Natural Language Processing Environment to Support Greek Language Educational Games

Aristides Vagelatos[1[0000-0001-7825-0550]], John Stamatopoulos[2], Maria Fountana[1], Monica Gavrielidou[1], Christos Tsalidis[2]

[1] Computer Technology Institute & Press, 10563 Athens, Greece
vagelat@cti.gr
[2] Neurocom S.A., 15124 Maroussi, Greece
info@neurocom.eu

Abstract. In this paper, we present the so far implemented infrastructure of the "Lexipaignio" project, a research project co-financed by EU and Greek national funds, where an innovative and state-of-the-art NLP (Natural Language Processing) environment is being developed for the creation of digital educational games for Greek students. An initial, brief presentation of the the position of digital games in the today's educational system is followed by a more detailed presentation of the implemented NLP infrastructure for the Greek language. Examples of the games that have already been implemented are also provided. The paper concludes with the current stage of the project and the pending steps towards its completion.

Keywords: Digital Educational Games, Natural Language Processing, Game-based Learning, Mobile Learning, Open and Distance Learning.

1 Introduction

Lately, with the integration of new digital accomplishments into both educational and everyday life of students, important changes are under progress in the educational and learning processes, where Information and Communication Technology (ICT) plays a critical role. The situation was boosted even more, during the COVID-19 pandemic, where more and more pupils started to utilize digital technology in order to participate in distance learning classes. From this perspective, the use of digital games to support (game-based) learning through an alternative, more attractive way is rapidly developing in both European and worldwide level.

"Digital games" is a rapidly developing field, as they are 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. According to related research, computer games provide a quick and interesting learning pace in contrast to the conventional teaching methods and in this perspective, they introduce modern alternatives in digital learning. Furthermore, the embedding of Artificial Intelligence (AI) solutions (in the form of machine learning, big data analysis, as well as “modern” NLP techniques) within the school setting is becoming more intense, providing a wide variety of tools and opportunities. In this context, the use of Natural Language Processing (NLP) can be very useful for the creation of educational language games.

Within a project, called “Lexipaignio”, an innovative and state-of-the-art computational environment is under development, targeting the creation of digital educational games for Greek 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 supports the automated production of questions (e.g., in the form of quizzes) related to the various levels of the Greek language structure and use as for example spelling, morphology, vocabulary, as well as terminology found in school textbooks which is integrated into the overall environment and narration of digital educational games.

Additionally, it enables teachers to introduce new content and define additional areas of interest regarding the Greek language structure. This will result to an automatic creation of an extended volume of questions/tasks, supervised by educators, through their intervention in educational games (crosswords, match games, multiple choice, games where students have to find the correct order of the mixed letters). At the same time, teachers are able to adjust specific game parameters taking under consideration aspects such as: a) educational level, b) school subject (biology, geography, literature etc.), c) school grade and teaching module, d) the general class of grammatical phenomena (conjugation, spelling, syntax, vocabulary).

In this paper, we present the so far implemented infrastructure of the “Lexipaignio” project. Firstly we discuss briefly the position of digital games in today’s educational system, then we present the implemented NLP infrastructure for the Greek language, following by a few examples of the games that have already been implemented and finally we give some conclusions.

2 Games in Education

Digital games are becoming a useful tool in the teaching process. The multiplicity of their qualities results in considering digital games as a pedagogical support to teach, train, solve, provide examples and raise awareness among students (EGDF, 2020).

Mechanics, one of the main constituents of digital games (Schell, 2008) determine the rules and procedures of the gameplay. Thus, player interaction, immersion and sense of engagement are made possible. In the context of Artificial Intelligence (AI), Natural Language Processing (NLP) is being increasingly used in digital game design. Although current research on educational digital games does not treat NLP techniques as distinct game features but as a facilitator of game objectives (Picca et al. 2015), nowadays NLP claim a pivotal role in the creation of educational games. In “Lexipaignio” project NLP is connected to the dynamic creation of a series of language games integrated in the Greek school curricula.

Effectively, the current curricula for the teaching of Modern Greek in primary and secondary education promote the study of grammatical phenomena through the text-centered approach (Iordanidou, 2007). Thus, grammatical phenomena are studied in textual environments where students are trained in recognizing the function of grammatical structures and language mechanisms. Similarly, in the context of vocabulary teaching, words are not treated individually, but are included in the text and are studied as structural elements of speech. With focus to the computer-human interactions through human language, NLP technologies can prove to be particularly useful in the development of educational language games where language challenges will be produced dynamically, utilizing a specific body of text and other resources.

3 NLP for the Greek Language

At the point of convergence of computer science and linguistics, Natural Language Processing (NLP) constitutes a challenging area for numerous applications in our everyday life. Focusing on education, the "Lexipaignio" project aims at the use and further development of a series of Natural Language Processing infrastructure tools (Morphological Lexicon, Lemmatizer, "Mnemosyne" language processing system, corpus of Greek school subjects, etc.), for the implementation of dynamically created gamified educational material.

In this section, the NLP infrastructure that has been developed is presented, alongside some basic Greek language characteristics, in order to exemplify the peculiarities that had to be addressed.

3.1 Some facts about the Greek language

Greek is the official language of Greece, as well as one of the 24 official languages of the European Union (EU). In the 23rd edition of Ethnologue - a language reference database published by SIL International (Eberhard, Simons, and Fennig 2019), it is described as having almost 13 million native language speakers[1], whereas it has been ranked at the 89th position among the 200 most common languages worldwide. The Greek alphabet consists of 24 letters and two diacritics, the stress mark (e.g., «ί») and the diaeresis, consisting of two dots above a vowel letter and indicating a separate syllable (e.g., «ϊ»).

Certain characteristics of the Greek language have challenged its computational processing, with the most significant being the following:

- (a) Historical orthography: Certain vowel phonemes are represented with multiple orthographic graphemes: /o/ can be spelled with either ο or ω, /i/ with *η, ι, υ, ει, οι,* or *υι;* and /e/ can be spelled with either *ε* or *αι*. This has an effect not only on lexemes (e.g., '*φύλλο*' (leaf), '*φίλο*' (friend), '*φύλο*' (gender) are all pronounced

[1] Besides, Greek is spoken by around five (5) million people, members of the Greek communities (the Diaspora) who live outside Greece.

/filo/), but also on inflectional affixes (e.g., '*πολλοί*' (many) (masculine plural), '*πολλή*' (much) (feminine singular), '*πολύ*' (neuter singular) are all pronounced /poli/)).

- (b) Morphology and inflection: Greek language is characterized by an extensive morphological structure, especially in the noun and verb inflectional systems (i.e., three genders, two numbers and four cases for nouns and adjectives, and a sheer number of discrete stems and endings for verbs which indicate mood, aspect, voice, tense, and person).
- (c) Morphosyntactic ambiguities: These are due to homographs, as in the case of most adverbs (e.g., '*καλά'* (well)) vs adjectives in plural ('*καλά'* (good)).
- (d) Word order and syntax: Greek deploys a subject-verb-object pattern (SVO) in syntactic structures. However, unlike languages like English, the Greek word order is not fixed and the constituents of a clause or of a nominal phrase are characterized by a great flexibility.

In the next subsections, we provide an overview of the standard handling methods developed for the Greek language at different levels of text processing: (a) segmentation and tokenization, (b) lemmatization, stemming, spell checking, and morphosyntactic tagging, (c) morphosyntactic disambiguation, syntactical analysis and/or parsing, d) named entity recognition (NER), and term extraction.

3.2 Segmentation and tokenization

Text analysis follows a two-phase process. In the first phase, the text is separated to paragraphs, sentences, and words or other tokens, such as the punctuation symbols. In the second phase, the synthesis, the tokens are compiled to syntactic structures that facilitate the recognition of named entities and meanings. Tokenization is usually the first component in every NLP pipeline. The main problem in sentence splitting is the recognition of dot ('.') as a sentence delimiter in distinction with other uses (e.g. in abbreviations). As far as Greek language is concerned, there is no secure way to deal with the task of abbreviations' recognition. The most effective approach has proved to be the use of techniques that combine dictionaries of abbreviations with heuristic patterns.

3.3 Word processing

One important feature in processing textual data, especially from social media platforms and chatbots, is the ability to recognize misspelled words. Spelling errors are divided in two main categories, the typographic and the orthographic ones (Hiadek et al. 2020). During a previous project, a spelling correction system for the Greek language has been implemented some years ago (Vagelatos et al. 1994). Based on this implementation, and for the needs of the current project, correction algorithms are implemented for the opposite task: to produce misspelled words from a correct one. Thus, typographic errors which are common to all languages and are usually created by mistake, are now introduced in correct word forms in order to produce misspelled ones e.g., wrong letter, missing letter, extra letter, wrong order on a pair of letters, etc. Notably, the orthographic errors are usually due to the lack of knowledge of the correct word. The misspelled word sounds or looks very similar to the correct one and the correction process must consider the similarity factor and order of the proposed words accordingly.

3.4 Sentence based techniques

Ambiguity is an inherent property of natural languages, as it expresses the level of uncertainty about the meaning of a word, phrase, or sentence, and is present in all phases of content analysis. Word ambiguity can be lexical, syntactic, and semantic. Lexical ambiguity is due to polysemy (i.e., the fact that there is more than one meaning for a word in a natural language). The process to resolve the correct meaning is called "word sense disambiguation" and is presented in advanced NLP pipelines. A subcase is considered the Part of Speech (POS) disambiguation, where the process is referred as tagging and the program that accomplishes it is called tagger. Usually, a Greek tagger uses a mixed scheme with three phases (Orphanos, and Christodoulakis 1999). In the first phase, the word is looked up in the morphological dictionary and, if it is found and it has a single entry, all morphological attributes are returned. In case no single entry is found, a set of rules is applied on the context of the word, trying to distinguish the correct one. If a word is unknown, a set of rules examine the suffix and other characteristics of the word, trying to guess its POS. To support this processing, a morphological dictionary has been developed (Tsalidis et al. 2004) with ~100.000 lemmas and 1.200.000 words, containing rich morphological and semantic information for lemmas and inflectional or derivational types.

3.5 Semantic analysis

Named Entity Recognition (NER) as well as term recognition, are widely used in many NLP applications, such as classification, semantic indexing and searching, virtual assistants, etc. The task of recognition of the syntactic structure of a phrase or a sentence is complex and nondeterministic. Although for multi-word terms the task is simpler because they follow specific morphosyntactic patterns, nevertheless a grammar formalism is necessary to describe these patterns, as well as a resolution algorithm that will apply the grammar to a text. A context sensitive grammar formalism has been developed, called Kanon (Vagelatos et al. 2011), along with an efficient resolution algorithm that applies the grammar to the text and recognizes syntactic structures. The resolution algorithm applies a surface parsing technique that permits the recognition of parts of a sentence without requiring the syntax recognition of the whole sentence.

3.6 NLP applications

Modern NLP emerges from the opportunity to better utilize the high volume of open data created in the Web (Eisenstein 2019). Using ML and deep learning technologies, this knowledge is encoded to multidimensional numeric vectors that are used in a plethora of classification applications. For the Greek language, the volume of data that are open and in appropriate format is restricted and cannot lead to efficient models. A combination of rule-based technologies utilizing the existing solid knowledge stored on prefabricated language resources, and ML technologies with auto-adaptation characteristics utilizing the knowledge residing inside texts, has proved to deliver the best results (González-Carvajal, and Garrido-Merchán 2020).

Using the technologies mentioned above, several useful components have been developed which can be used either as stand-alone applications, or in suites that implement complex NLP tasks and support the "Lexipaignio" infrastructure. Some of these components are presented below:

Stemming & lemmatization. Stemming is an important function in NLP applications, when about the main interest is the meaning/concept and less the nuances of morphology. Stemming is achieved by using a process called "Suffix Stripping", which returns the word stem by removing the end part of the word that bears morphological information (i.e., inflection or derivation). Stemming is not considered an efficient technique to deal with Greek words, since in many cases it produces non-existent words. Therefore, it is used only in cases of unknown words. Lemmatization is a similar function with the exception that the produced word, called lemma, is a correct word. Lemma is usually the headword in electronic or printed dictionaries and is used for citation. Accurate lemmatization for Greek language requires the existence of a morphological lexicon. For the medical sector, two additional problems arise: (a) unknown vocabulary that is not negligible, (b) the mixed morphology (Katharevousa vs Demotic forms) that virtually doubles the size of the words.

Keyword extraction. This component applies tokenization, lemmatization and stemming to the words of the documents and the result becomes the input to statistical algorithms as TF-IDF (Manning et al. 2008) and BM25 (Robertson et al. 1994). The output is the tagging of the documents in a collection with the most "important" words, i.e., keywords that identify the document in the collection. Keywords are useful in indexing, searching, and classifying documents as presented in components below.

Related documents. This component utilizes the "Keyword extraction" and NER components (Nadeau, and Sekine 2007) to find the attributes that characterize the document and through them to find related documents in a collection. The component can utilize dictionaries or vocabularies, named entities, terms and events with a customized weighting scheme to compute the level of relation between documents.

Clustering. For profiling documents and mining the knowledge therein contained, the clustering component implements known algorithms as K-Means (Manning et al. 2008) and Hierarchical (Rokach, and Maimon 2005). These algorithms are unsupervised learning techniques which can be successfully used for clustering collections with medical diagnoses or medicines without needing other resources.

Word Embeddings. Word representation with a vector of numbers opens a window to the utilization of modern technologies of machine and deep learning emerging from the application of mathematical sectors as linear algebra, statistics, probability theory, etc. These vectors are created from the processing of large corpora and encoding the semantic characteristics of the words examining the different context environments they appear. According to Distributional Hypothesis (Harris 1954), two words have similar meaning if they can appear in the same context. The word vectors map the words pointing in a high dimensional space; having these properties means that the Euclidean distance of the points corresponding to two words with similar contexts is small. In our systems, word embeddings are also used inside grammar rules to codify the meaning of a word.

4 Example games

4.1 Configuration

The above described NLP infrastructure was extended and configured appropriately in order to serve as the basis for dynamic production of linguistic games. More specifically the following steps were made:

- A corpus was compiled in order to serve as the input for the games. The collection of documents used includes more than 180 chapters from the educational books used in upper primary and lower secondary education in Greece, cleared from images, tricky typographical and layout related elements.
- The corpus was processed and annotated, using some of the NLP tools described earlier.
- Regarding each individual game preferences, data input was created dynamically. For example, in the case of “fill in the blanks” type of games, certain sentences were selected (see Table 1) from the corpus, based on a number of requirements (e.g. specific language textbook, school level, grade, etc.). For each of these sentences, various information has been retrieved and is ready to be presented whenever needed (e.g. input document, link to the digital book, etc.).
- Having all the morphosyntactic information for these sentences, and according to the game characteristics, it is easy to find out certain part of speech (e.g. verbs), hide them, and ask the user to put these words in the correct boxes (see Fig. 1). In a similar yet different game type, the educator might select to ask the user to find certain word types (e.g. verbs) in a sentence (see Fig. 2).
- Changing to a different game type like “find the correct word form”, through which the teacher seeks a way to motivate students to improve their language competency, for a certain grammatical phenomenon e.g. adjective inflection system and more specifically adjectives ending in «–ης» (επίθετα σε –ης), the infrastructure selects an adjective ending in “-ης”, removes it from the sentence, and produces three incorrect – although valid- word forms and present them to the user together with the correct form. In this manner, games of the type “who wants to be a millionaire” can be produced as can be seen in figures 3 and 4.

It has to be stressed once again, that the input data for the games are produced dynamically, based on a) the corpus collected, b) the type/needs of the selected game and c) on the educator’s preferences/needs.

1. The educator has to choose the grammatical/language phenomenon that he wants to teach with the utilization of the game (e.g. use of adjectives, ending in "–υς").
2. The appropriate processing of the corpus (e.g. language school textbooks) has to be realized. The output of this processing becomes the input of the selected game.
3. The teacher gets the games that are produced and integrates them in the way that he/she chooses in his/her class lesson.

In the following section, we present examples of the abovementioned procedure.

4.2 Input Data samples

Following the above described procedure, the elementary school's 6th grade language text book, has been processed in order to produce the appropriate output that can become the input data for the first experimental games.

The above described NLP infrastructure was extended and configured appropriately in order to serve as the basis for dynamic production of linguistic games.

A sample output is presented in Table 1, where the document title can be found in the first column, the selected sentence in the second one and a link to the actual textbook in the third column.

Table 1. Sample output of the preprocessing stage.

Document Title	Selected sentence(s)	Link to the textbook
Η χελώνα και ο Ρεβιθάκης	Άμα την είδε αυτός, τρελάθηκε από την ομορφιά της. Βλέπει και το πέπλο που είχε κεντήσει κι ήταν το καλύτερο απ' όλα, και είπε να την πάρει και αυτή αποκρίνεται πως είναι παντρεμένη μ' έναν ψαρά.	https://tinyurl.com/bdsj4c6y
Αιολική γη	Αργά τα δάχτυλα του γέροντα ανοίγουν το μαντίλι όπου είναι φυλαγμένο το χώμα. Ψάχνουν κει μέσα, ψάχνουν και τα δάχτυλα της γιαγιάς, σαν να το χαϊδεύουν. Τα μάτια τους, δακρυσμένα, στέκουν εκεί.	https://tinyurl.com/8kmax5p2
Σπίτι μας είναι η γη	Ένα πράγμα που ξέρουμε εμείς –και που ο λευκός άνθρωπος θα ανακαλύψει ίσως κάποτε– είναι ότι ο Θεός μας είναι ο ίδιος Θεός. Ίσως να σκέφτεστε να τον αποκτήσετε τώρα, όπως θέλετε να αποκτήσετε τη γη μας, αλλά δεν μπορείτε. Είναι ο Θεός του ανθρώπου, και το έλεός του είναι το ίδιο και για το λευκό και για τον ερυθρόδερμο άνθρωπο. Αυτή η γη είναι πολύτιμη για τον άνθρωπο, και όταν τη βλάπτει είναι σαν να δείχνει περιφρόνηση στον Δημιουργό.	https://tinyurl.com/8kmax5p2
Τα Χριστούγεννα του υπολογιστή	Επειδή ο Ντίνος ήταν πάντα πρόσχαρος και γελαστός, οι συνάδελφοί του νόμισαν ότι αστειευόταν, όπως συνήθως. Χαμογέλασαν μ' αυτά που έλεγε και συνέχισαν τη δουλειά τους.	https://tinyurl.com/2ndsnvx6

In the next table (table 2) a sample input for the “millionaire” game is presented: following the selected lemma (first column), the main sentence is presented followed by the correct answer as well as three incorrect answers that have been produced by the system. The incorrect answers in this game, have been produced by utilizing the thesaurus of the infrastructure where for each entry, synonyms, as well as antonyms are included. The algorithm selects only lemmas that have adequate synonyms, as well as antonyms in order to be able to produce incorrect answers.

Table 2. Sample output of the preprocessing output.

Selected Lemma	Selected sentence(s)	Correct answer	Incorrect answers
άβολος	«Αυτή η καρέκλα είναι μικρή και ... για το θείο Αλέκο» είπε ο Κώστας. Ο θείος έχει μεγάλα πόδια και δε βολεύεται.	άβολη	εξυπηρετική, άνετη, βολική
ανόητος	«Τι ... που είναι ο Νίκος» σκέφτηκε η Αθηνά. «Κυνηγάει τη Ροζαλία που δεν τον πείραξε ποτέ σε τίποτα».	ανόητος	εύστροφος, πονηρός, έξυπνος
βγάζω	Ο θείος Τάκης είχε ταξίδι. ... τα ρούχα του από την ντουλάπα και τα έβαλε στη βαλίτσα του.	έβγαλε	ενέταξε, σημείωσε, έβαλε
δειλός	Ο Κώστας είναι ... και φοβάται να πάει στον οδοντίατρο.	δειλός	θαρραλέος, γενναίος, ανδρείος
καθαρίζω	Ο κύριος Μιχάλης ... τα γυαλιά του κάθε μέρα για να βλέπει καλά.	καθαρίζει	ρυπαίνει, βρομίζει, λερώνει
μυστικός	Στον πύργο της κακιάς βασίλισσας υπήρχε ένα ... δωμάτιο που δεν το ήξερε κανείς.	μυστικό	φανερό, σαφές, εμφανές
συχνός	Οι ... επισκέψεις του θείου Τάκη έδωσαν μεγάλη χαρά στον κύριο Γιάννη: ήρθε τρεις φορές μέσα σ' ένα μήνα.	συχνές	σπάνιες, αραιές, νερουλές

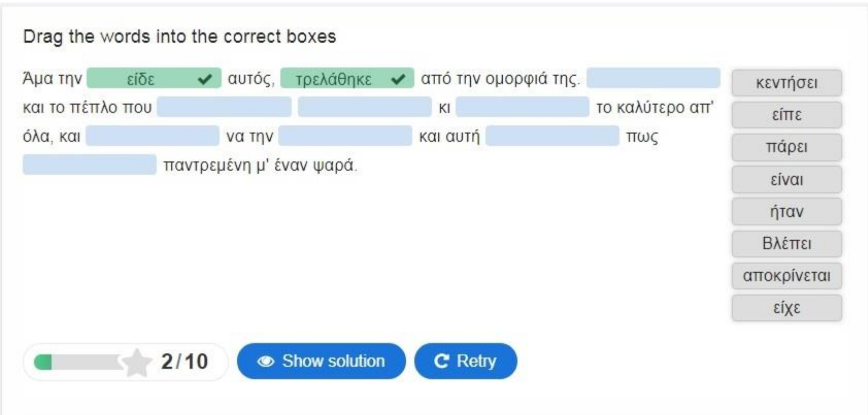

Fig. 1. An example of Drag and Drop tasks, showing the validation process of a correct selection.

Fig. 2. An example of Drag and Drop tasks, showing the validation process of both correct and incorrect selections along with the scoring mechanism.

Fig. 3. A sample input for the "millionaire" game is presented followed by the correct answer as well as three incorrect answers that have been produced by the system.

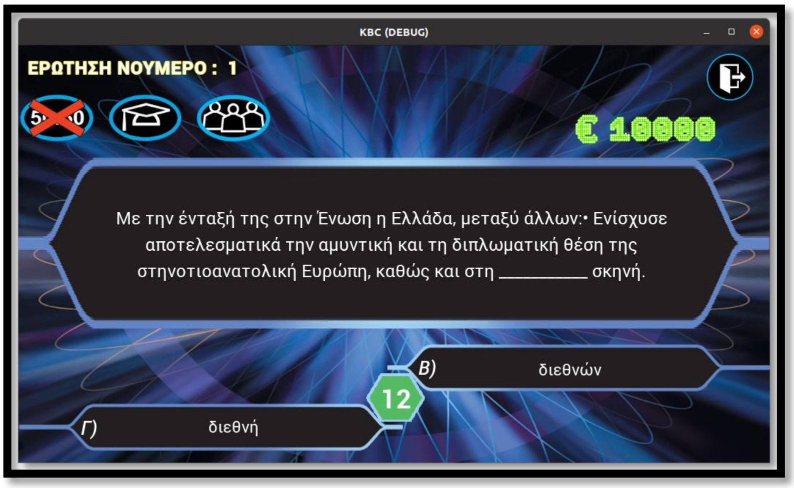

Fig. 4. A sample input for the "millionaire" game is presented followed by the correct answer and one incorrect answer after having applied the option of eliminating half of the possible answers.

5 Conclusions

The NLP environment that is being implemented within the "Lexipaignio" project can become a successful tool to support and enrich the educational process in an appealing and attractive way. As far as language discipline is concerned, the traditional approach, solely giving emphasis to the teaching of rules and the monotone problem-solving on the premise that 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.

It should be pointed out that the school textbooks, constitute the testing field for the previously described technology. Consequently, the produced results will almost immediately be useable by everybody involved in the learning process (teachers-parents-students), in various platforms (mobile devices, desktop/laptop computers, etc.). Last but not least, the outcomes of the project will be freely available under open source license.

The project is currently passing its final stages with really promising, initial results. The thus far implemented infrastructure has already enabled the development of pilot digital games for the Greek language, offering users the option to customize (by selecting / setting various parameters) or use them directly for beta-testing purposes. The challenge now is to finalize and introduce the outcomes to educators in order to receive feedback from the actual application both inside and outside the classroom.

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).

References

1. Eberhard, D. M., G. F. Simons, and C. D. Fennig, eds. (2020). *Ethnologue: Languages of the world (23rd edition)*. Dallas, Texas: SIL International. https://www.ethnologue.com/.
2. Eisenstein, J. 2019. *Introduction to Natural Language Processing.* Cambridge MA: The MIT Press.
3. European Games Developer Federation. (2020). Using Digital Games in Education. Retrieved from www.egdf.eu Using digital games in education (egdf.eu).
4. González-Carvajal, S., and E. C. Garrido-Merchán. (2020). *Comparing BERT against traditional machine learning text classification*. https://arxiv.org/abs/2005.13012.
5. Harris, Z. S. (1954). Distributional structure. *Word* 10, no. 2-3: 146–62. doi:10.1080/00437956.1954.11659520.
6. Iordanidou, A. (2007). Κειμενοκεντρικές Προσεγγίσεις του Σχολικού Εγγραμματισμού: Κείμενο, Συμφραζόμενα, Γραμματική. In Η. Ματσαγγούρας (Ed.), *Σχολικός Εγγραμματισμός. Λειτουργικός, Κριτικός, Επιστημονικός*. Athens, Grigori (In Greek).
7. Manning, C. D., P. Raghavan, and H. Schutze. (2008). Scoring, term weighting, and the vector space model. In *Introduction to Information Retrieval*, 100–23. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809071.007.
8. Nadeau, D., and S. Satoshi. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes* 30: 3–26.
9. Orphanos G. and D. Christodoulakis. (1999). Part-of-speech disambiguation and unknown word guessing with decision trees. In *Proceedings of the 9th EACL Conference, Bergen, Norway, June 8-12, 1999*, 134–41.
10. Picca, D., Jaccard, D., Eberle, G. (2015). Natural Language Processing in Serious Games: A state of the art. International Journal of Serious Games, 2(3), 77-97.
11. Reinders, H. (2012). Digital Games in Language Learning and Teaching. Palgrave Macmillan Publishing.
12. Robertson, S. E., S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. (1994). Okapi at TREC-3. In *Proceedings of the Third Text Retrieval Conference (TREC 1994), Gaithersburg, USA, November 2-4, 1994*. NIST Special Publication 500-225, 109–26. Gaithersburg: National Institute of Standards and Technology (NIST).
13. Rokach, L., and O. Maimon. (2005). Clustering Methods. In *Data Mining and Knowledge Discovery Handbook,* eds. O. Maimon, and L. Rokach, 321–52. Boston, MA: Springer.
14. Schell, J. (2008). The Art of Game Design – A Book of Lenses, Carnegie Mellon University, Elsevier.
15. 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.
16. Vagelatos, A., Mantzari, E., Pantazara, M., Tsalidis, Ch., Kalamara, C. (2011). Developing tools and resources for the biomedical domain of the Greek language. Health Informatics Journal, 17(2), 127-139.
17. 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.