UNDERSTANDING ROMANIAN TEXTS BY USING GAMIFICATION METHODS

Ştefania-Eliza BERGHIA[1], Bogdan PAHOMI[1],
Daniel VOLOVICI[1],

[1] *„Lucian Blaga" University of Sibiu, Engineering Faculty, Computer Science and Electrical and Electronics Engineering Department*

Abstract

In recent years, there has been increasing interest in the field of natural language processing. Determining which syntactic function is right for a specific word is an important task in this field, being useful for a variety of applications like understanding texts, automatic translation and question-answering applications and even in e-learning systems. In the Romanian language, this is an even harder task because of the complexity of the grammar. The present paper falls within the field of "Natural Language Processing", but it also blends with other concepts such as "Gamification", "Social Choice Theory" and "Wisdom of the Crowd". There are two main purposes for developing the application in this paper:

a) For students to have at their disposal some support through which they can deepen their knowledge about the syntactic functions of the parts of speech, a knowledge that they have accumulated during the teaching hours at school

b) For collecting data about how the students make their choices, how do they know which grammar role is correct for a specific word, these data being primordial for replicating the learning process

Keywords: Natural Language Processing, Gamification

1 Introduction

Natural Language Processing (NLP) offers one way to make human-computer interaction more interesting and accessible.[1]

The present paper describes a different way of tagging the grammatical functions, by using gamification methods. The main idea is for the students to use the developed application to increase their knowledge about the grammatical roles of the words. They will be able to choose from a list of various syntactic functions, the one that they think is truly eligible for a word or group of words they choose from a text. Their answers are stored and compared with the ones of an expert in the domain, in our case a Romanian language teacher. This comparison will result in a score. On the basis of the score, it will be calculated how much of the answers are correct and a trust will be given. All of this is done in a fun way by using gamification elements to increase user motivation. The goal is to use this data, about the choices of a

user, in implementing a system for automatically tagging the grammatical functions in the Romanian language texts.

2 The Romanian Grammar

As far as the Romanian language is concerned, efforts in the field of natural language processing are made in a few academic centres in Romania and the Republic of Moldova. Among the NLP applications developed, we mention the syllabus application for words, automatic flexing applications (FAVR modelling in the Mac-ELU environment, AnMorph system).

2.1 Parts of Speech in the Romanian Language

The grammar of the Romanian language is considered to be a very complex one. When talking about morphology, we actually refer to the classification of words into parts of speech. In Romanian, from a morphological point of view, the speech parts can be classified into ten types: noun, adjective, pronoun, article, numeral, verb, adverb, preposition, conjunction, interjection. There are also some parts of speech that can be divided into subtypes:

- pronouns: personal, polite, reflexive (which is separate)
- verbs: auxiliary and two other main types
- the article: determined, indefinite, possessive and demonstrative

2.2 Syntactic Functions within Romanian Grammar

According to GALR 2008, "syntactic functions are classes of terms linked by the same type of relationship and the same substitution class "(GALR 2008 II, p. 9) or "classes of substitutable terms in the same context, in other words, classes of functionally equivalent terms in the same position "(ibidem, p. 10).

The main grammatical roles are the following:

1. The subject
2. The predicate
3. The predicative
4. The attribute
5. The complement

Some of these can be of several kinds. For example, for the complement we have:

- object
- adverbial

Model: *Deodată un vânt mare a izbit ferestrele.* (E. Camilar)

S (*cine?*)   P (*ce a făcut?*)

un vânt   a izbit

A (*ce fel de?*)   C.d. (*ce?*)   C.mod (*cum?*)

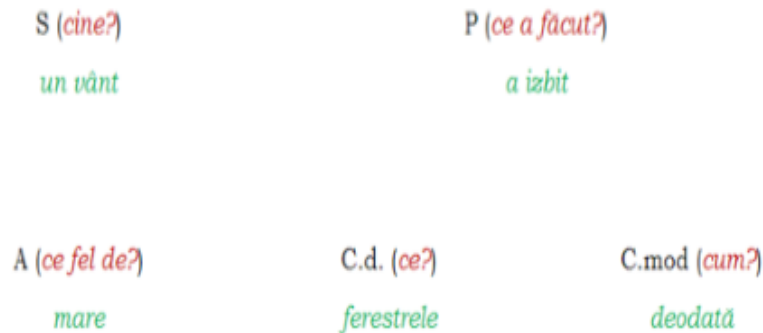mare   ferestrele   deodată

Figure 1: Syntax analysis of a sentence

3 Concept Presentation

In developing this paper we used concepts such as "Gamification", "Social Choice Theory" and "Wisdom of the Crowd".

3.1 1Gamification

In today`s society, students can be considered as digital natives. This means that because they grew up surrounded by technology, they have a different learning process. To keep up with this, the teachers need to adapt this process to the new needs, preferences and requirements of the students in order to maintain their motivation and desire to learn. Such adaptation is exactly the gamification which involves the use of game-specific elements and thoughts, such as giving points based on progress or giving special badges in the context of education.

Game-based approaches lead to a higher level of commitment and motivation of users towards the activities and processes they are involved in. Most of the consumers played or continue to play different games so game mechanics are familiar to them. These approaches based on games are not only true for companies and their employees but in education too.[1] That is precisely why we thought that implementing an application that uses some gamification methods will help the students to learn, test or deepen their knowledge. In this way, their engagement and motivation are enhanced.

3.2 Social Choice Theory

In the Romanian Language analyzing a word or group of words is not an easy task as everything depends very much on the context. In this application, the syntactic analysis of the sentence parts will be done using the Social Choice Theory concept. Basically, this theory explains how to make a decision in the case of voting. We know that elections are based on the majority rule, which classifies a candidate x above a candidate y if and only a majority of individuals do the same.[2]

Let`s see how this applies to our context. For example, for a specific word from a text, the "winning" syntactic function will be considered the one which was chosen most often.

3.3 Wisdom of the Crowd

This phenomenon has gained increasing attention lately. The idea of this is simple: two heads think better than one, and the more they are, the better. It uses the principle of group think, and the concept that the masses are better problem solvers, forecasters, and decision makers than any one individual. An example would be the well-known contest "Who wants to Be a Millionaire". It has been noticed that when one of the contestants was challenged by a particular question, aggregating all the opinions of the people who were left to vote resulted in the correct answer, even if some individuals did not respond properly.[3]
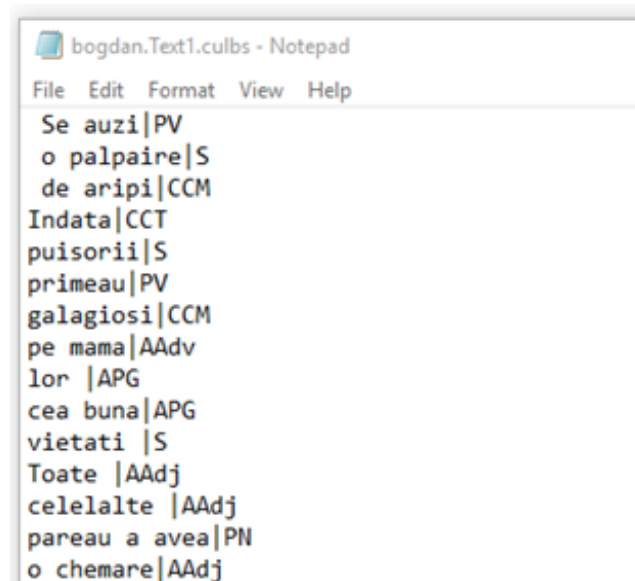
4 Application Implementation

The application from this paper is actually a website. It has a registration/login page, the main page, and a user profile page. In order for the user to access the main page, he will first need to create an account and then log in using the data for which he opted when registering his account. Once logged in, the user will be presented with the main page where he will have certain texts that he will be able to analyze. Also on the main page, the user will be able to view his own profile.

Firstly, for being able to analyze a text the user will need to choose one from a list he will be presented. After that, he will be able to split the text into sentences and take one or more words at a time and choose the syntactic function that seems the more eligible to them, from a list of given grammar functions. When a user finished up analyzing he will submit his answers that will be stored in a database. The answers will be automatically compared with the ones of an expert in domain and a score will be computed. Based on the scores a user gets he will be awarded some merit badges to increase the motivation to learn. He will be able to see his progress and the badges he received in the View Profile section. Also, the scores a user gets are important when computing his "trust rate". The rate will be used to decide if the user is

trustworthy or not. If he gets a high rate it means that his answers were mostly good and we can take them into consideration in a higher proportion. This information among other ones, like the time a user thinks between assigning grammar roles, are essential in replicating the learning process of the students.

5 Results

When submitting the analysis of a text, a file is also created. In this file, we can see the words or group of words and their assigned syntactic functions. This is basically a file with the tags.


```
bogdan.Text1.culbs - Notepad
File  Edit  Format  View  Help
 Se auzi|PV
 o palpaire|S
 de aripi|CCM
Indata|CCT
puisorii|S
primeau|PV
galagiosi|CCM
pe mama|AAdv
lor |APG
cea buna|APG
vietati |S
Toate |AAdj
celelalte |AAdj
pareau a avea|PN
o chemare|AAdj
```

Figure 2: Example of a tags file

These files have major importance since they can be used further on, along with the information about how the students assigned the syntactical function to set up a mechanism for understanding Romanian texts.

6 Conclusions and Future Work

This paper presents an application that may be used to analyze texts and also collect data on how the users made their choices. The analysis is made using some gamification methods, like awarding badges and accumulating scores. After the students analyze some texts the information gathered from them and the tags files resulted are to be used in the next step of the project.

As no application is perfect this one is no exception and it can use a series of further improvements and developments such as:

- adding more gamification elements (for example, a user may not be allowed to access some more complex texts until he managed to correctly analyze a series of x texts)

- giving the possibility of analyzing multiple texts
- using collected data to implement a syntactic recognition system for Romanian texts

7 References

[1] Gabriela Kiryakova, Nadezhda Angelova, Lina Yordanova, "GAMIFICATION IN EDUCATION ".

[2] K. Arrow, „The Principle of Rationality in Collective Decisions" (1952), apud A. Sen (2002: p. 328)

[3] Aidan Lyon, Eric Pacui, „The Wisdom of Crowds: Methods of Human Judgement Aggregation", University of Maryland, College Park.