4th World Conference on Future of Education (WCFEDUCATION)
17 - 19 December 2021 Berlin- Germany

Natural Language Processing as the basis for Digital Educational mini-games for the Greek Language

Aristides Vagelatos[1,*], Christos Tsalidis[2]

[1]Computer Technology Institute, Greece

[2]Neurocom S.A., Greece

*Corresponding author

Abstract

The use of digital games to support learning through an alternative, more engaging, playful way is growing rapidly both in Europe and worldwide (Majuri et al., 2018). Digital games are a rapidly developing field, and at the same time one of the most popular technologies that young people use to entertain themselves. In this context, the project "Lexipaignio" intends to develop an innovative Natural Language Processing (NLP) environment for the creation of digital educational games for students of primary and secondary education in Greece. These games are dynamically created by the teacher for the student, in order to improve vocabulary and more generally language skills, as well as improve their understanding of the general context of specific subject areas.

Keywords: ICT in education, game-based learning, natural language processing, digital educational games.

1. Introduction

The paper concerns the presentation of the digital educational games implemented in the framework of the project "Lexipaignio". This project aims at designing and creating digital language games with the support of natural language processing (NLP) technologies. Moreover, it aims to explore the potential of these technologies in the design of educational games.

The novelty of the games developed in the framework of Lexipaignio consists in the dynamic, automated way of generating the input data from a dedicated corpus that includes textbooks as well as other entities, e.g. generating questions and false alternative answers to use in multiple choice type games, with graded difficulty, for different cognitive subjects.

The basis for this dynamic production is the NLP infrastructure, implemented both as specific resources and as text processing applications, as well as machine learning algorithms, through which the appropriate input to the games is prepared.

In this way, the teacher can control a number of factors, intervening and adjusting the outcome of the chosen game in relation to, for example, the level of education, the class, the difficulty, the subject, the grammatical phenomenon, etc.

Considering that language teaching in school (both in primary and secondary education) is often limited to standardized, mechanistic ways of learning rules and vocabulary that may or may not sometimes be embedded in textual contexts, these games, aspire to be a tool - an

assistant in the hands of the teacher but also a playful alternative at the disposal of the students (Picca, Jaccard & Eberle, 2015).

In the rest of the paper, we present the results of project "Lexipaignio" (Vagelatos, et al., 2021): In the next section we describe the NLP resources that were utilized. Then, the tools that have been implemented based on these resources. Next, we describe the grammatical phenomena that were chosen to be "treated" and the types of games that were evaluated as being the appropriate ones. Finally, some examples are given as well as the conclusions.

2. Digital NLP resources

The digital resources that were improved/utilized to serve as the infrastructure for the implementation of the games within the project are the following.

1.2 Greek Morphological Dictionary

It consists of about 90,000 entries (Tsalidis, Vagelatos & Orphanos, 2004), with various information, such as:

- Orthographic: which letters and in which order constitute the correct spelling of the clitic type.
- Syllabification: which syllables make up the clitic type.
- Morphematic: which morphemes (prefix, subject, suffix, suffix) make up the clitic type.
- Morphosyntactic: which word the clitic type comes from and what are its morphosyntactic features (part of speech, gender, case, number, person, voice, tense, etc.).
- Terminological: if the clinical type is a special term, to which special vocabulary it belongs, e.g. the type abiogenesis is a term of biology.

In addition to common words, the Morphological Dictionary also includes words of special vocabulary, such as 10,000 Greek toponyms (names of Greek prefectures, municipalities, communities, provinces, cities, villages, etc.).

With the use of this dictionary, certain procedures can be automated: e.g. to reverse a word, to differentiate (spelling) a word, to "code" (in the way chosen, e.g. removal of vowels) a word, etc.

1.3 Thesaurus

The Thesaurus is a special type of dictionary which attempts to convey the meanings of words or expressions in Modern Greek through synonyms, opposites and examples of usage. The main features of the Thesaurus are that:

- It contains ~22,000 entries: 11,500 nouns, 5,150 adjectives, 3,500 verbs, 650 expressions and 1,200 adverbs, prepositions and conjunctions.
- Each entry has a distinction of meanings. Synonyms, opposites and examples of usage are grouped by meaning.
- The lemma types, their meanings and synonyms/opposites by meaning are accompanied by stylistic and pragmatic information.

- For any word or phrase that plays the role of synonym, antonym or relative expression within an entry, there is almost always (96%) a Thesaurus entry describing it.
- Each reference to a polysemous synonym or antonym is accompanied by a numerical index, which indicates the number of meanings within the entry of the synonym or antonym.
- Entries are accompanied by references to relevant expressions (if any).

1.4 Corpus

Within the projects' framework, a special corpus was constructed to serve as a basis for the rest of the tools as well as input for the educational games, after being annotated. The main part of the corpus are school textbooks (primary and secondary education) as well as text extracts from "Phrorodentro" the Greek national aggregator of educational content (http://photodentro.edu.gr/). The annotation process of the LAIXIPAIGNIO corpus involved almost all NLP components that existed and included: a tokeniser, a sentence splitter, a morphosyntactic tagger, a toponym gazetteer and a single and multi-word term recognizer.

1.5 Dictionary "Χτίζω Λέξεις" (Build Words)

The special dictionary "Χτίζω Λέξεις" (Vassiliou & Tseva, 2010) presents the ways in which Modern Greek words are constructed, i.e. what their components are and how they are combined with each other. It contains detailed annotated meanings, examples of usage, synonyms-antonyms and notes on spelling and etymology. It is an important source of information on the vocabulary of Modern Greek, and in particular it can be useful for planning and solving language exercises.

The dictionary contains about 220 entries, and more specifically 140 first constituents and 80 second constituents. The glossary has been derived from the indexing and editing of existing dictionaries of Modern Greek.

3. NLP tools that were utilized

Based on the digital resources that described above, a series of NLP tools that have been implemented in previous projects, were augmented and adapted for use in Lexipaignio.

1.6 Speller

Through the spell checker, words typed/read on the computer are checked for spelling problems, and if they are not in the dictionary (meaning that they are either incorrect or simply not entered in the dictionary), alternatives are suggested (Vagelatos et al, 1994). The Speller, can also be used 'in reverse': certain mistakes can be done and inserted into the text automatically, so that the learner can then try to correct these errors.

1.7 Lemmatizer

The Lemmatizer is the linguistic tool that takes as input any verbal type and returns the lemmatic type to which it corresponds (Tsalidis et al, 2009), e.g. for the type "έγινε" (done) it returns the lemmatic type "κάνω" (do). In case the given type corresponds to more than one lemma type, the Lemmatizer returns all of them.

1.8 Hyphenator

In order to achieve full alignment on the lines of a paragraph, electronic typesetting systems move to the next line any word that tends to fall outside the margin, while increasing or decreasing the spaces between words on each line evenly. This process cannot be applied when lines are short (e.g. lines in newspaper columns) or when there are long words at the end of lines, because this creates large gaps in some lines, something that affects negatively the aesthetics of the text and leaves a lot of printable space unused. In this case, the traditional practice of hyphenate words at the end of lines is irreplaceable (Tsalidis et al, 2009). The role of the Hyphenator is to indicate all possible hyphen points of a word.

1.9 Grammar Checker

It is an online tool (Gakis et al., 2015; Kokkinos et al., 2020) for annotating parts of text where appropriate, and more specifically when:

- there is a deviation from a rule of school grammar, e.g. in the use of the final "–ν" in writing or in intonation (e.g. «ο δάσκαλος μας» instead of the correct «ο δάσκαλός μας»),
- using the wrong grammatical form of a word, e.g. the form of the adjective does not match the noun it identifies (e.g. «το αγχώδη άτομο» instead of «το αγχώδες άτομο»),
- incorrect use of a word due to confusion with another word (e.g. «πραγματεύομαι – διαπραγματεύομαι»),
- there is incorrect use of a word within a phrase (e.g. «απορώ και εξανίσταμαι» instead of the correct word «εξίσταμαι»),
- there is a deviation from a typical syntactic structure (e.g. «αποποιήθηκε της περιουσίας αντί την περιουσία»),
- mismatch of a linguistic type to the style of the text (e.g. use of "θεωρούσανε" in a formal style instead of "θεωρούσαν").

Annotation helps the author of the text to identify possible errors and support him/her by indicating the correct grammatical rule. This electronic tool cannot access the meaning of the text, but "collaborates" with the editor (user), who decides on the final form of the text.

4. Examples of grammatical phenomena and types of games

In the following, indicative grammatical phenomena chosen to be analysed, the methods best suited to each of them, and the types of games that can be used for each are described.

1.10 Educational issues addressed

Based on pedagogical needs and priorities, the following phenomena were selected in principle:

- problematic inflections (e.g. "τριγενή και δικατάληκτα»),
- a- privative ("α στερητικό),
- certain suffixes (e.g. "-ώδης", "-ειδής"),

- analogue formations such as "αληθής – αληθινός",
- word families,
- other spelling issues,
- final "ν",
- punctuation marks,
- etymology of words,
- hyphenation of words,
- the adjective "πολύς",
- interrogative words "πού" - "πώς",
- loans from foreign languages,
- imperative,
- synonyms - antonyms.

1.11 Educational issues – technics and exercises

The table below lists the educational issues and the techniques/exercises that can be used for each of them, in mini games implementation:

Table 1: Training Methods - Techniques - Exercises

n	Types of Exercises - Techniques / Educational Issues	True – False	Multiple Choice	Word or List Composition	Fill in the blanks	Word or list generation
1.	difficult inflections	✓	✓	✓	✓	
2.	a- privative	✓	✓	✓	✓	✓
3.	certain suffixes	✓	✓	✓	✓	✓
4.	analogue formations such as "αληθής – αληθινός"	✓	✓	✓	✓	✓
5.	Word families	✓	✓	✓	✓	✓
6.	Homophones	✓	✓		✓	✓
7.	Final "ν"	✓			✓	
8.	Punctuation marks					
9.	Word etymology	✓	✓	✓	✓	✓
10.	Word hyphenation	✓			✓	
11.	Adjective "πολυς"	✓			✓	✓
12.	interrogative words "πού" – "πώς"	✓			✓	
13.	loans from foreign	✓			✓	

	languages					
14.	imperative	✓	✓			
15.	synonyms - antonyms	✓	✓	✓	✓	✓

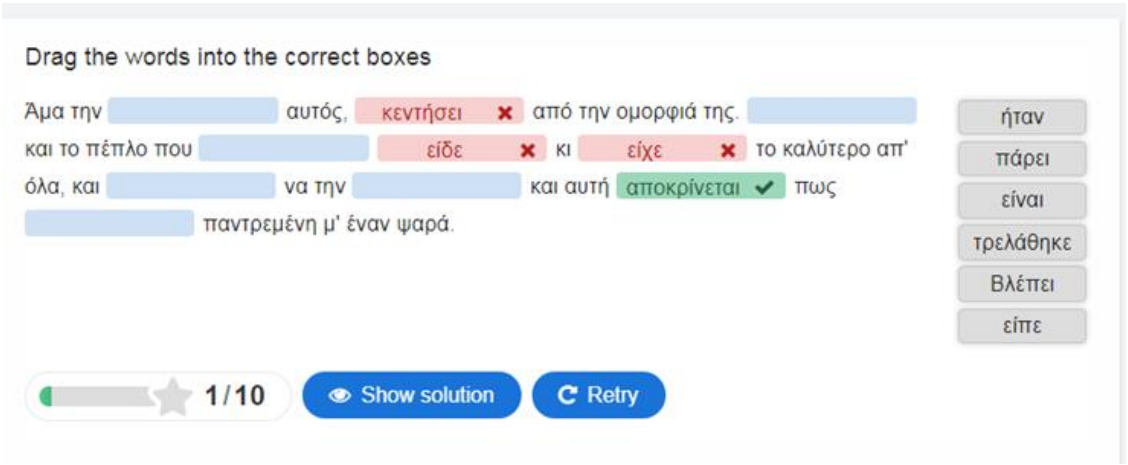
5. Game Examples

Some examples of games that were implemented based on the NLP infrastructure, the pedagogical choices and the educational needs that emerged, are presented below.

1.12 Drag the Words

In this game, a sentence from a textbook is presented where certain words have been removed (hide). The "player" (student) is asked to put the words shown on the right in the correct positions. At the end, he/she can check the answers.

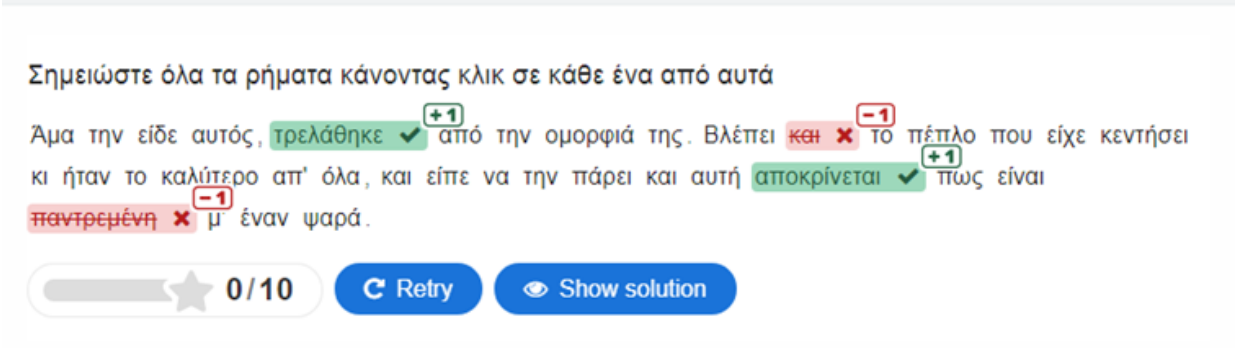
Figure 1: Drag the words into the correct boxes



1.13 Locate the verbs

In this game, a sentence from a textbook is presented. The "player" (student) is asked to identify the verbs in the sentence (the teacher can choose any part of speech). At the end he/she can check the answers.

Figure 2: Locate the Verbs



1.14 Millionaire

This is the well-known game: “who wants to be a millionaire”. In this game, a sentence is presented to the player, with a missing word and four alternatives to select. Among them, only one is the right one.

Figure 3: Millionaire

6. Conclusion

With the aim of using the natural language processing infrastructure for the creation of educational games in various school subjects (e.g., Modern Greek Language, Biology, Geography), the Lexipaignio project has so far given encouraging results (based on the feedback from the research presentations and discussions with teachers that have taken place so far). With the infrastructure completed and a number of games implemented, the next step is to test them in the classroom and win teachers’ (as well as pupils’) approval. Although this step is not straightforward, as it requires interdisciplinary collaboration as well as a rather extensive 'field' testing, the researchers are expecting it with great interest so that the results can provide a basis for improving/developing what has been done so far. Hopefully the project will allow teachers to easily create "minigames" according to their own needs and the needs of their students.

Acknowledgment

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-05094). More information you may find at: http://lexipaignio.cti.gr/

References

Gakis P., Panagiotakopoulos C., Sgarbas K., Tsalidis C., Verykios V. (2015). Construction of a Modern Greek Grammar Checker Through Mnemosyne Formalism. In Ronzhin A.,

Potapova R., Fakotakis N. (Eds.), Speech and Computer. Lecture Notes in Computer Science, v. 9319. Springer.

Kokkinos, T., Gakis, P., Iordanidou, A., & Tsalidis, C. (2020). Utilizing grammar checking software within the framework of differentiated language teaching. 9th International Conference on Educational and Information Technology (ICEIT 2020), St Anne's College, University of Oxford, United Kingdom

Majuri J., Koivisto J., Hamari J. (2018). Gamification of education and learning: A review of empirical literature. Proceedings of the 2nd International GamiFIN conference, (Pori, Finland, May 21-23, 2018), pp. 11-19.

Picca, D., Jaccard, D., Eberle, G. (2015). Natural Language Processing in Serious Games. International Journal of Serious Games, 2 (3) pp. 77-97. Retrieved from 988f1b221c1b8a1fd8c098c3addc6e930f2d.pdf (semanticscholar.org)

Tsalidis, C., Pantazara, M., Minos, P. & Mantzari, E. (2009). NLP tools for lexicographic applications. In Modern Greek, eLexicography in the 21st century: new challenges, new applications, Proceedings of elex 2009, Louvainla-Neuve 22-24 October 2009, Cahiers du CENTAL No 7, Université Catholique de Louvain, pp. 457-462.

Tsalidis, C., Vagelatos, A., Orphanos, G. (2004) An electronic dictionary as a basis for NLP tools: The Greek case. In Proc. Of 11th Conference on Natural Language Processing, Fez, Morocco.

Vagelatos, A., Triantopoulou, T., Tsalidis, C. & Christodoulakis, D. (1994). A Spelling Correction System for Modern Greek, International Journal on Artificial Intelligence & Tools, vol. 3, n. 4.

Vagelatos, A., Stamatopoulos, J., Fountana, M., Gavrielidou, M. & Tsalidis, C. (2021). Natural Language Processing Environment to Support Language Education Games for the Greek Language. Proceeding of International Conference on Interactive Mobile Communication, Technologies and Learning, Thessaloniki, Greece, 4-5 November 2021.

Βασιλείου, Β. & Τσεβά. Ε. (2010). Χτίζω Λέξεις: Πώς σχηματίζονται οι Νεοελληνικές λέξεις. Επιμέλεια: Ιορδανίδου, Α. & Πανταζάρα, Μ. Εκδόσεις Κοτύλι.