

Implementing Language Games with NLP Tools: The Greek Case

Christos
Tsalidis
Neurocom S.A.
Athens, Greece
tsalidis@neurocom.gr

Maria
Fountana
CTI&P
Athens, Greece
fountana@cti.gr

Monica
Gavrielidou
CTI&P
Athens, Greece
monica@cti.gr

John
Stamatopoulos
Neurocom S.A.
Athens, Greece
stamatop@neurocom.gr

Aristides
Vagelatos
CTI&P
Athens, Greece
vagelat@cti.gr

ABSTRACT

Digital games, as a popular technology in youth entertainment, constitute a fast-growing field which has been affecting various aspects of education for several years now. The research project “Lexipaignio” focuses on the development of an innovative and state-of-the-art NLP (Natural Language Processing) environment for the creation of digital educational games for Greek students. A variety of simple and easy-to-play mini-games has been specified aiming to improve students’ linguistic competence by developing a better understanding on various grammatical, morphological and vocabulary related phenomena in general, but also in the context of specific subjects (e.g. geology – geography, biology, etc.). In this paper, the main functionalities of the NLP environment will be presented towards the implementation of mini-games for the Greek language.

CCS CONCEPTS

• Computer games • Natural Language Processing

KEYWORDS

Educational Games, NLP, Game-based Learning

1 Introduction

Natural Language Processing (NLP) is not really a new research field since the first effort started in the 1950s with the so called “Turing test”. Nevertheless, it took more than three decades of research work in order to have real progress with substantial results. Nowadays NLP (which in fact is part of AI) is a research area that gains extreme interest mainly due to the enormous amount of data that are produced every single minute in digital format: the ability to process information and transform it to knowledge is of great value in today’s “information jungle” [1].

On the other hand, the use of digital games to support learning (game-based learning) through an alternative, more attractive way is rapidly developing in both European and worldwide level. Obviously, digital games is a fast developing field, as it is amongst the most popular technologies young people use to amuse themselves. The educational potential of digital games is correlated to the properties of motivation, amusement and the

trigger of interest, which are considered consistent with positive learning results [6].

In this paper, we examine the needed NLP infrastructure that can support the dynamic compilation of educational games for the Greek language, within the research project “Lexipaignio”.

2 Educational Games

Focusing on Education, the “Lexipaignio” project aims at the utilization and further development of a series of Natural Language Processing tools (Morphological Lexicon, Lemmatizer, Mnemosyne language editing system, corpus of Greek school subjects, etc.), for the implementation of dynamically created gamified educational material. The paper highlights the creation of mini-games related to the subject of Greek language in schools. Being part of an ongoing project, the development of language mini-games will provide us with useful feedback regarding the use of NLP for the development of dynamic gamified materials in many school subjects.

According to relevant research [2, 8], computer games provide a quick and interesting learning pace in contrast to the conventional teaching methods and in this perspective, they can affect the dynamics as far as digital learning is concerned. The purpose of the ongoing project is the development of an innovative and state-of-the-art computational environment through the creation of digital educational games for students (primary and secondary level) in order to: a) improve language competence and overall level of students’ knowledge and b) develop various vocabulary and linguistic skills, while understanding the context of specific school subjects (biology, geography etc.).

The new environment will support the automated production of questions related to different levels of competency as far as the Greek language structure and its use are concerned in terms of spelling, morphology, vocabulary, as well as terminology found in school textbooks which is integrated into the overall environment and narration of digital educational games. It will also enable teachers to automatically create a large volume of questions through their insertion in educational games (crosswords, match games, multiple choice, scrabble games, word search puzzles, etc.) and at the same time, it will be possible for them to control

parameters such as: a) school subject (biology, geography, literature etc.), b) grade, c) grammatical phenomena (conjugation, spelling, syntax, vocabulary). We believe that the proposed environment can become a successful tool in supporting and enriching the educational process in an appealing and attractive way. As far as language teaching is concerned, the traditional approach which is mainly restricted to the teaching of rules and is exhausted in monotonous exercises for the students on the assumption that the language is a one-dimensional teaching object, seems not to convey the expected results and should be redefined on the basis of modern functional and communicative teaching approaches.

3 NLP Infrastructure

For the needs of the “Lexipaignio” project various Language Resources and NLP technologies are used, to create a Web Services API to support the operational requirements of educational games [10, 11]. In the backend “Mnemosyne” platform [3] is utilized in pre-processing and runtime phase (see Fig. 1). Mnemosyne incorporates a vast number of language resources and technologies including a) many different dictionaries, e.g. spelling vocabularies, morphology, thesaurus, gazetteers, and b) “classic” NLP technologies like fuzzy matching engines, stemmers, taggers, syntax checkers. Besides the “standard” technologies, the environment offers “modern” NLP and machine learning functionality as *classification mechanisms* such as K-Means and Hierarchical clustering algorithms, *keyword extraction and indexing* using TF/IDF and BM25 algorithms [7], *text production* using n-gram language models [9]. At the top of the stack, Mnemosyne implements several supervised and unsupervised machine learning algorithms such as Naïve-Bayes and Multinomial Linear Regression, as well as word embeddings using CBOW & SKIPGRAM algorithms [4] as well as GLOVE algorithm [5].

The above models have been applied on a corpus of more than 1.5G words collected from electronic news, movies subtitles, literature books, legislation documents, etc. The NLP infrastructure is used in two consequent phases: 1) the preparation phase where the teacher must prepare the data for the gamified lessons and 2) the runtime phase when the games run and asks for data.

In the preparation phase the functionality supported includes:

- queries to language resources, e.g. morphosyntactic dictionary for adjectives ending in “-ης” and “-ες” (see following section)
- the incorporation of a document collection with educational material and extraction of:
 - n-gram language models,
 - clusters of similar documents (application of K-Means and hierarchical clustering algorithms)

- keyword extraction (application of TF/IDF algorithms)
- extraction of candidate terms (application of morphosyntactic patterns that followed by terms)

At the runtime, the system after modelling the functionality and the data needed by a rich set of games useful in education, offers several NLP functions that can feed the games through web services API. Examples of services supported are:

- predetermined word lists from preparation phase,
- fuzzy matching using spelling checker engine,
- synonyms and antonyms using thesaurus engine,
- inflection of nouns, adjectives, verbs, ...
- morphology of words with decomposition in hyphenation, formation using morphemes,
- grammatical checking using grammar checker, etc.

4 Language Games

The focus of Lexipaignio educational mini-games relates to the improvement of language competency level and linguistic abilities of upper primary and lower secondary Greek students. To this, an initial study of categorization of grammatical phenomena and common linguistic errors was conducted. A common mistake in Modern Greek relates to the application of conjugation rules in adjectives ending in “-ης” and “-ες” (πλήρης – πλήρες). These adjectives are of increased difficulty level due to a particularity in the formation of some masculine and feminine types. Such difficulties are noted in terms of spelling, word formation, as well as word use in sentences.

As a step forward, a Greek Language corpus was compiled, which, along with the NLP components served as a basis for the creation of the educational mini-games. The Greek Language corpus comprised of all the material included in the Greek Language books studied in upper primary and lower secondary Greek school in the context of the Modern Greek language course. The use of adjectives in “-ης” and “-ες” in NLP educational mini-games resulted from a thorough study on grammar exercise typologies and their possible applications to the suggested grammatical phenomena and common linguistic mistakes. Next, the above typologies were considered regarding mini-game alternative solutions.

In the case of adjectives in “-ης” and “-ες”, the above led to the construction of a series of mini-games regarding True/False, multiple choice, word creation, list creation, gap filling, text processing and text checking.

Based on the produced infrastructure, the teacher can create his/her own games by selecting a) the grammatical phenomenon of interest, b) the level of difficulty and c) a certain text that he/she is willing to use.

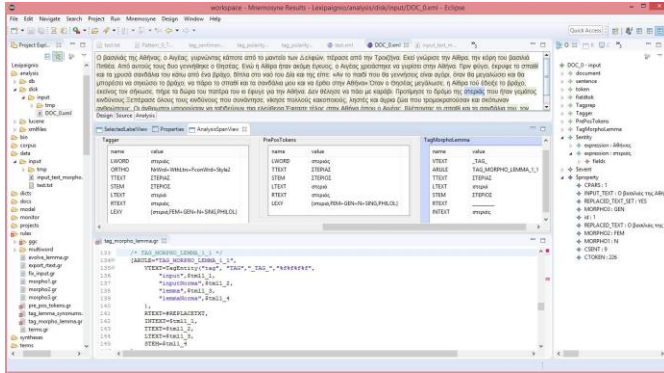


Figure 1: Mnemosyne platform, with morphosyntactic analysis of a text from the corpus.

5 Conclusions

With the aim to deploy Natural Language Processing infrastructure for the creation of educational games in a variety of school subjects (Geography, Modern Greek Language, Biology), so far, the language processing techniques applied in “Lexipaignio” project provide encouraging results. Regarding the implementation of the appropriate infrastructure for dynamic educational games, we hope that soon educators will be able to easily create mini-games according to their students’ needs, by regulating the game content.

ACKNOWLEDGMENTS

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).

REFERENCES

- [1] Alhawiti, KM.: Natural language processing and its use in education. *International Journal of Advanced Computer Science and Applications*, 5(12), (2014).
- [2] Gregory, S., Torsten, R., Wood, L., Henderson, M.: Gamification and digital games-based learning in the classroom. In: *Teaching and Digital Technologies: Big Issues and Critical Questions*. Henderson, M., Romeo, G. Editors, Cambridge University Press (2015).
- [3] Kokkinos, Th., Gakis, P., Iordanidou, A., Tsalidis, Ch.: Utilising Grammar Checking Soft-ware within the Framework of Differentiated Language Teaching. In *Proceedings of the 2020 9th International Conference on Educational and Information Technology*, Oxford, United Kingdom, (2020).
- [4] Mikolov, T., Chen, K., Corrado, G.S., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781.
- [5] Pennington, J., Socher, R., & Manning, C.D. (2014). Glove: Global Vectors for Word Representation. *EMNLP*.
- [6] Picca, D., Jaccard, D., Eberle, G.: Natural Language Processing in Serious Games: A state of the art. *International Journal of Serious Games*, 2(3), 77-97 (2015).
- [7] Ramos, J.: Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, 242, 133-142, (2003).
- [8] Reinders, H.: *Digital Games in Language Learning and Teaching*. Palgrave Macmillan Publishing (2012).
- [9] Singla, A., Karambir, M.: Comparative analysis & evaluation of euclidean distance function and manhattan distance function using k-means algorithm. *International Journal*, 2(7), (2012).

- [10] Tsalidis, C., Vagelatos, A., Orphanos, G.: An electronic dictionary as a basis for NLP tools: The Greek case. In *Proc. Of 11th Conference on Natural Language Processing*, Fez, Morocco (2004).
- [11] Vagelatos, A., Mantzari, E., Pantazara, M., Tsalidis, Ch., Kalamara, C.: Developing tools and resources for the biomedical domain of the Greek language. *Health Informatics Journal*, 17(2), 127-139 (2011).