

The impacts of Education are extensive and profound. Firstly, through education it is possible to combat poverty. In fact, the more people study, the more opportunities they will have in the job market. Secondly, Education makes the economy grow because good education improves a country's economy.

Education also promotes health. A parent who has had access to quality education is better able to take care of their children's health, as they are more sensitive to the importance of prevention, vaccination and hygiene habits, in addition to knowing how to seek treatment when necessary. UNESCO's Global Education Monitoring Report[1] showed that a child whose mother can read is 50% more likely to survive after the age of five. The OECD study *"What Are the Social Benefits of Education?"* [105] has shown that personal satisfaction among people who studied in higher education is greater than the satisfaction of people who stopped at high school. There are many other positive impacts that education provides such as violence decrease, environment protection (educates citizens to be more aware of the impact of their actions on the environment). Finally, it provides citizenship helping society to fulfill its civic duties.

Unfortunately, there are many problems and gaps related to Education, starting from poverty to family factors. Student's attitudes and behaviors also play a role in many schools such as apathy and disrespect for teachers, being a major problem faced by schools today. A poll from the National Center for Education Statistics[2] has cited that problems like apathy, tardiness, disrespect and absenteeism posed significant challenges for teachers. Another important challenge is the integration of technology in the current Education system.

---

[1] https://unesdoc.unesco.org/ark:/48223/pf0000245752
[2] https://nces.ed.gov/

The stimulus for this dissertation comes from one of the most major challenges that Education faces today: The effective use of Educational technology in teaching. Some of the questions that arise are:

> *What are the challenges of Education when it comes to integration with new technologies? How can technology improve the area of Education?*

In order to keep up with the learning demands of the 21st century, technology needs to come into the classroom. There is still a lot of work ahead since educational technology is still not being applied sufficiently. Today, more than ever, the role of educational technology in teaching has great importance mainly because of the use of information and communication technologies (ICT) [134]. Moreover, there are various applications for students and teachers such as distance education, instructional games and intelligent tutoring systems (ITS), and so on. Stosic [134] has defined education technology as a systematic and organized process of applying modern technology to improve the quality of education, regarding its efficiency in the teaching and learning processes. Besides, Stosic highlighted three domains of use: technology as a tutor (computer gives instructions and guides the user), technology as a teaching tool and technology as a learning tool.

The theme of this dissertation is related with these last two points, since certain techniques and technologies from Artificial Intelligence can be used in order to contribute to the educational process. Concretely, this work has focused on the development of a tool capable of generating questions from educational texts. In the next sections we will explain what is our motivation for conducting a study in this topic and then highlight the importance of asking questions in the educational context. Lastly, we focus on our objectives for this project.

## 1.1   Motivation

Nowadays, generally speaking, there are certain obstacles that traditional education systems have been encountering. The affected population are all those who are directly or indirectly involved in the field of education, such as students, parents, teachers or even education assistants. The obstacles that motivated us to work on this topic refer specifically to the problems that affect the students' educational process. In addition, our motivation also comes from the desire to streamline the teaching process from teachers to their respective students. One of the most used means to assess the students knowledge is by asking questions; by analysing their given answers it will be possible to know what are their main difficulties. Our desire to intervene begins precisely on this route, because we encounter specific problems, both on the teacher and student side, and they can be summarized this way:

- Student

    - **Lack of support** in the individual learning process: Despite the online tools that are available today, it is hard for a student to find specialized support according to their own difficulties. If there was a tool capable of creating questions specifically addressing their difficulties, it would be a good improvement.

– **Poor feedback and scarcity of questions**: There are already some tools that allow the student to obtain automatic feedback (e.g. Escola Virtual[3]) as well as other websites for different contents (e.g. Khan Academy[4]). Unfortunately, this type of feedback usually appears only when a student's answer is right or wrong and does not help to achieve the correct one. Furthermore, these systems normally have a limited number of questions from their respective databases. We believe that through the automatic generation of those questions, it may be possible to address these shortcomings.

- Teacher

  – **Excessive time** in elaborating questions: Teachers spend a large part of their time preparing questions for tests and other assessments. In other cases, they tend to reuse tests that have already been published. Automatic question generation can provide great assistance here since teachers would be able to, for example, pick up a particular news item and generate specific questions about it.
  – **Lack of resources** to accompany each student individually: Regardless of the number of students per class, it is still complicated to follow each individual student. Through an intelligent tool it might be possible to generate questions with different difficulty levels and assist effectively the teacher in that teaching process.

Indeed, there are already some intelligent tools that address the shortcomings stated previously (including some for the Portuguese language). Their main features consist in a set of questions that are divided by teaching modules. Then, the student only needs to connect to these applications and answer the provided questions. We want to reinforce that some limitations start right here. For example, the student is totally dependent on the questions provided by this kind of applications or tutors. The number of questions is often limited and the variety of question types is also scarce.

But after all, what is the real importance of asking questions? The fact that questions help in the educational process has been supported by research for some time now [11, 124]. Furthermore, we can see more recent studies that support the fact that this is a very important teaching principle. After an in-depth work on which educational strategies work (and do not work), Roediger and Pyc [122] organized their findings by enumerating three points: (1) the distribution of material and practice during learning, (2) the frequent assessment of learning, and (3) explanatory questioning. It is noticeable that all these points are directly or indirectly linked to the importance of questioning. The IES Practice Guide for Teachers has studied seven evidences based on the principles of education, two of which refer explicitly to questioning [111]. Bloom [16] has proposed a taxonomy that defines and describes learning levels categorized into six different cognitive domains. This is the current revised version of that taxonomy:

1. **Remembering**: List, recite, outline, define, name, match, quote, recall, identify, label, recognize;

---

[3]https://www.escolavirtual.pt/
[4]https://pt-pt.khanacademy.org/

2. **Understanding**: Describe, explain, paraphrase, restate, give original examples of, summarize, contrast, interpret, discuss;

3. **Applying**: Calculate, predict, apply, solve, illustrate, use, demonstrate, determine, model, perform, present;

4. **Analysing**: Classify, break down, categorize, analyze, diagram, illustrate, criticize, simplify, associate;

5. **Evaluating**: Choose, support, relate, determine, defend, judge, grade, compare, contrast, argue, justify, support, convince, select, evaluate;

6. **Creating**: Design, formulate, build, invent, create, compose, generate, derive, modify, develop.

These levels are arranged in hierarchical order, moving from the lowest to the highest level of understanding.

According to Pandey and Rajeswari [108] one of the advantages regarding question generation is based on Bloom's taxonomy, because it helps to assess the learning ability of the students. For example, if the tutor wants to assess the student's learning about some topic located in the first basic level, the questions to be asked should be taken from the remembering level. In fact, Bloom suggests that learners need to have knowledge even before understanding it. On the other hand, they need to understand the knowledge even before applying it in different contexts. Students need to be able to handle "lower order" skills such as knowledge, understanding, and application, before "higher" levels like analysing and creating.

Asking higher cognitive questions requires higher cognitive responses from students. Therefore, this dissertation aims to comprise (through the generated questions) the largest possible number of the listed educational objectives. The possible benefits of these objectives can be applied to both closed and open questions. Some of these benefits: Development of critical thinking, Motivation, and engagement; Progress (teachers can see progress over time). Finally, this progress over the hierarchy of educational goals helps pupils to move towards independence. Our ambition is to take advantage of the relevance of asking questions in order to implement a useful and applicable tool in a real context.

## 1.2   Problem and Research Questions

Lately, the need to automate certain processes that optimize the educational process has been growing. For example, regarding the Portuguese Language course (mother tongue), the traditional way for creating questions to assess student's knowledge is by choosing an appropriate text and pose questions accordingly. Other times, the teacher may take a test model and manually adapt it according to certain needs. We believe this is a clear example of an educational process that can be optimized. Moreover, these problems are affecting both teachers and students. In fact, different

students have different difficulties and if questions are the same for everyone it becomes difficult to cover the real difficulties for each student. For this reason, the importance of generating questions with different difficulties also emerges as an important need.

To improve the issues that were previously mentioned, we set out ourselves these research questions which all concern the Portuguese language:

1. *How can we reduce the time spent on generating questions?*

2. *How can we automatically generate grammar questions?*

3. *How can we automatically generate reading comprehension questions?*

4. *How can we automatically generate pronoun reference questions?*

5. *How can we control the difficulty levels of automatically generated questions?*

To answer the first question, we hypothesize that one possible way to mitigate this problem is to follow a methodology capable of generating questions automatically, given a text. For questions 2, 3 and 4 we need to follow one or more methodologies that allow us to generate those referred question-types. We plan to follow a different methodology for each question type. For the last question, we intend to apply heuristic functions to define the difficulty of the questions. These functions have different factors depending on the question type.

## 1.3  Goals

This dissertation covers these research areas: Education, Artificial Intelligence and Natural Language Processing. This proposal aims to materialize the application of these concepts by implementing automatic generation of questions. The main goal is to automatically generate questions from Portuguese educational texts and also ensure that these questions would have different difficulty levels. As follows, we can concretely divide the goals of this project into three main points:

- **Use of AI and NLP techniques** for bringing improvement in educational setting and academic performance: This is achieved by creating effective approaches for teachers and students, by providing assistance focusing on Portuguese subject (main focus on the 3rd cycle of studies);

- **Improve the question generation process** composed traditionally with multiple-choice and factual questions: Deepen and improve study in the field of Item Response Theory (IRT);

- **Enrich and refine ways to control the difficulty of the generated questions**.

All of these steps goals consummated in a final tool that will generate questions receiving as input a certain text. These questions have different difficulties. We apply modern and relevant technologies/tools/frameworks in order to solve all the challenges inherent to the project.

## 1.4   Main Contributions

We can summarize our main contributions as follows:

- **Methodologies** to generate questions, in Portuguese, providing **pertinent**, **well-formed** and **relevant** questions;

- **Methodologies** to generate the questions with different **difficulty levels**;

- **A tool**, with a graphical interface, that allows the user to generate questions (reading comprehension and grammar), from the text.

This dissertation has its basis from two studies that we have previously developed. The first one has been published and regards Question Generation and Question Answering for English [12]. The second [76] in under publication, and it is based on a study that regards Factual Question Generation for Portuguese.

## 1.5   Dissertation Structure

While in this chapter we started by elucidating the importance of Education and its relation to educational technologies, in (Chapter 2) we reinforce how AI is making an impact in the world of pedagogy. We also look on what is the applicability of ITS and then it is presented our main topic, question generation. In Chapter 3 there is a deep bibliographic review regarding our study. It is started by a general review of all the previous work over the past few years, pointing out what are the domains, question types and the language used in Automatic Question Generation (AQG) systems. Also, traditional approaches regarding AQG will be described. There is a section dedicated to the new paradigm of this area: Neural Question Generation. Controlling Item difficult and Feedback Generation will also be analysed. In Chapter 4 we explain and detail all the methodologies followed throughout our research. Our methodology is divided into three distinct approaches: (A) Grammar, (B) Reading Comprehension and (C) Pronoun Reference questions. While (A) is capable of generating multiple-choice questions, (B) aims to generate factual questions. For (C) we do not only generate our questions (multiple-choice) but also the text excerpts they are generated from. Also in this chapter, we describe what we did to control the difficulty of the questions and how we improved them by performing post-processing. In Chapter 5 we perform Evaluation by applying a survey and questionnaire for teachers and students, respectively. Finally, in Chapter 6 we will provide a retrospective regarding our work as well peeking a brief look into the future.

# Chapter 2

# AI in Education

Although not yet standard in schools, AI is now part of our daily lives. We are completely surrounded with several technologies from automatic parking systems to personal assistance. Likewise, AI in education has been gaining prominence over the world, and traditional methods are changing drastically. From administrative tasks to smart content creation, there are new possibilities. Also, the prospect of personalized learning is quickly becoming a reality. While AI in education might appear threatening for some people, the benefits are too great and can't be ignored. In this line of thought, we are now going to explain how AI is changing the world of pedagogy (Section 2.1) and then we'll analyse intelligent tutoring systems (Section 2.2), which have a lot to do with our work. Following ITS, we formally describe the question generation task (Section 2.3).

## 2.1 The impact of AI on Education

The first question is always the same: *What is AI?* There are several possible definitions from different authors, but in general, it can be said that AI is a field in computer science that deals with the intelligent behavior of machines. This is possible though specific algorithms that make AI applicable in a specified scope of activities. There are a lot of different scopes when using AI, for example, in organizations, security frameworks, energy and natural resource management. Even though AI advancement levels and usage may differ substantially according to each region, there are indicators pointing to the fact that more people are acknowledging solutions that these new technologies bring to us[1]. In this dissertation, our focus is to understand the importance of AI in Education and then apply specific techniques in this context to obtain a useful tool for teachers and students. So this brings us to our next question: *What is the real importance of AI in the educational context?*

---

[1]https://www.forbes.com/sites/cognitiveworld/2019/02/24/ai-transforming-the-world

There are several educational tools provided by AI that have recently attracted much attention for their potential to improve teaching quality. Specifically, the use of AI in Education can be summarized as follows:

- **New efficiencies**: IT processes are constantly improved by AI and that unleashes new efficiencies. Examples include the modeling of complex data to make forecasts, saving costs to schools by avoiding a lot of wastages caused by over-ordering (materials or food);

- **Simplifying Administrative Tasks**: AI can perform several routine tasks such as taking attendance, grading assignments and generating test questions [88]. This means that professors can spend more time with their students rather than spending long hours in the grading task. Several software providers are coming up with better means of grading written answers or normal essays. School admission boards is another department that is gaining a lot from AI as it allows automation of classification and paperwork processing.

- **Smart Content creation** [49]: This technology has already reached a classroom setting. It attempts to condense text books into a useful tool for test/exam preparation. "Smart content" creation goes from digitized guides of textbooks to customizable learning digital interfaces. Content Technologies Inc. (CTI)[2] is an artificial intelligence research and development company specializing in production automation, business process automation, instructional design and content application solutions. It has created a suite of smart content services for secondary education and beyond such as *Cram101* that uses AI do breakdown textbook content into "smart" study guides with chapter summaries, true-false, multiple choice tests and flashcards. *JustTheFacts101* is a similiar tool but has a more streamlined purpose – creating and highlighting text and chapter-specific summaries (they are archived and can be made available on Amazon). There are more companies which create smart digital platforms. *Netex Learning*[3], for example, allows teachers to design curriculum in a digital way and use several content across devices, integrating rich media like audio and video.

- **Virtual Facilitators and Learning Environments** [49]: Despite this is not yet a reality, the idea is to use virtual human guides and facilitators for use in a variety of educational environments. The ultimate goal is to create virtual human-like characters and configure them to act, react and interact through verbal and nonverbal communication. *Captivating Virtual Instruction for Training (CVIT)*[4] is a distributed learning strategy that integrates live classroom methods with virtual technologies such as virtual facilitators, augmented reality in the context of remote learning.

- **Universal access for all students**: AI tools can help make global classrooms which are available to all students, including those who have visual or hearing issues and those who speak different languages. *Microsoft Translator*[5] is a multilingual machine translation cloud

---

[2] http://contenttechnologiesinc.com/
[3] https://www.netexlearning.com/en/
[4] https://ict.usc.edu/prototypes/captivating-virtual-instruction-for-training-cvit/
[5] https://translator.microsoft.com/

service and it offers text and speech translation through cloud services. This makes it possible to create captions in real time to show what the teacher is saying.

- **Personalized Learning**: Adjusting learning methods based on the needs of each student has been an important aspect for Educators in recent years. ITS and adaptive tutors can tailor learning material, sequence, pace and difficulty to each student's needs [88]. For example, Carnegie Mellon University has developed *iTalk2Learn system16*[6], which is used to assess students learning of fractions and it applies a learner model that includes information about an individual's mathematics knowledge, emotional state and cognitive needs. Also, feedback is received after student's responses. We'll talk in more detail about ITS in Section 2.2.

In general, there are promising benefits of AI in classrooms such as improvements in student learning outcomes, academic performance and reduction in gaps between groups of students. Also, personalized learning can possibility increase student engagement, independence and motivation. Some drawbacks can appear, given that budget implications of using AI in education are unclear and some uncertainties regarding the cost-effectiveness of putting the technology on the ground. Other more sensitive issues can also be found when we try to understand how AI can be useful to important soft skills such as creativity, innovation, socializing, leadership, empathy, collaboration and communication [89]. Teacher preparation and product support can be a time and energy-intensive effort and policy considerations like student privacy also raise several issues about data security and how to maintain reasonable security measures. The last big question is: *What will be the role of the Teacher?* Naturally, there is a lot of hesitation when it comes to bringing AI to the classroom, but one thing seems to be certain: AI isn't anywhere close of doing a skilled educator's job, much less outperforming it, so the teacher will not be replaced. Instead, multiple forms of support will emerge to give an effective assistance to teaching activity.

In conclusion, researchers have yet to reach a consensus regarding AI-based instruction effectiveness, but everything indicates that AI's impact in education will continue to grow in the next years. A study published by eSchool News[7] predicts that by 2021, the application of AI in U.S education will be increase by 47,5%.

## 2.2 Intelligent Tutoring Systems

Artificially intelligent tutoring systems, or ITS, have made remarkable progress in recent years. They can be defined as being computer programs which use AI techniques in order to facilitate instruction. They help students by providing immediate and customized instruction or feedback, usually without requiring intervention from humans. One of the biggest advantages by using these intelligent tools is that instruction would be specialized according to each student's proficiency [59], providing a suitable and personal excursion for learners.

---

[6]https://www.italk2learn.com/
[7]https://www.eschoolnews.com/2017/05/22/brace-ai-set-explode-next-4-years/
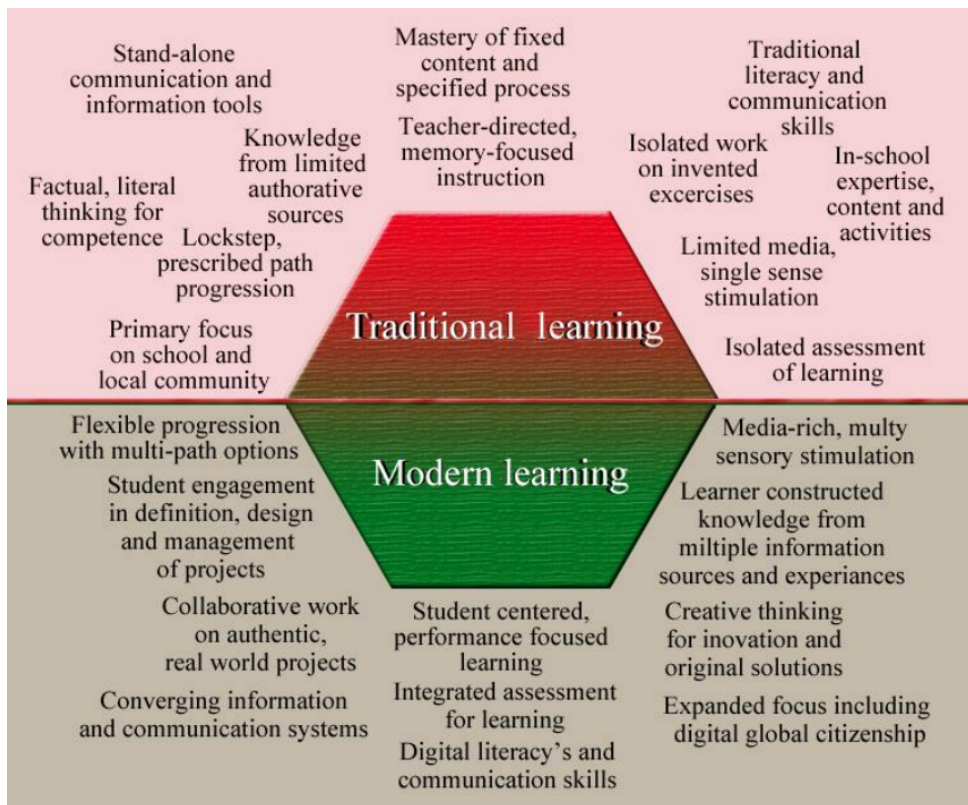
Figure 2.1: Transition from traditional to new advanced teaching paradigms [43]

Typically, an ITS system consists of 4 distinct modules [43]: Student module that contains a dynamic model to acquire student's knowledge and skills; Teacher module that is the unit which controls the student's learning process; Domain module as it holds the domain knowledge from which the students will communicate (during learning and teaching); Communication module or user interface (connects student-teacher-knowledge) from which students interact during the educational process.

While in traditional computer programs the student responds to specific computer prompts, by the use of intelligent systems the student or computer can ask a question to open a dialogue, posing problems or even provide explanations. Another interesting comparison is that by the use of traditional programs, questions area based on student's performance from prior questions, or on achievement benchmarks; with ITS, there is the possibility to incorporate prior student data form inside or outside the system. For that, it can use several engagement measures such as frequency of help requests and keystroke speed. Figure 2.1 conveys this transition from traditional to new advanced teaching paradigm.

Traditional systems use a limited script of questions (and responses) but ITS may use natural language processing (and AI), creating a new paradigm for human-machine interaction. This is precisely the point we wanted to highlight.

Today, more than ever, smart voice interactive interfaces and NLP are being applied in the

field of education with teaching contents such as language study, math, etc. Furthermore, NLP research is integrated with context analysis, linguistics and semantics. It "teaches" machines how to understand human language so they can effectively communicate and help to build social skills and good habits.

Language learning is very important in education and it allows younger children to be accompanied and entertained by intelligent machines powered by NLP technology. Thus, it engages children by rational and appropriate conversations. An example of this is the Turing Robot's NLP-powered DuoDuo[8] that is specialized in teaching children Chinese and English. These type of robots can help children to practice and develop both listening and speaking skills. E-rate and Text Adaptor [22] are software systems focused on developing educational strategies that can assist learners by the of using natural language. Many ITS also help students to improve reading comprehension skills [17] such as RSAT [91] and iSTART [98] which was designed to help students to become better readers using multimedia technology. Language Muse [90] is for instructional authoring and it is projected to help K12 instructors in the creation of English-language learners materials. Natural voice interaction through the technologies that use NLP allows the creation of several chatbots and smart toys. These tools allow interactive and playful experiences such as animated books and rich communications between children and characters from the books. Echo Dot Kids edition[9] has its own Alexa Echo smart speaker that provides several ludic materials for kids. There are other relevant examples such as the DoDo robot from RSVP[10] that will remind children to brush their teeth, to do their homework and let them know when it's bedtime. Anki's Cozmo[11] is more targeted to gauge children's ability to familiarize them with machines for natural interactions. They all have one thing in common: encouraging children to practice positive social behaviour like, for example, treating people with respect and speaking politely.

Alhawiti [66] has summarized the importance of using NLP techniques in Education as being a set of effective approaches for teachers, students, authors and educators. Thus, it is also referred its proven assistance for writing, analysis, assessment, procedures and understanding of various effective linguistics tools like syntax, sentence compositions and grammatical constructions.

From encyclopedic databases and question answering systems there is the possibility to automatically answer queries posed by humans in natural language. *What if we use these techniques not only to answer questions but also to generate them?* Let's see on what that task consists of.

## 2.3 Automatic Question Generation

*How can we define the process of automatic question generation?* Rus et al. [127] argue that question generation is primarily a dialogue and discourse task, drawing on both NLU (Natural Language Understanding) and NLG (Natural Language Generation), i.e., generating semantically and grammatically correct questions [123]. Considering AQG from raw text, e.g. a textbook or

---

[8]https://syncedreview.com/2018/12/15/natural-language-processing-in-early-education
[9]https://www.amazon.com/All-New-Echo-Kids-designed-Rainbow/dp/B07Q2MXPH6
[10]http://dodo.rsvp.ai/
[11]https://anki.com/en-us/cozmo.html

the dialogue history between a tutor and tutee, it will involve some sort of analysis, i.e., NLU of the input to select content and pattern detection. The main goal is to construct the output and it will involve NLG: a representation of the input is mapped to a certain sentence. Piwek and Boyer [116] concluded that question generation could also be viewed as an algorithm search to transform certain inputs to certain types of outputs. Mazidi et al. [97] observed that the majority of question generation systems input a text source, automatically parse sentences and transform those sentences into questions.

The first studies in AQG goes back to 1976, when Wolfe [141] demonstrated the feasibility of automatically generating questions from text and also affirmed that this task could be as effective as manual human-made questions [142]. In recent years, there has been a revived growing interest in the area, fueled by a series of question generation workshops [19]. Until the day when we can define with the certainty provided by the Turing test that computers really understand natural language, NLG systems are challenged to go beyond text processing and create illusions of understanding [97].

In fact, there is still a lot of work to do in order to apply AI into AQG. This type of task raises the possibility of generating creative, rich and interesting questions. AQG can be directly applied in many other domains such as automated help systems, instructional games, Massive open online courses (MOOC), search engines, chatbot systems (custom interaction in initiating conversations or obtaining specific information from humans) and healthcare for analyzing mental health. Research on Question Answering (QA) also benefits from AQG in a sense that it reduces the need from human labor to create large-scale datasets. AQG has practical importance in education: forming good and well-formed questions that are crucial for evaluating student knowledge and stimulate self-learning, generate classroom quizzes and assessments for course materials [63], it can be used in intelligent tutoring systems or as a component in adaptive learning and, finally, targeting grammatical categories and linguistic forms in a text supporting incidental focus-on form [86] by a meaning-focused reading task. These practical utilities demonstrate that AQG will continue to be a promising task in the next years.

# Chapter 3

# State of the Art in Automatic Question Generation

In this chapter we focus on summarizing the main findings in this area by referring the main tendencies and by providing our vision for those same tendencies (Section 3.1). We point out what are the domains, question types and language of AQG systems without leaving aside current limitations that are present in the literature (Section 3.2). Then, we go deeper on what are the traditional approaches (Section 3.3), generation tasks (Section 3.4) and we have reserved a section dedicated to the most recent studies in what is called neural question generation (Section 3.5). Ultimately, we express two important topics related to item difficulty (Section 3.6) and feedback (Section 3.7) in the context of this work.

## 3.1 A General Review on AQG

AQG research has largely increased in recent years (mainly since 2012). Workshops in AQG have been highlighted since there are more researchers working in the field. In addition, it should be noted that the year 2012 coincided with an increase in Massive open online courses (MOOC) providers like Udacity, Udemy, Coursera and edX, which all started around 2012 [14].

Two of the most notable reviews on AQG was made by Alsubait [9] and Kurdi et al. [70]. Alsubait characterized the AQG topic dividing it into the following sections: 1) purpose of the generated questions, 2) domain, 3) knowledge sources, 4) generation methods, 5) question type, 6) response format and, finally, 7) evaluation type. The results of this discovery can be seen in Table 3.1 and the first conclusion is that the main purpose of using AQG is assessment. Kurdi et al. also came to the same conclusion. Furthermore, questions are generated for other purposes such as in tutoring of assisted learning. Most of the study focuses on experimental settings but, for example, Zavala et al. [74] reported its use in the classroom. Let us now observe what were the general observations made from Kurdi et al. regarding domains, knowledge sources, and

limitations of AQG. Then, in Section 3.2 we'll go even deeper into some aspects (also based on the study by Kurdi et al. but with a greater focus in language learning).

Table 3.1: Alsubait's review regarding AQG topic [70].

| **Dimension** | **Categories** | No. of studies |
|---|---|---|
| Purpose | Assessment | 51 |
| | Knowledge acquisition | 7 |
| | Validation | 4 |
| | General | 3 |
| Domain | Domain-specific | 35 |
| | Generic | 30 |
| Knowledge source | Text | 38 |
| | Ontologies | 11 |
| | Other | 16 |
| Generation Method | Syntax based | 26 |
| | Semantic based | 25 |
| | Template based | 12 |
| | Other | 5 |
| Question type | Factual wh-questions | 21 |
| | Gap-fill questions | 17 |
| | Math word problems | 4 |
| | Other | 28 |
| Response format | Free response | 33 |
| | multiple-choice | 31 |
| | True/false | 1 |
| Evaluation | Expert-centred | 20 |
| | Student-centered | 15 |
| | Other | 12 |
| | None | 18 |

Generating questions for a specific domain is more common when compared to unspecified domains and one of the most studied domains in the literature is language learning, followed by mathematics and medicine. These are the most addressed domains since there are standardized tests prepared by professional organizations (e.g. Test for English International Communication (TOEIC) for language, Test of English as a Foreign Language (TOEFL), International English Language Testing System (IELTS) and Scholastic Aptitude Test (SAT) for mathematics field and several examinations for medicine). These tests need to be continuously receiving new types of questions and furthermore, regarding language learning, it is clearly noticeable that it can be more practical to generate questions that belong to other domains. The ease in creating language questions is also explained by the amount of natural language processing tools allowing an in-depth analysis of the texts by performing, for example, Syntactic Analysis and Part-of-Speech (PoS) tagging. On the other hand, it is worth noting that there are currently multiple textual sources that are in the public domain, such as educational texts, news, books, or even databases.

Regarding knowledge sources, the most used source of information is textual content. This

trend was also verified by Rakangor and Ghodasara [120]. In the specific case of multiple-choice questions, it might be necessary to use additional inputs such as WordNet [52], textual corpora and thesaurus because one of the most important phases of generating multiple-choice questions is the creation of distractors. Although texts are used to generate language questions, where distractors can be generated using simple techniques such as selecting words having a certain PoS tag or another syntactic property, text does not usually contain distractors, hence the need to search in external structured sources.

The limitations on AQG are mainly found in controlling the difficulty of the questions and in generating informative feedback to the learner. Few authors have focused on these tasks [70]. Controlling item difficulty is only applied to a specific type of question and it becomes a complex task to validate the real difficulty for each generated question. Feedback is used to explain the correctness/incorrectness of the answer given by the learner and it is used to motivate students to try again and to provide extra reading material. An in-depth explanation of what has been done in these two areas can be analyzed in the Sections 3.6 and 3.7.

## 3.2 Domain, Question Types and Language of AQG systems

As it was possible to observe in Section 3.1, language learning is the most studied domain by researchers on AQG. In fact, questions about language learning target multiple important skills such as reading comprehension, knowledge of grammar and vocabulary. Since this dissertation focuses on these contents, it is important to highlight which types of questions can be generated when learning a language, whether that language is the student's mother tongue or not. According to Chinkina and Meurers [29], in foreign language teaching and learning (FLTL) questions can be asked to serve a broad range of different goals:

- Questions about the **learner's experience** (e.g., "What do you know about Portugal?");

- Questions that focus on **comprehension** and recall skills to check whether the learner has understood a text;

- Questions with the **goal of eliciting a linguistic form** from the learner (e.g., "What would you do if you won in a game?" requires the student to work with conditionals);

- Meta-linguistic questions to check learner's **explicit knowledge** of the language system (e.g., "From which verb is the noun decision derived?" or "What is the synonym of sleep?").

More relevant work related to language learning can be seen in the work of Susanti et al., [139] and Chinkina and Meurers [29]. Furthermore, there is relevant study regarding vocabulary and grammar questions [129, 112]. Reading Comprehension was also recently studied [55]. AQG has been performed in several other domains including history [3], general science [136, 32], biology [133, 118] and also in more generic domains.

Regarding the type of questions that can be asked, they can be divided into the following categories: *wh*-questions, gap-fill, multiple-choice and true/false. Factual questions (or *wh*-questions)

ask for answers that are shorter facts and which are explicitly mentioned in the input. In gap-fill there is only the need to hide a word or set of words in order to fill an incomplete sentence. Kurdi et al. [70] concluded that gap-fill and *wh*-questions were the most common types of generated questions. One of the possible reasons is the ease to generate them since there is no need to generate distractors (as in multiple-choice questions). Additionally, in gap-fill questions, there are few concerns regarding linguistic aspects (e.g. grammaticality) because the sentence is constructed only by removing a word or set of words from the source text. Concerning multiple-choice questions there is also a substantial study that can be observed within this format. A MCQ is mainly composed of three components: *stem*, *key* and *distractor*. The stem is basically the phrase from which a question was generated. The key is the word (or target word) that belongs to the stem and defines the questionable term of the question. It is also the correct answer of the question. Distractors are the set of wrong choices along with the correct answer. Multiple-choice questions have some advantages over many other categories such as non-human evaluation and consistent scoring. They have also some weaknesses such as discredit of partial knowledge. The questions of true/false are very important especially when the Teacher intends to understand the student's ability to assess facts from the textual content. In addition, through these types of questions it is possible to achieve higher levels in Bloom's taxonomy (see Section 1.1).

Regarding the language of the generated questions, most of the studies can be found in English; however, questions in French [112], Basque [5, 6], Russian [71], Chinese [31, 83] and Thai [72] have also been generated. The questions generated from our methodology are in the European Portuguese language. One of the reasons for this decision is precisely the need to fill the gap in the literature concerning this language. Our research found a master's thesis made in 2010 by Curto [38] and two studies by Correia et al. [35, 34]. Pirovani et al. [48] has made research for Brazilian Portuguese. In summary, we can say that the AQG community has been growing and, therefore, we anticipate that there will be greater coverage of languages in the near future.

## 3.3   Traditional Approaches for AQG

Yao et al. [144] has divided generation of questions into three main categories. They are: *Template-based*, *Syntax-based* and *Semantic-based*. Through the study of literature, this trend is confirmed, but in more recent papers there is a particular technique that has been studied with some depth and, therefore, we believe that it should be considered as a distinct category: *Dependency Analysis*. An important note is that all of these approaches have already been explored individually or even used together. We believe that there is no "best one", but sometimes an approach is simply better suited to a particular type of question. In addition, there is a new dimension to the whole task of generating questions with the emergence of Machine Learning techniques, namely those that use Neural Networks. In fact, there is an extensive study in what is today called *Neural Question Generation (NQG)*. This area has a lot of content to be explained and deepened and, therefore, we decided to make a more detailed explanation of what is being done in Section 3.5. Regarding the other approaches, we will now make a detailed description for each one of them.

### 3.3.1   Template-based

One of the most popular methods in AQG involves analyzing the structure of a sentence with a structure parser and then creating questions using predefined templates [84, 85, 125]. These templates are basically fixed text (set of interrogative sentences) with some placeholders witch are then populated with the term or set of terms extracted from the input text. Typically, these methods do not use any transformation rules. The intention is to extract as much relevant information and make inferences as possible. From there, the goal is to use pre-defined templates and then create the questions. One of the biggest advantages of this approach is the possibility to generate well-formulated and very specific questions on a given subject. On the other hand, it will be always limited to the number of templates that were defined in a previous instance. We can see an illustrative example in Table 3.2, where the informational content is "activation energy" [75].

Table 3.2: An example of question template for the question class Definition [75].

| Type | Question template | Question |
|---|---|---|
| Definition | What is ⟨X⟩? | What is **activation energy**? |
| | What do you have in mind when you think about ⟨X⟩ ? | What do you have in mind when you think about **activation energy**? |
| | What does ⟨X⟩ remind you of? | What does **activation energy** remind you of? |

### 3.3.2   Syntactic Analysis

Syntax-based methods make use of syntactic constituency trees and a set of syntactic transformations to generate questions. This approach applies a parser to discover the syntactic structure of the sentence (e.g, constituency parsing) in order to apply syntactic transformation rules [8, 57, 56] and question word placements (e.g., "When","Where","Who"). The key idea is to convert the declarative sentence and transform into an interrogative form. For that, it usually uses the manipulation of derived syntactic tree which is typically parsed by using a Context Free Grammar parser (CFG), such as Tregex and Tsurgeon [78]. Usually, information from PoS is also very useful in this type of approaches. These methods do not require semantic analysis or, in other words, the meaning of the given input (i.g., meaning of the words). Guy Danon and Mark Last [40] present a more recent example of the syntax-based approach. We can see an illustrative example in Figure 3.1 which basically parses the input phrase: *Michelangelo has sculpted the statue of David*. After that, when applying a specific pattern learning algorithm, the authors perform certain transformation rules in order to convert the sentence into question. The question will be: ***Who sculpted the Statue of David?***. As Heilman [62] noted, these purely syntactic approaches do not allow higher-level abstractions like those ones used in semantic analysis. Anyway, AQG systems show that syntactic analysis can be productive and a robust method for generating fact-based questions.
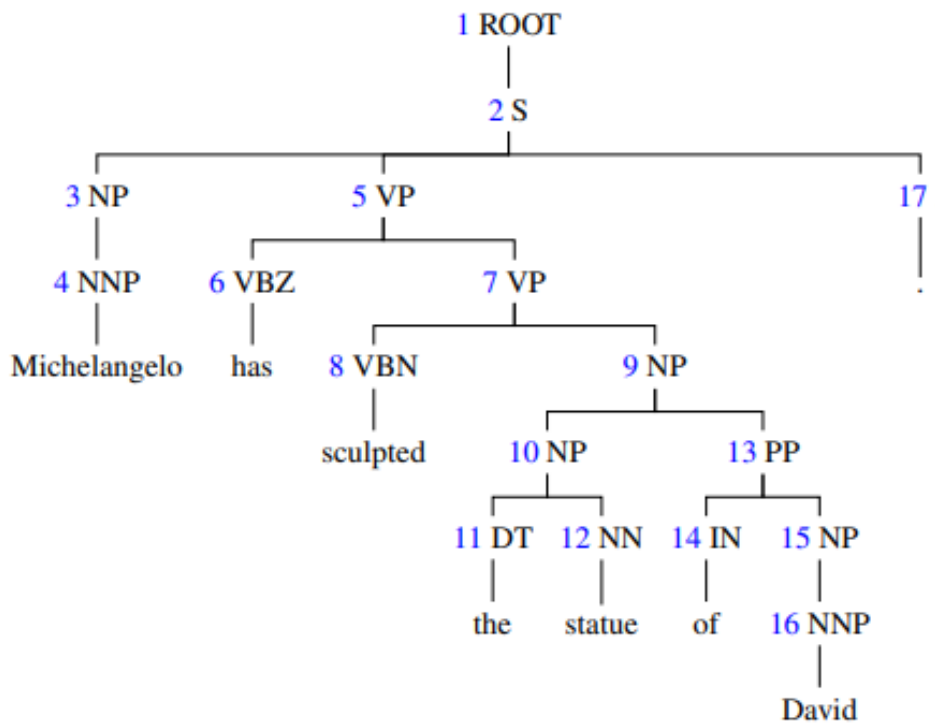
Figure 3.1: Parse tree of the sentence *Michelangelo has sculpted the statue of David [39]*

### 3.3.3   Semantic Analysis

Semantic-based approach focus on semantic parsing for generating questions from text. Basically, for each predicate in a sentence, the goal is to discover its associated arguments and modifiers, and specify their semantic roles. This technique provides a deeper understanding of the input and goes beyond lexical and syntactic understanding. In fact, it provides a deeper level of analysis when compared to syntactic parse. All of this is possible because this approach performs semantic role labeling (SRL) as the main driver of linguistic analysis, with which it identifies the semantic roles of each argument and modifier in a sentence.

We can see the identification of these labels (by performing SRL) in the following examples [53]:

- ***A0*** (Joe) has sold ***A1*** (his house).

- ***A0*** (Peter) called ***AM-TMP*** (for six hours).

- ***A0*** (They) can fly from here ***A4*** (to any country).

The label *A0* means proto-agent, *A1* means proto-patient, *AM-TMP* means time and *A4* means end point or state. The types of questions that could be raised would be: ***Who has sold the house? What did Joe sell? For how long did Peter call? Where can they fly to?***. Selecting the appropriate *wh*-word is encouraged by the identity of the focused argument. For example, label

AM-MNR which means manner, invites the word *How* and location (AM-LOC) invites *Where*. There are a lot of possibilities when we observe different types of labels. For example, in addition to those shown above, there are some of the other possible labels[1] that can be obtained from the *SENNA*[2] software.

AQG systems use this type of information (arguments and modifiers) to perform question construction [25, 80, 93, 96]. These types of systems are not as close to the sentence structure as in phrase-structure parse, therefore, there are not so many concerns regarding syntactic information, although they typically also use some syntactic analysis. They allow AQG systems to create a wide range of questions but this method is dependent on the performance of the model that assigns the labels. In other words, if the identification of the semantic role fails, then the generated question may be incorrect. This type of event happens when, for example, this technique is applied to less studied languages (like Portuguese), where the tools (pre-trained models) are not so improved. For these reasons, it is quite interesting to combine this technique with syntactic analysis.

### 3.3.4 Dependency Analysis

Another type of analysis is with the use of dependency parsing which is the task of extracting a dependency parse of a sentence that represents its grammatical structure. This procedure identifies a set of typed relations (called dependency labels) between words and it connects those words in a graphical structure based on their grammatical and functional relations. We can ask about *What?* or *To whom?* figuring out what is the sentence's object. For instance, we can observe the dependency labels from the sentence *John offered a motorcycle to his uncle.* in Figure 3.2. This sentence could lead to the following questions: *What did John offer his uncle?* or *To whom did John offer a motorcycle?*.
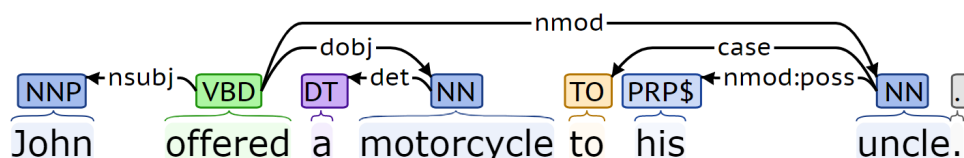


Figure 3.2: Dependency parse of the sentence *John offered a motorcycle to his uncle.*

Mazidi et al. [97] revealed that it is possible to find several patterns through dependency labels, e.g, extract the meaning of these relationships and, from there, generate questions accordingly. Some of these patterns and meanings can be seen in Table 3.3.

The utility of dependency analysis led to a number of applications related to NLP tasks such as information extraction [143], textual entailment [126] or even extraction of casual relations from biological texts to find analogies for biomimetic engineering [28]. Afzal and Mikov [1] used this technique in sentence transformation and simplification to create multiple-choice questions.

---

[1] https://ufal.mff.cuni.cz/conll2009-st/task-description.html
[2] https://ronan.collobert.com/senna/

Table 3.3: Meaning of the patterns found with dependency labels [97].

| Pattern | Meaning |
|---|---|
| S-V-acomp | adjectival complement that describes the subject |
| S-V-attr | nominal predicative complement defining the subject |
| S-V-ccomp | clausal complement indicating a proposition of subject |
| S-V-dobj | indicates the relation between two entities |
| S-V-iobj-dobj | indicates the relation between three entities |
| S-V-parg | phrase describing the how/what/where of the action |
| S-V-xcomp | non-finite clause-like complement |
| S-V | indicates an action of the entity |
| other | combinations of constituents |

Certain dependency relations give greater insight into semantics than the SRL parse, proving (the dependency analysis) to be a powerful technique to understand the sentence's meaning. However, the problems are similar to those mentioned for SRL parse, that is, we are totally dependent on the accuracy of the models when identifying the dependency labels.

## 3.4 Traditional Tasks for AQG

If we look carefully, when a question is manually elaborated from a certain text, the first step is to acquire the information that is contained in that same text. In the general process of automatically generating a question, the first step also goes through that way, because an AQG system primarily demands one (or more) informative sentence and, for that purpose, it also tries to identify the sentences that contain any possible relevant content (i.e, questionable facts) for generating questions. The main task is to identify the word (or words) or sentence that acts as the source key, or set of keys (depending on the question type). From that key, it is possible to form a suitable question. This way we can divide this task into six distinct phases: **(1) Pre-Processing**, **(2) Sentence Selection**, **(3) Key Selection**, **(4) Question Formation**, **(5) Distractor Generation** (only for multiple-choice), **(6) Post-Processing** and **(7) Evaluation**. Although this strategy varies slightly from system to system, we can define this as being the generic flow. Let us now understand what was done in each of these phases by the different authors. Ch and Saha [24] did an in-depth study of the first six tasks (with a greater focus on multiple-choice questions). The next subsections (from 3.4.1 to 3.4.6) present a summary of their findings plus ours.

### 3.4.1 Pre-Processing

Standard pre-processing is common to diverse NLP tasks and it is used to prepare input for the forthcoming tasks. The following tasks are pre-processing stages that are commonly performed for AQG. They are sentence splitting, tokenization (the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens), coreference resolution, PoS tagging, named entity recognition (seeks to locate and classify entities in text such as

person names, organizations, locations, medical codes, time expressions, monetary values and percentages), or even relation extraction. Several pre-processing steps have been implemented by researchers: Text Normalization, which removes unnecessary content and converts the input text into the desired format [62, 15]; Sentence Simplification, which involves converting complex sentences into simple ones [113, 42]; Structure Analysis, that identifies the sections, paragraphs and some relevant tags in the input text [113, 33]; Lexical Analysis, that splits up the text into a set of words, numbers and symbols. Lexical Analysis also serves to produce inflections or to perform stemming to find the root form of the words [15, 62]; Syntactic Analysis, that includes named entity recognition, PoS tagging and syntactic parsing [113, 81]; Coreference and pronoun resolution, that has the goal of mapping pronouns to the corresponding nouns [103, 21]; and finally, word sense disambiguation as the task of identifying the exact sense of a given word in a certain sentence [81, 21].

### 3.4.2 Sentence Selection

Not all phrases are suitable for generating questions as not all of them have questionable facts. Huang and He [64] had defined three main characteristics for selecting sentences that contain important content by proposing the following metrics: keyness, completeness and independence (that covers multiple and different aspects from text content). That said, several techniques have been created to choose a particular sentence that contains relevant information to create appropriate questions. The techniques that were found are: PoS information [7] to find certain occurrences of a verb or patterns along the sentence and parse information [100, 99] to filter the sentences according to their structure (e.g pronoun-verb-noun) using parse tree structures; Sentence length is used to filter short sentences or even a long sentence [113, 135]; Finally, summarization techniques have been also applied, for example, by Shah [132] or by Narendra et al. [103] that used an extractive summarizer called MEAD.

### 3.4.3 Key Selection

The selection of the key is the step responsible for determining the word or set of words in the selected sentence that will originate the question. The main techniques that have been used are: Predicate extraction used by Fattoh [51]; Frequency count of the words as a selection metric [33, 100]. Normally this technique is accompanied with Tf-Idf; Pattern matching was also used as a procedure for selecting the key, for exemple, Gates et al. [58] applied syntactic patterns that were based on the parse structure of the phrases to find the keys; Lastly, PoS and parse information were again used to find a particular PoS tag or to search for a specific keyword given some specific applications and domains [99]. Sumita et al. [135] have chosen the verbs as being the keys.

### 3.4.4 Question Formation

Following key selection, this is the main task and it involves dissimilar processes that are based on the type of questions to generate as well as its response format. In the case where there isn't

any transformation, the sentence stays in its original form and we have a gap-fill question. In the remaining cases, we see several approaches to perform this task, many of them consisting in rules, pattern based, templates or statistical methods. The main sub-processes are characterized by: 1) Transform the declarative sentence into an interrogative sentence; 2) Determine what is the question type such as the *wh*-word; 3) Select position of the gap word that is only relevant for gap-fill questions. The approaches that have been studied can be summarized as follows: Syntactic transformations were performed by Heilman [62] by decomposing main verbs, inverting subject-auxiliary terms and by marking unmovable phrases; Majumder and Saha [92] performed *Wh*-word selection by identifying the suitable *wh*-word and then form the questions with specific rules; Subject-Verb-Object relationship was used by Mitkov and Ha [99] in order to understand the target sentence and certain terms occuring along it; Sentence knowledge was used by Pabitha et al. [106] to extract knowledge labels, for example, concept, definition, example, procedure, calculation and result. They are used as the beginning terms of the questions; Afzal and Mitkov [1] performed dependency parse to extract the dependency tree, selected the necessary portion from sentence and then picked the appropriate *wh*-word; Semantic Role Labeling was used by Lindberg et al. [80], and by Mazidi and Nielsen [95] to identify semantic patterns; Discourse connectives were used in order to discover discourse relations (e.g, causal, temporal, elaboration, result, contrast) and to determine of what type the question should be.

### 3.4.5 Distractor Generation

The process of creating incorrect options is a very important task in the context of multiple-choice questions as the quality of those options (also called distractors) directly influences the question's difficulty. Actually, if the incorrect options are not capable to create reasonable doubts, then it makes the question easy and irrelevant. For this purpose, several approaches have been used: Word Frequency (e.g, the number of times a distractor appears in a corpus, considering that this distractor is similar to the key) is one of the most used techniques [33, 21]; WordNet [52] is also very used because it can group the words into a set of synonyms and record relations among these synonyms [100, 99]; Ontologies were used by Papasalouros et al. [109] with the applicability of the Web Ontology Language in order to find words in the context of the key; Distributional Hypothesis (words that occur in the same contexts tend to have similar meanings) [61, 140]; Semantic similarity is another common approach used by Pino et al. [113] and by Aldabe et al. [7] to discover words that are semantically close to the keys.

### 3.4.6 Post-Processing

Post-processing is the final stage and its main objective is to improve the quality of the generated questions. There are several types of errors that can be found in AQG systems such as improper question words, punctuation errors, lengthy stems and the poor quality of distractors for multiple-choice questions. The system should perform this step in order to minimize errors in the final output. In some cases the tool may be able to rectify these errors, otherwise, questions may even be

deleted. Covered methods can be divided into three (Question ranking, filtering and post-editing): Question Ranking is an approach in which the authors pick the previously generated questions and then, rank and filter them taking into account a certain priority related to their quality. The ultimate goal is to prioritise good quality questions and that was achieved by Mannem et al. [93] employing an heuristics-based ranking module related to the depth of the predicate and the number of pronouns. Heilman and Smith [63] employed a ranking system in order to develop a discriminative question ranker. This way, questions could be ordered by the predictions of a logistic regression model for question acceptability; Question Filtering (also called question selection) was used by Sumita et al. [135] for reducing test size by selecting effective items. Heilman [62] used this method to filter out questions that contained noun phrases consisting only by determiners, unresolved pronouns or for questions that contained more than 30 tokens; Question post-editing was made by Miltok et al. [99] in order to perform a set of manual post-editing steps (minor, fair or major steps) such as spell correction, punctuation, re-orderings, deletions or insertions of several words, replacement of distractors and substantial rephrasing of the stems. Heilman [62] applied some post-editing steps such as the addition of final periods and the removal of white spaces and symbols. Some authors created a final post-editing step where they asked humans to make changes to the generated questions.

### 3.4.7 Evaluation

After studying AQG literature it becomes clear that there in no standard way to evaluate the automatic generation of questions. Anyway, since AQG contains multiple tasks and components, different types of evaluation have been proposed for assessing the quality of these individual components. Kurdi et al. [70] pointed out several types of evaluation that have been developed over time, namely **expert-based evaluation**, **student evaluation**, **comparing machine-generated questions to manually human-made** and **crowd-sourcing**. In the next paragraphs, we will explain what the conclusions were regarding each of these evaluation types.

Expert-based evaluation consists of presenting a sample of the generated questions to reviewers. Aldabe et al. [4] used one expert language teacher to assess the quality of the questions. The same happened with Pino et al. [113] and Satria et al. [129] that asked five English teachers to do the same tasks. Expert rating is the most common evaluation approach since it is very similar to a real context in which questions are manually selected for exams. The main point is that invalid questions must be filtered whenever they are ambiguous, guessable, or do not require any knowledge. By using this type of evaluation it is important to ensure that there is an appropriate question set in order to keep participants involved, motivated and interested along the process. One of the main precautions that must be assured when doing something like this is to ensure that the evaluators do not interpret criteria in distinct ways. For example, Mostow et al. [102] reported different interpretations of the instructions for rating the overall quality.

Students also have been asked to review questions in several studies. The goal is to obtain further evidence of their quality by checking empirical difficulty, reliability and discrimination. Anyways, invalid questions must have been filtered in a previous instance as explained in the last

paragraph. Susanti et al. [139] have applied this evaluation with 79 undergraduate students (and also eight English students).

Comparing machine-generated questions to manually human-made ones is also commonly employed for evaluation. Zhang and VanLehn [145] evaluated their approach by counting the number of questions in common (human and machine). The justification (from the authors) is that humans typically create questions that delve deeper into certain topics, that is, questions that require a high cognitive level (recall in 1.1).

Crowd-sourcing has also been used to perform the evaluation. Chinkina et al. [30] produced test questions in order to make sure that a person could understand the task at hand and be able to distinguish good and bad questions (low vs high quality).

In terms of criteria and metrics, there are several that have been suggested over the past few years. Most of these criteria consist of verifying linguistics quality in questions, such as grammatical correctness, semantic ambiguity, fluency or even distractor readability. Other types of criteria can be more related with educational goals such as relevance to the domain, usefulness and learning outcome. We make a synthesis of the various metrics we found [70]:

- **Question itself**: Statistical difficulty, reviewer rating of difficulty, acceptability, grammatical correctness, semantic ambiguity, fluency, number of errors;

- **Distractor**: Quality, plausibility, correctness, functionality, homogeneity, readability;

- **Educationally oriented**: Usefulness, input relevance, domain relevance, cognitive level or depth, learning outcome.

## 3.5 Neural Question Generation

With recent advances in deep learning, AQG research has also begun to take advantage of the new possibilities. Neural networks are now used in order to try to generate deeper questions [27] by training these networks to generate questions from sentences of word tokens [131]. Using substantial amounts of sentences accompanied by human-generated questions and an encoder-decoder framework, it becomes possible to make these systems learn how to map declarative sentences into questions in a manner very similar to the approaches used in neural machine translation. The name adopted in this new paradigm is called Neural Question Generation.

These neural models encourage the use of end-to-end architectures, that is, they work by following a sequence-to-sequence framework which uses a singular representation. This representation receives a certain content to be learned through an encoder and outputs the result by a decoder. In this specific case, the input is a paragraph and the output is the question.

Given an input sentence $\mathbf{x}$, the goal is to generate a natural question $\mathbf{y}$ of arbitrary length $\mathbf{k}$, such that it has the maximum conditional likelihood of being generated. The AQG task is defined as finding y:

$$y = argmax P(y|x)$$

where *P(y|x)* is the conditional log-likelihood from the predicted question sequence **y**, given **x** [45]. Methodologies differ essentially in factors such as question word generation, answer encoding and paragraph-level contexts. Pan et al. [107] did an in-depth study of these factors. In the next paragraphs, we will present a summary of their findings regarding each of these factors as well as do a brief reference (also based in Pan et al.) to input modality and evaluation metrics.

Duan et al. [46] implemented question word generation by a model design in a separate way. A more flexible model was proposed by Sun et al. [137] by introducing a decoding mode that generates the specific question word. Answer encoding relies on models that must consider answer's position as being the input [147] (feature) and in other cases, by just encoding answers with a Recurrent Neural Network (RNN) [68]. To leverage rich paragraph-level contexts, Zhao et al. [146] implemented an encoder that refined the encoded context. This was done by fusing relevant information from the context of its representation.

Regarding input modality, there are some differences when compared with traditional methods. In addition to textual inputs, declarative sentences, and other textual sources, NQG has widened several possibilities by the use of knowledge bases and images. NQG takes advantage of these new possibilities given the huge success in the use of neural networks for feature representation such as images [101] and knowledge representations [18]. The existence of substantial corpora allows promising results (mainly factoid questions) in datasets such as in MS MARCO [104] and SQuAD [119], but performance decreases when the goal is to obtain deeper questions, for example, with the LearningQ [27] dataset.

Evaluation metrics are widely used in NQG such as BLEU [110], METEOR [13] and ROUGE [79], but Liu et al. [82] claim that these metrics do not prove fluency, coherence, and adequacy of the questions since they only compare the similarity between the source text and the generated questions.

By the use of NQG models, the task of building questions becomes completely data-driven and requires little effort compared to those ones which uses transformation rules. Additionally, traditional AQG architectures are roughly limited by their intermediate representations, templates, or transformation rules. On the other hand, traditional approaches still allow generating deeper and specific questions given a specific content because current datasets for NQG have a very general scope, which prevents their use in an educational context. Fortunately, some efforts are already being made in that direction, such as in LearningQ and in RACE [73].

## 3.6 Controlling Difficulty in AQG

Controlling difficulty is a fundamental property of AQG systems. The lack of approaches responsible for controlling question difficulty raises some problems such as questions that are too easy, questions that are too difficult, or even nonsensical questions. Furthermore, it is not feasible for the examiner to manually search from generated questions which of them are more effortless or complex. For these reasons, we considered that it would be especially relevant to study what

were the solutions proposed in the literature. Despite the continuous growth in AQG, the study of difficulty control has not been much explored.

Regarding the studied domains, Kurdi et al. [70] found significant work on vocabulary by Susanti et al. [138] and reading comprehension questions by Gao et al. [55]. Susanti et al. used the following metrics: **Reading passage difficulty**, **contextual similarity between key and distractors** and **distractor word difficulty level**. Here, the evaluation was made by checking agreement between student performance and predicted difficulty. Gao et al. used **question word proximity hint** (the distance of all nonstop sentence words to the answer in the sentence) as a metric for difficulty. The evaluation was performed by the employment of automatic solvers. For other domains, the most used approach for evaluation is by checking agreement between expert predictions and predicted difficulty.

Bearing in mind that the focus of this dissertation will stand on grammar and reading comprehension questions, we are convinced that the study of difficulty on AQG is extremely important for our work. Our goal is to extrapolate, from these previous techniques, an approach that would be appropriate for questions in the Portuguese language.

It is important to keep in mind that this issue is not trivial because some assumptions must be made when implementing this module for an AQG system. We have to assume, for example, that the generated questions are well-formulated and do not have any irregularities. What is the point of studying the difficulty of a question when it is not grammatically correct? Besides, we must assume that the difficulty of a question is subjective to the students' proficiency or even to education level, therefore, some external factors must have taken into account. In conclusion, there are several challenges regarding difficulty for AQG, but there is no doubt about its potential and relevance, so it should continue to be studied and explored.

## 3.7   Feedback Generation

Feedback generation should provide an explanation regarding learner's answers weather they are correct or not. Also, it should suggest a way forward to arrive at the right answer. Usually, feedback is justified in the scope of the information provided regarding the response to a set of questions. It is a technique widely used in the context of Intelligent Tutoring Systems (Section 2.2) and electronic environments since interaction between instructors and students is restricted. Thus, feedback plays an important role in emending learner's misconceptions and, at the same time, it guides them to the knowledge they should acquire, possibly with the aid of additional materials.

It is a fact that, regarding AQG literature, this aspect has not been studied and deepened by the researchers. Kurdi et al. [70] found two studies that looked into this matter and the conclusions were that: Das and Majumder [41] have implemented answer hints for the examinees in order to reduce the number of possible answers that make assessment easier. They have used syntactic features like the number of words that compose the key, the first letters of a key or even the second word of a key formed by two words. Leo et al. [77] have used feedback as an axiom verbalisation

to select options. Thus, in the cases of incorrect options for multiple-choice, axioms were used to generate both distractors (included in the feedback) and the key.

We believe that there is still a lot of work that can be done in this area, considering that feedback is a very important functionality. Actually, it becomes quite difficult to take advantage of these tools if students are unable to understand their mistakes.

## 3.8   Conclusions and Lines of Research

In this chapter we have conducted a detailed study regarding what has already been done in AQG. We found that AQG is a constantly expanding activity and has an increasingly active community. Throughout this study we want to highlight what are the main conclusions that can be drawn and, in addition, to show what can be done in the future. The first major conclusion is that the most studied domain is language learning and the main source of information (input) are raw texts. The generation of multiple-choice and free answer questions (with a greater focus on factual questions) are extensively investigated. It is also notable the growing interest for generating questions for other languages. There is a variety of tasks (from pre-processing to post-processing) that serve to address the entire pipeline of question generation.

Many improvements can be made regarding question level such as enrich template construction, forms and structures. Feedback generation is a very important aspect that should be studied with greater depth. There is a need to improve evaluation methods because using an expert review or student's mock exams is an expensive and time-consuming process. Finally, we believe that the study of item control difficulty has still much room for improvement.

# Chapter 4

# Automatic Question Generation

This chapter aims to explain with detail all the implemented approaches that covered our research. To achieve this, we will start by providing a general overview (Section 4.1) and then, we will expound what was done concerning Pre-Processing (Section 4.2). Regarding the methodologies for generating our questions, we present, in the first place, grammar questions (Section 4.3). In second place, the factual questions for reading comprehension (Section 4.4). Finally, our last approach, to explain how to generate pronoun reference questions (Section 4.5). A summary will be made in Section 4.6.

## 4.1 System Overview

We reveal a general overview of our methodology. In Figure 4.1 we can observe its general flow.

Our system undergoes all the conventional tasks (as mentioned in Section 3.4) needed for generating questions. We start with **Pre-Processing**, which is common to all modules, being a necessary and fundamental step for treating the input text before we move on into the next steps. This stage includes tokenization, lemmatization, PoS tagging, dependency parsing and named entity recognition. After that, were carried out two different tasks. Firstly, we performed the necessary changes so that the morphological analysis of the input words is as faithful as possible with the European Portuguese. Second and last, we use this phase to perform a difficulty analysis (to be explained in the next section) regarding our texts. This analysis will be important for later, when generating questions, use it as a factor within the heuristic function that will define the difficulty of our questions. The type, number, and intended difficulty of our questions are input variables, concerning our approaches.

In order to generate questions, each of the proposed approaches has a unique module so that we can handle all their specific cases. The first one, called **Generation of Grammar questions** is responsible of generating questions about the Portuguese grammar. The second module called **Generation of Reading Comprehension questions** produces questions about the facts present in the
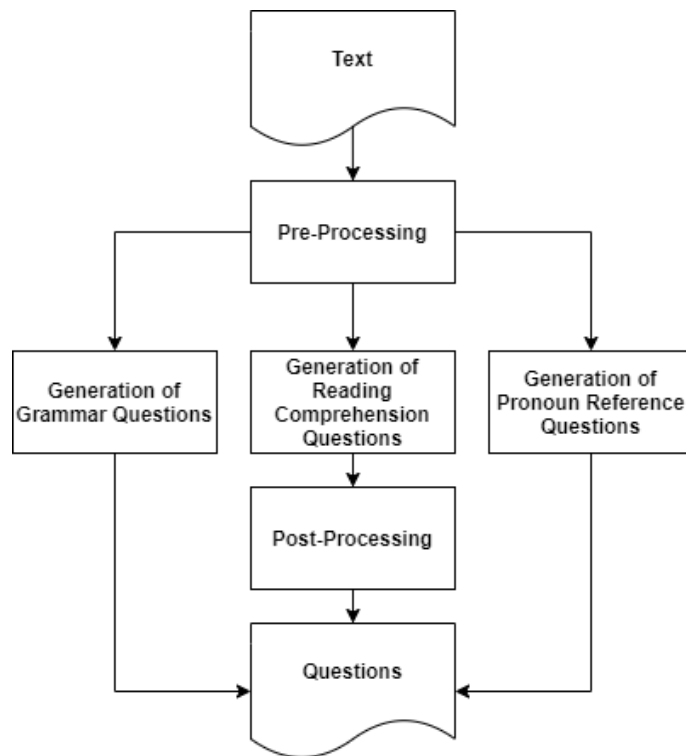
Figure 4.1: General overview of our methodology.

text. Finally, the **Generation of Pronoun Reference questions** which has the goal of generating questions which had been largely used in TOEFL iBT exams. They are also commonly used in Portuguese tests and national exams.

To improve the quality of the generated questions and to minimize errors in the the final output, we implemented the last step, called **Post-Processing** that has its effect with reading comprehension (factual) questions.

## 4.2 Pre-Processing

As previously mentioned, the Pre-Processing phase is divided into two phases. The first will be explained in Subsection 4.2.1 and focuses on Syntactic Analysis. The second will be explained in Subsection 4.2.2 and focuses on Morphological Analysis and Difficulty Analysis.

### 4.2.1 Pre-Processing: 1st Phase

Essentially, we use this Pre-Processing phase to carry out **Syntactic Analysis** which includes the following NLP techniques: Tokenization, Lemmatization, PoS Tagging, Dependency Parsing and Named-entity Recognition. Taking into account the great availability of tools that contain these implemented techniques, we decided to test as many as possible to make a good decision. The criteria we follow to choose the intended tool was:

- The tool must have support for the Portuguese language;

- The tool must have implemented as many of the NLP tasks listed previously as possible;

- The tool must be commonly accepted by the scientific community and properly updated.

The tools that we have found and fulfill these requirements are: *TreeTagger*[1], *FreeLing*[2], *spaCy*[3], *LX-Suite*[4] and *StanfordNLP*[5]. For the latter a new version was launched, called Stanza[6]. A first analysis was performed by analysing the values presented in the Table 4.1. In addition to this first exploration, and also to make us feel more confident with our final decision, we went through manual observations. To do this, we wrote a dozen of sentences and carefully compared the largest possible number of specific cases according to the Portuguese language. We put our main focus on the PoS results since this module would be crucial to our work. For example, regarding a grammar question, if a determinant is wrongly classified as a preposition, then the exercise would be immediately considered as incorrect and unusable. We did our best to choose the tool which reveled the best performance values, but we have the sense that none of them is infallible. The main goal was always to minimize the occurrence of these undesirable cases.

Table 4.1: Comparison of the tools that implement NLP tasks for Portuguese (Acc for accuracy and P for precision).

| Tool/Task | Tokenization | Lemmatization | PoS Tagging | Dep Parser | NER |
|---|---|---|---|---|---|
| **TreeTagger** | - | - | - | (not supported) | (not supported) |
| **FreeLing** | - | - | - | - | - |
| **spaCy** | - | - | Acc: 79,94% | Acc: 85,97% | P: 89,03% |
| **LX-Suite** | - | - | Acc: 96,24% | (not available) | (not available) |
| **StanfordNLP** | Acc: 99.71 % | Acc: 96,88 % | Acc: 96.71 % | - | (not supported) |

Through our observations, we started by excluding *TreeTagger* and *LX-Suite* because they contain little information about their performance and they are the most incomplete. *FreeLing* and *spaCy* are the most complete tools we had access to, but there are significant differences that are worth noting. *FreeLing* requires a very complex and time-consuming configuration, making its use impractical. Thus, it was slightly slower when comparing it to *spaCy*. If we take *spaCy* and compare it to *StanfordNLP*, the first impression is that *spaCy* is a better choice since it is faster and more complete (includes NER). Although this is true, our experiments showed us that *StanfordNLP* seems to be much better when performing lemmatization and PoS. Also, the information

---

[1]https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
[2]https://freeling-user-manual.readthedocs.io/en/latest/
[3]https://spacy.io/
[4]http://lxcenter.di.fc.ul.pt/
[5]https://stanfordnlp.github.io/stanfordnlp/
[6]https://stanfordnlp.github.io/stanza/

provided by the latter is richer than *spaCy*'s (concerning tenses and verbal modes). Another important advantage is that *StanfordNLP* contains a *MWTProcessor* that expands multi-word tokens into multiple words. Finally, its internal structure seemed to us to be far superior. Given this, we considered that we had the necessary information to decide that *StanfordNLP* would be our choice but we still needed to find a Portuguese NER model.

For the Portuguese language, we find the following models and tools (regarding NER) whose performance results are available in Table 4.2. All of them were trained using the same training set (HAREM [23]), except *spaCy*, which was trained on WikiNER[7] explaining the discrepancy between the performance results. HAREM contains 10 mentioned entities but WikiNER only contains 4 mentioned entities. Besides, after testing the *spaCy* NER, we realized that there were many times it classified people as being an organization, which would further undermine the quality of the generated questions. In fact (and after trying the other models from Table 4.2), we concluded that the existence of NER models for Portuguese is scarce and their performance is low when comparing to English models. Anyway, our goal was to use a model which included as many entities as possible and, at the same time, choosing the one with the best performance. We chose to use *Ner-Re-Pt*[8] since it is the most recent (2017), has the best average performance (within those that use HAREM) and contemplates several entities. It resulted from a Masters Dissertation [114] developed at University of Porto. The NER model was trained using the *Second Harem* dataset [23], which comprises 129 annotated documents with texts in both native Portuguese (pt-PT, 60%) and Brazilian Portuguese (pt-BR, 40%). Its main categories are *ABSTRACTION*, *EVENT*, *THING*, *LOCAL*, *ORGANIZATION*, *PERSON*, *TIME*, *VALUE*, *WORK OF ART* and *OTHER*.

Table 4.2: Comparison of the performance results regarding NER for Portuguese [10]. All of them were trained using the same training set which is HAREM, except *spaCy*, which was trained on WikiNER.

| Tools/Metrics | Precision | Recall | F-score |
|---|---|---|---|
| **NERP-CRF** | 53/% | 53% | 53% |
| **LanguageTasks** | 50% | 62% | 55% |
| **FreeLing** | 47% | 64% | 54% |
| **PALAVRAS** | 52% | 61% | 57% |
| **spaCy** | 89% | 89% | 89% |
| **Ner-Re-Pt** | 59% | 54% | 61% |

### 4.2.2   Pre-Processing: 2nd Phase

Along the second phase of Pre-Processing, as stated previously, these tasks were performed:

---

[7]https://hackage.haskell.org/package/chatter-0.9.1.0/docs/NLP-Corpora-WikiNer.html

[8]https://github.com/arop/ner-re-pt/wiki/Stanford-CoreNLP

- **Morphological Analysis** according to the European Portuguese designations. From European Portuguese designations, morphological analysis refers to the process that assigns each word in a sentence its corresponding class and subclass (e.g, negation adverb);

- **Text Difficulty Analysis** in which we implement an automatic classifier for our texts (as a whole) and also an automatic classifier for each sentence. These classifications are obtained through heuristic functions.

**Morphology**, **Morphological** or **Morphic Analysis** is the act of studying each of the different words in a sentence independently, aiming at their grammatical class. There are ten grammatical classes according to the Portuguese grammar: Noun, Adjective, Verb, Adverb, Determinant, Pronoun, Preposition, Interjection, Conjunction and Quantifier. The latter is not contemplated in *StanfordNLP* universal POS (UPOS) tags[9] and, consequently, no specific question about this grammatical class is asked. Although all other grammatical classes are covered by *StanfordNLP*, our focus, for this specific phase, is to analyze their grammatical subclasses. In fact, *StanfordNLP* also includes universal morphological features[10] (UFeats), which provide information about grammatical subclasses, but there are two problems. Firstly, these subclasses are universal and, therefore, some of them are not specific to European Portuguese. The second problem is that they are incomplete, that is, many of the Portuguese subclasses are not contemplated. These problems then made us do perform this morphological analysis in which we could analyze word by word, and correctly assign, whenever possible, its corresponding Portuguese subclass. The list from Appendix A contains all classes and subclasses resulting from this second verification. When deciding what the subclass of a particular word will be, there may exist some cases in which, for that, we would need to disambiguate its context. For example, the words *onde (where)* and *como (how)* might be interrogative or relative adverbs. While in the first case they are inserted in an interrogative sentence, in the second they belong to a declarative one. Therefore, disambiguation in this case is simple to perform but there are cases in which this process is not so trivial. For example, the word *pois (because)* may be a conclusive or explicative coordinating conjunction, depending on the meaning of the sentence. It can also be considered a causal subordinating conjunction. For this last case, if disambiguation isn't performed by *StanfordNLP*, we do not assign any grammatical subclass, leaving that kind of disambiguation for future work.

**Text Difficulty Analysis** is an important topic in applied linguistics. Understanding and evaluating text complexity is relevant to the teacher, who wishes to select appropriate materials for students with different proficiency levels. In addition, it is useful for test developers, in order to select appropriate texts according to the desired proficiency level and include them in the reading sections of the examinations. In general, everyone involved in writing texts for various audiences needs certain guidelines regarding the factors that make a text more or less accessible. For these reasons, we consider to be of the utmost importance to study the potential of this component and include it in our approach. By incorporating this component, we hypothesize that the difficulty

---

[9]https://universaldependencies.org/u/pos/
[10]https://universaldependencies.org/u/feat/index.html

of a given text will have direct implications in the difficulty of the questions that will arise from it. Susanti et al. [138] have already used reading passage difficulty as a factor to control item difficulty for automatic vocabulary question generation (in English). For that, they have applied three measures to calculate the difficulty (or readability) of their reading passage sources, namely the Flesh-Kincaid Grade Level, Fresh-Kincaid Reading Ease [69] and Dale-Chall readibility formula [26]. We present a brief explanation for each of these measures:

- **Flesch–Kincaid Reading Ease**: Here, higher scores indicate material that is easier to read; lower numbers mark passages that are more difficult to read. The formula for the Flesch reading ease score test is:

$$206.835 - 1.015 * (\frac{total\ words}{total\ sentences}) - 84.6 * (\frac{total\ syllables}{total\ words})$$

- **Flesch–Kincaid Grade Level**: It presents a score as a U.S. grade level. It can also mean the number of years of education generally required to understand a text. The grade level is calculated with the following formula:

$$(0.39 * \frac{total\ words}{total\ sentences}) + (11.8 * \frac{total\ syllables}{total\ words}) - 15.59$$

- **Dale–Chall readability formula**: It is a readability test that provides a numeric gauge of the comprehension difficulty that readers come upon when reading a text. It uses a list of 3000 words that groups of fourth-grade American students could reliably understand, considering any word not on that list to be difficult. The formula for calculating the raw score of the Dale–Chall readability score is:

$$0.1579 * (\frac{difficult\ words}{total\ words} * 100) + 0.0496 * (\frac{total\ words}{total\ sentences})$$

For the Portuguese language, some study has also been done but not in the scope of question generation. Branco et al. [20] presented a tool that could support human experts in their task of classifying texts excerpts suitable to be used in quizzes and items of exams. This tool is expected to help the productivity and consistency when classifying excerpts for students taking courses of Portuguese as a second language. One of the conclusions from that study was precisely the importance of the Flesch–Kincaid Reading Ease metric in order to assess this categorization task. Given this, we decided to use this metric as being useful for classifying our input texts regarding their readability. The adaptation of the Flesch Reading Ease to Portuguese (BR) resulted in the formula [130]:

$$248.835 - 1.015 * (\frac{total\ words}{total\ sentences}) - 84.6 * (\frac{total\ syllables}{total\ words})$$

This adaptation corresponds to the original formula with the addition of 42 which, according to Martins et al. [94], is, on average, the number that differentiates texts in English from Portuguese

texts. The values can vary between 100-75 (very easy), 75-50 (easy), 50-25 (difficult) and 25-0 (very difficult). To count the number of syllables in the sentences (taking into account that this count varies from language to language) we used a tool called *LX-Syllabifier* [121], which performs syllabification following a rule-based approach and is implemented according to *Acordo Ortográfico da Língua Portuguesa (1990)*.

In addition to analyzing text complexity (as a whole) we also intended to find a way to analyze the difficulty of an individual sentence. We assume that two sentences that belong to the same text do not necessarily have the same difficulty. More deeply, if different questions arise from two distinct sentences, our hypothesis also concerns the possibility of these questions to have different complexities. That is the reason why we concluded it would be useful to implement a local heuristic for each sentence, given the input text. To achieve this goal and, within the same scope of the study made by Branco et al. (as mentioned above), we found another research [37] whose objective was to assist the selection of adequate reading materials for European Portuguese Teaching, especially as second language. The developed system made use of existing NLP tools to extract linguistic features from texts, which are then used by an automatic readability classifier. 52 features were extracted such as PoS, syllables, words, chunks, averages, frequencies and some extra features. The authors did some experiments to assess the contribution of the features extracted for the readability classification. We focus on the experience whose goal was to perceive the feature contribution for a three-level scale classification (A, B and C). The following features are pointed out as being some of the most important for that classification task and, therefore, we use them to assign a difficulty value for our sentences:

- **Number of words**: We understand this metric as referring to the length of a sentence. The more words it has, the more complex it will be.

- **Number of different words**: This measure is an indicator for the number of distinct words within a sentence, the more distinct words, more complex the sentence will be.

- **Number of dependencies**: The more dependencies a sentence has, the more complex it will be.

- **Numerals**. The more numerals a sentence has, the more complex it will be.

- **Average of coordinating relations chains**: The more coordinating relations a sentence has, the more complex it will be.

- **Adverbs**: The more adverbs a sentence has, the more complex it will be.

Our heuristic can be represented as follows (the weights indicated in the article had been normalized and adapted):

*diff_sentence = 0.27\*NR_WORDS + 0.24\*NR_DIFFERENT_WORDS + 0.24\*NR_DEPENDENCIES + 0.12\*NR_NUMERALS + 0.07\*NR_COORD + 0.06\*NR_ADVERBS*

We highlight the fact that we are using these two factors, **the difficulty of a text as a whole** and **individually for each sentence**, to classify the difficulty of a question. The combination of these factors is explained in Section 4.4.6. These are not the only factors that will define the difficulty of a question. Later on, when we explain the AQG process, we will detail the other factors based on the type of the generated question. To conclude, we believe that this can be a substance of future work, by including new features or by adjusting the factors.

## 4.3    Automatic Generation of Grammar Questions

Knowing grammar rules is very useful for language learning. In fact, it is the knowledge of these rules that allows us to master the mother tongue or even understand the way words are articulated in a sentence. This way, teaching the functioning of the language has a systematic and disciplinary role in the process of learning Portuguese. By mastering a set of language problems using serious linguistics, it is possible to give students effective communication skills. This normative character organizes thinking in a logical way, improving the capacity for comprehension, written expression and communication skills. Consequently, there is a need to continually encouraging the teaching of grammar which should take a form of reflection on the structure and functioning of the language, guided by a teacher. This activity of discovery is essential for mastering Portuguese, for improving and diversifying the use of language and for learning foreign languages. It is also necessary for the development of values and the training of cognitive abilities[11].

Given these reasons, we decided to implement a system capable of generating grammar questions for the Portuguese language. As far as we know, there is only one study [112] from 2012 in which the goal was to generate questions for the French grammar, therefore, we consider this as being a pioneering occurrence within AQG field. In order to achieve the objective of generating quality grammar questions, we considered necessary and prudent to consult some of the most modern Portuguese grammars/manuals [65, 87, 50, 44] and also, whenever necessary, consult teachers from the field to clarify any doubts. In fact, the process of reviewing the grammatical content and collecting teacher opinions proved to be extremely important to ensure the reliability of our research. It should be noted that the Portuguese grammar suffers, with some regularity, changes in its grammatical terminology and, therefore, it is necessary to be as rigorous as possible in order to avoid some misconceptions and fallacies. The selection criteria for choosing grammar exercise types were:

- The exercises respect the established rules from *Acordo Ortográfico da Língua Portuguesa (1990)*[12];

- The exercises are in the form of multiple-choice or gap-fill type;

---

[11]https://ciberduvidas.iscte-iul.pt/artigos/rubricas/ensino/a-importancia-da-gramatica/3813

[12]http://www.portaldalinguaportuguesa.org/acordo.php?action=acordo&version=1990

- The exercises are within the content foreseen for 3rd cycle of studies for the Portuguese subject, which corresponds to the 7th, 8th and 9th grades of the Portuguese educational system;

- The exercises allow us to create different situations of difficulty depending on their complexity[13].

We would like to highlight the fact that, in addition to the goal of automatically generating these questions, we also had the objective of ensuring they would have different difficulty levels. The notion of difficulty for grammar questions makes use of specific factors based on teachers' opinions. As such, we emphasize the importance of the tables present in the next sections, where several specific cases are listed and classified according to their complexity. Thus, over the next sections, we will explain the implementation steps of each question type, and also explaining what we did in order to control the difficulty. Over the next sections, we will follow this scheme: We will start by showing an example of an exercise entirely taken from the manuals. Then, we will be showing two generated questions with different difficulties. All generation steps will be explained in between these same illustrative examples.

### 4.3.1 Morphological Sequence

The following example (4.1) is a real question from the referred manuals (the correct answer is in bold):

---

**Example 4.1.** *Indique a frase que contém a sequência: determinante - nome - pronome - verbo - preposição - nome - adjetivo - verbo - advérbio - adjetivo - conjunção - adjetivo.*

    *A* **As crianças que gostam de histórias fantásticas são, provavelmente, destemidas e criativas.**

    *B As crianças que leem poesia também costumam ler outros géneros de literatura.*

    *C O livro que te emprestei foi comprado ontem na Feira do livro.*

    *D As histórias contadas pelas crianças são, às vezes, fruto da sua imaginação.*

---

This type of grammar question is typically composed of:

- **Instruction**: Presents a morphological sequence and asks the respondent to select the sentence whose words correspond exactly to that same sequence. This morphological sequence presents only the grammatical class for each word, it does not specify its subclass;

---

[13]This notion of difficulty was also obtained from the opinions that we have collected from the teachers.

- **Correct answer**: It consists of a sentence that fully corresponds to the morphological sequence requested by the instruction;

- **Distractors**: Each option consists of a sentence whose morphological sequence does not fully match the sequence requested from the instruction;

- **Notion of difficulty**: The complexity of these questions is directly related to the proximity of the correct morphological sequences to distractors. The closer a distractor (sequence) is to the correct option (sequence), the more difficult the question will be. On the contrary, the further apart these sequences are, the easier it will be to identify the correct answer.

The implementation steps are described in Algorithm 1.

---

**Algorithm 1** Create MCQs for Grammar - Morphological Sequence

---

1: **function** MCQSEQUENCE(*nrquest*: number of desired questions, *diff*: desired difficulty degree, *all_sents*: list of sentences)
2:     *mcqs* ← []                                         ▷ Create empty list for multiple-choice questions
3:     *all_pairs* ← []                                     ▷ Create empty list for storing pairs of sentences
4:     *sents_samesize* ← GETSENTSSAMESIZE(*all_sents*)
5:     **for** each *sents* in *sents_samesize* **do**
6:         *pairs* ← COMBINATIONS(*sents*, 2)                                         ▷ $^{len(sents)}C_2$
7:         **for** each *pair* in *pairs* **do**
8:             *pair.common* ← POSCOMMON(*pair.sent1*, *pair.sent2*)   ▷ Nr. of PoS in common
9:         **end for**
10:        **if** *diff* == "MAX" **then**
11:            sort list *pairs* in descending order of *pair.common*
12:        **else if** *diff* == "MIN" **then**
13:            sort list *pairs* in ascending order of *pair.common*
14:        **else**
15:            sort list *pairs* randomly
16:        **end if**
17:        *all_pairs*.EXTEND(*pairs*)
18:     **end for**
19:     **while** *nrquest* > 0 **and** *len(all_pairs)* >= 2 **do**
20:         *pairs_2* ← GETTWOPAIRS(*all_pairs*)
21:         *mcq* ← CREATEMCQ(*pairs_2*)                                         ▷ Create MCQ with 2 pairs (4 sentences)
22:         *mcqs*.APPEND(*mcq*)
23:         *nrquest* ← *nrquest* − 1
24:     **end while**
25:     **return** *mcqs*
26: **end function**

---

The main idea for this algorithm is to join pairs of sentences that have the same size (Line 4), and count the number of PoS they have in common (Line 8). The MCQ will, therefore, contain 2 pairs of sentences: 4 sentences in total (Line 21). Only one of them (the correct answer) will correspond exactly to the requested morphological sequence. Distractors will have a close

morphological sequence (if the difficulty is maximized, Line 11) or vice versa (if the difficulty is minimized, Line 13). The heuristic that calculates the difficulty of the question is:

$$diff\_quest = avg\_count$$

where *avg_count* is the average count of the same PoS tags between the correct answer and its distractors. The bigger the *avg_count*, the more complex the question. In the Examples 4.2 and 4.3, we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.2.** *Indique a frase que contém a sequência: nome-conjunção-determinante-nome-verbo-preposição-nome-preposição-nome.*

   A  *Foi baptizado de urgência recebendo o nome de Sirena.*

   B  **Hans e a família moravam no interior da ilha.**

   C  *A língua estrangeira fechava em sua roda um círculo.*

   D  *A vida de Hans mais uma vez tinha virado.*

---

Option (B) is the correct answer. Option (A) has 5 PoS tags in common with the correct answer. Option (C) has 2 PoS tags in common. Option (D) has 2 PoS tags in common. All sentences have 9 words. Thus,

$$diff\_quest = ((5/9) + (2/9) + (2/9))/3 = 0.33$$

---

**Example 4.3.** *Indique a frase que contém a sequência: determinante-nome-verbo-preposição-determinante-nome.*

   A  *A renúncia endurecia os seus músculos.*

   B  *A praia estava cheia de gente.*

   C  **A tarde corria sobre o rio.**

   D  *Agora as gaivotas recolhiam a terra.*

---

Option (C) is the correct answer. Option (A) has 5 PoS tags in common with the correct answer. Option (B) has 4 PoS tags in common. Option (D) has 1 PoS tags in common. All sentences have 6 words. Thus,

$$diff\_quest = ((5/6) + (4/6) + (1/6))/3 = 0.56$$

### 4.3.2   Determinants, Pronouns, Adverbs and Conjunctions

Example 4.4 is a real question from the referred manuals (the correct answer is in bold and the important words are underlined):

---

**Example 4.4.** *Assinale a única frase que contém um advérbio de tempo.*

   A  ***Ontem fui ao cinema.***

   B  *Eis a minha comida favorita!*

   C  *Eles não levaram a chave.*

   D  *Vai devagar!*

---

This type of grammar question is typically composed of:

- **Instruction**: Asks to choose the unique sentence that contains a word of a certain grammatical class and subclass;

- **Correct answer**: It consists of the unique sentence that contains the grammatical class and subclass as requested in the instruction of the question;

- **Distractors**: Each of the options is a sentence which contains a word of the same grammatical class as that of the instruction, but the subclass is necessarily different;

- **Notion of difficulty**: The difficulty of the question is defined by the complexity of the correct answer, that is, the more complex the class and subclass requested, the more difficult the question will be. For example, students often find it easier to identify a definite article determinant than a demonstrative determinant. Also, the complexity is also related to the difficulty of identifying the classes and subclasses of the words from distractors.

The implementation steps can be found in Algorithm 2.

The main idea for this algorithm is to obtain sentences that contain a word with the desired grammatical class (Line 3). If the goal is to maximize the difficulty of these questions, then the algorithm will sort the sentences in decreasing order by the difficulty (Line 8) of the grammatical class (of the word contained therein). If the goal is to minimize the difficulty of these questions, then the algorithm will sort them (the sentences) in an increasing (Line 10) order. All specific cases can be seen analyzed in Table 4.3. The correct answer (Line 15) is a sentence that contains a word with the required grammatical class. Distractors (Line 19) are sentences that also contain

---

**Algorithm 2** Create MCQs for Grammar - Determinants, Pronouns, Adverbs and Conjunctions

---

1: **function** MCQDETPRONADVCONJ(*nrquest*: desired number of questions, *diff*: desired difficulty degree, *all_sents*: list of sentences, *class*: desired word class)
2:   *mcqs* ← [] ▷ Create empty list for multiple-choice questions
3:   *sents_sameclass* ← GETSAMECLASS(*all_sents*, *class*) ▷ Get all sentences that contain a word from the desired class
4:   **for** each *sent* in *sents_sameclass* **do**
5:     *sent.diff* ← CALCDIFF(*sent*) ▷ Assign difficulty according to Table 4.3
6:   **end for**
7:   **if** *diff* == "MAX" **then**
8:     sort *sents_sameclass* in decreasing order of *sent.diff*
9:   **else if** *diff* == "MIN" **then**
10:     sort *sents_sameclass* in ascending order of *sent.diff*
11:   **else**
12:     sort *sents_sameclass* randomly
13:   **end if**
14:   **while** *nrquest* > 0 **and** *len*(*sents_sameclass*) >= 4 **do**
15:     *correct_answer* ← *sents_sameclass*.*pop*(0) ▷ Extract the first sentence from top
16:     *distractors* ← [] ▷ Create empty list for distractors
17:     **for** each *sent* in *sents_sameclass* **do**
18:       **if** *sent.word.subclass* != *correct_answer.word.subclass* **then**
19:         *distractors*.APPEND(*sent*)
20:       **end if**
21:       **if** *len*(*distractors*) == 3 **then**
22:         *break*
23:       **end if**
24:     **end for**
25:     **if** *len*(*distractors*) == 3 **then**
26:       *mcq* ← CREATEMCQ(*correct_answer*, *distractors*)
27:       *mcqs*.APPEND(*mcq*)
28:       *nrquest* ← *nrquest* − 1
29:     **end if**
30:   **end while**
31:   **return** *mcqs*
32: **end function**

---

a word from that grammatical class, but the subclass is necessarily different. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = 0.6 * diff\_ans + 0.4 * diff\_dist$$

where *diff_ans* is the difficulty value attributed to the correct answer and *diff_dist* is the average difficulty value attributed to all distractors. All difficulty values are listed in Table 4.3. Again, these values were obtained experimentally through the Teachers' opinions. The bigger the *diff_ans* and *diff_dist*, the more complex will the question. 0.6 and 0.4 were also obtained experimentally. The value of 0.6 allows to give greater importance to the difficulty of the correct answer and 0.4 also intends to give importance (but not as prominent with the last one) to the multiple-choice distractors.

In the examples 4.5 and 4.6, we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.5.** *Assinale a única frase que contém uma conjunção subordinativa causal.*

   *A* **Manda-me o teu pai que te diga que não voltes a Vig <u>pois</u> não te receberá.**

   *B* *Unido ao balanço, Hans, enquanto lavava o convés, polia os metais <u>ou</u> enrolava os cabos, aspirava a 4 veemência da vasta respiração marítima.*

   *C* *Assim é desde o tempo antigo das guerras quando os invasores que ocupavam a ilha penetravam nas casas de cabeça erguida <u>mas</u> exigiam que a gente da ilha se curvasse para os saudar.*

   *D* *No entanto parecia a Hans que algo em sua vida, <u>embora</u> fosse já tão tarde, era ainda espera e espaço aberto, possibilidade .*

---

Option (A) contains a causal subordinating conjunction which is the correct answer. Its value for the difficulty is equal to 2/7. Option (B) contains a disjunctive coordinating conjunction. Its value for the difficulty is equal to 5/5. Option (C) contains an adversative coordinating conjunction. Its value for the difficulty is equal to 4/5. Finally, option (D) contains a concessive subordinating conjunction. Its value for the difficulty is equal to 6/7. Thus,

$$diff\_quest = 0.6 * (2/7) + 0.4 * ((5/5 + 4/5 + 6/7)/3) = 0.53$$

---

**Example 4.6.** *Assinale a única frase que contém um determinante indefinido.*

Table 4.3: Attributed difficulty values per Class and Subclass.

| Class and Subclass | Attributed Difficulty Value | Class and Subclass | Attributed Difficulty Value | Class and Subclass | Attributed Difficulty Value |
|---|---|---|---|---|---|
| Adverb - Negation | 1/7 | Determinant - definite article | 1/5 | Pronoun - personal | 1/5 |
| Adverb - Affirmation | 1/7 | Determinant - indefinite article | 1/5 | Pronoun - indefinite | 5/5 |
| Adverb - Degree | 4/7 | Determinant - possessive | 2/5 | Pronoun - possessive | 2/5 |
| Adverb - Mode | 3/7 | Determinant - demonstrative | 3/5 | Pronoun - demonstrative | 3/5 |
| Adverb - Time | 2/7 | Determinant - relative | 5/5 | Pronoun - relative | 5/5 |
| Adverb - Local | 2/7 | Determinant - interrogative | 4/5 | Pronoun - interrogative | 4/5 |
| Adverb - Doubt | 4/7 | Determinant - indefinite | 5/5 | | |
| Adverb - Inclusion | 5/7 | | | | |
| Adverb - Exclusion | 5/7 | | | | |
| Adverb - Designation | 6/7 | | | | |
| Adverb - Relative | 7/7 | | | | |
| Adverb - Interrogative | 1/7 | | | | |
| Adverb - Connective | 7/7 | | | | |
| | | | | | |
| Coord. Conj. copulative | 1/5 | Subord. Conj. completive | 7/7 | | |
| Coord. Conj. adversative | 4/5 | Subord. Conj. causal | 2/7 | | |
| Coord. Conj. disjunctive | 5/5 | Subord. Conj. final | 6/7 | | |
| Coord. Conj. conclusive | 5/5 | Subord. Conj. temporal | 3/7 | | |
| Coord. Conj. explanative | 5/5 | Subord. Conj. concessive | 6/7 | | |
| | | Subord. Conj. comparative | 6/7 | | |
| | | Subord. Conj. concecutive | 7/7 | | |

> *A* <u>*Este*</u> *estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de arrebatado e selvático, tornou-se depressa um dos monumentos famosos da cidade e vinha gente das redondezas para o ver.*
>
> *B* **Escolhe <u>outra</u> coisa.**
>
> *C* *No entanto parecia a Hans que algo em <u>sua</u> vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade.*
>
> *D* *Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe <u>as</u> roupas de que precisava e também papel e caneta.*

Option (A) contains a demonstrative determinant. Its value for the difficulty is equal to 3/5. Option (B) contains a indefinite determinant which is the correct answer. Its value for the difficulty is equal to 5/5. Option (C) contains a possessive determinant. Its value for the difficulty is equal to 2/5. Finally, option (D) contains a determinant definite article. Its value for the difficulty is equal to 1/5. Thus,

$$diff\_quest = 0.6 * (5/5) + 0.4 * ((3/5 + 2/5 + 1/5)/3) = 0.76$$

### 4.3.3 Determinants, Pronouns and Prepositions

Example 4.7 is a real question from the referred manuals (correct answer is in bold and the important words/letters are underlined).

> **Example 4.7.** *Assinale a única frase em que a letra "a" é uma preposição:*
>
> *A* **Gosto de ir <u>a</u> casa da minha tia.**
>
> *B* *Penso que <u>a</u> casa fica muito longe.*
>
> *C* *Ele comprou uma casa nova e remodelou-<u>a</u>.*
>
> *D* *Esta casa é grande, mas <u>a</u> que visitamos ontem ainda é maior.*

This type of grammar question is typically composed of:

- **Instruction**: Asks to identify the unique sentence in which the word *a* is a determinant, pronoun or preposition.

- **Correct answer**: It consists of the unique sentence in which the word *a* corresponds to the grammatical class requested by the instruction of the question.

- **Distractors**: Each of the options contemplates sentences with the word *a* as being a deter-
  minant, pronoun or preposition. This grammatical class cannot be the same as the correct
  answer.

- **Notion of difficulty**: The notion of difficulty in this question is defined by the complexity of
  the right answer. The more complex it is to identify the word *a* given its grammatical class,
  the more difficult the question becomes. For example, students often find it more difficult
  to identify *a* as a preposition than as a determinant. In addition, the difficulty in identifying
  the class for word *a* from distractors also contributes to define question's complexity.

The implementation steps can be found in Algorithm 3.

---

**Algorithm 3** Create MCQs for Grammar - Prepositions, Determinants and Pronouns for word "a"

---

1: **function** MCQPREPDETPRON(*nrquest*: desired number of questions, *diff*: desired difficulty
   degree, *all_sents*: list of sentences)
2:     *mcqs* ← []           ▷ Create empty list for multiple-choice questions
3:     *sents* ← GETPREPDETPRON(*all_sents*)    ▷ Get sentences that have the word "a" as a
   preposition, determinant or pronoun
4:     **for** each *sent* in *sents* **do**
5:         *sent.diff* ← CALCDIFF(*sent*)       ▷ Assign difficulty according to Table 4.4
6:     **end for**
7:     **if** *diff* == "MAX" **then**
8:         sort *sents* in decreasing order of *sent.diff*
9:     **else if** *diff* == "MIN" **then**
10:        sort *sents* in ascending order of *sent.diff*
11:     **else**
12:        sort *sents* randomly
13:     **end if**
14:     **while** *nrquest* > 0 **and** *len*(*sents*) >= 4 **do**
15:        *correct_answer* ← *sents.pop*(0)       ▷ Extract the first sentence from top
16:        *distractors* ← []           ▷ Create empty list for distractors
17:        **for** each *sent* in *sents* **do**
18:           **if** *sent.word.class* != *correct_answer.word.class* **then**
19:             *distractors*.APPEND(*sent*)
20:           **end if**
21:           **if** *len*(*distractors*) == 3 **then**
22:             *break*
23:           **end if**
24:        **end for**
25:        **if** *len*(*distractors*) == 3 **then**
26:           *mcq* ← CREATEMCQ(*correct_answer, distractors*)
27:           *mcqs*.APPEND(*mcq*)
28:           *nrquest* ← *nrquest* − 1
29:        **end if**
30:     **end while**
31:     **return** *mcqs*
32: **end function**

---

The main idea for this algorithm is to obtain sentences that contain the word "a" as preposition, determinant or pronoun (Line 3). If the goal is to maximize the difficulty of these questions, then the algorithm will sort the sentences in decreasing (Line 8) order by the difficulty of the grammatical class (concerning the word "a"). If the goal is to minimize the difficulty of these questions, then the algorithm will sort them (the sentences) in an increasing (Line 10) order. All specific cases can be seen analyzed in Table 4.4. The correct answer (Line 15) is a sentence that contains the word "a" as preposition, determinant or pronoun. Distractors (Line 19) are sentences that also contain the same word, but its class is necessarily different. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = 0.6 * diff\_ans + 0.4 * diff\_dist$$

where *diff_ans* is the difficulty value attributed to the the correct answer and *diff_dist* is the average difficulty value attributed to all distractors. All possible values are listed according to the Table 4.4. The bigger the *diff_ans* and *diff_dist*, the more complex the question. 0.6 and 0.4 were obtained experimentally. The value of 0.6 allows to give greater importance to the difficulty of the correct answer and 0.4 also intends to give importance (not as prominent with the last one) to the multiple-choice distractors. In the next examples (4.8, 4.9), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.8.** *Assinale a frase em que a letra "a" é uma determinante:*

   A **Hans fitou a̲ toalha.**

   B *De repente, começou a̲ chover.*

   C *Porque deles se desprendia cheiro a̲ mar.*

   D *À noite relatava a̲ Hoyle as conversas que tivera, as decisões que tomara.*

---

Option (A) contains the word **a** as a determinant and it is the correct answer. Its value for difficulty is equal to (1/4). Option (B), (C) and (D) contain the word **a** as a preposition. Its value for difficulty is equal to 4/4. Thus,

$$diff\_quest = 0.6 * (1/4) + 0.4 * ((4/4 + 4/4 + 4/4)/3) = 0.55$$

---

**Example 4.9.** *Assinale a frase em que a letra "a" é um pronome:*

   A *De repente, começou a̲ chover.*

   B *Porque deles se desprendia cheiro a̲ mar.*

---

> C  *À noite relatava <u>a</u> Hoyle as conversas que tivera, as decisões que tomara.*
>
> D  **Pois algo na sua cara <u>a</u> fascinava e inquietava.**

Option (A), (B) and (C) contain the word ***a*** as a preposition. Its value for difficulty is equal to (4/4). Finally, option (D) contains the word ***a*** as a pronoun (before verb), which is the correct answer. Its value for difficulty is equal to (2/4). Thus,

$$diff\_quest = 0.6*(2/4)+0.4*((4/4+4/4+4/4)/3) = 0.7$$

Table 4.4: Attributed difficulty value by Grammatical Class (only when comparing prepositions, determinants and pronouns for the word "a").

| Grammatical Class | Attributed difficulty value |
|---|---|
| Pronoun (before verb) | 2/4 |
| Pronoun (couple within verb) | 3/4 |
| Determinant | 1/4 |
| Preposition | 4/4 |

### 4.3.4  About Verbs

This section presents a set of generated questions that focus on testing knowledge about verbs. This way, the first 3 subsections (4.3.4.1, 4.3.4.2 and 4.3.4.3) share a first common step. The purpose of this common step is to find verbs, located along with the sentences, and to keep all the information about those verbs regarding their tense, mood, person and number. As such, Table 4.5 contains useful information for these three types of questions, in relation to their difficulty.

#### 4.3.4.1  Choosing the Correct Tense and Mood Given a Verb

The following example (4.10) is a real question from the referred manuals (correct answer is in bold):

> **Example 4.10.** *"Eles compuseram uma nova música." A forma verbal "compuseram" encontra-se no*
>
> A  **pretérito perfeito do indicativo.**
>
> B  *pretérito imperfeito do indicativo.*
>
> C  *pretérito mais-que-perfeito do indicativo.*
>
> D  *pretérito imperfeito do conjuntivo.*

Table 4.5: Attributed difficulty values per verb's tense and mood.

| Mood | Attributed difficulty value |
|---|---|
| Indicative | 1/3 |
| Conjunctive | 2/3 |
| Conditional | 3/3 |
| **Simple Tense** | - |
| Present | 1/6 |
| Past Perfect | 2/6 |
| Imperfect | 2/6 |
| Pluperfect | 4/6 |
| Future | 3/6 |
| **Compound Tense** | - |
| Past Perfect | 5/6 |
| Past (Compound Conditional) | 6/6 |
| Pluperfect | 6/6 |
| Future | 6/6 |

This type of grammar question is typically composed of:

- **Instruction**: Asks to identify the tense and mood a given verbal form.

- **Correct answer**: It is the set which contemplates the tense and mood requested from the instruction of the question;

- **Distractors**: Each of the options has a different verbal tense or mood in relation to the correct answer;

- **Notion of difficulty**: The difficulty of these types of questions is identified with the greater or lesser ease for identifying certain tenses and verbal moods. There are tense and moods that students find it easier to identify. For example, students find it more difficult to identify Pluperfect tense (Indicative mood) than the Present (Indicative mood). Again, from Table 4.5, we organize all possible cases.

The implementation steps are described Algorithm 4.

The main idea of this algorithm is to obtain pairs of **sentence, verb** (Line 3) from all input sentences. If the goal is to maximize the difficulty of these questions, then the algorithm will sort the sentences in decreasing (Line 8) order by the difficulty of their verbs. If the goal is to minimize the difficulty of these questions, then the algorithm will sort them (the sentences) in an increasing (Line 10) order. All specific cases can be seen analyzed in Table 4.5. The correct answer (Line 15) is a set with a specific tense and mood. Distractors (Line 16) are composed of different sets of tenses and moods. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = (1/2) * diff\_tense + (1/2) * diff\_mood$$

---

**Algorithm 4** Create MCQs for Grammar - Choosing The Correct Tense and Mood Given a Verb

---

1: **function** MCQTENSEMOOD(*nrquest*: desired number of questions, *diff*: desired difficulty degree, *all_sents*: list of sentences, *tenses_moods*: list of all tenses and moods)

2:      *mcqs* ← []                         ▷ Create empty list for multiple-choice questions

3:      *sents* ← GETSENTSVERBS(*all_sents*)   ▷ Get pairs of **sentence-verb** from all sentences

4:      **for** each *sent* in *sents* **do**

5:          *sent.diff* ← CALCDIFF(*sent*)              ▷ Assign difficulty according to Table 4.5

6:      **end for**

7:      **if** *diff* == "MAX" **then**

8:          sort *sents* in decreasing order of *sent.diff*

9:      **else if** *diff* == "MIN" **then**

10:         sort *sents* in ascending order of *sent.diff*

11:      **else**

12:         sort *sents* randomly

13:      **end if**

14:      **while** *nrquest* > 0 **and** *len*(*sents*) >= 0 **do**

15:         *correct_answer* ← *sents.pop*(0)          ▷ Extract the first element from top

16:         *distractors* ← GETDISTRACTORS(*correct_answer*,*tenses_moods*)   ▷ Get tenses and moods that can act as distractors given the correct answer

17:         *mcq* ← CREATEMCQ(*correct_answer*,*distractors*)

18:         *mcqs.*APPEND(*mcq*)

19:         *nrquest* ← *nrquest* − 1

20:      **end while**

21:      **return** *mcqs*

22: **end function**

---

where *diff_tense* is the difficulty value attributed to verb's tense and *diff_mood* to its mood. In the next examples (4.11, 4.12), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.11.** *"Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...)." A forma verbal "ouvia" encontra-se no*

   A  *futuro do modo indicativo.*

   B  **pretérito imperfeito do modo indicativo.**

   C  *futuro composto do modo conjuntivo.*

   D  *presente do modo condicional.*

---

Options (A), (C) and (D) are defined by the different combinations of verbal tenses and moods. None of them is the correct answer. Option (B) is the correct answer. Its difficulty value is calculated as follows:

$$diff\_quest = (1/2) * (2/6) + (1/2) * (1/3) = 0.33$$

---

**Example 4.12.** *"Embora, em rigor tudo tivesse sido possível." A forma verbal "tivesse sido" encontra-se no*

   A  *futuro composto do modo indicativo.*

   B  *pretérito imperfeito do modo conjuntivo.*

   C  *pretérito imperfeito do modo indicativo.*

   D  **pretérito mais-que-perfeito composto do modo conjuntivo.**

---

Options (A), (B) and (C) are defined by the different combinations of verbal tenses and moods. None of them is the correct answer. Option (D) is the correct answer. Its difficulty value is calculated as follows:

$$diff\_quest = (1/2) * (6/6) + (1/2) * (2/3) = 0.83$$

### 4.3.4.2  Complete with the indicated tenses and moods

The following examples (4.13) are real questions from the referred manuals:

---

**Example 4.13.** *Completa cada uma das frases seguintes com a forma do verbo indicado entre parênteses, no tempo e modo referidos.*

A **Pretérito imperfeito do modo indicativo**
   *Em criança, eu _____ (admirar) os livros com muitas imagens.*

B **Pretérito perfeito composto do modo conjuntivo**
   *Os alunos _____ (ler) muito por incentivo da professora de Português.*

C **Pretérito perfeito simples do conjuntivo**
   *Se _____ (ter) muito dinheiro, ajudaria crianças de todo o mundo.*

---

This type of grammar question is typically composed of:

- **Instruction**: Ask to complete the sentence with the requested verbal form. The verb must be completed with a certain tense and mood.

- **Correct answer**: The correct option will be the conjugation of the verb in the tense and mood requested from the instruction.

- **Notion of difficulty**: The difficulty of this type of question is related (as in Section 4.3.4.1) with the greater or lesser ease for identifying certain tenses and verbal moods.

In order to generate these types of questions, we adapted Algorithm 4. Over here, the difference is that we reveal the complete sentence and then we hide the place where the verb is located, placing it ahead in parentheses and lemmatized. Hence, the heuristic function that calculates the difficulty of the question is:

$$diff\_quest = (1/2) * diff\_tense + (1/2) * diff\_mood$$

where *diff_tense* is the difficulty value attributed to the verb's tense and *diff_mood* to its mood. In the next examples (4.14, 4.15), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.14.** *Completa a seguinte frase com a forma do verbo indicado entre parênteses, no tempo e modo referidos.*
**Pretérito perfeito do modo indicativo**
*No entanto, o navio _____ (naufragar) quando a experiência e o cálculo não mediram exactamente a força e a proximidade do temporal.*

---

$$diff\_quest = (1/2) * (2/6) + (1/2) * (1/3) = 0.33$$

---

**Example 4.15.** *Completa a seguinte frase com a forma do verbo indicado entre parênteses, no tempo e modo referidos.*

**Pretérito mais-que-perfeito composto do modo indicativo**

*Dizia-se mesmo que nesse dia _____ (chicotear) o mar.*

---

$$diff\_quest = (1/2) * (6/6) + (1/2) * (1/3) = 0.67$$

### 4.3.4.3 Choosing which verbal forms belong to the same verbal mood

The following example (4.16) is a real question from the referred manuals (the correct answer is in bold):

---

**Example 4.16.** *O conjunto constituído apenas por formas verbais que pertencem ao mesmo modo verbal é:*

   A  **lêssemos - tenha escrito - falem - pintarem**

   B  *haja - intervenham - tiver discutido - conversara*

   C  *interveio - propuseram - leríamos - tivesse estudado*

   D  *corríamos - saltávamos- brincaríamos - rirmos*

---

This type of grammar question is typically composed of:

- **Instruction**: Asks to choose the unique option that contains the set of verbal forms that belong to the same mood.

- **Correct answer**: It is the option that includes the set of verbal forms that belong to the same mood.

- **Distractors**: Each option contains different verbal forms. Some of them can belong to the same tense and mood, but there is no set which has all its verbs within the same mood.

- **Notion of difficulty**: The difficulty of this type of question is related (as in Section 4.3.4.1) with the greater or lesser ease for identifying certain tenses and verbal moods.

In order to generate these types of questions, we adapted Algorithm 4. Over here, the difference is that we now create groups of verbal forms. These groups, therefore, constitute our multiple-choice options. The option that contains a group composed of verbal forms that belong to the same mood, is the correct answer. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = avg\_diff$$

where *avg_diff* is the average difficulty values of the verbal forms that compose the correct answer. The difficulty of the verbal forms are calculated according to the values from Table 4.5. In the next examples (4.17, 4.18), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.17.** *O conjunto constituído apenas por formas verbais que pertencem ao mesmo modo verbal é:*

   A  **dera - abria - naufragou**

   B  *perdera - estendesse - sacudisse*

   C  *parecera - tivesse - diga*

   D  *fizesse - proteja - estendiam*

---

Each of the groups from options (B), (C) and (D) contain verbal forms that do not share the same verbal mood. They are distractors. Option (A) has 3 verbal forms that share the same mood. Thus, it is the correct answer. To calculate the difficulty of this question, we first need to calculate the difficulty of each of the verbal forms that make up the correct answer:

$$diff\_form1 = (1/2) * (4/6) + (1/2) * (1/3) = 0.50$$
$$diff\_form2 = (1/2) * (2/6) + (1/2) * (1/3) = 0.33$$
$$diff\_form3 = (1/2) * (2/6) + (1/2) * (1/3) = 0.33$$

Thus, the difficulty of the question is calculated as follows:

$$diff\_quest = (0.50 + 0.33 + 0.33)/3 = 0.39$$

---

**Example 4.18.** *O conjunto constituído apenas por formas verbais que pertencem ao mesmo modo verbal é:*

> A **tinha chicoteado - olhara - levarei**
>
> B  tinha atirado - tinha projetado - estendesse
>
> C  tinha virado - tinha encalhado - fizesse
>
> D  tinha sido - navegara - sacudisse

Similarly to the previous example:

$$diff\_form1 = (1/2) * (6/6) + (1/2) * (1/3) = 0.67$$

$$diff\_form2 = (1/2) * (4/6) + (1/2) * (1/3) = 0.5$$

$$diff\_form3 = (1/2) * (3/6) + (1/2) * (1/3) = 0.42$$

Thus, the difficulty of the question is calculated as follows:

$$diff\_quest = (0.67 + 0.5 + 0.42)/3 = 0.53$$

### 4.3.4.4   Choosing the subclass of the indicated verbal forms

In the Portuguese language, the main verbs are the nucleus, the most important element of the verbal group, of the sentences in which they occur. Like the main verbs, copulative verbs are the core of the sentence but serve as a link between the subject and the subject's predicative. These are (for Portuguese): *ser, estar, parecer, permanecer, ficar, continuar, tornar-se, revelar-se*. Auxiliary verbs precede the main or copulative verbs, forming with them the compound and the passive tenses.

The following example (4.19) is a real question from the referred manuals (correct answer is in bold):

> **Example 4.19.** *"Só aquele que estou a escrever é feito por mim, os restantes ..."*
> *Identifique corretamente as subclasses dos verbos presentes na frase.*
>
> A  *um verbo copulativo e um verbo principal.*
>
> B  *dois verbos auxiliares.*
>
> C  **um verbo auxiliar e um verbo principal.**
>
> D  *dois verbos principais.*

This type of grammar question is typically composed of:

- **Instruction**: Asks to choose the unique option that composes a sentence according to its verbal forms regarding its classes and subclasses;

- **Correct answer**: It is the option that specifies the case where the class and subclass of the verbal forms correspond to the ones presented form the sentence;

- **Distractors**: They specify specific cases that do not represent the subclasses presented in the target verbs.

- **Notion of difficulty**: The difficulty of this question type is related with the difficulty of the correct answer. For example, it can be more difficult to identify a sentence that has 1 auxiliary verb and 1 copulative verb than a sentence that has 2 main verbs.

In order to generate these types of questions, we can observe the steps from algorithm 5.

---

**Algorithm 5** Create MCQs for Grammar - Choosing The Subclass of the indicated verbal forms

---

1: **function** MCQVERBSSUBCLASS(*nrquest*: desired number of questions, *diff*: desired difficulty degree, *all_sents*: list of sentences, *main_cop_aux*: list of all cases as specified in Table 4.6)
2:     *mcqs* ← [] ▷ Create empty list for multiple-choice questions
3:     *sents_verbs* ← GETSENTSCASES(*all_sents*) ▷ Get sentences that have a specific set of 2 verbal forms according to the cases specified in Table 4.6
4:     **for** each *sent* in *sents_verbs* **do**
5:         *sent.diff* ← CALCDIFF(*sent*) ▷ Assign difficulty according to Table 4.6
6:     **end for**
7:     **if** *diff* == "MAX" **then**
8:         sort *sents_verbs* in decreasing order of *sent.diff*
9:     **else if** *diff* == "MIN" **then**
10:         sort *sents_verbs* in ascending order of *sent.diff*
11:     **else**
12:         sort *sents_verbs* randomly
13:     **end if**
14:     **while** *nrquest* > 0 **and** *len*(*sents_verbs*) >= 0 **do**
15:         *elem_top* ← *sents_verbs.pop*(0) ▷ Extract the first element from top
16:         *sent, case* ← GETSENTCASE(*elem_top*) ▷ Get sentence and its specific case
17:         *distractors* ← GETDISTRACTORS(*elem_top, main_cop_aux*) ▷ Get distractors given the correct answer
18:         *mcq* ← CREATEMCQ(*sent, case, distractors*)
19:         *mcqs.*APPEND(*mcq*)
20:         *nrquest* ← *nrquest* − 1
21:     **end while**
22:     **return** *mcqs*
23: **end function**

---

The main idea for this algorithm is to obtain sentences that have a specific set of 2 verbal forms (Line 3), according to the cases specified in Table 4.6. Regarding this table, both cases 1 and 2 contain 1 auxiliary verb and 1 main verb. The difference is, for case 2, that the auxiliary

Table 4.6: Attributed difficulty values for each specific case - main, copulative and auxiliary verbs.

| Grammatical Class and Subclass | Attributed difficulty value |
|---|---|
| 2 Main Verbs | 1/5 |
| 1 Copulative Verb and 1 Main Verb | 2/5 |
| 1 Auxiliary Verb and 1 Main Verb - Case 1 | 3/5 |
| 1 Auxiliary Verb and 1 Copulative Verb | 4/5 |
| 1 Auxiliary Verb and 1 Main Verb - Case 2 | 5/5 |

verb is the verb *ser (to be)* and it is in a passive sentence. If the goal is to maximize the difficulty of these questions, then the algorithm will sort the sentences in decreasing (Line 8) order by the difficulty of their specific cases. If the goal is to minimize the difficulty of these questions, then the algorithm will sort them (the sentences) in an increasing (Line 10) order. The correct answer (Line 16) has 2 verbal subclasses that comply with the requested 2 verbal forms. Distractors (Line 17) are composed of different verbal subclasses. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = diff\_case$$

where *diff_case* is the difficulty value for the case that composes the correct answer. In the next examples (4.20, 4.21), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.20.** *"A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si próprio envolvia."*
*Identifique corretamente as subclasses dos verbos presentes na frase.*

   *A  dois verbos principais*

   *B  **um verbo copulativo e um verbo principal***

   *C  um verbo auxiliar e um verbo principal*

   *D  um verbo auxiliar e um verbo copulativo*

---

Options (A), (C) and (D) describe cases that do not correspond to the ones represented from

the correct answer, therefore, they are distractors. Option (B) represents the correct answer as it has the case specified by the verbal forms from the instruction. Thus,

$$diff\_quest = 2/5 = 0.4$$

---

**Example 4.21.** *"Embora, em rigor tudo tivesse sido possível."*
*Identifique corretamente as subclasses dos verbos presentes na frase.*

    A  *um verbo copulativo e um verbo principal*

    B  *dois verbos principais*

    C  *um verbo auxiliar e um verbo principal*

    D  **um verbo auxiliar e um verbo copulativo**

---

Similarly to the previous example,

$$diff\_quest = 4/5 = 0.8$$

### 4.3.4.5   Simple and Complex Sentences

In the Portuguese language, simple sentences are composed of the main verb or a copulative verb (combined or not with auxiliary verbs). Complex sentences integrate more than one main or copulative verb (combined or not with auxiliary verbs). The following example (4.22) is a real question from the referred manuals (the correct answer is in bold and the important words are underlined):

---

**Example 4.22.** *Assinale a única frase complexa das quatro apresentadas.*

    A  *Ao longo de algumas horas, ele <u>observou</u> aquela linda rapariga de olhar triste e cansado.*

    B  *Os simpáticos concorrentes daquele concurso de televisão <u>vieram</u> de longe.*

    C  *<u>Come</u> mais uma fatia deste maravilhoso bolo de laranja com cobertura de chocolate.*

    D  **<u>Lê</u> o texto que <u>escreveste.</u>**

---

- **Instruction**: Asks to identify the unique sentence that is simple or complex among the listed alternatives;

- **Correct answer**: It is a simple or complex sentence, depending on what is requested from question's instruction;

- **Distractors**: Each of the options is a complex sentence, if the correct option is a simple sentence. In the case of the correct option being a complex sentence, all distractors will be simple sentences;

- **Notion of difficulty**: The notion of difficulty for this question is related to the complexity of identifying specific cases. For example, students can better identify a simple sentence composed of a single main verb than a simple sentence composed of an auxiliary verb and a copulative verb (for this case, they appear consecutively in the sentence).

The implementation steps are described from algorithm 6.

The main idea for this algorithm is to obtain simple or complex sentences (Line 3), regarding the cases specifies in Table 4.7. If the goal is to maximize the difficulty of these questions, then the algorithm will sort the sentences in decreasing (Line 8) order by the difficulty of their specific cases. If the goal is to minimize the difficulty of these questions, then the algorithm will sort them (the sentences) in an increasing (Line 10) order. The correct answer (Line 15) is the option that has the sentence's case (simple or complex) requested by the instruction. Distractors (Line 19) are complex sentences (if the correct answer is simple), or vice versa. The heuristic that calculates the difficulty of the question is:

$$diff\_quest = diff\_simple\_complex$$

where *diff_simple_complex* is the difficulty value assigned to the specific case of the correct answer. In the next examples (4.23, 4.24), we can observe two generated questions with different complexities. The second is more difficult than the first.

---

**Example 4.23.** *Assinale a única frase simples das quatro apresentadas.*

  A  *O mar do Norte, verde e cinzento, rodeava Vig, a ilha, e as espumas varriam os rochedos escuros.*

  B  *Havia nesse começo de tarde um vaivém incessante de aves marítimas, as águas engrossavam devagar, as nuvens empurradas pelo vento sul acorriam e Hans viu que se estava formando a tempestade.*

  C  ***Hans concentrava o seu espírito para a exaltação crescente do grande cântico marítimo.***

  D  *Mas ele não temia a tempestade e, com os fatos inchados de vento, caminhou até ao extremo do promontório.*

---

---

**Algorithm 6** Create MCQs for Grammar - Simple or Complex

---

1: **function** MCQSIMPLECOMPLEX(*nrquest*: desired number of questions, *diff*: desired difficulty degree, *all_sents*: list of sentences)
2:     *mcqs* ← []                             ▷ Create empty list for multiple-choice questions
3:     *sents* ← GETSIMPLECOMPLEX(*all_sents*)         ▷ Get Simple or Complex sentences according to Table 4.7
4:     **for** each *sent* in *sents* **do**
5:         *sent.diff* ← CALCDIFF(*sent*)           ▷ Assign difficulty according to Table 4.7
6:     **end for**
7:     **if** *diff* == "MAX" **then**
8:         sort *sents* in decreasing order of *sent.diff*
9:     **else if** *diff* == "MIN" **then**
10:        sort *sents* in ascending order of *sent.diff*
11:     **else**
12:        sort *sents* randomly
13:     **end if**
14:     **while** *nrquest* > 0 **and** *len*(*sents*) >= 4 **do**
15:        *correct_answer* ← *sents.pop*(0)          ▷ Extract the first sentence from top
16:        *distractors* ← []                  ▷ Create empty list for distractors
17:        **for** each *sent* in *sents* **do**
18:           **if** *sent.type* != *correct_answer.type* **then**
19:             *distractors*.APPEND(*sent*)
20:           **end if**
21:           **if** *len*(*distractors*) == 3 **then**
22:             *break*
23:           **end if**
24:        **end for**
25:        **if** *len*(*distractors*) == 3 **then**
26:           *mcq* ← CREATEMCQ(*correct_answer*, *distractors*)
27:           *mcqs*.APPEND(*mcq*)
28:           *nrquest* ← *nrquest* − 1
29:        **end if**
30:     **end while**
31:     **return** *mcqs*
32: **end function**

---

Table 4.7: Attributed difficulty value per case regarding simple or complex sentences.

| Case | Attributed Difficulty Value |
|---|---|
| Simple<br>- 1 main verb<br>- 0 auxiliary verbs | 1/4 |
| Simple<br>- 1 main verb<br>- 1 auxiliary verb<br>(main and auxiliary verbs<br>are consecutive) | 2/4 |
| Simple<br>- 0 main verbs<br>- 1 copulative verb<br>- 0 auxiliary verbs | 1/4 |
| Simple<br>- 0 main verbs<br>- 1 copulative verb<br>- 1 auxiliary verb<br>(copulative and auxiliary<br>verbs are consecutive) | 3/4 |
| Complex | 4/4 |

Options (A), (B) and (D) are complex sentences, therefore, they are distractors. Option (C) is the unique simple sentence composed of a main verb and it is the correct answer.

$$diff\_quest = 1/4 = 0.25$$

---

**Example 4.24.** *Assinale a única frase complexa das quatro apresentadas.*

A  *A vida de Hans mais uma vez tinha virado.*

B  *Hans e a família moravam no interior da ilha.*

C  ***Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos.***

D  *Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).*

---

Option (A), (B) and (D) are all simple sentences, therefore, they are distractors. Option (C) is the unique complex sentence and it is the correct answer.

$$diff\_quest = 4/4 = 1$$

## 4.4 Automatic Generation of Reading Comprehension Questions

For reading comprehension, it is possible to develop several important skills such as reading fluency and vocabulary understanding. To test student's abilities regarding reading comprehension, it is quite common to produce questions focusing on interpretation from narrative, poetry, or dramatic texts. In fact, several examples of these questions can be observed from the National Exams for Portuguese - 3rd Cycle repository[14]. These are often called factual *wh*-questions and, as previously mentioned in Section 3.1, factual questions are those that have a greater study in the literature. They can actively serve to attest reading comprehension skills because questions are asked based on text passages. Thus, by giving a correct answer to these questions, students prove to understand its content. Given this, in the next sections, we will explain what were our methods for generating this type of questions in Portuguese. We will present a total of 5 distinct methods.

Here, our goals were not only be able to cover a wide range of question types but also make an effort so that they would have different difficulties. For that, we've used the factors explained in Section 4.2.2 (Difficulty Analysis). We also intended to generate questions that could possibly cover various educational levels from Bloom's taxonomy (as described in Section 1.1).

With this in mind, the rationale for the next sections will be as follows: Firstly, explain in detail our implementation of each approach and give illustrative examples. In second place, explain how we performed difficulty analysis. Lastly, point out how we manage to improve the quality of the generated questions with post-processing.

### 4.4.1 Generating Questions using PoS and NER Information

This methodology is based on Syntactic Analysis (see in Section 3.3.2) where the main goal is to analyze the structure of a sentence and, from there, apply a series of transformation rules. Here, we propose to use PoS Tagging to obtain relevant information from the sentence (namely its morphological sequence) and then, to use NER to extract all the entities from those sentences. From there, we intend to find patterns (using PoS and NER information) for identifying questionable facts. We have established well-defined rules that help us find these patterns. Finally, the last step is to choose the initial interrogative term that defines how the question will begin.

We now present all the implementation steps and an illustrative example:

1. **Select Sentence**: The sentence is analyzed and if one or more entities is identified within that sentence, then proceed to the next steps.

    - E.g, sentence: *Francisco Pizarro descobriu o Império Inca na América do Sul.*

    - E.g, entitie(s): PER: *Francisco Pizarro*, LOC: *América do Sul*

2. **Produce PoS Sequence**: For this step, the morphological sequence of the sentence is produced using PoS Tags.

---

[14]http://iave.pt/index.php/avaliacao-de-alunos/arquivo-de-provas-exames/exemplo-basico

- E.g, PoS sequence: *[NOUN, NOUN, VERB, DET, NOUN, NOUN, PREP, NOUN, PREP, NOUN, PUNCT]*

3. **Combine NER with PoS Sequence**: All entities which have been recognized by Named Entity Recognition are combined with the previous morphological sequence. The tags that represent the nouns have been replaced by the corresponding entity tag. Thus, a new sequence is obtained.

   - E.g, NER combined with PoS: *[**PER**, **PER**, VERB, DET, NOUN, NOUN, PREP, **LOC, LOC, LOC**, PUNCT]*

4. **Search for Patterns**: All established rules are compared with the sequence obtained from PoS and NER. If one or more matches are found, this is an indication that the sentence may have one or more questionable facts, so we proceed to question formulation.

   - E.g, expression used as a rule: *[PER][VERB | AUX].\*?[PUNCT]*
   - E.g, other expression used as a rule: *[PER][VERB | AUX].\*?[LOC][PUNCT]*

5. **Formulate Question**: The initial interrogative term will be chosen according to the questionable fact that will be asked. Also, the necessary changes are made to the initial declarative sentence so that it can be transformed into an interrogative form.

   - E.g, question 1: ***Quem*** *descobriu o Império Inca na América do Sul?*
   - E.g, question 2: ***Onde é que*** *Francisco Pizarro descobriu o Império Inca?*

From the previous steps, we would like to make the analogy with those traditional tasks enumerated in Section 3.4. While Pre-Processing is done previously (Section 4.2), Sentence Selection is performed from Step 1. Then, Key Selection from Step 2 to 4. Finally, Question Formulation from Step 5. We will now present all possible cases within the scope of this methodology and provide examples[15].

### 4.4.1.1   Questionable term: *Quem*

**Expression rule**: [PER][VERB | AUX].\*?[PUNCT]

**Description**: Tries to find a sequence where the first element is a per-type entity, followed by a main or auxiliary verb. It consumes all words until the first punctuation mark. This pattern can be found at any position within the sentence.

**Question formulation**: To formulate the question replaces the per-type entity with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule.

**Example 4.25.** *Sentence: A verdade dessa teoria é conhecida pelo menos desde os anos 1950, quando o psicólogo norte-americano Harry Harlow estudou o desenvolvimento de macacos. **Question**: Quem estudou o desenvolvimento de macacos?*

---

[15]The following examples may have undergone further changes through the post-processing stage, to be explained in the section 4.4.7.

**Example 4.26.** *Sentence: Meros quatro anos depois , a capital asteca estava praticamente em ruínas, o Império Asteca era coisa do passado, e Hernán Cortés dominava um novo e vasto império espanhol no México.* **Question**: *Quem dominava um novo e vasto império espanhol no México?*

#### 4.4.1.2 Questionable term: *Que pessoas*

**Expression rule**: [PER][CONJ][PER][VERB | AUX].*?[PUNCT]

**Description**: Tries to find a sequence where the first element is a per-type entity, followed by a conjunction and again a per-type. These last elements must be followed by a main or auxiliary verb. Then, consumes all words until the first punctuation mark.

**Question formulation**: To formulate the question, replaces both per-type entities (including conjunction) with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule.

**Example 4.27.** *Sentence: Em 20 de julho de 1969, Neil Armstrong e Buzz Aldrin aterraram na superfície da Lua.* **Question**: *Que pessoas aterraram na superfície da Lua?*

**Example 4.28.** *Sentence: Porém, para determinar quanto os pastores tinham de pagar a fim de que o fundo tivesse dinheiro suficiente para honrar suas obrigações, Webster e Wallace precisavam ser capazes de prever quantos pastores morreriam a cada ano, quantas viúvas e órfãos eles deixariam e quantos anos as viúvas viveriam a mais do que os maridos.* **Question**: *Que pessoas precisavam ser capazes de prever quantos pastores morreriam a cada ano?*

#### 4.4.1.3 Questionable terms: *Que organização | Que acontecimento*

**Expression rule**: [ORG | EVENT][VERB | AUX].*?[PUNCT]

**Description**: Tries to find a sequence where the first element is an organization-type or event-type entity, followed by a main or auxiliary verb. It consumes all words until the first punctuation mark.

**Question formulation**: To formulate the question, replaces the org-type entity with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule.

**Example 4.29.** *Sentence: A VOC usou o dinheiro que obteve com a compra de ações para construir navios, enviá-los à Ásia e trazer de volta produtos chineses, indianos e indonésios.* **Question**: *Que organização usou o dinheiro que obteve com a compra de ações para construir navios?*

**Example 4.30.** *Sentence: A fim de controlar o comércio no importante rio Hudson, a Companhia das Índias Ocidentais fundou uma colônia chamada Nova Amsterdã numa ilha na foz do rio.* **Question**: *Que organização fundou uma colônia chamada Nova Amsterdã numa ilha na foz do rio?*

**Example 4.31.** *Sentence: No ano seguinte, o envolvimento de Portugal na Primeira Guerra Mundial justificaria a censura prévia da imprensa, que começou a sair cheia de cortes.* **Question**: *Que acontecimento justificaria a censura prévia da imprensa?*

**Example 4.32.** *Sentence*: *O clima da Reforma e da Contra-Reforma faz-se agora sentir por toda a Europa, e os seus efeitos chegaram também à Península e a Portugal, em particular.* **Question**: *Que acontecimento se faz agora sentir por toda a Europa?*

### 4.4.1.4   Questionable terms: *Quando é que | Onde é que*

**Expression rule**: [TIME | LOC][VERB | AUX].*?[PUNCT]

**Description**: Tries to find a sequence where the first element is a time-type or loc-type entity, followed by a main or auxiliary verb. It consumes all words until the first punctuation mark.

**Question formulation**: To formulate the question, replaces the time-type or loc-type entity with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule.

**Example 4.33.** *Sentence*: *Na passagem da década de 1970 para a de 1980 surgiu o novo rock urbano (Rui Veloso, Ar de Rock, 1980) , renovou-se o humor televisivo (com Herman José e o seu programa O Tal Canal, 1983) e o desporto ganhou imensa audiência, graças aos maratonistas vitoriosos nos Jogos Olímpicos.* **Question**: *Quando é que surgiu o novo rock urbano?*

**Example 4.34.** *Sentence*: *Em particular, as pesquisas genéticas realizadas após 1945 demonstraram que as diferenças entre as várias linhagens humanas são muito menores do que os nazistas postulavam.* **Question**: *Quando é que demonstraram que as diferenças entre as várias linhagens humanas são muito menores do que os nazistas postulavam?*

**Example 4.35.** *Sentence*: *Nos Estados Unidos aboloriam a escravatura.* **Question**: *Onde é que aboliram a escravatura?*

**Example 4.36.** *Sentence*: *Em 2001, no conjunto das duas grandes áreas metropolitanas de Lisboa e do Porto viviam cerca de 4,4 milhões de pessoas, o equivalente a 44% da população residente, acumulada em 4% da superfície do país.* **Question**: *Onde é que viviam cerca de 4,4 milhões de pessoas?*

### 4.4.1.5   Questionable term: *Onde é que* - Case 2

**Questionable term:** *Onde é que*

**Expression rule**: [PER][VERB | AUX].*?[LOC][PUNCT]

**Description**: This rule tries to find relationships between people and places. The first element has to be a per-type entity, followed by a main or auxiliary verb. It consumes all words until the loc-entity appear followed by a punctuation mark.

**Question formulation**: To formulate the question, removes the loc-type from the sentence. The question, after the initial term, will be followed by the per-type entity. The rest of the question will be composed of the matched words from our rule.

**Example 4.37.** *Sentence*: *Francisco Pizarro descobriu o Império Inca na América do Sul.* **Question**: *Onde é que Francisco Pizarro descobriu o Império Inca?*

**Example 4.38.** *Sentence: Meros quatro anos depois, a capital asteca estava praticamente em ruínas, o Império Asteca era coisa do passado, e Hernán Cortés dominava um novo e vasto império espanhol no México. **Question**: Onde é que Hernán Cortés dominava um novo e vasto império espanhol?*

#### 4.4.1.6 Questionable terms: *Que número | Que percentagem*

**Expression rule**: [VAL][VERB | AUX].*?[PUNCT]

**Description**: Tries to find a sequence where the first element is a val-type entity, followed by a main or auxiliary verb. It consumes all words until the first punctuation mark. Finally, to distinguish possible values types, performs disambiguation to assess whether it is a numerical or percentage value.

**Question formulation**: To formulate the question, replaces the val-type entity with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule.

**Example 4.39.** *Sentence: Do século XVI ao século XIX , por volta de 10 milhões de escravos africanos foram importados para a América. **Question**: Que número de escravos africanos foram importados para a América?*

**Example 4.40.** *Sentence: Revelou-se que de 1% a 4% do DNA das populações modernas no Oriente Médio e na Europa são DNA de neandertal. **Question**: Que percentagem do DNA das populações modernas no Oriente e na Europa são DNA de neandertal?*

#### 4.4.1.7 Questionable terms: *Quem | Que organização | Que acontecimento*

> Note: Although applicable for these mentioned entities, the following examples are only for the person-type entity.

**Expression rule 1**: [PER][VERB|AUX][PUNCT][VERB|AUX].*?[PUNCT]
.*?[PUNCT]

**Description**: Tries to find a sequence where the first element is per-type entity, followed by a main or auxiliary verb. Then, verifies if the following term is an expression delimited by commas. This is an indicator that, between commas, may exist additional information about the entity person.

**Question formulation**: To formulate the question, replaces the per-type entity with the appropriate interrogative term and uses the information delimited by commas as part of the question.

**Example 4.41.** *Sentence: A Andorinha ouviu, atenta como a boa educação ordena, e ficou triste. **Question**: Quem ouviu atenta como a boa educação ordena?*

**Expression rule 2**: [PER][VERB|AUX][PUNCT][VERB|AUX].*?[PUNCT][CONJ]
.*?[PUNCT]

**Description**: Tries to find a sequence where the first element is per-type entity, followed by a main or auxiliary verb. Then, verifies if the following term is an expression delimited by commas.

Finally, verifies if what is the first word following the expression between commas is a conjunction. If so, that is an indicator that may exist over there additional information about the entity person.

**Question formulation**: To formulate the question, replaces the per-type entity with the appropriate interrogative term and then uses both information (before and after the commas) to generate the interrogative sentence.

**Example 4.42.** *Sentence: A Andorinha ouviu, atenta como a boa educação ordena, e ficou triste.*
*Question: Quem ouviu e ficou triste?*

**Expression rule 3**: [PER][VERB|AUX].*?[CONJ][PUNCT].*?[PUNCT][VERB|AUX]
.*?[PUNCT]

**Description**: Tries to find a sequence where the first element is per-type entity, followed by a main or auxiliary verb. Then, tries to find a conjunction that is followed by an expression enclosed in commas. Following that same expression, a verbal form must exist. This is an indicator that there are, for the same person entity, two actions that have been performed. The first action is before the commas and the second one is after the commas.

**Question formulation**: To formulate the question, replaces the per-type entity with the appropriate interrogative term and then uses both information (before and after the commas) to generate two distinct questions.

**Example 4.43.** *Sentence: O Papagaio serviu de sacristão e, à noite, embriagou-se.* ***Question 1:***
*Quem serviu de sacristão?* ***Question 2****: Quem se embriagou?*

**Expression rule nr. 4** [PER][VERB|AUX].*?[CONJ][PUNCT].*?[PUNCT][VERB|AUX]
.*?[PUNCT]

**Description**: Tries to find a sequence where the first element is per-type entity, followed by a main or auxiliary verb. Then, tries to find a completive subordinating conjunction *que* that is followed by an expression enclosed in commas (typically a time modifier). Following that same expression, a verbal form must exist. This is an indicator of a certain action being described and is interrupted by some extra information enclosed between commas.

**Question formulation**: To formulate the question, replaces the per-type entity with the appropriate interrogative term. The rest of the question will be composed of the matched words from our rule except the words enclosed by commas.

**Example 4.44.** *Sentence: Otis lembrou-se de repente que, uns dias antes, dera licença a um bando de ciganos para acamparem no parque.* ***Question****: Quem se lembrou de repente que dera licença a um bando de ciganos para acamparem no parque?*

### 4.4.2 Generating Questions using Semantic Analysis

This methodology is a direct adaptation and implementation of what was previously explained in Section 3.3.3. Here, the main goal is to perform semantic parsing within the Portuguese sentences. By doing so, we convert a natural language sentence into a logical form: a machine-understandable

representation of its meaning. Semantic parsing can thus be presumed as extracting the meaning of a given sentence. All of this is possible by performing semantic role labeling for Portuguese, which is the process of assigning semantic labels to words in a sentence. Therefore, here the main driver linguistics focuses on identifying the semantic roles for each argument and modifier in a sentence. This method is an alternative to the previous phrase-structure parsing in which the main focus is the information provided by the PoS tags and NER entities.

To perform SRL in Portuguese we have used the *nlpnet*[16] tool which was inspired by SENNA[17], but has some conceptual and practical differences. Most of the architecture is language independent, but some functions were especially tailored for working with Portuguese. Given a sentence, this tool produces its semantic role labels (see Table 4.8) according to the *PropBank-Br* [47]. The verbs in a sentence are considered predicates, so the semantic roles include core arguments of the verbs and a specific set of modifiers.

We propose to use these labels to obtain relevant information from the sentence. Next, we intend to find patterns that would help us to identify questionable facts. For that, we have established well-defined rules to find these patterns. Finally, we choose the convenient initial interrogative term and apply the necessary transformations to produce the sentence in its interrogative form. We now present all the implementation steps and an illustrative example:

1. **Select Sentence**: The sentence is analyzed and if one or more semantic labels are identified, then proceed to the next steps.

   - E.g, sentence: *Com um beijo, a Manhã apaga cada estrela enquanto prossegue a caminhada em direção ao horizonte.*

   - E.g, semantic labels: *[(AM-MNR, [Com, um, beijo]), (A0, [a, Manhã]), (V, [apaga]), (A1, [cada, estrela])].*

2. **Produce SRL Sequence**: The SRL sequence is produced using, for that, the output labels from semantic role labeling.

   - E.g, SRL sequence: *[AM-MNR][A0][V][A1]*

3. **Search for Patterns**: All the established rules are compared with the sequence obtained from the semantic role labels. If one or more matches are found, this is an indication that the sentence may have one or more questionable facts.

   - E.g, expression used as a rule: *[AM-MNR][A0][V][A1]*

4. **Formulate Question**: The initial interrogative term will be chosen according to the questionable fact that will be asked. The rest of the question will be composed according to the defined template.

   - E.g, pre-defined question template: *"Como é que" + [A0] + [V] + [A1] + ?*

---

[16]https://pypi.org/project/nlpnet/
[17]https://ronan.collobert.com/senna/

- E.g, question: ***Como é que*** *a Manhã apaga cada estrela?*

Table 4.8: Semantic roles according PropBank 1.0 [53]

| Label | Role |
|---|---|
| A0 | Proto-agent (often grammatical subject) |
| A1 | Proto-patient (often grammatical object) |
| A2 | Instrument, attribute, benefactive, amount, etc |
| A3 | start point or state |
| A4 | end point or state |
| AM-ADV | Adverbial |
| AM-CAU | Cause |
| AM-DIS | Discourse |
| AM-EXT | Extension |
| AM-LOC | Location |
| AM-MNR | Manner |
| AM-NEG | Negation |
| AM-PNC | Purpose |
| AM-TMP | Time |

From the previous steps, we would like to make the analogy with those traditional question generation steps enumerated in Section 3.4. While Pre-Processing step is done previously (Section 4.2), Sentence Selection is performed in Step 1. Then, Key Selection in Step 2 and 3. Finally, Question Formulation in Step 4. We will now present all possible cases within the scope of this methodology and provide examples:

### 4.4.2.1   Questionable term: *Como é que*

**Expression rule 1**: [AM-MNR][A0][V][A1]
**Description**: Tries to find a sequence where the first element is a manner (AM-MNR), followed followed by an agent (A0) who performs a certain action (V and A1).
**Question template**: Como é que + [A0] + [V] + [A1] + ?

**Example 4.45.** ***Sentence***: *Entre 1676 e 1681, com sucesso, o papa suspendeu a atuação da Inquisição portuguesa por causa da questão do testemunho singular (condenação na base de um único testemunho incriminatório e secreto), ao que parece, com forte apoio de grupos de pressão de cristãos-novos de Roma e até com alguma intervenção do padre António Vieira.* ***Question***: *Como é que o papa suspendeu a atuação da Inquisição portuguesa?*

**Expression rule 2**: [A0][V][A1][AM-MNR]
**Description**: Tries to find a sequence where the first element is an agent (A0) who is performing some action (V and A1) with a specific manner (AM-MNR).
**Question template**: Como é que + [A0] + [V] + [A1] + ?

**Example 4.46.** **Sentence**: *Spínola compôs uma personagem característica, com monóculo e pingalim, e começou a lembrar o presidente de uma república africana.* **Question**: *Como é que Spínola compôs uma personagem característica?*

#### 4.4.2.2 Questionable term: *Quando é que*

**Expression rule 1**: [AM-TMP][A0][V][A1]

**Description**: Tries to find a sequence where the first element is a time expression (AM-TMP), followed by an agent (A0) who performs a certain action (V and A1).

**Question template**: Quando é que + [A0] + [V] + [A1] + ?

**Example 4.47.** **Sentence**: *Depois da morte de Afonso VI e de Raimundo, a monarquia leonesa viveu um período de grave crise e de confrontos vários.* **Question**: *Quando é que a monarquia leonesa viveu um período de grave crise?*

**Expression rule 2**: [A0][V][AM-TMP]

**Description**: Tries to find a sequence where the first element is an agent (A0), who is performing some action (V) in a given time space (AM-TMP).

**Question template**: Quando é que + [A0] + [V] + ?

**Example 4.48.** **Sentence**: *O Gato Malhado dormia quando a Primavera irrompeu , repentina e poderosa.* **Question**: *Quando é que o Gato Malhado dormia?*

**Expression rule 3**: [A0][V][A1][AM-TMP]

**Description**: Tries to find a sequence where the first element is an agent (A0), who is performing some action (V and A1) in a given time space (AM-TMP).

**Question template**: Quando é que + [A0] + [V] + [A1] + ?

**Example 4.49.** **Sentence**: *Don Juan de Rhode Island engasgou-se e um silêncio total cobriu todo o parque naquela hora da chegada da Primavera.* **Question**: *Quando é que Don Juan Rhode Island se engasgou e um silêncio total cobriu todo o parque?*

#### 4.4.2.3 Questionable term: *Onde é que*

**Expression rule 1**: [A0][V][AM-LOC]

**Description**: Tries to find a sequence where the first element is an agent (A0), who is performing some action (V) in a given location (AM-LOC).

**Question template**: Onde é que + [A0] + [V] + ?

**Example 4.50.** **Sentence**: *E foi neste contexto que Inês de Castro se fixou em Portugal, provavelmente em 1339 ou 1340, integrada no séquito de D. Constança Manoel, que casou com D. Pedro, o herdeiro do trono.* **Question**: *Onde é que Inês de Castro se fixou?*

**Expression rule 2**: [A0][V][A1][AM-LOC]

**Description**: Tries to find a sequence where the first element is an agent (A0), who is performing some action (V and A1) in a given location (AM-LOC).

**Question template**: Onde é que + [A0] + [V] + [A1] + ?

**Example 4.51.** *Sentence: Os exilados formavam um embrião de Estado, que em 1830 arranjou território, na ilha Terceira, uma das maiores dos Açores, onde resistiram e estabeleceram um Governo.* **Question**: *Onde é que os exilados formavam um embrião de Estado?*

**Expression rule 3**: [AM-LOC][V][A1]

**Description**: Tries to find a sequence where the first element is a location (AM-LOC), followed by an action being performed (V and A1).

**Question template**: Onde é que + [V] + [A1] + ?

**Example 4.52.** *Sentence: Na Bélgica, com uma população equivalente à portuguesa, votaram 1 667 000 eleitores nas eleições de 1912 e o partido vencedor teve 851 000 votos.* **Question**: *Onde é que votaram 1 667 000 eleitores?*

### 4.4.3  Generating Questions using Dependency Analysis

This methodology is based on dependency parsing (see in Section 3.3.4) where the main goal is to identify a set of typed relations between words by connecting them in a graphical structure based on their grammatical and functional relations. Here, we propose to use the output labels from dependency parsing to obtain relations within the words that compose our sentences. From there, we intend to find patterns for identifying questionable facts. We have established well-defined rules that help us find these patterns. While some rules have already been explored and applied for English [97], we propose rules that are specific to the Portuguese language. After applying these rules and find specific patterns, we use the question formulation step to choose the initial interrogative term that defines how the question will begin. Finally, all necessary transformations are applied so that the declarative sentence is converted into the interrogative form.

We now present all the implementation steps and an illustrative example:

1. **Select Sentence**: The sentence is analyzed and, if one or more desired dependency labels are found, then proceed to the next steps.

   - E.g, sentence: *O ano de 1917 foi difícil para todos os beligerantes.*
   - E.g, dependency parsing: *(O, 2, det), (ano, 6, nsubj), (de, 4, case), (1917, 2, nmod), (foi, 6, cop), (difícil, 0, root), (para, 10, case), (todos, 10, det), (os, 10, det), (beligerantes, 6, obl), (., 6, punct).*

2. **Produce Sequence**: The sequence is produced using the output labels from dependency parsing.

   - E.g, *[det][nsubj][case][nmod][cop][root][case][det][det][obl][punct]*

3. **Search for Patterns**: All the established rules are compared with the sequence obtained from the dependency labels. If one or more matches are found, the sentence may have one or more questionable facts.

   - E.g, expression used as a rule: *[det][nsubj][case][nmod][cop][root].\*?[punct]*

4. **Formulate Question**: The initial interrogative term will be chosen according to the questionable fact that will be asked. The rest of the question will be composed according to the defined template.

   - E.g, pre-defined question template: *"Como caracteriza"* + *[det]* + *[nsubj]* + *[case]* + *[nmod]* + *?*

   - E.g, question: ***Como caracteriza*** *o ano de 1917?*

From the previous steps, we would like to make the analogy with those traditional tasks enumerated in Section 3.4. While the Pre-Processing step is done previously (Section 4.2), Sentence Selection is performed in Step 1. Then, Key Selection in Step 2 and 3. Finally, Question Formulation in Step 4. We will now present all possible cases within the scope of this methodology and provide examples.

### 4.4.3.1 Questionable term: *Como caracteriza*

**Expression rule 1**: [det][nsubj][cop][root].\*?[punct]
**Description**: Tries to identify a subject (nsubj) that is followed by a copulative verb (cop). After the copulative verb, an adjective (root) must appear. This rule informs that there is a subject being characterized in some way. The correct answer (for the requested characterization) will be the adjective (root).
**Question template**: Como caracteriza + [det] + [nsubj] + ?

**Example 4.53.** *Sentence: Mesmo assim, os capítulos são diferentes, e não apenas por causa dos autores. Question: Como caracteriza os capítulos?*

**Expression rule 2**: [det][nsubj][cop][advmod][root].\*?[punct]
**Description**: Tries to identify a subject (nsubj) that is followed by a copulative verb (cop). After the copulative verb, an adverb (advmod) and an adjective (root) must appear. This rule informs that there is a subject being characterized in some way. The correct answer (for the requested characterization) will be the adverb plus adjective.
**Question template**: Como caracteriza + [det] + [nsubj] ?

**Example 4.54.** *Sentence: Embora se fale de «decadência» desde meados do século XVI, o diagnóstico era claramente excessivo. Question: Como caracteriza o diagnóstico?*

**Expression rule 3**: [det][nsubj][case][nmod][cop][root].\*?[punct]

**Description**: Tries to identify a certain characteristic/attribute (nsubj) from a person/object/number (nmod) that is being characterized in some way (cop + root). The correct answer (for the requested characterization) will be the adjective (root).

**Question template**: Como caracteriza + [det] + [nsubj] + [case] + [nmod] + ?

**Example 4.55.** *Sentence*: *A honestidade de Hans era célebre e a sua palavra era de oiro.* ***Question***: *Como caracteriza a honestidade de Hans?*

### 4.4.3.2   Questionable term: *Como é que*

**Expression rule 1**: [det][nsubj].*?[root][xcomp]
**Description**: Tries to identify a subject (nsubj) that is being described through a certain action with a verb (root) followed by an adjective (xcomp). The correct answer will be the adjective (xcomp).
**Question template**: Como é que + [det] + [nsubj] + [root] + ?

**Example 4.56.** *Sentence*: *Os juízes ficaram dependentes, nos seus julgamentos, de jurados eleitos que decidiam sobre matéria de facto.* ***Question***: *Como é que os juízes ficaram?*

   **Expression rule 2**: [det][nsubj].*?[root][advmod][xcomp]
**Description**: Tries to identify a subject (nsubj) that is being described through a certain action with a verb (root) followed by an adverb (advmod) and an adjective (xcomp). The correct answer will be the adverb and adjective (advmod + xcomp).
**Question template**: Como é que + [det] + [nsubj] + [root] + ?

**Example 4.57.** *Sentence*: *Os poderes formais que serviram de contraponto aos do centro político mantiveram-se largamente estáveis entre o século XVI e o início do XIX.* ***Question***: *Como é que os poderes se mantiveram?*

### 4.4.3.3   Questionable term: *O que é que*

**Expression rule**: [det][nsubj][root][det][obj].*?[punct]
**Description**: Tries to identify, as denominated from the Portuguese language, the syntactic function of direct complement. With this syntactic function, there is an indication of the subject (nsubj) on which the action expressed by the verb (root) falls directly (obj).
**Question template**: O que é que + [det] + [nsubj] + [root] ?

**Example 4.58.** *Ou seja, o tratado formalizava o reconhecimento dos reinos como entidades independentes e cujos monarcas renunciavam de forma explícita a qualquer tentativa de impor pela força a unidade das Coroas.* ***Question***: *O que é que o tratado formalizava ?*

### 4.4.4   Generating Questions using Relative Pronouns and Adverbs

This approach presents a proposal to produce questions from relative pronouns and relative adverbs regarding Portuguese sentences. Our procedure is implemented by using the information

from dependency parsing, combined with the information obtained from PoS tags which identify pronouns and adverbs. Given this proposal, we intend to improve some aspects, such as the syntactic correction, semantic suitability and fluency of our questions. In addition, we also intend to reduce the chance of misidentifying the initial interrogative term that starts a question. Beyond the study from Khullar et al., [67] whose study is for English questions, the research that concerns relative pronouns and adverbs is novel for AQG field. Our goal is to adapt the method proposed by the authors specifically to the Portuguese language.

### 4.4.4.1 Motivation behind Relative Clauses

Relative pronouns/adverbs have a dual function: pronoun and connector. As for pronouns, their role is to represent a noun/nominal group, assuming the same syntactic function that the nominal name/nominal group would represent. They refer to nouns that were already previously mentioned. Thus, an interesting thing about using relative pronouns and adverbs is that they contain unique information on their syntactic relationship between the two parts of the sentence. For example:

**Example 4.59.** *O Gato tomou a direção dos estreitos caminhos que conduzem à encruzilhada do fim do mundo.*

The relative pronoun *que (which)* refers to the noun ***caminhos (paths)***, introducing a relative subordinate clause ***que conduzem à encruzilhada do fim do mundo*** representing the syntactic function of restrictive name modifier. The noun ***caminhos (paths)*** is said to be an antecedent of the relative pronoun *que (which)*. In addition, we can also analyze the potential of the relative adverb *onde (where)*, which provides relevant information about a particular location. For example:

**Example 4.60.** *Além de tudo, o seu olhar já está de novo fixo na árvore onde a Andorinha pousara na véspera.*

Here, the relative adverb *onde (where)* refers to the noun *árvore (tree)* introducing again a relative subordinate clause ***onde a Andorinha pousara na véspera***. The noun *árvore (tree)* is said to be an antecedent of the relative adverb *onde (where)*. Finally, it is also possible to have another case (which is a variation of the previous one) where the relative adverb is separated by a comma in relation to its antecedent. For example,

**Example 4.61.** *Só 20 por cento dos cerca de 6000 padres aceitaram as pensões, sobretudo no Sul, onde o clero era menos numeroso e estava menos apoiado.*

For this case, the relative adverb *onde (where)* refers to the noun *Sul (South)* introducing a relative subordinate clause ***onde o clero era menos numeroso e estava menos apoiado***. The noun *Sul (South)* is said to be an antecedent of the relative adverb *onde (where)*. In essence, we intend to use these structural relationships to generate new questions and thus, adding them to the set of questions that can be generated for a given sentence.

**4.4.4.2   Procedure**

Regarding our procedure, we cover three different cases to generate questions. For the first one, the system checks for the presence of one or more relative pronouns *que (which)*. For the second, it tries to identify characterizations/descriptions. Third and last, it tries to identify one or more relative adverbs *onde (where)*. After comparing the morphological sequence of the sentence and its dependencies (from dependency parsing) with the rules that have been established, we proceed to question formulation. We will go through the next subsections to explain each of the rules with an illustrative example.

**4.4.4.3   Case 1 - Using Relative Pronouns**

The implementation steps can be described as follows:

1. **Select Sentence**: The sentence is analyzed and, if it contains one or more relative pronouns *que (which)*, then proceeds to the next steps.

   - E.g, sentence: *O Gato tomou a direção dos estreitos caminhos que conduzem à encruzilhada do fim do mundo.*

   - E.g, relative pronouns(s): *que (which)*

2. **Produce Sequence**: The sequence is produced from the PoS tags.

   - E.g, PoS sequence: (DET)(NOUN)(VERB)(DET)(NOUN)(PREP)(ADJ)(NOUN) (PRON)(VERB)(PREP)(NOUN)(PREP)(NOUN)(PREP)(NOUN)(PUNCT)

3. **Search for Pattern**: The established rule is compared with the sequence obtained from PoS tags. If one or more matches are found, proceeds to the next steps.

   - **Expression rule**: [NOUN][PRON].*?[PUNCT]

4. **Formulate Question**: The initial interrogative term will be chosen (verifies if there are entities from NER) according to the noun being analyzed. The rest of the question will be composed according to the match starting from [PRON] until [PUNCT] (not included).

   - E.g, question: *O que é que conduz à encruzilhada do fim do mundo?*

From the previous steps, we would like to make the analogy with those traditional tasks enumerated in Section 3.4. While the Pre-Processing step is previously done in Section 4.2, Sentence Selection is performed in Step 1. Key Selection in Step 2 and 3. Finally, Question Formulation in Step 4. For this step, a disambiguation step is performed, which goal is to verify whether the noun belongs to one of the entities identified by the NER module. If so, the initial interrogative term is used accordingly. For example:

**Example 4.62.** *Sentence: E de tal maneira o plano foi bem urdido que logo se lançou a ideia , rapidamente espalhada por Lisboa, de que era o mestre de Avis que corria perigo de vida.* ***Question**: Quem corria perigo de vida?*

#### 4.4.4.4  Case 2 - Using Relative Pronouns

This approach is analogous to the one described in Case 1 but the goal is to generate a question about a characterization/description of the noun that is being analyzed. So it is using a different expression rule within the pattern search stage. The new rule is:

$$[DET][NOUN][PRON][VERB|AUX].*?[PUNCT]$$

where *PRON* refers to the pronoun **que (which)**. If we verify that the verb (main or auxiliary verb) is copulative, then we conclude that it is being done a certain characterization/description of the noun entity. For example:

**Example 4.63.** *Existia o reino que estava ainda a formar-se e se encontrava muito longe da estabilização.* **Question***: Como estava o reino?*

#### 4.4.4.5  Case 3 - Using Relative Adverbs

This approach is analogous to the one described in Case 1 but it uses the relative adverb **onde (where)** which provides information about localities, positions, or areas. From the pattern search stage, we are using the [ADV] tag rather than [PRON] and we verify if the adverb is actually relative. Regarding the question formulation stage, as the initial term, we are using **Onde é que (Where)**. Thus, we can see the following examples:

**Example 4.64.** *Além de tudo, o seu olhar já está de novo fixo na árvore onde a Andorinha pousara na véspera.* **Question***: Onde é que a Androinha pousara na véspera?*

**Example 4.65.** *Só 20 por cento dos cerca de 6000 padres aceitaram as pensões, sobretudo no Sul, onde o clero era menos numeroso e estava menos apoiado.* **Question***: Onde é que o clero era menos numeroso e estava menos apoiado?*

### 4.4.5  Generating Questions using Discourse Connectors

For this approach, we propose to generate questions using the information provided from Discourse Connectors (DC). To achieve this, we intend to explore the utility that DC have for AQG. In linguistics, a connector is an operator capable of joining two sentences into a single transformed sentence. It is a designation for the words being used to connect, link or unite several linguistic segments. Since they can interconnect two clauses by displaying several discourse relations, several functions can be presented, such as *result, elaboration, causal, temporal, contrast, etc*. The use of DC for generating questions had already demonstrated its usefulness [117, 132], but identifying these relationships in the text is a tough task [115], most often because it is necessary to understand and disambiguate the connector's meaning [2].

Our procedure follows the hypothesis that it is possible to create questions from the analysis of three Portuguese DC: **porque (why)**, **pois (why)** and **quando (when)**. While the first two have the purpose of explaining a certain topic, the third is to situate some temporal action.

To explain this approach, we will divide the generation pipeline into three distinct parts. The first, responsible for identifying the question type. The second, responsible for identifying the target arguments for DC. Finally, the phase that will make the necessary transformations to formulate the question.

### 4.4.5.1 Question type identification

The value/function of the discourse connector will influence the question type. It is therefore important to deeply understand the connector under analysis. In our case, we have:

*porque/pois (why)*: These connectors manifest an explanation of a particular topic. They introduce a specific reason for a particular event. To validate that we are under the intended word, we have a validation step where we confirm that its morphological classification corresponds to a causal subordinating conjunction. For example,

**Example 4.66.** *Os portugueses estariam em superioridade numérica, **porque** as forças ocupantes tinham-se dispersado por Alcácer e outras povoações.*

**Example 4.67.** *O reino mergulhou na maior consternação, **pois** sabia-se que o reinado do cardeal-infante seria um intervalo antes de outra solução mais definitiva.*

*quando (when)*: This connector locates an action/event in a given time-space. To validate that we are under the intended word, we have a validation step where we confirm that its morphological classification corresponds to a subordinating temporal conjunction. For example,

**Example 4.68.** *O governador de armas do Minho foi derrotado perto do rio Coura, **quando** estava a cercar as forças espanholas em Monção, quase na mesma altura em que as tropas do Sul levantavam o cerco a Badajoz e se recolhiam a Elvas.*

For the first case, the initial interrogative will be ***Qual o motivo pelo qual (Why)*** whereas in the second, the question will start with ***Quando é que (When)***.

### 4.4.5.2 Arguments definition for DC

After identifying the connector type of the sentence and the corresponding question type, it is necessary to delimit the two arguments that make up the sentence. From the discourse connector, it is possible to discover its two arguments, ***Arg1*** and ***Arg2***. We do so by analyzing the structure of the sentence. We assume that the clause appearing before the connector, separated by a comma, is the argument 1. The clause that appears after the connector (including the connector) is the argument 2.

**Example 4.69.** *[**Arg1** Os portugueses estariam em superioridade numérica], [**Arg2 porque** as forças ocupantes tinham-se dispersado por Alcácer e outras povoações.]*

*[**Arg1** O reino mergulhou na maior consternação], [**Arg2 pois** sabia-se que o reinado do cardeal-infante seria um intervalo antes de outra solução mais definitiva.]*

*[**Arg1** O governador de armas do Minho foi derrotado perto do rio Coura], [**Arg2 quando** estava a cercar as forças espanholas em Monção], quase na mesma altura em que as tropas do Sul levantavam o cerco a Badajoz e se recolhiam a Elvas.*

### 4.4.5.3 Question formulation

To generate the questions, the goal is to take the interrogative terms discovered in Section 4.4.5.1 and associate them with the discovered arguments (**Arg1**) from 4.4.5.2. This way, the questions, in their final format, will take the following forms:

**Example 4.70.** *Qual o motivo pelo qual os portugueses estariam em superioridade numérica?*
   *Qual o motivo pelo qual o reino mergulhou na maior consternação?*
   *Quando é que o governador de armas do Minho foi derrotado perto do rio Coura?*

We believe that this methodology has great potential to be extended and improved in future work. On the one hand, to expand the range of discourse connectors to be analyzed. On the other, to take into account new cases where the arguments are not in the same sentence but in two different sentences.

### 4.4.6 Question Difficulty Analysis

All previous questions that are generated through the implemented system, regarding all methods, have an assigned difficulty. This difficulty is calculated as follows:

$$diff\_quest = (1/3) * diff\_text + (1/3) * diff\_target\_sentence + (1/3) * cognitive\_level$$

We decided that all factors should have the same weights since we consider it would be correct that all of them have an equal contribution to defining the question's difficulty.

*diff_text* is calculated as explained in Section 4.2.2. The goal of this factor is to distinguish the difficulty of two different questions taking into account their source texts, which had originated them. For example, the following excerpt has three sentences and was extracted from a children's book. It has a Flesch reading ease score of **42.57**.

**Example 4.71.** *A Noite é uma apavorada, tem horror às trevas. Com um beijo, a Manhã apaga cada estrela enquanto prossegue a caminhada em direção ao horizonte. Semiadormecida, bocejando, acontece-lhe esquecer algumas sem apagar.*

The next excerpt, with also 3 sentences, is taken from a history book. To notice that it has longer and more labored sentences. Moreover, its Flesh reading ease score is **33.15**.

**Example 4.72.** *Embora a repressão inquisitorial tenha registado outros picos ulteriores, a verdade é que de permeio teve lugar uma viragem decisiva. Entre 1676 e 1681, com sucesso, o papa suspendeu a actuação da Inquisição portuguesa por causa da questão do testemunho singular*

*(condenação na base de um único testemunho incriminatório e secreto), ao que parece, com forte apoio de grupos de pressão de cristãos-novos de Roma e até com alguma intervenção do padre António Vieira. A interferência papal acabou por ter a oposição do então regente D. Pedro, que a reputou de uma violação da jurisdição real.*

From the last two excerpts, these two possible questions can be generated:

**Example 4.73.** ***Question 1****: Como é que a Manhã apaga cada estrela?* ***Question 2****: Como é que o papa suspendeu a atuação da Inquisição portuguesa?*

Although the previous questions are of the same type, we highlight for the fact that the second is more elaborate and requires greater attention from the reader. Also, according to Flesch readability test, lower values indicate greater complexity, therefore, the second question will have a higher value for this factor.

***diff_target_sentence*** is calculated as explained in Section 4.2.2. Here, the key idea is to look only at the target sentence from which the question was generated. The more complex that target sentence is, the more difficult the question will be. This complexity takes into account the sentence length, number of dependencies, numerals, coordinating relations, adverbs and the number of different words.

***cognitive_level*** intends to distinguish the difficulty of two questions taking into account their cognitive position in which they find themselves, according to Bloom's Taxonomy (Section 1.1). At this moment, we consider that our questions fit the two first levels of this taxonomy, therefore, we contemplate two difficulty levels. We now reveal these two levels by referring to the types of questions that are associated:

- **Level 2 (Describe, explain, contrast, interpret, discuss, etc...)**: *Como é que?, Qual o motivo pelo qual? Como caracteriza?*

- **Level 1 (List, recite, outline, name, recall, identify, etc...)**: *Quem? Que pessoas? Que organização? Que acontecimento? Que número? Que percentagem? Onde é que? Quando é que? O que é que?*

### 4.4.7 Post-Processing

Some errors can arise during the generation process. Therefore, the post-processing phase is important to improve the quality of the questions. The big goal is to minimize the number of errors for the final output. To achieve this goal, we did two main options: Question Post-Editing and Question Filtering. While the first adds, removes and modifies certain aspects of a given question, the second is responsible for rejecting ill-formed and unacceptable questions. We will now list all the handled cases:

- **Reverse the pronoun position**. Method: Identifies a verb that has a pronoun attached to its right. Changes the pronoun position.

**Example 4.74.** ***Before***: *Quem* *dirigiu-se* *para Brockley?* ***After:*** *Quem* *se dirigiu* *para Brockley?*

- **Change verbs conjugation, from plural to singular**. Method: The lemma of the verb is obtained and the verb is conjugated in the third person, singular.

   **Example 4.75.** ***Before***: *O que é que* *conduzem* *à encruzilhada do fim do mundo?* ***After:*** *O que é que* *conduz* *à encruzilhada do fim do mundo?*

- **Remove unnecessary parts**. Method: Search for pattern [PUNCT][PRON].*?[PUNCT] being that the first PUNCT is a comma, the PRON is a relative pronoun *que*, and the last PUNCT is a question mark. If it matches, then removes the matched tokens from the question (except the question mark).

   **Example 4.76.** ***Before***: *Onde é que viviam vários clãs patrilineares, que se baseavam na descendência por parte de pai?* ***After:*** *Onde é que viviam vários clãs patrilineares?*

- **Remove prepositions or conjunctions at the end of the question**. Method: Search for pattern [CONJ][PUNCT] being that CONJ is a conjunction and PUNCT a question mark. If it matches, then removes the conjunction from the question.

   **Example 4.77.** ***Before***: *Quando é que soava a meia-noite no relógio da torre* *e*? ***After:*** *Quando é que soava a meia-noite no relógio da torre?*

- **Change words that are not nouns to lowercase**:

   **Example 4.78.** ***Before***: *Como é que* *Os* *neandertais caçavam?* ***After:*** *Como é que* *os* *neandertais caçavam?*

- **Reject questions that contain personal pronouns**:

   **Example 4.79.** ***Rejected***: *Quem deseja vivamente que o senhor* *a* *autorize a guardar para* *ela* *o cofre?*

- **Reject questions that contain demonstrative pronouns**:

   **Example 4.80.** ***Rejected***: *Quem foi por* *este* *doado para financiar a edificação da própria?*

- **Reject too short (or extremely large) questions**:

   **Example 4.81.** ***Rejected***: *O que é que vestia?*

For future work, we leave open the possibility of creating more post-editing operations or even more rules that identify ill-formed questions. Besides, it would also be interesting to have a Question Ranking module in which the goal is to order/rank the questions given certain acceptability factors.

## 4.5   Automatic Generation of Pronoun Reference Questions

This study presents a method that allows the generation of pronoun reference questions from Portuguese texts. An illustrative example of a generated question using this approach can be observed from Figure 4.2. A pronoun reference question is a multiple choice question consisting of a reading passage from the text, a correct answer (which is the antecedent of the target pronoun) and three wrong options (distractors). It asks for test-takers to choose the correct antecedent for the target pronoun from the reading passage. This questions-type are commonly used in English exams like TOEFL iBT. Regarding Portuguese exams and tests, we can also observe that it is an exercise that frequently appears.

Our motivation comes from the fact that these types of questions are very important for measuring the student's ability for reading comprehension. First of all, it is a frequent exercise-type in exams. Secondly, as Gordon and Scearce [60] had concluded, people are able, naturally, to resolve pronominalizations when reading a text and finding a certain pronoun. Then, if a student is able to discover the antecedent for a pronoun when reading the text excerpt, it is possible to validate his/her ability to understand the reading passage.

Our proposal has its bases from the study carried out by Satria and Tokunaga [128] that proposed the elaboration of this questions-type, for English. Our goal is to adapt the method proposed by the authors specifically to the Portuguese language. While this is our last proposal for AQG in Portuguese, we point out its great potential for taking into account not only one but multiple sentences from texts, in order to build our questions.



Figure 4.2: Example of a Portuguese generated pronoun reference question. (1) Reading Passage. (2) Pronoun. (3) Correct Answer. (4) Distractors.

### 4.5.1   Procedure

One possible way to generate reference questions is to apply a coreference resolver directly into the text [128]. For that case, the role of the coreference resolver would be to find a pair composed of antecedent-pronoun. The problem with this approach is that the system would be totally dependent

on the performance of the coreference resolver. Although there was (and there is) some work in progress regarding coreference resolution for Portuguese [54, 36], we decided to follow a different approach, which would not leave us dependent on that task to solve our problem. Still, as we will mention at the end of this section, we are going to leave open the possibility of integrating a coreference resolver with our approach for future work.

Our procedure takes advantage of the appositive modifiers to find pronoun-antecedent pairs. For example, here we can observe a sentence that contains an appositive modifier:

**Example 4.82.** *Os concelhos**, que tinham sido comunidades autónomas lidando com um poder distante e indirecto,** eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.*

To notice that, analogously to the approach explained in Section 4.4.4, here we also have the same relative pronoun *que (which)*. The difference is that, in the context of the previous approach, the modifier was restrictive but now, it is an appositive. An appositive modifier modifies the name so as not to limit/restrict its reference. They introduce an explicative relative adjective subordinate clause (providing additional information). Moreover, they appear to the right of the name, being delimited by commas. A very important aspect of the appositive modifier (unlike the restrictive one) is that it can be removed from the sentence, without the sentence losing its meaning. That can be verified by looking at the previous (transformed) sentence:

**Example 4.83.** *Os concelhos eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.*

To summarize, the main idea is to divide the sentence at the relative pronoun and then introduce/replace it with a new suitable pronoun so that we can obtain both pronoun and antecedent. More deeply, we will be able to generate our own references within text passages. The details of this process can be analyzed along with the following subsections.

### 4.5.2 Correct Answer Generation

The first step taken by our system in order to achieve its goal is to extract, from paragraphs, sentences containing appositive modifiers. For such, we use pattern matching that searches for paragraphs that include this string pattern: *", que ...."* (*", which ...."*). If a sentence contains an appositive modifier, we extract it from that paragraph. The expression rule being used to match this pattern can be express as:

$$[PUNCT][PRON].*?[PUNCT]$$

where *PRON* refers to the pronoun (we then check if it is a relative one) and *PUNCT* refers to punctuation mark (we then verify if it is a comma). If we take the previous sentence, the match will be as follows:

Figure 4.3: Example of the dependency relation: acl:relc.

**Example 4.84.** *Os concelhos**, que tinham sido comunidades autónomas lidando com um poder distante e indirecto,** eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.*

Then, we proceed to the next step, where we discover the antecedent of the relative pronoun. This antecedent is the correct answer for the multiple-choice that will be generated. To find what the antecedent is, we use dependency analysis. More specifically, we search for a dependency label called ***acl:recl*** [18]. This relation starts in the antecedent and points to inside of the clause. It is possible to better observe this relation in Figure 4.3. The remaining dependency labels had been omitted to simplify the example. We highlight the importance of this step since the antecedent may consist of more than one word. By using this label we manage to extract all the words that could compose it. With regard to our example, the correct answer was already found and is composed of one word, being ***Os concelhos (The counties)***.

### 4.5.3  Reading Passage Generation

After discovering the appositive modifier and the antecedent of the relative pronoun, we proceed to the phase which is responsible for modifying the sentence and constructing a new excerpt. For that, we can divide the process into two distinct steps: sentence splitting and creation of the reference pronoun. Through sentence splitting, the original sentence is divided into two distinct sentences. There are two ways to do this. Observe the following example:

**Example 4.85.** ***Os concelhos*** *eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.* ***Os concelhos*** *tinham sido comunidades autónomas lidando com um poder distante e indirecto.*

This example consists of two new sentences that were obtained from the original sentence. Essentially, a new sentence was created using the appositive modifier. Then, the relative pronoun ***que (which)*** was replaced by its antecedent ***Os concelhos (the counties)***. There is another way to do this arrangement, looking at the next example:

**Example 4.86.** ***Os concelhos*** *tinham sido comunidades autónomas lidando com um poder distante e indirecto.* ***Os concelhos*** *eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.*

---

[18]https://universaldependencies.org/en/dep/acl-relcl.html

For this new example, the essence is to remove the commas that delimit the appositive modifier. Then, for the new sentence, insert the subject which in this case is, again, the antecedent ***Os concelhos (the counties)***. In order to choose which of the two methods will be applied, we consider the number of available candidate distractors before the pronoun (to be explained in Section 4.5.4).

The generation of the reference pronoun will be performed by analyzing its antecedent. There are four Portuguese demonstrative pronouns that are possible to used. They are:

- *Aquele*: Demonstrative pronoun in male gender and singular number;

- *Aquela*: Demonstrative pronoun in feminine gender and singular number;

- *Aqueles*: Demonstrative pronoun in male gender and plural number;

- *Aquelas*: Demonstrative pronoun in feminine gender and plural number.

The appropriate pronoun is chosen by analyzing the morphological features of the antecedent, such as its number and gender. In our example, the antecedent has masculine gender and plural number. For these reasons, the result will be:

**Example 4.87.** *Os concelhos eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.* ***Aqueles*** *tinham sido comunidades autónomas lidando com um poder distante e indirecto.*

To finish the generation of the reading passage, the last step is to add the sentence consecutively before and after the original sentence. More specifically, the final result will consist of four sentences. The sentence before and after the original sentence plus two sentences (which resulted from the explained transformations). The purpose of doing this is to give a context for the excerpt. See the final result:

**Example 4.88.** *Devido à limitação de recursos materiais e humanos do Estado, o Governo acabou por confiar nos concelhos para recolher impostos e organizar o recrutamento. Os concelhos eram agora órgãos administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante.* ***Aqueles*** *tinham sido comunidades autónomas lidando com um poder distante e indirecto. Como reconheceu o duque de Palmela em 1844, «em outras eras exerciam as nossas municipalidades poderes que actualmente não têm no sentido político».*

### 4.5.4 Distractor Generation

Distractors are intended to be the wrong multiple-choice options, so there is the need to make them plausible. This avoids the risk of the question being too easy. To achieve our goal, we have defined a series of rules for choosing distractors given our final reading passage.

1. The distractor word class (PoS tag) must be the same as the correct answer. The three possibilities are to be a common, proper or collective noun. This is because the correct answer will always be a noun.

2. The distractor must be of the same gender and number as the correct answer. If it were not so, the distractor would be too obvious.

3. The distractor should preferably precede the antecedent. Even so, there may not be enough candidates. If this happens, we consider using candidates who proceed the correct answer. It should be noted that the number of candidates available will influence how the text passage will be generated, as we saw earlier.

4. The distractor cannot be the same as the correct answer. Despite this is an obvious rule, it is not trivial to ensure it. A distractor could be in the same coreference chain of the correct answer, without sharing the same word/noun. For this reason, we present a line for future work. That would be a complement within this coreference verification, using a Portuguese coreference resolver.

Bearing in mind that we need three distractors in addition to the correct answer, the possible options made available by our example are here underlined:

**Example 4.89.** *Devido à limitação de <u>recursos</u> materiais e humanos do Estado, o Governo acabou por confiar nos concelhos para recolher <u>impostos</u> e organizar o recrutamento. Os concelhos eram agora <u>órgãos</u> administrativos de um Estado nacional, submetidos diretamente a um Governo vigilante. **Aqueles** tinham sido <u>comunidades</u> autónomas lidando com um poder distante e indirecto. Como reconheceu o duque de Palmela em 1844, «em outras eras exerciam as nossas <u>municipalidades</u> poderes que actualmente não têm no sentido político».*

The system finds precisely three candidates for distractors, all of them precede the correct answer. As they all meet the requirements previously established, they will be used as multiple-choice options. Two other possible candidates appear after the antecedent. These would only be used if there were not enough precedent candidates.

The difficulty of these question types is calculated as previously explain in Section 4.4.6. The only difference is that it takes into account more than one target sentence.

To conclude, we present another line of future work for this type of question. It would be interesting, for cases where there are more than three candidates, to sort them by semantic similarity to the correct answer. Doing that, it would be possible to use that kind of similarity as a difficulty factor.

## 4.6   Automatic Question Generation for Portuguese - A Summary

At this point, we can observe the Mind Map present in Figure 4.4. In this scheme, we summarize the work developed throughout this research. The generated questions are diverse. On the one hand, we have the grammar questions that allow us to generate 8 different question types. Then, we have the reading comprehension questions in which five different methods are studied. Each of them addresses several types of questions. Finally, we have the pronoun reference questions.

For this, it is possible to generate questions concerning pronouns that can refer to proper, common or collective nouns. In fact, from the beginning, we decided to give priority to the number of approaches and methods to implement because, by doing this, we were able to obtain diverse and rich questions. It is possible to observe a Graphical User Interface developed for this application in Appendix B. Over there, we have our methodologies combined in one place so that they can be experimented with by the final user. To conclude, we now link to our 2 final chapters: Chapter 5, to assess the results and Chapter 6 to conclude our work.
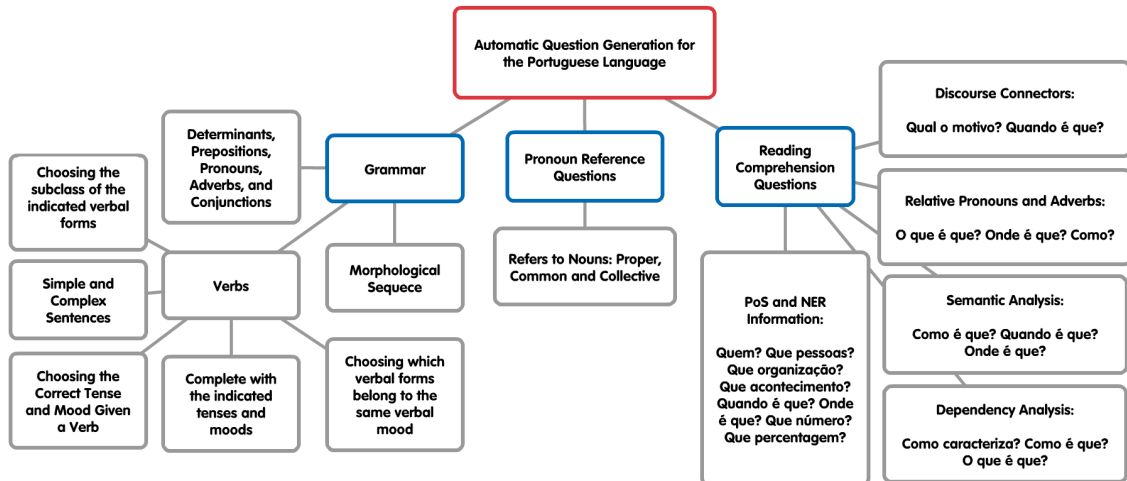


Figure 4.4: Mind Map - Automatic Question Generation for the Portuguese Language.

# Chapter 5

# Results and Evaluation

This chapter aims to present the results of the evaluation that can be divided into two distinct parts. The first (Section 5.2) is a survey that was answered by several Portuguese teachers where the objective was to draw conclusions about the formulation, objectivity, relevance and the difficulty of the generated questions. The second part (Section 5.3) is a questionnaire applied to students of the Portuguese subject. Here, the goal was to conclude about the practical utility that our system has in a real context, when applied with the pupils. Also, the main focus was once again to validate the difficulty within our questions.

## 5.1 Evaluation - Survey and Questionnaire

As mentioned and explained in Section 3.4.7 there is no standard methodology to evaluate the automatic generation of questions. What is usually done is an expert-based evaluation which consists of presenting a sample of the generated question to reviewers. Not so common but can also be observed from the literature (again from Section 3.4.7), some evaluations are performed using questionnaires that are applied with students, also with the generated questions. This second form of evaluation has the great advantage of being able to draw conclusions in a real context, with the students themselves, who after all are the end-users of an application capable of automatically generating questions. However, there is substantial risk in this type of assessment. Such risk is related to the fact that students can realize that the questionnaire is not important for final marks and provide answers that do not correspond to their knowledge. The opposite may also happen, students may find that the test is really important for final evaluation and, in case the questions are too easy or too difficult, they are suggested as to what the next tests will be (raising pedagogical issues with their teachers).

Taking all of the previous factors into account, we decided to elaborate a survey for teachers and a questionnaire for students. Both have the same generated questions but the first focuses on well-established metrics to draw further conclusions. The second (the questionnaire) simulates the structure of a real test where the students had to answer reading comprehension, pronoun reference and grammar questions.

In Section 5.2 we will present the results for the survey (which is included in Appendix C.1). It was answered by **27 Portuguese Teachers** and contains a total of **18 generated questions**. They were all generated from the Portuguese tale *Saga, Sophia de Mello Breyner Andresen* and selected according to the following criteria:

- Questions are fully generated by the developed system;

- Questions are representative of all implemented approaches;

- Questions have different difficulties.

The metrics, depending on each approach, are as follows:

---

For Reading Comprehension (**7 questions**):

---

- **Question Formulation** - *Do you consider the question as being well formulated? (Yes/No);*

- **Objectivity** - *Do you consider the question objective? (scale from 1 to 5);*

- **Answerability** - *How many answers do you think this question may have? (One, Ambiguous: Two or more, None: The answer is not in the excerpt);*

- **Suitability for study cycle** - *You consider this question suitable for (select one or more options) (1st cycle, 2nd cycle, 3rd cycle, no cycle).*

There is still one last question that asks: *Overall, which questions do you find to be the **most difficult for students**? (select one or more options).*

---

Pronoun Reference Questions(**1 question**):

---

Regarding Pronoun Reference Questions, the metrics are the same as those used for grammar (see below) except for the last metric (that compares two grammar questions). The results for this approach will be coupled with reading comprehension questions.

---

For Grammar (**10 questions**):

---

- *Taking into account the above question, you consider that*

  - *both the instruction and the MC options* *are suitable;*

  - *only the instruction* *is suitable;*

  - *only the MC options* *are suitable;*

  - *nothing* *is suitable within this question.*

- *You consider this* ***question suitable*** *for the (select one or more options)*

  - *1st cycle*: *1st, 2nd, 3rd and 4th years;*

  - *2nd cycle*: *5th and 6th years;*

  - *3rd cycle*: *7th, 8th and 9th years;*

  - *no cycle*.

- *Regarding the difficulty of the previous (2 grammar) questions, you consider that*

  - *question* *x* *is more difficult than question* *y*.

  - *question* *y* *is more difficult than question* *x*.

  - *both have the same degree of difficulty.*

Regarding the 10 grammar questions, we want to highlight the following: half of the 10 questions are easier (according to heuristics) than the other half. That is, there are 2 grammar questions for each type, one is easier and one is more difficult (according to heuristics). The purpose of doing this is to **validate if Teachers will consider the most difficult questions as effectively being the most difficult ones**. On the other hand, **students are expected to score lower on the most difficult questions**.

In Section 5.3, we will present the results for the questionnaire (which is included in Appendix C.2). It contains the same **18 generated questions** but follows the scheme of a conventional test/exam. It was answered by **163 students** of the following school years:

- **7th Year** (24 students)

- **8th Year** (76 students)

- **9th Year** (49 students)

- **PFL - Portuguese as a Foreign Language** (14 students): 4 from A1-A2 level, 2 from A2-B1 level, 6 from B1-B2 level, 1 from B2-C1 level and 1 from C1-C2 level.

## 5.2   Evaluation - Survey Applied to Teachers

---
Reading Comprehension (including Pronoun Reference) Questions
---

We consider that, on average, the results for **Question Formulation** (see Figure 5.1) are good (84.275%). The not so good results are from the Pronoun Reference approach which is responsible for generating a certain passage from the text. This is due to the fact that with this approach, the text passage can, sometimes, be unnatural due to the way its sentences are generated and positioned within the text. On the contrary, the best results go to PoS and NER as they use short, clear and concise initial interrogative terms ( *Quem, Que acontecimento*).

Regarding **Objectivity** (see Figure 5.2), the lowest results are from the Dependency Analysis and Discourse Connectors approaches. This is expected since these approaches use deeper questions, which are not so general and may require a higher cognitive level (Section 1.1) to be answered. We recall that these approaches generate questions that ask for characterizations or reasons for certain events. Thus, we take this opportunity to make a relation with the results presented in Figure 5.3 regarding **Answerability**. Its results, also good, show a significant percentage (12,69%) of teachers who affirm there is some ambiguity. We find that these questions (where there may be ambiguity according to some opinions) are precisely generated from the previous approaches (dependency analysis and discourse connectors). Again, since they are not so direct questions and require a more extensive answer (and not just a simple fact), they can contemplate more than one acceptable answer.

For Question **Difficulty** (see Figure 5.4) we asked teachers to order all the previous reading comprehension questions by difficulty level (difficulty from a student viewpoint). The questions that were considered as being the most difficult are from the Discourse Connectors and Dependency Analysis approaches. Again, following the reasoning of the previous paragraphs, these questions ask for deeper answers and are not limited to questions about simple facts as people, locations or organizations (PoS and NER approach). Therefore, these results are expected.

The **Suitability** results (see Figure 5.5) validate the relevance of our questions to be applied in a real context with the pupils. Results demonstrate that the questions can be mostly applied to students of the 3rd cycle without excluding the other cycles (1st and 2nd). We look at this indicator as a promising future for generating questions within the Portuguese language.

---
Grammar Questions
---

The **overall results for grammar questions** (see Figure 5.6), show a positive trend in which both instructions and their multiple-choice options are appropriate. However, there is a small percentage of Teachers who consider that some questions are not suitable (7,77%) or partially suitable (5,18% and 2,96%). While we have not been able to give an obvious reason for this to happen, we can point to the fact that there is often disagreement between Portuguese Teachers in relation to grammatical nomenclatures for certain contents. In addition, as we indicated earlier, Portuguese grammar changes regularly, which does not contribute to the good stability of opinions
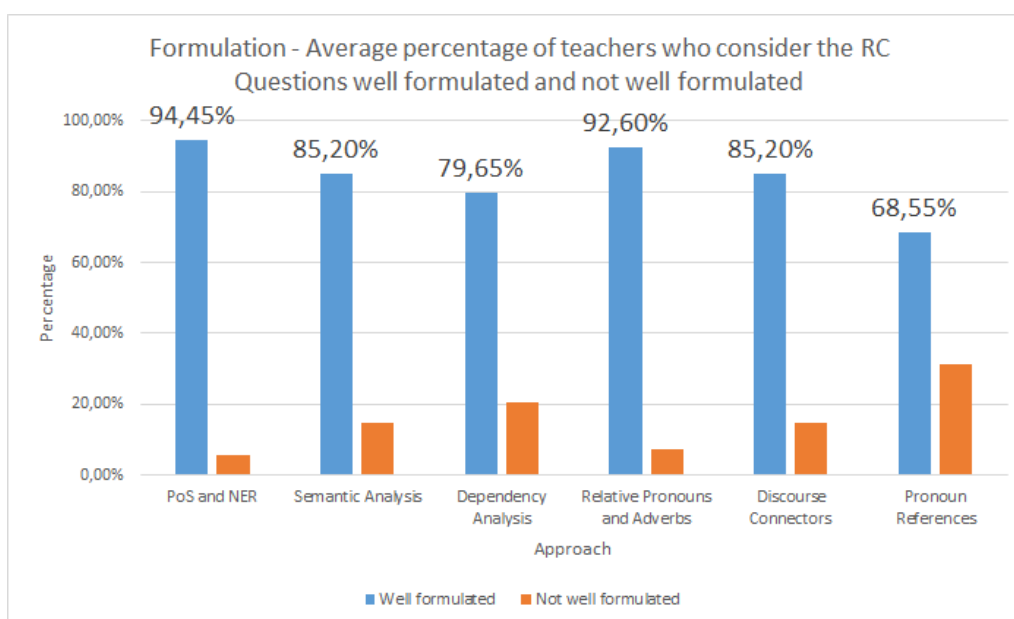
Figure 5.1: Average percentage of Teachers who consider the RC questions well formulated and not well formulated.

in this field. As far as we are concerned, we consulted the most current and modern grammars and in case of doubt, we clarify with Portuguese Teachers.

Regarding **Difficulty** for grammar questions, the results can be observed from Figure 5.7. We would expect a higher percentage of Teachers to consider that the most difficult questions (according to the heuristics from the methodologies) were in fact the most difficult questions. Anyway, there is a higher percentage of Teachers who consider the contrary (30,36%). The biggest slice (37,78%) goes for those Teachers who find grammar questions to have the same difficulty. Fortunately, we did a double check for this metric, through the questionnaire detailed in 5.3 for students. Over there, the results are more positive in the sense that they demonstrate that the most difficult (according to heuristics) questions were in fact those in which students scored less.

The **Suitability** results (see Figure 5.8) validate over again the great relevance of these questions. We emphasize that grammar questions are considered to be more relevant (84,09%) to the 3rd cycle when comparing them to reading comprehension questions (73,64%). In addition, for the other cycles (1st and 2nd) it reveals less relevance (1,48% and 23,39%). This can be explained by the fact that this type of questions is more focused on the program content for the 3rd cycle, although many of them are part of the content for other cycles. To conclude, we present a new line of future work. The system could be perfectly extensible with more questions that can better cover other content and, in turn, more study cycles.
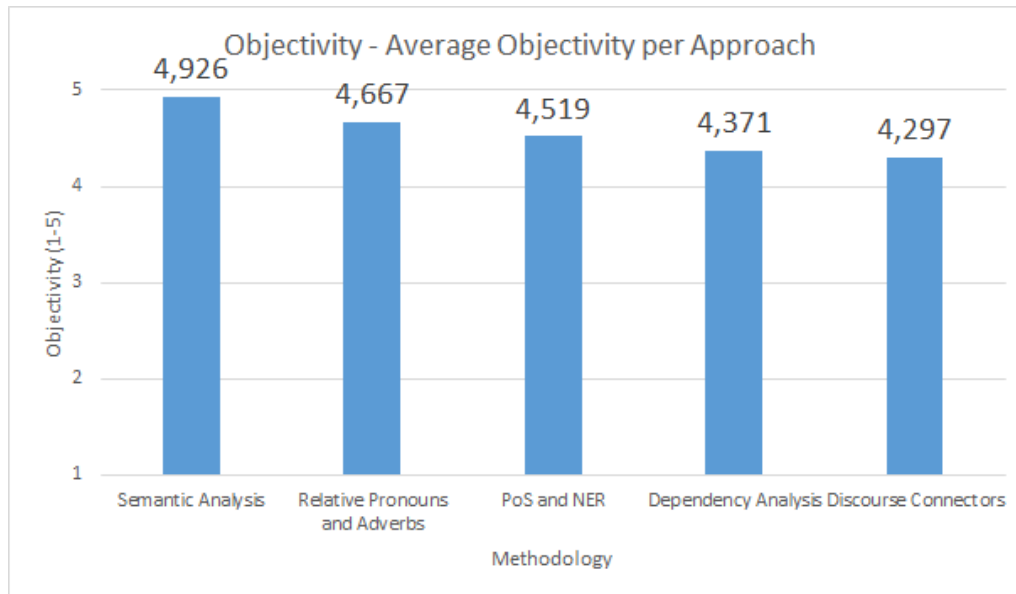
Figure 5.2: Average objectivity per approach (1-5).
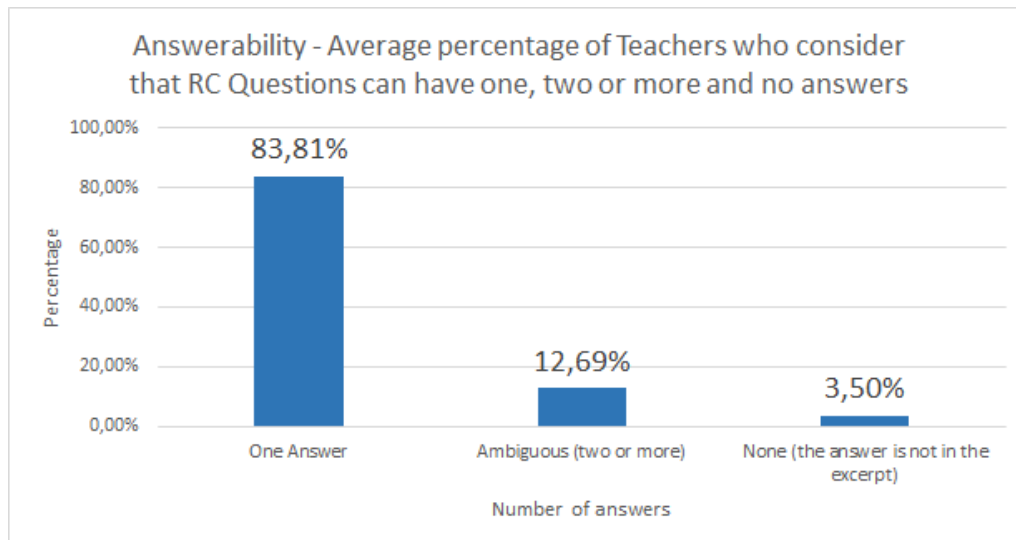


Figure 5.3: Average percentage of Teachers who consider that RC questions can have one, two or more and no answer.
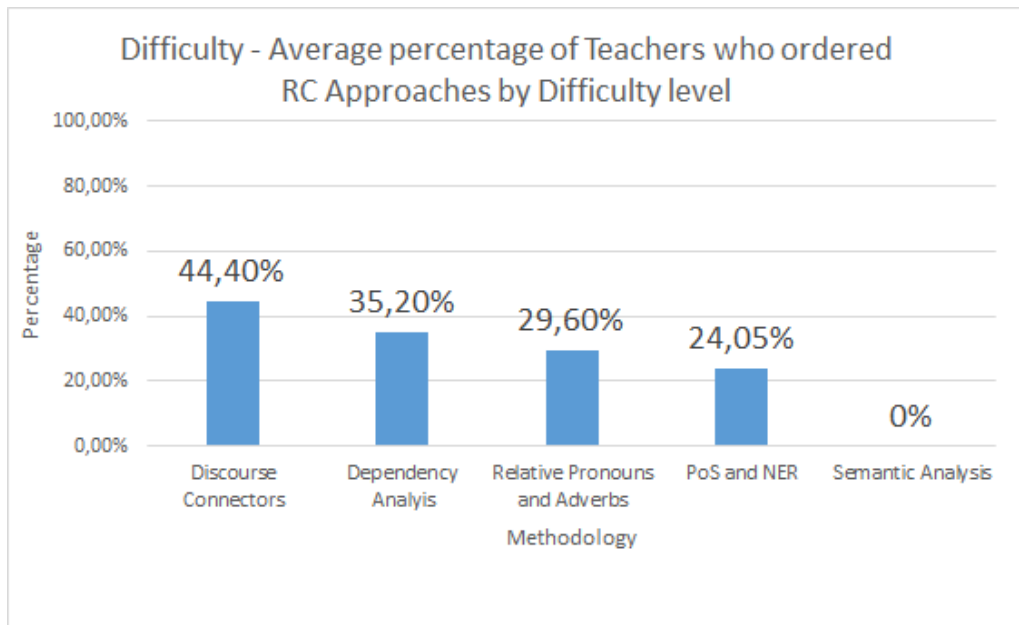
Figure 5.4: Average percentage of Teachers who ordered RC approaches by their difficulty level. Note: For this metric, it was possible to choose more than one option, hence the sum of the percentages can be greater than 100%.
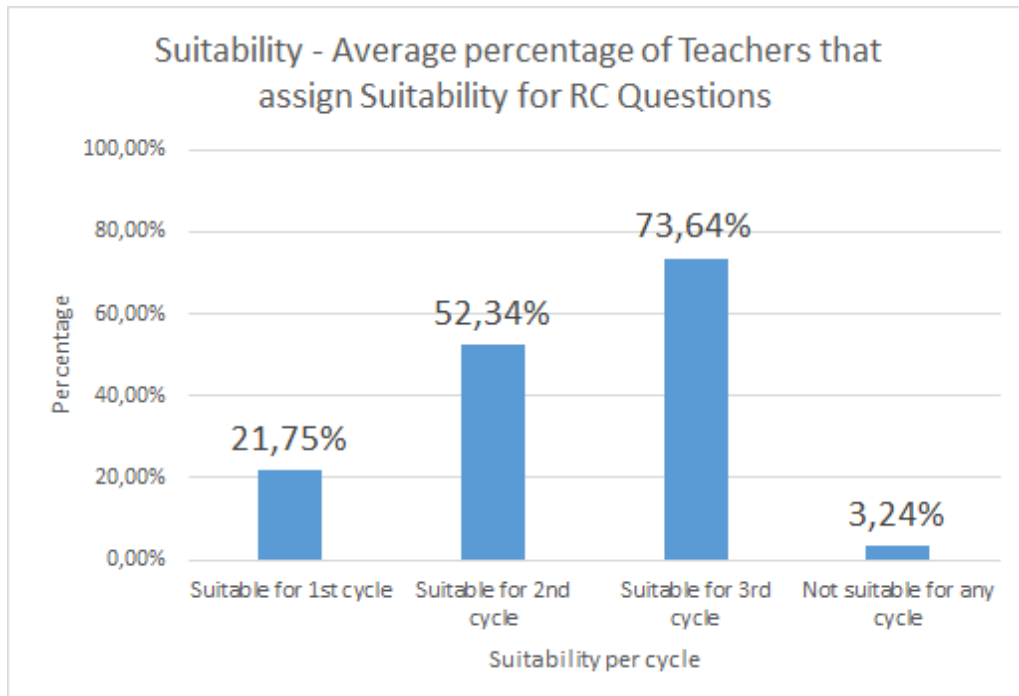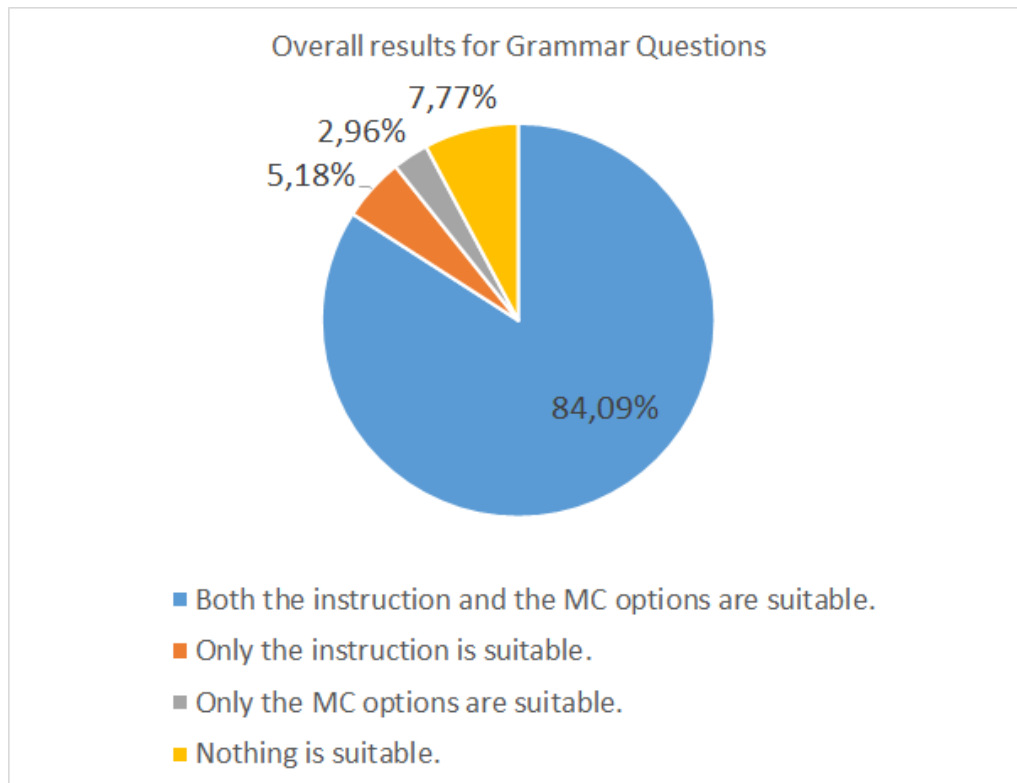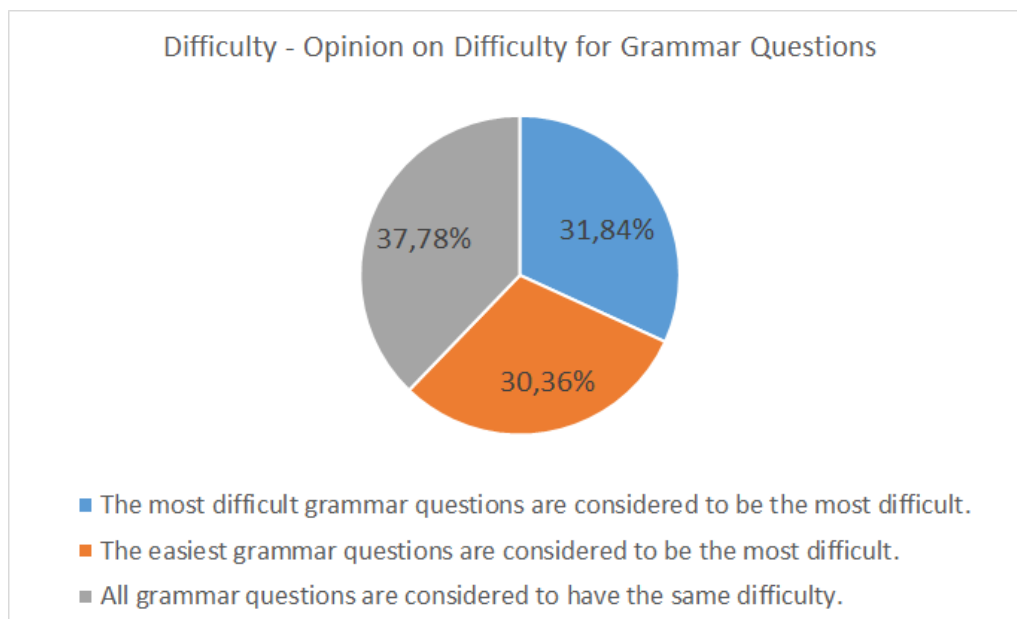


Figure 5.5: Average percentage of Teachers that assign suitability for RC questions. Note: For this metric, it was possible to choose more than one option, hence the sum of the percentages can be greater than 100%.

Figure 5.6: Overall results for Grammar questions.



Figure 5.7: Opinion on difficulty for grammar questions.
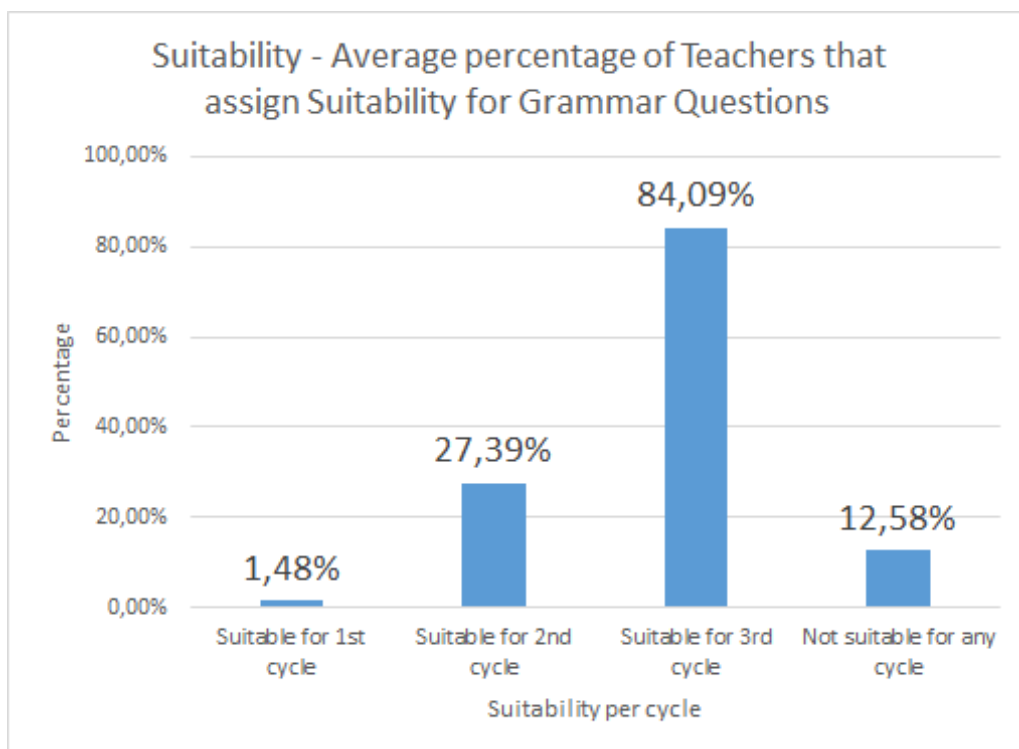
Figure 5.8: Average percentage of Teachers that assign suitability for grammar questions. Note: For this metric, it was possible to choose more than one option, hence the sum of the percentages can be greater than 100%.

## 5.3   Evaluation - Questionnaire for Students

The questionnaire for students followed a scheme similar to a real test/exam. They first read an excerpt extracted from a Portuguese tale, as previously mentioned. After reading that text, they answered the reading comprehension questions. Since these types of questions cannot be automatically evaluated, a Portuguese Teacher kindly evaluated all the answers. There are no intermediate scores, so either an answer is considered correct or incorrect. The results can be observed in Table 5.1. As expected, students attending the 9th grade have been able to answer more questions correctly (84,95%) than in other years. Still, the difference is not very significant. The percentage of correctly answered questions from the students of 8th grade has a lower value (74,67%). In addition to the results presented in the referred table, we can advance that the questions from which the students failed most, were those generated from Dependency and Discourse Connectors approaches. Again, that is an expected event, since those questions require a higher cognitive level.

Finally, we present the results for the answers given to the grammar questions in the Table 5.2. Contrary to what was observed from the teacher's survey, here we can observe a clear conclusion showing that students had more difficulties in questions that are more difficult according to the defined heuristics. All students, from all school levels, scored lower, on average, in the questions that are defined as being more difficult. As such, we see promising practical utilities by using our approaches for grammar questions. At the same time, controlling their difficulty reveals to be truly useful in order to help identify and overcome existing difficulties from students.

Table 5.1: Average percentage of students who answered correctly the reading comprehension questions, per school level.

| School Level | 7th Year | 8th Year | 9th Year | PFL | Overall |
|:---:|:---:|:---:|:---:|:---:|:---:|
| Avg. Percentage of students who answered correctly | 80,19% | 74,67% | 84,95% | 79,47% | 79,82% |

Table 5.2: Average percentage of students who answered correctly the grammar questions (easier vs harder), per school level.

| School Level/ Difficulty Level | Easier Questions | Harder Questions | Overall (Easier + Harder) |
|:---:|:---:|:---:|:---:|
| 7th year | 71,6% | 53,3% | 62,5% |
| 8th year | 58,42% | 43,42% | 50,92% |
| 9th year | 73,06% | 62,04% | 67,55% |
| PFL | 65,71% | 51,43% | 58,57% |
| 7th + 8th + 9th + PFL | 67,22% | 52,56% | 59,89% |

# Chapter 6

# Conclusions

This chapter is a summary of what was covered throughout the document and where we highlight the importance of this theme. Also, we intend to reinforce what is our motivation which leads us to the proposal. We will recall the problem, the one we set out to solve. Then, we will explain why our work is important and how it can contribute positively to the target population. To this end, we will recall our research questions and initial goals. We will strengthen the reasons that made us follow the implemented approaches and what we hoped to achieve with them. Besides, we will explain how effective/useful our methodology was to solving our research questions. We will recall the obtained results and conclude how they corresponded to our initial expectations. Finally, we will give a concrete answer to our research questions and emphasize what were our scientific contributions. We will, after all, present new lines of research for future work.

The inclusion of new technologies in the world of pedagogy is taking place very slowly, so it is necessary to reinforce this inclusion with effective approaches that make life easier for both teachers and students. One of the most important aspects of teaching is the elaboration of rich questions since it helps teachers to understand what are the main difficulties of each student. In addition, the student is able to develop crucial skills such as critical thinking.

If we look carefully, one of the main problems regarding the teaching environment is the excessive time spent on question elaboration. Teachers spend a large part of their time preparing questions for evaluations. In other cases, they tend to reuse previous tests. Furthermore, in general, questions are always the same for all students, preventing questions to be rich and variate.

The contributions of this dissertation have their greatest relevance and importance to the educational process. For teachers, a new mechanism to produce questions, automatically, covering varied contents, avoiding repetition of the previous examinations. The fact that these questions can contemplate different difficulties is another major point, making it possible to address the difficulties for each class, or individually for each student.

Our research questions, raised at the beginning of this document, sought concrete answers on how would it be possible to automatically generate meaningful questions, for the Portuguese subject. To address that, we defined our goals by using AI and NLP techniques to generate questions, bringing improvement in the educational setting and academic performance. Moreover, we enrich and refine ways to control the difficulty of the generated questions. All of this was achieved by creating effective approaches to assist teachers, students and educators.

For grammar questions we follow a rule-based approach. To achieve this, we have established well-defined rules, according to the current Portuguese grammar. We control the difficulty of these questions during their generation. In order to define difficulty, we established specific factors and difficulty values.

For reading comprehension (and factual) questions, we tested five distinct approaches, allowing us to observe which ones are the most advantageous depending on the context. The first one (1) performs a syntax-based analysis by using the information obtained from PoS and Named Entity Recognition. The second (2) approach carries out a semantic analysis of the sentences, through Semantic Role Labeling. The third (3) method extracts the inherent dependencies within sentences using Dependency Parsing. The fourth (4) takes advantage of the relative pronouns and adverbs found throughout the sentences, and generates questions accordingly. The fifth (5), explores the usefulness and practicality of the discourse connectors.

Finally, we implement our last approach that builds pronoun reference questions, in which we do not only generate our questions but also their text passages. We are using heuristic functions that assign difficulty values for each question. For that, we take into account 3 factors: cognitive level, complexity of the source text and complexity of the sentence that originates the question. We follow all these methodologies since we believe it would be important to test several hypotheses that allow questions to be generated for the Portuguese language. Furthermore, we believe that all of these methodologies were very imperative to give concrete answers for our research questions.

The evaluation was performed in two parts: a survey applied to teachers and a questionnaire for students. In the first, we draw conclusions about the formulation, objectivity, relevance, and the difficulty of the generated questions. For the second, we conclude about the practical utility that our questions have in a real context when applied with students. The results corresponded to our initial expectations. Overall, teachers considered questions to be well-formulated, objective and relevant, with greater prominence for the 3rd cycle of studies. The questions that use the dependency analysis (3) and discourse connectors (5) approaches were considered the most difficult, which fits with our heuristics, as they require a higher cognitive level. It was also possible to validate difficulty control for grammar questions since students scored less on their most difficult questions.

Taking into account the previous results, we consider being in a position to answer the research questions, initially proposed. The answer is positive and promising. It is possible to reduce the time spent on the manual elaboration of the questions, by using an automatic mechanism that generates questions from texts. Grammar questions can be generated, through the methodology that was followed and explained in our proposal. Also, it is possible to generate factual questions,

for reading comprehension, using any of the five proposed approaches, providing rich questions that can serve various educational needs. Finally, we conclude that it is possible to create questions within different difficulty levels, before or after their generation, using appropriate heuristics.

We emphasize our scientific contribution in the AQG field, for the Portuguese language. This contribution is confirmed because, when returning to the exposed problem (Section 1.2), we concluded that we managed to optimize the educational process with regard to question elaboration. Beyond this, we were able to fill gaps from the literature (Section 3.2) that concerns the very little study in AQG for Portuguese. We also fill gaps for the AQG community, in general, by proposing innovative manners to control the difficulty of the generated questions. Our findings outline a promising future for AQG, not only in Portuguese but for other languages that contain little study.

Following the developed study, we identify several paths for future work. The first thing would be to improve the morphological analyzer for Portuguese, from the Pre-Processing phase. As previously indicated, improvements can be made to more accurately classify the classes and subclasses for the Portuguese words, namely by performing disambiguation. Still in that phase, it would be possible to enrich our heuristics which classify a text concerning its difficulty. To do that, it would be interesting to readjust factor weights, or even consider new ones. In relation to our approaches, grammar questions can be extended, by establishing new rules. The same can be applied for the reading comprehension approaches, namely 1, 2 and 3 (above-numbered), as these can be enriched by searching for more patterns. More specifically, in 1 and 2, better models may be used, in order to perform NER and SRL, respectively, for Portuguese. From the approach that takes advantage of relative pronouns and adverbs (4), it will be interesting to admit new pronouns and adverbs, allowing the generation of more questions in different contexts. Similarly, for the methodology that makes use of discourse connectors (5), more discourse connectors could be considered and more than one sentence as arguments. Pronoun reference questions can be improved by using a coreference resolver, in the step of distractor generation, reducing the possibility of considering distractors in the same coreference chain. Still, on these questions, it would be useful to calculate the semantic similarity within their multiple-choice options, helping to control the difficulty. Finally, in the post-processing phase, it would be compelling to add more post-editing operations that reduce the number of ill-formed questions. Also, to add a question-raking phase within our pipeline, making it possible to sort the questions according to their formulation and grammatical quality. Regarding the weights of the functions that define the difficulty of the generated questions, it would be interesting to calibrate them according to the experiences made with the students - we would have a system that fits the student's community reality in which it is being used.

We want to highlight, as the last possible line of investigation, the study of neural question generation for Portuguese. For that, we would need to work towards building a dataset (in Portuguese), with paragraph-answer pairs and questions generated based on the pairs. While this can be a promising approach for the future, it is important to keep in mind the continuous need of generating usefulness questions for an educational context.

We close this dissertation by reinforcing precisely this last point. Whatever approaches are

used to generate questions, we consider that the most important of all is that they assure educational relevance.

*If not, what would be the real purpose of generating questions?*

# Appendix A

# Portuguese Words Classes and Subclasses

- DET_ARTICLE_DEFINITE = [o,a,os,as]

- DET_ARTICLE_INDEFINITE = [um,uma,uns,umas]

- DET_POSSESSIVE = [meu,teu,seu,nosso,vosso,seu, minha, tua, sua,nossa, vossa, sua, meus, teus, seus, nossos, vossos, seus, minhas, tuas, suas, nossas, vossas, suas]

- DET_DEMONSTRATIVE = [este, esse, aquele, esta, essa, aquela, estes, esses, aqueles, estas, essas, aquelas]

- DET_INDEFINITE = [certo, outro, certa, outra, certos, outros, certas, outras]

- DET_RELATIVE = [cujo, cuja, cujos, cujas]

- DET_INTERROGATIVE = [qual, quais, que]

- PRON_POSSESSIVE = [meu,teu,seu,nosso,vosso,seu, minha, tua, sua,nossa, vossa, sua, meus, teus, seus, nossos, vossos, seus, minhas, tuas, suas, nossas, vossas, suas]

- PRON_DEMONSTRATIVE = [este, esse, aquele, esta, essa, aquela, estes, esses, aqueles, estas, essas, aquelas, o, a, os, as, isto, isso, aquilo]

- PRON_INDEFINITE = [algum, nenhum, todo, muito, pouco, tanto,outro,alguma, nenhuma, toda, muita, pouca, tanta, outra, qualquer, alguns, nenhuns, todos, muitos, poucos, tantos, outros, algumas, nenhumas, todas, muitas, poucas, tantas, outras, quaisquer, alguém, ninguém, outrem, tudo, nada, algo, cada]

- PRON_RELATIVE = [qual, quais, que, quem]                    ▷ Does not include "onde".

- PRON_INTERROGATIVE = [quanto, qual, quanta, quantos, quais,quantas, que, quê, quem, porque, porquê, como, onde ]

- ADV_MODE = [assim, bem, debalde, mal, depressa, devagar, alegremente, simpaticamente, agradavelmente, fortemente, velozmente, carinhosamente]

- ADV_TIME = [agora, ainda, amanhã, anteontem, antigamente, cedo, então, frequentemente, hoje, já, jamis, nunca, ontem, sempre, tarde]

- ADV_LOCAL = [abaixo, acima, acolá, adiante, aí, além, algures, ali, aquém, aqui, atrás, cá, defronte, dentro, fora, junto, lá, longe, perto]

- ADV_DEGREE = [bastante, demais, demasiadamente, deveras, extremamente, mais, menos, muito, pouco, quase, tanto, tão]

- ADV_AFFIRMATION = [sim, certamente, decerto, efetivamente, realmente]

- ADV_NEGATION = [não]

- ADV_INCLUSION = [até, inclusivamente, mesmo, também]

- ADV_EXCLUSION = [apenas, exceto, exclusivamente, salvo, senão, só, somente, unicamente]

- ADV_DOUBT = [provavelmente, possivelmente, talvez, porventura, acaso, quiçá]

- ADV_DESIGNATION = [eis]

- ADV_INTERROGATIVE = [aonde, donde, quando, porquê, porque]      ▷ Does not include "como" and "onde".

- ADV_CONNECTIVE = [porém, contudo, todavia, primeiramente, seguidamente, consequentemente]

- ADV_RELATIVE = [onde, como]                    ▷ Does not include "como" and "onde".

- CCONJ_COPULATIVE = [e,nem]

- CCONJ_ADVERSATIVE = [mas, porém, todavia, contudo]

- CCONJ_DISJUNCTIVE = [ou, quer]

- CCONJ_CONCLUSIVE = [logo, portanto]                    ▷ Does not include "pois".

- CCONJ_EXPLICATIVE = [que, porquanto]                    ▷ Does not include "pois".

- SCONJ_CAUSAL = [porque, pois, porquanto]        ▷ Does not include "que" and "como".

- SCONJ_FINAL = [para]                    ▷ Does not include "que".

- SCONJ_TEMPORAL = [quando, enquanto, mal, apenas]       ▷ Does not include "que" and "como".

- SCONJ_CONCESSIVE = [embora, conquanto] CommentDoes not include "que".

- SCONJ_COMPARATIVE = [conforme, segundo, qual, quanto] ▷ Does not include "que" and "como".

# Appendix B

# Application - Graphical User Interface

In order to integrate all approaches in a unique application, we developed a graphical user interface. This interface allows the user to choose the desired type of questions, as well the desired number of questions and difficulty degree. Finally, the user can introduce any textual source in order to generate the questions. Figure B.1 presents the first page of the application where we can observe all the available options.

The first option allows to **choose one or more question types** for grammar (8 question types), reading comprehension (4 question themes) and pronoun reference (1 question type). The second option allows to choose the **desired number of questions** (per question type). The available options are: up to 5, 10, 20, 50 questions or as many as possible. The system will attempt to generate the indicated number of questions. If this is not possible, then it will present only those which were generated. The third option allows the user to choose the **intended difficulty for the questions** that will be generated. There are three possible options: a higher degree of difficulty, random degree of difficulty and a lesser degree of difficulty. Depending on the chosen option, the application will try to maximize, randomize or minimize the difficulty of the generated questions. Finally, the user is asked to enter a text. This is the text from which the questions will eventually be generated. Figure B.2 presents the first page with all the options filled out. For this example, we aim to generate as many questions as possible, from all topics. In addition, we chose the option that will try to maximize the difficulty of the questions.

After clicking on the *Gerar Questões (Generate Questions)* option, the application will start the generation process. When this process is over, the questions are presented on another page. Figure B.3 shows the results for the grammar questions, Figure B.4 for the reading comprehension questions, and finally, Figure B.5 for the pronoun reference questions.

# Geração Automática de Questões em Português

**Que tipo de questões pretende gerar?**

- ☐ Gramática
  - ☐ Sequência morfológica
  - ☐ Determinantes, pronomes, advérbios e conjunções
  - ☐ Determinantes, pronomes e preposições (a)
  - ☐ Verbos - Identificar os tempos e modos verbais indicados
  - ☐ Verbos - Completar com os tempos e modos verbais indicados
  - ☐ Verbos - Identificar as formas verbais pertencentes ao mesmo modo verbal
  - ☐ Verbos - Identificar a subclasse das formas verbais indicadas
  - ☐ Frases simples e complexas
- ☐ Compreensão da Leitura
  - ☐ Identificação (pessoas, organizações, acontecimentos, valores)
  - ☐ Localização no tempo e no espaço
  - ☐ Caracterização/Descrição
  - ☐ Causa/Motivo
- ☐ Referenciação de Pronomes

**Quantas questões pretende gerar (por tipo de pergunta) ?**

Selecione uma opção ⌄

**Prentende obter questões com...**

Selecione uma opção ⌄

**Coloque aqui o seu texto**

Escreva aqui uma ou mais frases.

**Gerar Questões**

Figure B.1: Graphical User Interface - The first page with all the options to fill.

Figure B.2: Graphical User Interface - The first page with all the options filled.

**Gramática**

**Indica a frase que contém a sequência: determinante-nome-verbo-preposição-determinante-nome**

A. A tarde corria sobre o rio.
B. A renúncia endurecia os seus músculos.
C. A praia estava cheia de gente.
D. Agora as gaivotas recolhiam a terra.

**Assinale a única frase que contém um determinante indefinido.**

A. Escolhe outra coisa.
B. Este estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de arrebatado e selvático, tornou-se depressa um dos monumentos famosos da cidade e vinha gente das redondezas para o ver.
C. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade.
D. Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe as roupas de que precisava e também papel e caneta.

**Assinale a única frase que contém um determinante indefinido.**

A. Escolhe outra coisa.
B. Este estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de arrebatado e selvático, tornou-se depressa um dos monumentos famosos da cidade e vinha gente das redondezas para o ver.
C. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade.
D. Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe as roupas de que precisava e também papel e caneta.

**Assinale a única frase complexa das quatros apresentadas.**

A. Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos.
B. A vida de Hans mais uma vez tinha virado.
C. Hans e a família moravam no interior da ilha.
D. Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).

**"Embora, em rigor tudo tivesse sido possível." A forma verbal "tivesse sido" encontra-se no**

A. Pretérito mais-que-perfeito composto do modo conjuntivo
B. Futuro composto do modo indicativo
C. Pretérito imperfeito do modo conjuntivo
D. Pretérito imperfeito do modo indicativo

**"Embora, em rigor tudo tivesse sido possível." Identifique corretamente as subclasses dos verbos presentes na frase.**

A. um verbo auxiliar e um verbo copulativo
B. um verbo auxiliar e um verbo principal
C. dois verbos principais
D. um verbo copulativo e um verbo principal

Figure B.3: Graphical User Interface - Grammar questions.

**Compreensão da Leitura**

**Questão: Em que local Hans e a família moravam?**

Hans e a família moravam no interior da ilha.

**Questão: Como é que se tornaram as suas viagens?**

As suas viagens tornaram-se rápidas e espaçadas.

**Questão: Quem compreendeu que a sua vida não seria a sua própria vida?**

E Hans compreendeu que, como todas as vidas, a sua vida não seria a sua própria vida, a que nele estava impaciente e latente, mas um misto de encontro e desencontro, de desejo cumprido e desejo fracassado.

**Questão: O que é que Hans nunca tinha projetado?**

Associado ao inglês, começou a construir uma fortuna que Hans nunca tinha projetado.

**Questão: Qual o motivo pelo qual era um homem de negócios hábil?**

Era um homem de negócios hábil, pois apercebia-se da natureza das coisas e da natureza das pessoas.

**Questão: Que acontecimento originou o desembarque de muitos liberais no Mindelo?**

De facto, a Guerra Civil Portuguesa originou o desembarque de muitos liberais no Mindelo.

**Questão: Como caracteriza a honestidade de Hans?**

A honestidade de Hans era célebre e a sua palavra era de oiro.

Figure B.4: Graphical User Interface - Reading Comprehension questions.

**Referenciação de Pronomes**

*Do dever cumprido, da liberdade assumida, não esperava sucesso nem prosperidade, nem mesmo paz. Os seus irmãos mais novos tinham morrido no naufrágio de um veleiro construído com o melhor material, o seu barco era um bom barco. Aquele era familiar. Ele sabia que os seus jovens irmãos eram perfeitos homens do mar.*

**No excerto anterior, o pronome "Aquele" refere-se a**

A. *veleiro.*
B. *naufrágio.*
C. *material.*
D. *barco.*

Figure B.5: Graphical User Interface - Pronoun Reference questions.

# Appendix C

# Survey and Questionnaire

## C.1   Survey for Teachers

Secção 1 de 16

# [Professores] Geração Automática de Perguntas em Português

Olá! Sou um estudante do Mestrado Integrado em Engenharia Informática e Computação da Faculdade de Engenharia da Universidade do Porto.

O questionário que se segue decorre no âmbito da minha Dissertação orientada pelo Professor Henrique Lopes Cardoso e coorientada pelo Professor Luís Paulo Reis.

O objetivo da minha Dissertação é (entre outros) a criação de uma ferramenta (técnicas de Inteligência Artificial) que permite gerar perguntas de forma automática a partir de frases declarativas extraídas de textos (artigos, notícias, livros, biografias, etc...). As perguntas geradas são de interpretação e de gramática.

O objetivo deste questionário é avaliar a qualidade das questões geradas.

Leia atentamente, por favor, as questões geradas e responda consoante a sua opinião.

A obra que serviu de base para a geração automática de questões é:
"Saga" de Sophia de Mello Breyner Andresen.

Obrigado desde já! Espero que num futuro próximo possam usufruir da nossa aplicação.
A sua opinião é muito importante!

por: Bernardo Leite

Após a secção 1    Continuar para a secção seguinte

---

Secção 2 de 16

# Assinale a sua situação profissional.

Descrição (opcional)

**Sou... ***

◯ Professor(a) de Português (Língua Materna)

◯ Professor(a) de Português (Língua Não Materna)

◯ Outro

---

Se respondeu "Outro" escreva aqui a sua situação profissional.

Texto de resposta longa

Após a secção 2    Continuar para a secção seguinte

---

Secção 3 de 16

# 1.1. Em que local Hans e a família moravam?

[Excerto Original] Hans e a família moravam no interior da ilha. Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).

---

**Formulação da Pergunta - Considera que a pergunta está bem formulada? ***

◯ Sim, está bem formulada.

◯ Não, não está bem formulada.

---

**Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) ***

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ◯ | ◯ | ◯ | ◯ | ◯ | Muito objetiva |

---

**Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? ***

◯ Uma.

◯ Ambígua (duas ou mais).

◯ Nenhuma (a resposta não se encontra no excerto).

---

**Considera esta pergunta adequada para (selecione uma ou mais opções) ***

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

## 1.2. Como é que se tornaram as suas viagens?

[Excerto Original] Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos. As suas viagens tornaram-se rápidas e espaçadas.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

## 1.3. Quem compreendeu que a sua vida não seria a sua própria vida?

[Excerto Original] E Hans compreendeu que, como todas as vidas, a sua vida não seria a sua própria vida, a que nele estava impaciente e latente, mas um misto de encontro e desencontro, de desejo cumprido e desejo fracassado.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

## 1.4. O que é que Hans nunca tinha projetado?

[Excerto Original] Associado ao inglês, começou a construir uma fortuna que Hans nunca tinha projetado.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Após a secção 6    Continuar para a secção seguinte

## 1.5. Qual o motivo pelo qual era um homem de negócios hábil?

[Excerto Original] Era um homem de negócios hábil, pois apercebia-se da natureza das coisas e da natureza das pessoas. A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si próprio envolvia.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Após a secção 7    Continuar para a secção seguinte

## 1.6. Que acontecimento originou o desembarque de muitos liberais no Mindelo?

[Excerto Original] Algum tempo depois casou com a filha de um general liberal que desembarcara no Mindelo e cuja espada, mais tarde, transitando de herança em herança, se conservou na família. De facto, a Guerra Civil Portuguesa originou o desembarque de muitos liberais no Mindelo.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Após a secção 8    Continuar para a secção seguinte

## 1.7. Como caracteriza a honestidade de Hans?

[Excerto Original] Era agora um homem rico e também respeitado e escutado. A honestidade de Hans era célebre e a sua palavra era de oiro.

Formulação da Pergunta - Considera que a pergunta está bem formulada? *

○ Sim, está bem formulada.

○ Não, não está bem formulada.

Objetividade da Pergunta - Considera a pergunta objetiva? (1- Nada objetiva, 5 - Muito objetiva) *

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Nada objetiva | ○ | ○ | ○ | ○ | ○ | Muito objetiva |

Existência de Resposta - Quantas respostas considera que esta pergunta poderá ter? *

○ Uma.

○ Ambígua (duas ou mais).

○ Nenhuma (a resposta não se encontra no excerto).

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Após a secção 9    Continuar para a secção seguinte

## Atente novamente nas questões anteriores.

1.1. Em que local Hans e a família moravam?
1.2. Como é que se tornaram as suas viagens?
1.3. Quem compreendeu que a sua vida não seria a sua própria vida?
1.4. O que é que Hans nunca tinha projetado?
1.5. Qual o motivo pelo qual era um homem de negócios hábil?
1.6. Que acontecimento originou o desembarque de muitos liberais no Mindelo?
1.7. Como caracteriza a honestidade de Hans?

No geral, qual (ou quais) as questões que considera serem mais difíceis para os alunos? *
(selecione uma ou mais opções)

☐ 1.1

☐ 1.2

☐ 1.3

☐ 1.4

☐ 1.5

☐ 1.6

☐ 1.7

Após a secção 10    Continuar para a secção seguinte

## Título da secção (opcional)

Descrição (opcional)

Atente no seguinte excerto e na pergunta respetiva.

*Do dever cumprido, da liberdade assumida, não esperava sucesso nem prosperidade, nem mesmo paz. Os seus irmãos mais novos tinham morrido no naufrágio de um veleiro construído com o melhor material, o seu barco era um bom barco. Aquele era familiar. Ele sabia que os seus jovens irmãos eram perfeitos homens do mar.*

**No excerto anterior, o pronome "Aquele" refere-se a**

**(a) veleiro.**
(b) naufrágio.
(c) material.
(d) barco.

Considera o excerto bem redigido? *

○ Sim

○ Não

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Após a secção 11    Continuar para a secção seguinte

## Título da secção (opcional)

Descrição (opcional)

Atente na seguinte questão de gramática.

1. Indique a frase que contém a sequência:
determinante-nome-verbo-preposição-determinante-nome
   A. A renúncia endurecia os seus músculos.
   B. A praia estava cheia de gente.
   C. **A tarde corria sobre o rio.**
   D. Agora as gaivotas recolhiam a terra.

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Atente na seguinte questão de gramática.

2. Indique a frase que contém a sequência: nome-conjunção-determinante-nome-verbo preposição-nome-preposição-nome
   A. Foi baptizado de urgência recebendo o nome de Sirena.
   **B. Hans e a família moravam no interior da ilha.**
   C. A língua estrangeira fechava em sua roda um círculo.
   D. A vida de Hans mais uma vez tinha virado.

Tendo em conta a pergunta acima transcrita, considera que *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Atente novamente nas mesmas questões.

1. Indique a frase que contém a sequência:
determinante-nome-verbo-preposição-determinante-nome
   A. A renúncia endurecia os seus músculos.
   B. A praia estava cheia de gente.
   **C. A tarde corria sobre o rio.**
   D. Agora as gaivotas recolhiam a terra.

2. Indique a frase que contém a sequência: nome-conjunção-determinante-nome-verbo preposição-nome-preposição-nome
   A. Foi baptizado de urgência recebendo o nome de Sirena.
   **B. Hans e a família moravam no interior da ilha.**
   C. A língua estrangeira fechava em sua roda um círculo.
   D. A vida de Hans mais uma vez tinha virado.

Em relação à dificuldade das questões anteriores, considera que *

○ a pergunta 1 é mais difícil que a pergunta 2.

○ a pergunta 2 é mais difícil que a pergunta 1.

○ ambas apresentam o mesmo grau de dificuldade.

Após a secção 12    Continuar para a secção seguinte                               ▾

Secção 13 de 16

## Título da secção (opcional)

Descrição (opcional)

Atente na seguinte questão de gramática.

3. Assinale a única frase que contém uma conjunção subordinativa causal.
   **A. Manda-me o teu pai que te diga que não voltes a Vig pois não te receberá.**
   B. Unido ao balanço, Hans, enquanto lavava o convés, polia os metais ou enrolava os cabos, aspirava a 4 veemência da vasta respiração marítima.
   C. Assim é desde o tempo antigo das guerras quando os invasores que ocupavam a ilha penetravam nas casas de cabeça erguida mas exigiam que a gente da ilha se curvasse para os saudar.

D. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade .

---

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente na seguinte questão de gramática.

4. Assinale a única frase que contém um determinante indefinido.
   A. Este estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de arrebatado e selvático, tornou-se depressa um dos monumentos famosos da cidade e vinha gente das redondezas para o ver.
   B. **Escolhe outra coisa.**
   C. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade.
   D. Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe as roupas de que precisava e também papel e caneta.

---

Tendo em conta a pergunta acima transcrita, considera que *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente novamente nas mesmas questões.

3. Assinale a única frase que contém uma conjunção subordinativa causal.
   A. **Manda-me o teu pai que te diga que não voltes a Vig pois não te receberá.**
   B. Unido ao balanço, Hans, enquanto lavava o convés, polia os metais ou enrolava os cabos, aspirava a 4 veemência da vasta respiração marítima.
   C. Assim é desde o tempo antigo das guerras quando os invasores que ocupavam a ilha penetravam nas casas de cabeça erguida mas exigiam que a gente da ilha se curvasse para os saudar.
   D. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade .

4. Assinale a única frase que contém um determinante indefinido.
   A. Este estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de arrebatado e selvático, tornou-se depressa um dos monumentos famosos da cidade e vinha gente das redondezas para o ver.
   B. **Escolhe outra coisa.**
   C. No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço aberto, possibilidade.
   D. Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe as roupas de que precisava e também papel e caneta.

---

Em relação à dificuldade das questões anteriores, considera que *

○ a pergunta 3 é mais difícil que a pergunta 4.

○ a pergunta 4 é mais difícil que a pergunta 3.

○ ambas apresentam o mesmo grau de dificuldade.

---

Após a secção 13    Continuar para a secção seguinte                              ▼

---

Secção 14 de 16

# Título da secção (opcional)

Descrição (opcional)

Atente na seguinte questão de gramática.

5. Assinale a única frase complexa das quatro apresentadas.
   A.  A vida de Hans mais uma vez tinha virado.
   B.  Hans e a família moravam no interior da ilha.
   C.  **Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos.**
   D.  Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Atente na seguinte questão de gramática.

6. Assinale a única frase simples das quatro apresentadas.
   A.  O mar do Norte, verde e cinzento, rodeava Vig, a ilha, e as espumas varriam os rochedos escuros.
   B.  Havia nesse começo de tarde um vaivém incessante de aves marítimas, as águas engrossavam devagar, as nuvens empurradas pelo vento sul acorriam e Hans viu que se estava formando a tempestade.
   C. **Hans concentrava o seu espírito para a exaltação crescente do grande cântico marítimo.**
   D.  Mas ele não temia a tempestade e, com os fatos inchados de vento, caminhou até ao extremo do promontório.

Tendo em conta a pergunta acima transcrita, considera que *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

Atente novamente nas mesmas questões.

5. Assinale a única frase complexa das quatro apresentadas.
   A.  A vida de Hans mais uma vez tinha virado.
   B.  Hans e a família moravam no interior da ilha.
   C.  **Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos.**
   D.  Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).

6. Assinale a única frase simples das quatro apresentadas.
   A.  O mar do Norte, verde e cinzento, rodeava Vig, a ilha, e as espumas varriam os rochedos escuros.
   B.  Havia nesse começo de tarde um vaivém incessante de aves marítimas, as águas engrossavam devagar, as nuvens empurradas pelo vento sul acorriam e Hans viu que se estava formando a tempestade.
   C. **Hans concentrava o seu espírito para a exaltação crescente do grande cântico marítimo.**
   D.  Mas ele não temia a tempestade e, com os fatos inchados de vento, caminhou até ao extremo do promontório.

Em relação à dificuldade das questões anteriores, considera que *

○ a pergunta 5 é mais difícil que a pergunta 6.

○ a pergunta 6 é mais difícil que a pergunta 5.

○ ambas apresentam o mesmo grau de dificuldade.

## Título da secção (opcional)

Descrição (opcional)

---

Atente na seguinte questão de gramática.

7. "Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...)."
A forma verbal "ouvia" encontra-se no
    A. Futuro do modo indicativo
    B. **Pretérito imperfeito do modo indicativo**
    C. Futuro composto do modo conjuntivo
    D. Presente do modo condicional

---

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente na seguinte questão de gramática.

8. "Embora, em rigor tudo tivesse sido possível."
A forma verbal "tivesse sido" encontra-se no
    A. Futuro composto do modo indicativo
    B. Pretérito imperfeito do modo conjuntivo
    C. Pretérito imperfeito do modo indicativo
    D. **Pretérito mais-que-perfeito composto do modo conjuntivo**

---

Tendo em conta a pergunta acima transcrita, considera que *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente novamente nas mesmas questões.

7. "Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...)."
A forma verbal "ouvia" encontra-se no
    A. Futuro do modo indicativo
    B. **Pretérito imperfeito do modo indicativo**
    C. Futuro composto do modo conjuntivo
    D. Presente do modo condicional

8. "Embora, em rigor tudo tivesse sido possível."
A forma verbal "tivesse sido" encontra-se no
    A. Futuro composto do modo indicativo
    B. Pretérito imperfeito do modo conjuntivo
    C. Pretérito imperfeito do modo indicativo
    D. **Pretérito mais-que-perfeito composto do modo conjuntivo**

---

Em relação à dificuldade das questões anteriores, considera que

Escolha múltipla ▾

○ a pergunta 7 é mais difícil que a pergunta 8.  ✕

○ a pergunta 8 é mais difícil que a pergunta 7.  ✕

○ ambas apresentam o mesmo grau de dificuldade.  ✕

○ Adicionar opção ou adicionar "Outra"

Após a secção 15   Continuar para a secção seguinte

Secção 16 de 16

## Título da secção (opcional)
Descrição (opcional)

---

Atente na seguinte questão de gramática.

9. "A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si próprio envolvia." Identifique corretamente as subclasses dos verbos presentes na frase.
   A. dois verbos principais
   B. **um verbo copulativo e um verbo principal**
   C. um verbo auxiliar e um verbo principal
   D. um verbo auxiliar e um verbo copulativo

---

Tendo em conta a pergunta acima transcrita, considera que: *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente na seguinte questão de gramática.

10. "Embora, em rigor tudo tivesse sido possível." Identifique corretamente as subclasses dos verbos presentes na frase.
   A. um verbo copulativo e um verbo principal
   B. dois verbos principais
   C. um verbo auxiliar e um verbo principal
   D. **um verbo auxiliar e um verbo copulativo**

---

Tendo em conta a pergunta acima transcrita, considera que *

○ tanto a instrução da pergunta como as opções de escolha múltipla são adequadas.

○ apenas a instrução da pergunta é adequada.

○ apenas as opções de escolha múltipla são adequadas.

○ nada é adequado nesta pergunta.

---

Considera esta pergunta adequada para (selecione uma ou mais opções) *

☐ o 1º ciclo

☐ o 2º ciclo

☐ o 3º ciclo

☐ nenhum ciclo

---

Atente novamente nas mesmas questões.

9. "A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si próprio envolvia." Identifique corretamente as subclasses dos verbos presentes na frase.
   A. dois verbos principais
   B. **um verbo copulativo e um verbo principal**
   C. um verbo auxiliar e um verbo principal
   D. um verbo auxiliar e um verbo copulativo

10. "Embora, em rigor tudo tivesse sido possível." Identifique corretamente as subclasses dos verbos presentes na frase.
   A. um verbo copulativo e um verbo principal
   B. dois verbos principais
   C. um verbo auxiliar e um verbo principal
   D. **um verbo auxiliar e um verbo copulativo**

---

Em relação à dificuldade das questões anteriores, considera que *

○ a pergunta 9 é mais difícil que a pergunta 10.

○ a pergunta 10 é mais difícil que a pergunta 9.

○ ambas apresentam o mesmo grau de dificuldade.

## C.2   Questionnaire for Students

**Secção 1 de 3**

# Português

Vamos testar os teus conhecimentos!

Assinale o ano de escolaridade que frequenta. *

○ 5ºAno

○ 6ºAno

○ 7ºAno

○ 8ºAno

○ 9ºAno

Após a secção 1   Continuar para a secção seguinte

**Secção 2 de 3**

## Grupo I - Compreensão da Leitura

Descrição (opcional)

1. Leia atentamente o seguinte texto e responda corretamente às perguntas.

### TEXTO

Hans e a família moravam no interior da ilha. Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia (...).

A vida de Hans mais uma vez tinha virado. Já não eram as longas navegações até aos confins dos continentes, o avançar aventuroso ao longo das costas luxuriantes e de costas desérticas, de povo em povo, de baía em baía. Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, controlava as cargas e descargas, discutia negócios e contratos. As suas viagens tornaram-se rápidas e espaçadas.

E Hans compreendeu que, como todas as vidas, a sua vida não seria a sua própria vida, a que nele estava impaciente e latente, mas um misto de encontro e desencontro, de desejo cumprido e desejo fracassado. Embora, em rigor tudo tivesse sido possível. E compreendeu que as suas grandes vitórias seriam as que não tinha desejado e que, por isso, nem sequer seriam vitórias.

Escreveu ao Pai. Disse-lhe que não era mais um navegador entre as ondas e o vento. Que era um homem estabelecido, em terra firme e que queria voltar a Vig. Foi a Mãe que respondeu à sua carta dizendo que o pai não o receberia.

Associado ao inglês, começou a construir uma fortuna que Hans nunca tinha projetado. Era um homem de negócios hábil, pois apercebia-se da natureza das coisas e da natureza das pessoas. A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si próprio envolvia. Enriquecia porque a sua perceção e os seus cálculos estavam certos.

Algum tempo depois casou com a filha de um general liberal que desembarcara no Mindelo e cuja espada, mais tarde, transitando de herança em herança, se conservou na família. De facto, a Guerra Civil Portuguesa originou o desembarque de muitos liberais no Mindelo.

Escolheu Ana porque tinha a cara redonda e rosada e cheirava a maçã como a primeira mulher criada e como a casa onde ele nascera, e porque o seu loiro de minhota lhe lembrava as tranças das mulheres de Vig. Pouco antes do seu casamento Hoyle morrera e Hans fundara a sua própria firma cuja prosperidade crescia. Era agora um homem rico e também respeitado e escutado. A honestidade de Hans era célebre e a sua palavra era de oiro.

Sophia de Mello Breyner Andresen, "Saga" in *Histórias da Terra e do Mar*, Texto Editora [adaptado]

1.1. Em que local Hans e a família moravam? *

Texto de resposta longa

1.2. Como é que se tornaram as suas viagens? *

Texto de resposta longa

1.3. Quem compreendeu que a sua vida não seria a sua própria vida? *

Texto de resposta longa

1.4. O que é que Hans nunca tinha projetado? *

Texto de resposta longa

1.5. Qual o motivo pelo qual era um homem de negócios hábil? *

Texto de resposta longa

**1.6.** Que acontecimento originou o desembarque de muitos liberais no Mindelo? *

Texto de resposta longa

**1.7.** Como caracteriza a honestidade de Hans? *

Texto de resposta longa

**2.** Leia atentamente o seguinte excerto: *

*Do dever cumprido, da liberdade assumida, não esperava sucesso nem prosperidade, nem mesmo paz. Os seus irmãos mais novos tinham morrido no naufrágio de um veleiro construído com o melhor material, o seu barco era um bom barco. Aquele era familiar. Ele sabia que os seus jovens irmãos eram perfeitos homens do mar.*

**No excerto anterior, o pronome "Aquele" refere-se a**

○ veleiro.

○ naufrágio.

○ material.

○ barco.

Após a secção 2   Continuar para a secção seguinte ▼

Secção 3 de 3

# Grupo II - Gramática

Descrição (opcional)

**1.** Indique a frase que contém a sequência: determinante-nome-verbo-preposição determinante-nome *

○ A tarde corria sobre o rio.

○ A renúncia endurecia os seus músculos.

○ A praia estava cheia de gente.

○ Agora as gaivotas recolhiam a terra.

**2.** Indique a frase que contém a sequência: nome-conjunção-determinante-nome-verbo-preposição-nome-preposição-nome *

○ Hans e a família moravam no interior da ilha.

○ Foi baptizado de urgência recebendo o nome de Sirena.

○ A língua estrangeira fechava em sua roda um círculo.

○ A vida de Hans mais uma vez tinha virado.

**3.** Assinale a única frase que contém uma conjunção subordinativa causal. *

○ Manda-me o teu pai que te diga que não voltes a Vig pois não te receberá.

○ Unido ao balanço, Hans, enquanto lavava o convés, polia os metais ou enrolava os cabos, aspirava a veem...

○ Assim é desde o tempo antigo das guerras quando os invasores que ocupavam a ilha penetravam nas cas...

○ No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço ab...

**4.** Assinale a única frase que contém um determinante indefinido. *

○ Escolhe outra coisa.

○ Este estranho jazigo que entre lápides, bustos, anjos de pedra, canteiros e piedosas cruzes tinha algo de a...

○ No entanto parecia a Hans que algo em sua vida, embora fosse já tão tarde, era ainda espera e espaço ab...

○ Dois dias depois de ter recolhido Hans, Hoyle levou-o ao centro da cidade e comprou-lhe as roupas de que...

**5.** Assinale a única frase complexa das quatro apresentadas. *

○ Agora verificava a ordem dos armazéns, o bom estado dos navios, a competência das equipagens, control...

○ A vida de Hans mais uma vez tinha virado.

○ Hans e a família moravam no interior da ilha.

○ Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia ( ... ).

6. Assinale a única frase simples das quatro apresentadas. *

○ Hans concentrava o seu espírito para a exaltação crescente do grande cântico marítimo.

○ O mar do Norte , verde e cinzento, rodeava Vig, a ilha, e as espumas varriam os rochedos escuros.

○ Havia nesse começo de tarde um vaivém incessante de aves marítimas, as águas engrossavam devagar, a...

○ Mas ele não temia a tempestade e, com os fatos inchados de vento, caminhou até ao extremo do promont...

7. "Ali, o rumor marítimo só em dias de temporal, através da floresta longínqua, se ouvia ( ... ). " *
A forma verbal "ouvia" encontra-se no

○ Pretérito imperfeito do modo indicativo

○ Futuro do modo indicativo

○ Futuro composto do modo conjuntivo

○ Presente do modo condicional

8. "Embora, em rigor tudo tivesse sido possível." *
A forma verbal "tivesse sido" encontra-se no

○ Pretérito mais-que-perfeito do modo conjuntivo

○ Futuro composto do modo indicativo

○ Pretérito imperfeito do modo conjuntivo

○ Pretérito imperfeito do modo indicativo

9. "A fortuna não era nem a sua ambição, nem a sua aventura nem o seu jogo e nela nada de si *
próprio envolvia." Identifique corretamente as subclasses dos verbos presentes na frase.

○ um verbo copulativo e um verbo principal

○ dois verbos principais

○ um verbo auxiliar e um verbo principal

○ um verbo auxiliar e um verbo copulativo

10. "Embora, em rigor tudo tivesse sido possível." Identifique corretamente as subclasses dos *
verbos presentes na frase.

○ um verbo auxiliar e um verbo copulativo

○ um verbo copulativo e um verbo principal

○ dois verbos principais

○ um verbo auxiliar e um verbo principal

# References

[1] Naveed Afzal and Ruslan Mitkov. Automatic generation of multiple choice questions using dependency-based semantic relations. *Soft Computing*, 18:1269–1281, 07 2014.

[2] Manish Agarwal, Rakshit Shah, and Prashanth Mannem. Automatic question generation using discourse cues. In *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9. Association for Computational Linguistics, 2011.

[3] Maha M. Al-Yahya. Ontoque: A question generation engine for educational assesment based on domain ontologies. *2011 IEEE 11th International Conference on Advanced Learning Technologies*, pages 393–395, 2011.

[4] Itziar Aldabe, Maddalen Lopez De Lacalle, Montse Maritxalar, Edurne Martinez, and Larraitz Uria. Arikiturri: an automatic question generator based on corpora and nlp techniques. In *International Conference on Intelligent Tutoring Systems*, pages 584–594. Springer, 2006.

[5] Itziar Aldabe and Montse Maritxalar. Automatic distractor generation for domain specific texts. In *International Conference on Natural Language Processing*, pages 27–38. Springer, 2010.

[6] Itziar Aldabe and Montse Maritxalar. Semantic similarity measures for the generation of science tests in basque. *IEEE Transactions on Learning Technologies*, 7:375–387, 10 2014.

[7] Itziar Aldabe, Montse Maritxalar, and Ruslan Mitkov. A study on the automatic selection of candidate sentences distractors. In *AIED*, pages 656–658, 2009.

[8] H. Ali, Y. Chali, and S. A. Hasan. Automation of Question Generation from Sentences. In *Proceedings of QG2010: The Third Workshop on Question Generation*, Pittsburgh, USA, 2010.

[9] Tahani Alsubait. *Ontology-based multiple-choice question generation*. PhD thesis, School of Computer Science, The University of Manchester, 12 2015.

[10] Daniela Amaral, Evandro Fonseca, Lucelene Lopes, and Renata Vieira. Comparative analysis of Portuguese named entities recognition tools. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2554–2558, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[11] Richard C Anderson and W. Barry Biddle. On asking people questions about what they are reading. *Psychology of Learning and Motivation - Advances in Research and Theory*, 9(C):89–132, 1 1975.

[12] Pedro Azevedo, Bernardo Leite, Henrique Lopes Cardoso, Daniel Castro Silva, and Luís Paulo Reis. Exploring nlp and information extraction to jointly address question generation and answering. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 396–407. Springer, 2020.

[13] Satanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72, 2005.

[14] Meltem Huri Baturay. An overview of the world of moocs. *Procedia - Social and Behavioral Sciences*, 174:427–433, 02 2015.

[15] Laszlo Bednarik and Laszlo Kovacs. Implementation and assessment of the automatic question generation module. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 687–690. IEEE, 2012.

[16] Benjamin S Bloom. Taxonomy of educational objectives. vol. 1: Cognitive domain. *New York: McKay*, 20:24, 1956.

[17] Chutima Boonthum-Denecke, Irwin B Levinstein, Danielle S McNamara, Joseph P Magliano, and Keith K Millis. Nlp techniques in intelligent tutoring systems. In *Encyclopedia of Artificial Intelligence*, pages 1253–1258. IGI Global, 2009.

[18] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2787–2795, Red Hook, NY, USA, 2013. Curran Associates Inc.

[19] Kristy Elizabeth Boyer and Paul Piwek. Proceedings of qg2010: The third workshop on question generation, 2010.

[20] A. Branco, J. Rodrigues, F. Costa, J. Silva, and R. Vaz. Assessing automatic text classification for interactive language learning. In *International Conference on Information Society (i-Society 2014)*, pages 70–78, 2014.

[21] Jonathan Brown, Gwen Frishkoff, and Maxine Eskenazi. Automatic question generation for vocabulary assessment. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 819–826, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

[22] Jill Burstein. Opportunities for natural language processing research in education. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5449 LNCS:6–27, 2009.

[23] Paula Carvalho, Hugo Gonçalo Oliveira, Diana Santos, Cláudia Freitas, and Cristina Mota. Segundo harem: Modelo geral, novidades e avaliação, 01 2008.

[24] Dhawaleswar Rao Ch and Sujan Kumar Saha. Automatic multiple choice question generation from text: A survey. *IEEE Transactions on Learning Technologies*, 2018.

[25] Yllias Chali and Sadid A. Hasan. Towards topic-to-question generation. *Computational Linguistics*, 41(1):1–20, March 2015.

[26] Jeanne Sternlicht Chall and Edgar Dale. Readability revisited: The new dale-chall readability formula, 1995.

[27] Guanliang Chen, Jie Yang, Claudia Hauff, and Geert-Jan Houben. Learningq: a large-scale dataset for educational question generation. In *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[28] Hyunmin Cheong and LH Shu. Automatic extraction of causally related functions from natural-language text for biomimetic design. In *ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 373–382. American Society of Mechanical Engineers Digital Collection, 2012.

[29] Maria Chinkina and Detmar Meurers. Question generation for language learning: From ensuring texts are read to supporting learning. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 334–344, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[30] Maria Chinkina, Simón Ruiz, and Detmar Meurers. Automatically generating questions to support the acquisition of particle verbs: evaluating via crowdsourcing. *CALL in a climate of change: adapting to turbulent global conditions*, page 73, 2017.

[31] Min-Huang Chu, Wen-Yu Chen, and Shou-De Lin. A learning-based framework to utilize e-hownet ontology and wikipedia sources to generate multiple-choice factual questions. In *2012 Conference on Technologies and Applications of Artificial Intelligence*, pages 125–130. IEEE, 2012.

[32] Ricardo Conejo, Eduardo Guzmán, and Mónica Trella. The siette automatic assessment environment. *International Journal of Artificial Intelligence in Education*, 26, 11 2015.

[33] David Coniam. A preliminary inquiry into using corpus word frequency data in the automatic generation of english language cloze tests. *CALICO Journal*, 14, 01 1997.

[34] Rui Correia, Jorge Baptista, Maxine Eskenazi, and Nuno Mamede. Automatic generation of cloze question stems. In *International Conference on Computational Processing of the Portuguese Language*, pages 168–178. Springer, 2012.

[35] Rui Pedro dos Santos Correia, Jorge Baptista, Nuno Mamede, Isabel Trancoso, and Maxine Eskenazi. Automatic generation of cloze question distractors. In *Second language studies: acquisition, learning, education and technology*, 2010.

[36] Andre Cruz, Gil Rocha, and Henrique Lopes Cardoso. Exploring spanish corpora for portuguese coreference resolution. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 290–295. IEEE, 2018.

[37] Pedro Curto, Nuno Mamede, and Jorge Baptista. Automatic text difficulty classifier - assisting the selection of adequate reading materials for european portuguese teaching, 01 2015.

[38] Santos Lopes Curto. Automatic generation of multiple-choice tests. Master's thesis, Instituto Superior Técnico, 2010.

[39] Sérgio Curto, Ana Mendes, and Luisa Coheur. Question generation based on lexico-syntactic patterns learned from the web. *Dialogue & Discourse*, 3(2):147–175, 2012.

[40] Guy Danon and Mark Last. A syntactic approach to domain-specific automatic question generation. *CoRR*, abs/1712.09827, 2017.

[41] Bidyut Das and Mukta Majumder. Factual open cloze question generation for assessment of learner's knowledge. *International Journal of Educational Technology in Higher Education*, 14, 12 2017.

[42] Bidyut Das, Mukta Majumder, and Santanu Phadikar. A novel system for generating simple sentences from complex and compound sentences. *International Journal of Modern Education and Computer Science*, 1:57–64, 01 2018.

[43] Predrag Dašić, Jovan Dašić, Bojan Crvenković, and Veis Serifi. A review of intelligent tutoring systems in e-learning. *Annals of the Oradea University – Fascicle of Management and Technological Engineering*, 15:85–90, 12 2016.

[44] Ana Paula Dias. *Preparar a Prova Final 3º Ciclo - Português*. Texto Editora, 2013.

[45] Xinya Du, Junru Shao, and Claire Cardie. Learning to ask: Neural question generation for reading comprehension. *CoRR*, abs/1705.00106, 2017.

[46] Nan Duan, Duyu Tang, Peng Chen, and Ming Zhou. Question generation for question answering. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 866–874, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[47] Magali Sanches Duran and Sandra Maria Aluísio. Propbank-br: a brazilian treebank annotated with semantic role labels. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1862–1867, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA).

[48] Juliana Pirovani e Marcos Spalenza e Elias Oliveira. Geração automática de questões a partir do reconhecimento de entidades nomeadas em textos didáticos. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 28(1):1147, 2017.

[49] Daniel Faggella. Examples of artificial intelligence in education. https://emerj.com/ai-sector-overviews/examples-of-artificial-intelligence-in-education, 2019.

[50] Cristina Falcão and Maria do Rosário Luís. *Pratica e Aprende Gramática - Gramática Prática de Português - 3º Ciclo Ensino Básico 7º. 8º e 9º*. Porto Editora, 2016.

[51] Ibrahim Eldesoky Fattoh. Automatic multiple choice question generation system for semantic attributes using string similarity measures. *Computer Engineering and Intelligent Systems*, 5(8):66–73, 2014.

[52] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[53] Michael Flor and Brian Riordan. A semantic role-based approach to open-domain automatic question generation. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 254–263, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[54] EB Fonseca, Renata Vieira, and Aline Vanin. Corp: Coreference resolution for portuguese. In *12th International Conference on the Computational Processing of Portuguese, Demo Session (PROPOR)*, 2016.

[55] Yifan Gao, Jianan Wang, Lidong Bing, Irwin King, and Michael R. Lyu. Difficulty controllable question generation for reading comprehension. *CoRR*, abs/1807.03586, 2018.

[56] D Gates. Generating look-back strategy questions from expository texts. In *The Workshop on the Question Generation Shared Task and Evaluation Challenge, NSF, Arlington, VA. http://www. cs. memphis. edu/~ vrus/questiongeneration//1-Gates-QG08.pdf*, 2008.

[57] Donna M Gates. Automatically generating reading comprehension look-back strategy: Questions from expository texts. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 2008.

[58] Donna Marie Gates. How to generate cloze questions from definitions: A syntactic approach. In *2011 AAAI Fall Symposium Series*, 2011.

[59] Farhad Soleimanian Gharehchopogh and Zeynab Abbasi Khalifelu. Using Intelligent Tutoring Systems in Instruction and Education. *2nd International Conference on Education and Management Technology*, 13(5):250–254, 2011.

[60] Peter C Gordon and Kimberly A Scearce. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*, 23(3):313–323, 1995.

[61] Takuya Goto, Tomoko Kojiri, Toyohide Watanabe, Tomoharu Iwata, and Takeshi Yamada. An automatic generation of multiple-choice cloze questions based on statistical learning. In *Proceedings of the 17th International Conference on Computers in Education*, pages 415–422. Asia-Pacific Society for Computers in Education, 2009.

[62] Michael Heilman, Vincent Aleven, William W Cohen, Diane J Litman, and Noah A Smith. *Automatic Factual Question Generation from Text Thesis Committee Thesis Committee*. PhD thesis, Carnegie Mellon University, 2011.

[63] Michael Heilman and Noah A. Smith. Good question! statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 609–617, Los Angeles, California, June 2010. Association for Computational Linguistics.

[64] Yan Huang and Lianzhen He. Automatic generation of short answer questions for reading comprehension assessment. *Natural Language Engineering*, 22(3):457–489, 2016.

[65] Noémia Jorge, Maria Teresa, Brocardo, and Sofia Rente. *Gramática de Português - 3.º Ciclo*. Porto Editora, 2019.

[66] Dr Khaled. Natural language processing and its use in education. *International Journal of Advanced Computer Science and Applications*, 5, 12 2014.

[67] Payal Khullar, Konigari Rachna, Mukul Hase, and Manish Shrivastava. Automatic question generation using relative pronouns and adverbs. In *Proceedings of ACL 2018, Student Research Workshop*, pages 153–158, 2018.

[68] Yanghoon Kim, Hwanhee Lee, Joongbo Shin, and Kyomin Jung. Improving neural question generation using answer separation. *CoRR*, abs/1809.02393, 2018.

[69] J. Peter Kincaid, Robert P. Fishburne, Richard Lawrence Rogers, and Brad S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel, 1975.

[70] Ghader Kurdi, Jared Leo, Bijan Parsia, and Salam Al-Emari. A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 11 2019.

[71] Andrey Kurtasov. A system for generating cloze test items from Russian-language text. In *Proceedings of the Student Research Workshop associated with RANLP 2013*, pages 107–112, Hissar, Bulgaria, September 2013. INCOMA Ltd. Shoumen, BULGARIA.

[72] Chonlathorn Kwankajornkiet, Atiwong Suchato, and Proadpran Punyabukkana. Automatic multiple-choice question generation from thai text. In *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 1–6. IEEE, 2016.

[73] Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. Race: Large-scale reading comprehension dataset from examinations. *arXiv preprint arXiv:1704.04683*, 2017.

[74] Benito Mendoza Laura Zavala. On the use of semantic-based aig to automatically generate programming exercises. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, page 14–19, New York, NY, USA, 2018. Association for Computing Machinery.

[75] Nguyen-Thinh Le and Niels Pinkwart. Evaluation of a question generation approach using semantic web for supporting argumentation. *Research and Practice in Technology Enhanced Learning*, 10, 06 2015.

[76] Bernardo Leite, Henrique Lopes Cardoso, Luís Paulo Reis, and Carlos Soares. Factual question generation for the portuguese language, (accepted) 2020.

[77] Jared Leo, Ghader Kurdi, Nicolas Matentzoglu, Bijan Parsia, Uli Sattler, Sophie Forge, Gina Donato, and Will Dowling. Ontology-based generation of medical, multi-term mcqs. *International Journal of Artificial Intelligence in Education*, 01 2019.

[78] Roger Levy and Galen Andrew. Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *LREC*, pages 2231–2234. Citeseer, 2006.

[79] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[80] David Lindberg, Fred Popowich, John Nesbit, and Phil Winne. Generating natural language questions to support learning on-line. In *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 105–114, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[81] Chao-Lin Liu, Chun-Hung Wang, Zhao-Ming Gao, and Shang-Ming Huang. Applications of lexical information for algorithmically composing multiple-choice cloze items. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP*, pages 1–8, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

[82] Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2122–2132, Austin, Texas, November 2016. Association for Computational Linguistics.

[83] M. Liu, V. Rus, and L. Liu. Automatic chinese factual question generation. *IEEE Transactions on Learning Technologies*, 10(2):194–204, April 2017.

[84] Ming Liu, Rafael Calvo, and Vasile Rus. G-asks: An intelligent automatic question generation system for academic writing support. *Dialogue and Discourse*, 3:101–124, 01 2012.

[85] Ming Liu, Rafael A. Calvo, and Vasile Rus. Automatic question generation for literature review writing support. In Vincent Aleven, Judy Kay, and Jack Mostow, editors, *Intelligent Tutoring Systems*, pages 45–54, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[86] Shawn Loewen. Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, 27, 09 2005.

[87] Carmo Azeredo Lopes, Isabel Freitas M. Pinto, and Olga Azeredo. *Da Comunicação à Expressão, Gramática Prática de Português - 3º Ciclo do Ensino Básico e Ensino Secundário*. Raiz Editora, 2018.

[88] JJ Lu and LA Harris. Artificial intelligence (ai) and education. *Congressional Research Service in Focus*, 2018.

[89] Yizhi Ma and Keng L Siau. Artificial intelligence impacts on higher education. *MWAIS Proceedings*, page 42, 2018.

[90] Nitin Madnani, Jill Burstein, John Sabatini, Kietha Biggers, and Slava Andreyev. Language muse: Automated linguistic activity generation for english language learners. In *Proceedings of ACL-2016 System Demonstrations*, pages 79–84, 2016.

[91] Joe Magliano, Keith Millis, The Team, Irwin Levinstein, and Chutima Boonthum. Assessing comprehension during reading with the reading strategy assessment tool (rsat). *Metacognition and Learning*, 6:131–154, 08 2011.

[92] Mukta Majumder and Sujan Kumar Saha. A system for generating multiple choice questions: With a novel approach for sentence selection. In *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications*, pages 64–72, Beijing, China, July 2015. Association for Computational Linguistics.

[93] Prashanth Mannem, Rashmi Prasad, and Aravind Joshi. Question generation from paragraphs at upenn: Qgstec system description. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 84–91, 2010.

[94] T B F Martins, C M Ghiraldelo, Maria das Graças Volpe Nunes, and Osvaldo Novais de Oliveira Junior. Readability formulas applied to textbooks in brazilian portuguese, 1996.

[95] Karen Mazidi and Rodney D. Nielsen. Linguistic considerations in automatic question generation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 321–326, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[96] Karen Mazidi and Rodney D. Nielsen. Pedagogical evaluation of automatically gener-
ated questions. In Stefan Trausan-Matu, Kristy Elizabeth Boyer, Martha Crosby, and Kitty
Panourgia, editors, *Intelligent Tutoring Systems*, pages 294–299, Cham, 2014. Springer In-
ternational Publishing.

[97] Karen Mazidi and Paul Tarau. Automatic question generation: From nlu to nlg. In *Interna-
tional conference on intelligent tutoring systems*, pages 23–33. Springer, 2016.

[98] Danielle McNamara, Irwin Levinstein, and Chutima Boonthum. istart: Interactive strategy
training for active reading and thinking. *Behavior research methods, instruments, comput-
ers : a journal of the Psychonomic Society, Inc*, 36:222–33, 06 2004.

[99] RUSLAN MITKOV, LE AN HA, and NIKIFOROS KARAMANIS. A computer-aided
environment for generating multiple-choice test items. *Natural Language Engineering*,
12(2):177–194, 2006.

[100] Ruslan Mitkov and Le An Ha. Computer-aided generation of multiple-choice tests. In
*Proceedings of the HLT-NAACL 03 Workshop on Building Educational Applications Using
Natural Language Processing*, pages 17–22, 2003.

[101] Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and
Lucy Vanderwende. Generating natural questions about an image. In *Proceedings of the
54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long
Papers)*, pages 1802–1813, Berlin, Germany, August 2016. Association for Computational
Linguistics.

[102] Jack Mostow, YI-TING HUANG, HYEJU JANG, ANDERS WEINSTEIN, JOE VALERI,
and Donna Gates. Developing, evaluating, and refining an automatic generator of diagnostic
multiple choice cloze questions to assess children's comprehension while reading. *Natural
Language Engineering*, -1:1–50, 04 2017.

[103] Annamaneni Narendra, Manish Agarwal, and Rakshit Shah. Automatic cloze-questions
generation. In *Proceedings of the International Conference Recent Advances in Natural
Language Processing RANLP 2013*, pages 511–515, Hissar, Bulgaria, September 2013.
INCOMA Ltd. Shoumen, BULGARIA.

[104] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder,
and Li Deng. MS MARCO: A human generated machine reading comprehension dataset.
*CoRR*, abs/1611.09268, 2016.

[105] OCDE. What are the social benefits of education? *Education Indicators in Focus*, 10, 2013.

[106] P Pabitha, M Mohana, S Suganthi, and B Sivanandhini. Automatic question generation
system. In *2014 International Conference on Recent Trends in Information Technology*,
pages 1–5. IEEE, 2014.

[107] Liangming Pan, Wenqiang Lei, Tat-Seng Chua, and Min-Yen Kan. Recent advances in
neural question generation. *CoRR*, abs/1905.08949, 2019.

[108] Shivank Pandey and KC Rajeswari. Automatic question generation using software
agents for technical institutions. *International Journal of Advanced Computer Research*,
3(13):307–311, 2013.

[109] Andreas Papasalouros, Konstantinos Kanaris, and Konstantinos Kotis. Automatic generation of multiple choice questions from domain ontologies. In *e-Learning*, pages 427–434. Citeseer, 2008.

[110] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[111] Harold Pashler, Patrice Bain, Brian Bottge, Arthur Graesser, Kenneth Koedinger, Mark Mcdaniel, and Janet Metcalfe. Organizing instruction and study to improve student learning. ies practice guide. ncer 2007-2004. *National Center for Education Research*, 01 2007.

[112] Laura Perez-Beltrachini, Claire Gardent, and German Kruszewski. Generating grammar exercises. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 147–156. Association for Computational Linguistics, 2012.

[113] Juan Pino, Michael Heilman, and Maxine Eskenazi. A selection strategy to improve cloze question quality. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains. 9th International Conference on Intelligent Tutoring Systems, Montreal, Canada*, pages 22–32, 2008.

[114] André Pires. Named entity extraction from portuguese web text. Master's thesis, University of Porto, Faculty of Engineering, R. Dr. Roberto Frias, 4200-465 Porto, 2017.

[115] Emily Pitler, Annie Louis, and Ani Nenkova. Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 683–691. Association for Computational Linguistics, 2009.

[116] Paul Piwek and Kristy Boyer. Varieties of question generation: Introduction to this special issue. *Dialogue and Discourse*, 3:1–9, 01 2012.

[117] Rashmi Prasad and Aravind Joshi. A discourse-based approach to generating why-questions from texts. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*, pages 1–3. National Science Foundation Arlington, VA, 2008.

[118] Debra Pugh, Andre F De Champlain, Mark Gierl, Hollis Lai, and Claire Touchie. Using cognitive models to develop quality multiple-choice questions. *Medical teacher*, 38 8:838–43, 2016.

[119] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[120] Sheetal Rakangor and YR Ghodasara. Literature review of automatic question generation systems. *International journal of scientific and research publications*, 5(1):1–5, 2015.

[121] J. Rodrigues, F. Costa, J. Silva, and A. Branco. Automatic syllabification of portuguese, 2016.

[122] Henry L Roediger III and Mary A Pyc. Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, 1(4):242–248, 2012.

[123] Anselm Rothe, Brenden M. Lake, and Todd M. Gureckis. Question asking as program generation. *CoRR*, abs/1711.06351, 2017.

[124] Ernst Rothkopf. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal - AMER EDUC RES J*, 3:241–249, 11 1966.

[125] Vasile Rus, Zhiqiang Cai, and Arthur C. Graesser. Experiments on generating questions about facts. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 444–455, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[126] Vasile Rus, Philip M McCarthy, Danielle S McNamara, and Arthur C Graesser. A study of textual entailment. *International Journal on Artificial Intelligence Tools*, 17(04):659–685, 2008.

[127] Vasile Rus, Brendan Wyse, Paul Piwek, Mihai Lintean, Svetlana Stoyanchev, and Moldovan. A detailed account of the first question generation shared task evaluation challenge how to cite. *Dialogue and Discourse*, 3, 01 2012.

[128] Arief Yudha Satria and Takenobu Tokunaga. Automatic generation of english reference question by utilising nonrestrictive relative clause. In *CSEDU (1)*, pages 379–386, 2017.

[129] Arief Yudha Satria and Takenobu Tokunaga. Evaluation of automatically generated pronoun reference questions. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 76–85, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[130] Carolina Scarton and Sandra Aluisio. Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do coh-metrix para o português. *Linguamática*, 2, 04 2010.

[131] Iulian Vlad Serban, Alberto García-Durán, Caglar Gulcehre, Sungjin Ahn, Sarath Chandar, Aaron Courville, and Yoshua Bengio. Generating factoid questions with recurrent neural networks: The 30M factoid question-answer corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 588–598, Berlin, Germany, August 2016. Association for Computational Linguistics.

[132] Rakshit Shah. Automatic question generation using discourse cues and distractor selection for cloze questions. *Language Technology and Research Center (LTRC), International Institute of Information Technology, Hyderabad*, 2012.

[133] Katherine Stasaski and Marti A. Hearst. Multiple choice question generation utilizing an ontology. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 303–312, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[134] Lazar Stosic. The importance of educational technology in teaching. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, 3:111–114, 06 2015.

[135] Eiichiro Sumita, Fumiaki Sugaya, and Seiichi Yamamoto. Measuring non-native speakers proficiency of english by using a test with automatically-generated fill-in-the-blank questions. In *Proceedings of the Second Workshop on Building Educational Applications Using*

*NLP*, pages 61–68, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

[136] B. Sun, Y. Zhu, Y. Xiao, R. Xiao, and Y. Wei. Automatic question tagging with deep neural networks. *IEEE Transactions on Learning Technologies*, 12(1):29–43, Jan 2019.

[137] Xingwu Sun, Jing Liu, Yajuan Lyu, Wei He, Yanjun Ma, and Shi Wang. Answer-focused and position-aware neural question generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3930–3939, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[138] Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Controlling item difficulty for automatic vocabulary question generation. *Research and Practice in Technology Enhanced Learning*, 12:25, 12 2017.

[139] Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Evaluation of automatically generated english vocabulary questions. *Research and Practice in Technology Enhanced Learning*, 12, 03 2017.

[140] Goto Takuya, Kojiri Tomoko, Toyohide Watanabe, Iwata Tomoharu, and Yamada Takeshi. Automatic generation system of multiple-choice cloze questions and its evaluation. *Knowledge Management and E-Learning : an International Journal*, 2, 09 2010.

[141] John H. Wolfe. Automatic question generation from text - an aid to independent study. *SIGCUE Outlook*, 10(SI):104–112, February 1976.

[142] John H Wolfe. Reading retention as a function of method for generating interspersed questions. Technical report, NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN DIEGO CALIF, 1977.

[143] Fei Wu and Daniel S. Weld. Open information extraction using Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[144] Xuchen Yao and Yi Zhang. Question generation with minimal recursion semantics. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 68–75. Citeseer, 2010.

[145] Jianwei Zhang and Jun Takuma. A kanji learning system based on automatic question sentence generation. In *2015 international conference on asian language processing (IALP)*, pages 144–147. IEEE, 2015.

[146] Yao Zhao, Xiaochuan Ni, Yuanyuan Ding, and Qifa Ke. Paragraph-level neural question generation with maxout pointer and gated self-attention networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3901–3910, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[147] Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, and Ming Zhou. Neural question generation from text: A preliminary study. *CoRR*, abs/1704.01792, 2017.