Towards the development of a grammar checker and its utilization in the teaching of Modern Greek as mother tongue

Panagiotis Gakis
"Neurolingo" Company, Greece

Theodoros Kokkinos
University of Thessaly, Greece

Christos Tsalidis
"Neurolingo" Company, Greece

Abstract

The goal of the current study is the exploration of the standardization of grammatical errors through templates, a process that ultimately led to the development of a Grammar Checker for Modern Greek, an electronic tool for the automatic recognition and correction of grammatical errors. Moreover, the study investigates the utilization of such a tool in a classroom setting and its contribution to the teaching of mother tongue. The participants of the study were secondary school students and main results showed that errors they made in text production activities could be categorized as errors of mechanics, grammar, and usage through certain templates, which then lead to the development of a Grammar Checker for Modern Greek. Moreover, we also found other errors that could not be analysed and categorized using the theory of taxonomy of errors by Ho, i.e. style: forms—learned forms and semantic. Additionally, the participants highlighted that the specific tool has advantages for the teaching of mother tongue. Implications for the methodology of Grammar Checkers' development and their use in the teaching of mother tongue, generally, are discussed.

Correspondence:
Panagiotis Gakis,
"Neurolingo" Company,
Giannitson 68, Patras 26223,
Greece.
E-mail:
gakis@sch.gr

1 Introduction/Theoretical background

The current study consists of four distinct parts. First, in the theoretical background, we theoretically discuss the basic elements of Grammar Checkers and their general features. Additionally, we analyse the infrastructure of grammar checking software, such as templates on which errors are standardized. Then, we present specific templates of the Greek Grammar Checker, which set it apart from other relevant tools. Moreover, we mention the utilization of Grammar Checkers in educational settings. In the second part of the article, we present the results of the current study, concerning the engagement of students in template creation and, after the Greek Grammar Checker's development, their beliefs for its role in

Digital Scholarship in the Humanities, Vol. 36. No. 1, 2021.
doi:10.1093/llc/fqz095 Advance Access published on 27 January 2020

the teaching of mother tongue. Finally, we discuss and analyse the results of the study in relation to the findings of other studies.

1.1 Grammar checkers

Information and Communication Technology (ICT) has been widely used since the past few decades in language teaching. An example of ICT tool that has been widely utilized in language teaching is grammar checkers. Given the modern complex socio-economic and work circumstances, the necessity of grammar checkers for improving the quality of electronic texts is nowadays more than evident (Ehsan and Faili, 2010). Typically, grammar checkers work by scanning through a text and provide immediate feedback on grammar, spelling, and punctuation errors. Grammar checkers can highlight special linguistic/grammar issues such as subject–verb non-agreement, split infinitives, double negatives, run-on sentences, and incorrect use of prepositions. If the checker finds an error, it will explain the grammar rule and may also offer a solution which the user can accept or ignore (Cavaleri and Dianati, 2016). This is a very important element, since it could enhance the capacity of school students to produce a written text with fewer errors without the mediation of their teacher but through self-evaluation. The checkers also highlight spelling errors and words that may have been confused. Some grammar checkers also offer feedback on style and vocabulary usage. An important point to note is that grammar checkers do not claim to teach grammar; they are a tool to bring potential problems to the writer's attention.

1.2 Features of the grammar checker

The Grammar Checker examines and verifies each word not only separately but also in relation to its context, e.g. the preposition with the noun, the article with the adjective and the noun, the verb with pronouns and complements, etc. The grammatical test focuses primarily on the detection of the words or/and phrases described as problematic (morphological or stylistic). If the word or phrase has the required characteristics, it is considered incorrect. In this case, the grammar checker suggests to the user to use either the correct form (e.g. the same word in another case, e.g. <πληγέντων περιοχών (affected areas)> is corrected to <πληγεισών περιοχών (affected areas)> or another word with the correct morphological form, e.g. <πιο καλύτερος (more beter)> will be corrected to <πιο καλός (better)>.

This tool is designed to mark the stylistic differentiation of word forms. This means that the grammar checker is based on the Modern Greek electronic lexicon, a general language electronic lexicon and is not appropriate for direct usage by humans, but for computational applications such as morphosyntactic tagging, parsing, semantic tagging, and machine translation. The lexicon was designed specifically for Modern Greek and contains all the morphological information necessary for all, a rich set of morphological attributes as Part of Speech, Case, Number, Gender, etc. (Gakis *et al.*, 2012). In this way, the user can choose in hindsight the style she/he prefers and, therefore, the proper style checked by the grammar checker. The grammar checker handles words in which misunderstanding the correct meaning creates mistakes in written speech. Consequently, the word <απλά> with the sense of (no composite things) is used in sentences, where the word <απλώς (only)> must be used (Iordanidou, 2013). Moreover, the checker extends the functionality of the word level to the sentence level, a type of data that increases the complexity of such a tool (Gakis *et al.*, 2015).

1.3 Error standardization in a grammar checker

A language error is any deviation from the linguistic norm, the traditional rules the linguistic community sets as a model (Papanastasiou, 2008; Petrounias, 2013). The linguistic norm is the domination of a language against the linguistic forms or dialects and, ultimately, constitutes a standard, traditional linguistic form as it is described and encoded by the grammar and the dictionary of a language. However, this standardization is unclear, as the coding described in the grammars and dictionaries corresponds to a specific period, while it also differs in the goals and criterions they lay. There is a difference between the everyday use and the static image of the language displayed in textbooks, an image in which the vitality, variety, and discontinuities of the oral speech are not included (Setatos, 1991).

The formalistic and standardized listing of the errors in spoken and written Greek language does not aspire to regulate this. It simply points out any deviations from existing rules (Iordanidou, 1999, 2004), inadequate or incorrect usage and proposes suggestions close to the common linguistic criteria. It thus contributes to the elimination of some wrong forms and contributes to a more accurate written speech. In addition, ignorance and negligence of basic grammar rules pose the risk of establishing wrong forms.

The Grammar Checker is designed to underline the styling differentiation of verbal forms. This means that the Grammar Checker is based on the morphological lexicon in which the word production of each lemma allows the morphological variation of each form (spoken language, oral, learned), e.g. <*γυρνάνε, γυρνάν, γυρνούν, γυρνούνε* (walk around)> or <*αγαναχτούν, αγαναχτούνε and αγανακτούν, αγανακτούνε* (be indignant)>. In this way, the user will be able—through an option menu—to previously select the language type or styling text (formal, informal, or oral).

The Grammar Checker also covers the most specialized user needs. It is a valuable tool for those who are taught the Greek language (either as a first or second language), since it not only detects grammatical errors, but it also includes corrective suggestions for each error as well as a detailed explanation.

2 Templates of Grammars Checkers: The Case of the Greek Grammar Checker

In order for such a tool to be built, though, certain templates should be created. Templates allow the linguist to write Grammar Checker rules incorporating different forms of linguistic information from a wide range of parsing approaches, covering dependency grammar, phrase-structure grammar, and unification grammar (Bick and Didriksen, 2015). Templates, also, by ordering rules in batches with more heuristic rules, enhance the Grammar Checker's reliability, as Natural Language Processing (NLP) tools[1] are based on probabilistic methods. By using templates, we can also model a grammar checker, transforming the grammar rules from an erroneous sentence or word-form to the correct sentence or word-form.

The form of the templates for the Greek Grammar Checker is based on the context-sensitive grammar (Nava and Heshaam, 2013) and have the following structure:

```
[H1], [H2], . . . , [hν] =>
            [L1], [L2], . . ., [Lk]
            \
                        [C1], [C2], . . ., [Cm]
            /
                        [R1], [R2], . . ., [Rl]
            ;
```

where H, L, C, R^2 represent text of one or more words. The K is the head of the section created by the rule implementation, A and D are sections of the left and right context, respectively, and P are sections describing the grammatical phenomenon. The section is described by a sequence of text attributes. Examples of templates are listed are following.

Example 1:
[RULE= 'oral_noun_3', STATUS = 'Info', MESSAGE = The lemma (**δημαρχίνα**) **is oral. Replace (δημαρχίνα) with (δήμαρχος).] =>**

```
\
   [LEXY->Has Lemma(δημαρχίνα)]
/
;
```

Example 2:
[RULE=FINAL_N,Status: Level_1,4, TTEXT ='__τον__'] =>
1st LEVEL

```
\
  [Word='τον',MORPHOLOGY={ARTICLE,
ACCUSATIVE, MASCULINE, SINGULAR}]
/
     (
    [MORPHOLOGY={ADJECTIVE, MAS
    CULINE, SINGULAR, ACCUSATIVE},
    ORTHOSTYLE={GROUP_2}] [3] |
    [MORPHOLOGY={PARTICLE, MASC
    ULINE, SINGULAR, ACCUSATIVE},
    ORTHOSTYLE={GROUP_2}]
     ),
       [MORPHOLOGY={PRONOUN, GEN
       ETIVE}]?,
        [MORPHOLOGY = {NOT: NOUN}]
    ;
```

```
[RULE='FINAL_N_1', STATUS='Error',
MESSAGE=The word 'το' must have a final
-ν. Replace 'το' with 'τον'.] =>
2nd LEVEL
\
   [WORD='το']
/
(
    [MORPHOLOGY={ADJECTIVE, MALE,
    SINGULAR, ACCUSATIVE},
    ORTHOSTYLE={GROUP_1}] |
    [MORPHOLOGY={PARTICLE, MALE,SI
    NGULAR, ACCUSATIVE},
    ORTHOSTYLE={GROUP_1}]
),
(
     [MORPHOLOGY={ARTICLE, MALE,
     SINGULAR, ACCUSATIVE}] |
     [MORPHOLOGY={PRONOUN, MALE,
     ACCUSATIVE}]
)?,
     [MORPHOLOGY={PRONOUN,
     GENITIVE}]?,
    [MORPHOLOGY={NOUN, MALE, SING
    ULAR, ACCUSATIVE}]
;
```

This tool is innovative and necessary for Modern Greek, both for students and common users. It examines and verifies each word not only separately but also in relation to its context. The grammatical test focuses primarily on the detection of the words or/and phrases described as problematic (morphological or stylistic). If the word or phrase has the required characteristics, it is considered correct. The template's necessity arises in the fact that errors are gathered through an authentic body of specialized corpora (Granger *et al.*, 2002; Hunston, 2002). Finally, it should be noted that the standardization of grammatical errors in templates is the most important part in the design. (Fig. 1)

3 Utilization of Grammar Checkers in Educational Settings

Written corrective feedback permits teachers to promote accuracy and prevent fossilization (Bitchener, 2008; Ferris *et al.*, 2013). The Greek Grammar Checker, additionally, seems to be of particular interest within educational settings, as it has the potential to facilitate teachers' workload, i.e. providing feedback is time-consuming and at the same time boosting students' linguistic competence (Kokkinos *et al.*, 2018).

Previous studies have often adopted a narrow focus, evaluating the Grammar Checker only on articles/determiners, prepositions, and collocations (Leech, 1998; Han *et al.*, 2006; De Felice and Pulman, 2008) and not in a classroom setting. Teachers using grammar checkers provide comprehensive written corrective feedback on students' texts and also encourage them to scrutinize their own writing for errors that the Grammar Checker might have overlooked (John and Woll, 2018). Moreover, other studies suggest that it is necessary for teachers to teach students how to use the electronic spelling and grammar checkers and dictionaries in any language at schools so that the students can use them to correct their own texts (Bourjaili, 2014).

According to previous research focusing specifically on a possible utilization of a Grammar Checker for Modern Greek by educators, the findings highlighted the elements of the systematic and collaborative character of the reviewing process (Kokkinos *et al.*, 2018).

Current trends in Linguistics and Applied Linguistics suggest that language use must happen in an authentic context. There have been grammar checkers, developed for other languages. Grammatifix (Arppe, 2000) is a grammar checker for Swedish, which has the same approach as the Greek Grammar checker. In this grammar checker, the templates that they describe the errors are collected first and those which result in high precision are chosen for implementation with this argument that the precision is important for grammar checkers (Bernth, 2000). GramCheck (Bustamante and León, 1996) is a grammar checker for Spanish and Greek, including rules for templates concerning only for agreement errors and certain head-argument relation issues. Other Grammar checkers (Park *et al.*, 1997) have the same approach as the Greek Grammar Checker and do not attempt to handle templates for all the possible grammar errors. The English Grammar checker handles templates similar to the Greek

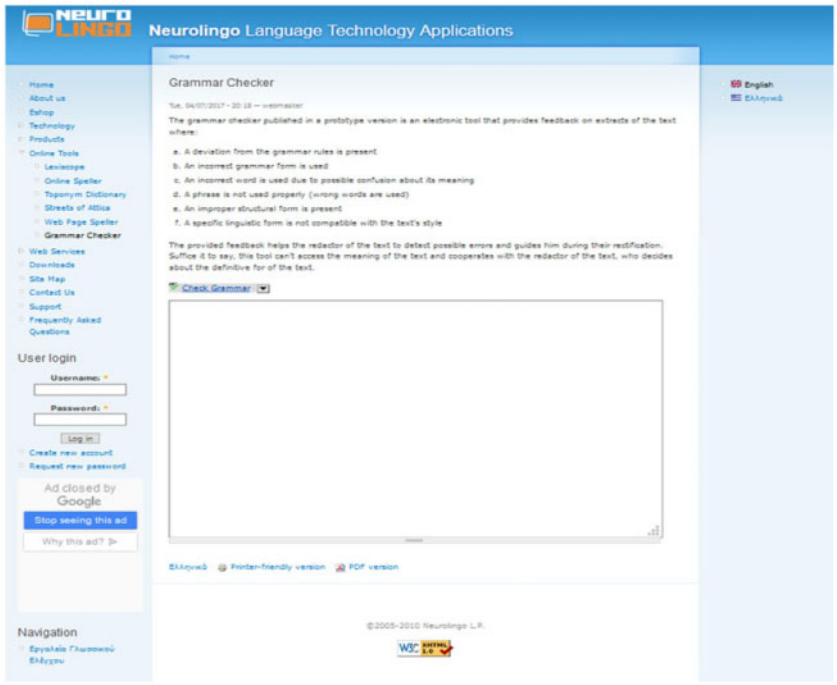


Fig. 1 The environment of Neurolingo's Greek Grammar Checker

Grammar Checker grammatical mistakes. The English Grammar Checker detects the following kinds of mistakes: wrong capitalization (sentence initial and wrong lowercase/uppercase initial letter), missing fragments (subjects, objects, some prepositions, complements, articles, clauses, the, than, etc.), some extra elements (e.g. the infinitive marker after auxiliary verbs), wrong agreement (number, case, etc.), wrong verb form, and various mismatches (verb tense with adverbs, etc.). It is beyond the scope of a grammar checker (Greek and other checkers) to identify templates concerning mistakes along with missing fragments, run-on sentences, wrong expressions, and wrong paragraph boundaries (Gakis *et al.*, 2016).

Therefore, it can be understood that Grammar Checkers' templates should derive from an authentic background, meaning that they should come from authentic communicative circumstances and mainly by native speakers. Given that Grammar Checkers can be utilized in the classroom, students would be an audience of native speakers and the classroom an authentic language setting through which grammar checker templates could be investigated.

4 The Current Study

4.1 Method

The ultimate goal of the research team is the design and implementation of a Grammar Checker for Modern Greek, which will carry out morphological and syntactic analysis of sentences, phrases, and words in order to correct syntactic, grammatical, and stylistic errors.

Given that such tool is not fully available yet for Modern Greek, the development of the grammar checker is based on the detailed recording, analysis, and standardization of errors of written speech. Such an effort, though, can only be done through the recording and classification of grammatical error templates in native speakers' text production and can be described in a formalistic and unambiguous manner by Unification Grammars. A classification of grammatical errors based on templates is a clear, systematic, and detailed process, constituting the main axis of the design of the grammar checker for Modern Greek. Overall, the goal of the current study is not only to define the Grammar Checker templates of the

student's most common mistakes but also to point out the possible value of the Greek Grammar Checker's incorporation into the classroom (Mountifield, 2006).

Specifically, the research questions that drove the current study were:

(1) Which templates can derive from a group of native speakers (e.g. students) in the frame of an authentic communicative context in order for the Greek Grammar Checker to be developed?
(2) Which are the participant students' beliefs related to the utilization of the Greek Grammar Checker in language teaching (after it was created)?

The participants of this study were secondary school students ($N = 20$) of the Experimental School of Patras (third biggest city) in Greece. The initial phase of data collection started during the school year 2015–16. The participants at the time of the study were 16–18 years old ($M = 16{,}9$) and had a considerably high linguistic competence[4] ($M = 17{,}2/20$), since a prerequisite for entering an Experimental School in Greece is passing demanding exams, which include extensive and rigorous testing of linguistic competence in mother tongue. The majority of the participants were female students (62%).

For the first research question, i.e. error templates to be investigated, the researchers used the statistical approach in scoring the participating students' fluency of the erroneous sentence or word-type. The standardization of grammatical errors and their categorization by creating templates were realized after the grammatical errors were collected through an authentic specialized corpus of texts. The resulting grammar checker will highlight and control the categories of these errors—which are the most typical and common—and will not do a complete syntactical analysis of the sentence. The syntactic checking will mainly focus on identifying the forms described as problematic, based on the aforementioned categories. If the word or phrase actually has the attributes that these templates impose, then it will be considered incorrect. Otherwise, the syntactical analysis will not extract any information and the sentence will be considered grammatically correct (Pravec, 2002). Data were collected through a text production activity given to the participant students. Specifically, they were given an exercise in which the students had to find out the grammatical errors or stylistic forms that were inappropriate in a formal text. The students marked—using track changes—these forms ($n = 289$) and, optionally, suggested the most accepted type.

Data were then analysed using SPSS (v. 23) and are presented at a descriptive level. They are also tabulated according to the participants' gender. Because of the participants' overall high achievement, data were not tabulated according to their linguistic competence.

For the second research question to be explored, the second phase of the research procedure was to gather data related to the participant students' beliefs for the utilization of the Greek Grammar Checker in the teaching of mother tongue. This phase was realized after the creation of a first experimental version of the Greek Grammar Checker. The participants' group was engaged in a text production activity within a specific communicative context. Then, they reviewed their texts using the Greek Grammar Checker (Fig. 2), which was developed based on the templates that were produced by their work during the previous phase. Afterwards, they answered a questionnaire with two open-ended questions about how they experienced the advantages and disadvantages of the specific tool in language teaching in the classroom. Finally, the participants' answers were categorized and analysed in relation to their gender (Table 1).

5 Results

5.1 First phase: template categories

Following the aforementioned analysis, we categorized error recognition in nine major/main categories, which are listed in Table 2. The main areas of the grammatical errors in which the grammar checker will interfere are (1) punctuation, (2) final -n, (3) stylistic, (4) standardization (stereotyped phrases, words of literary origin), (5) inflection (incorrect conjugation of names or verbs because of ignorance or of confusion), (6) vocabulary (cases of conceptual confusion, Greek translation of foreign words, redundancy, and use of incorrect word or phrase), (7) orthographic confusion (homonymous words), (8) agreement (cases of elements of nominal or verbal phrase disagreement), (9) syntax issues (verbs), and

Table 1 Error recognition categories in relation to the participants' gender ($N = 20$)

Grammar Checker's templates	Boys				Girls			
	Recognition		No recognition		Recognition		No recognition	
	f	%	*f*	%	*f*	%	*f*	%
Punctuation	4	14.2	22	85.7	8	31.2	17	68.7
Tonal signs	2	7.1	24	92.8	9	16.6	42	83.3
Standardization	2	4.9	42	95	8	17.6	36	82.3
Morphology Style	15	17.2	74	82.7	23	25.4	66	74.5
Oral forms	10	26.1	27	73.8	13	36.1	24	63.8
Learned forms	4	12.8	28	87.1	6	19.1	26	80.8
Semantics	7	12.8	48	87.1	20	36.6	35	63.3
Phrase agreement	42	33	84	66.9	74	58.8	52	41.1
Syntax	13	23.8	43	76.1	16	28.4	40	71.5

Table 2 Categories derived by the participants' error recognition

Category number	Category description
Category 1	Punctuation
Category 2	Tonal signs
Category 3	Standardization
Category 4	Morphology style
Category 5	Oral forms
Category 6	Learned forms
Category 7	Semantics
Category 8	Phrase agreement
Category 9	Syntax
Category 10	Other cases

(10) cases of errors that require more specialized management of the speller (Gakis *et al.*, 2016).

Consequently, according to Table 2, the major areas of grammar errors on which the grammar checker under development then would support the user are the following.

5.1.1 *Punctuation—stress diacritics*

This category describes: (1) mistakes concerning the incorrect presence or absence of punctuation (comma, dot, and hyphen) or typographic mistakes or inaccuracies that occur due to acquired speed when the user presses the wrong key (e.g. presence or absence of space after punctuation) and (2) mistakes that occur due to: (a) the incorrect application of monotonic [since 1982 Modern Greek language has only one stress diacritic over vowels (Holton *et al.*, 2012)] (13.8%), (b) the use of the apostrophe (verb forms that must be without stress after the removal of the apostrophe) (42.1%), (c) the use of a hyphen (21%), and (d) the use of personal or possessive pronouns that require the presence or absence of stress (15.7%).

5.1.2 *Final -n*

The most common error in written speech is described in this template and deals with almost all cases referring to the final -n. The specific template includes rules that work at four levels, because of the number of cases and their difficulties. Originally, letters are grouped into two groups: group (style 1) that contains the letters of the alphabet or symphonic complexes {$\mu\pi$, $\nu\tau$, etc.} which demand the presence of the final -n and the group (style 2) that contains the letters of the alphabet whcih emit the final -n. The rules are defined in four levels in order to describe more complex linguistic situations (possessive adjectives, learned participles, prepositional, etc.).

5.1.3 *Style (oral forms—learned forms)*

This category notes problems regarding the written language that are related to the more functional semantic marking of the oral forms in the following subcategories (Table 1):

(1) Entries (nouns, verbs, and adjectives) with an oral stylistic attribute,
(2) individual oral forms of noun,
(3) individual oral forms of verbs,
(4) individual oral forms of adjectives,
(5) individual oral forms of adverbs, and

(6) morphological forms of special characters (apostrophe) that are not acceptable in written speech.

In all of these categories, there will be matching with the acceptable (in written) morphological forms of entries or other acceptable in the spoken words.

Similarly, in this category, we mark and characterize entries as learned or the endings of nouns, verbs, adjectives, participles, and adverbs that are not acceptable in the official written form of the Modern Greek language due to their learned form.

5.1.4 *Standardization*

The rules of this template handle lemmas of Modern Greek that are misspelled (one word instead of two or two instead of one). The wrong forms are described by a context-free grammar. Standardization involves the creation of a 'commonly spoken' language and the introduction of spelling conventions. In this category, errors recognized concerning (1) the level of writing, where it is proposed to write in one or two words for noun or verb phrases that originate from an earlier period of the Greek language and have been incorporated in the Modern Greek language; (2) the level of the vocabulary, where the majority of the participants recognized the most used and in Modern Greek learned phrases, by matching the wrong form with the correct form; (3) the level of spelling, where modern writing is proposed mainly for words with non-Greek etymological background/origin.

The error rates for the participants are shown in Table 3.

5.1.5 *Morphology (inflection)*

In this category, all morphological and inflectional issues are classified with the description of the incorrect, unacceptable formation of the nouns, verbs, and adjectives. As we can see in Table 4, most of the participants as regards the category of morphology recognized issues concerning noun inflection (wrong plural of Greek words, 73.6%) and verb inflection (mistaken use of the past simple in active voice, 78.9%). Regarding the category of agreement in verb phrases, most of the participants recognized cases concerning noun–verb clarification (87.3%), while in noun phrases noun–noun agreement (89.4%).

Table 3 Standardization error recognition rates in relation to participants' gender ($N = 20$)

Category of errors	Boys		Girls	
	f	%	*F*	%
Writing level	20	16.6	32	50
Vocabulary level	17	26.5	24	28.5
Spelling level	21	28.5	31	33.3
Phonology level	11	2.6	17	9

5.1.6 *Semantics (vocabulary)*

In this category, we include all semantic issues, such as issues of conceptual and/or spelling confusion as well as lexical equivalent (foreign words and Latin phrases). The purpose is to facilitate the user by understanding the meaning of a word that is homophone or 'result' of conceptual confusion and misuse and to remove the difficulties in using it.

Table 4 Frequencies of error recognition categories ($N = 20$)

	f	%
Morphology		
Noun inflection		
Nouns without plural	4	15.7
Nouns with rare plural	3	21
Nouns without plural genitive	2	5.2
Wrong plural of Greek words	13	73.6
Wrong plural of foreign words	11	63.1
Verb inflection		
Mistaken use of present simple of active voice	8	42.1
Mistaken use of present simple of passive voice	11	78.9
Mistaken use of past simple of active voice	8	78.9
Mistaken use of past continuous of active voice	9	47.3
Adjective inflection		
Neutral	1	5.2
Agreement		
Verb phrase		
Periphrastic tense formation	11	42.1
Verb–subject agreement	15	73.6
Past simple formation	11	57.8
Noun–verb clarification	13	87.3
Imperative formation	17	68.4
Noun phrase		
Article–noun agreement	6	21
Noun–noun agreement	17	89.4
Adjective–noun agreement	16	57.8
Participle–noun agreement	15	84.2
Noun entities	4	15.7

5.1.7 *Semantic (orthographic confusion)*

In this template, grammar checker handles words in which misunderstanding the correct meaning creates mistakes in written speech. Consequently, the word [απλά] with the sense of [no composite things] is used in sentences where the word <απλώς (only)> must be used (Iordanidou, 2013). Homophones with different spelling words belong to this template (e.g. the <λήμμα (word)> and <λύμα (waste)>. Depending on the part of speech, the rules that have been created refer to adjectives, adverbs, pronouns, verbs, and homophone words that belong to different parts of speech.

5.1.8 *Syntax*

This template only describes the wrong formations of agreement of noun or verb phrase. Therefore, grammar checker checks the agreement between: (1) art with noun, adjective, or pronoun; (2) noun with noun; (3) particle or adjective with noun; (4) subject and verb; (5) parts of adverb phrase; (6) parts of preposition phrase; and (7) parts of verb phrase (correct verb syntax: with nominal, genitive, accusative, with preposition phrase, secondary sentence, and deponent verbs).

5.1.9 *Other case*

The errors included in this template are very frequent. First of all, the parser handles with simple cases of redundancy, e.g. <ευθύς αμέσως (right away)> and <από ανέκαθεν (every time)>. The template includes (1) bad declaration of comparative or superlative <πιο (most)> suffixes of comparative or superlative and (2) more complex cases such as the verb's syntax, e.g. verbs with supplements that are unnecessary <εμπεριέχω μέσα (contained within)> and <επαναλαμβάνω πάλι (repeat again)>. These are cases of errors that require more specialized management of the Speller.

5.2 Students' beliefs for the Greek Grammar Checker

As we can see in Table 5, most of the participant students believe that the Greek Grammar Checker definitely has advantages for the teaching of mother tongue, mainly concerning its modernization. By this, it is meant that a modern approach is followed in the teaching of mother tongue supported by the use of the Greek Grammar Checker, an approach that does not focus on explicit language teaching, grammar drills, and manual text reviews by the teacher. In relation to gender, most of the male students (40%) think that the Greek Grammar Checker offers a modern aspect of language teaching, while most of the female students (27.3%) think also that it contributes to a more solid linguistic understanding. As far as the Grammar Checker's disadvantages are concerned, most of the female students (60%) stressed out the fact that it cannot track all mistakes, pointing towards its next, more elaborate versions.

Table 5 Students' beliefs for the Greek Grammar Checker in relation to gender ($N = 16$)

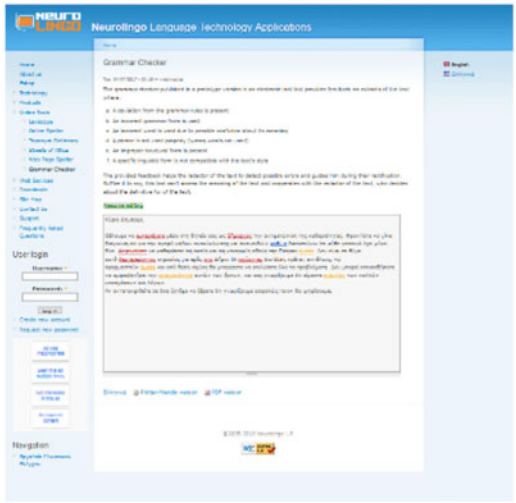
Students beliefs	Boys		Girls	
	f	%	*f*	%
Advantages of the Greek Grammar Checker				
Modernization of language teaching	2	40	3	27.3
Psychological boost	1	20	2	18.2
Better understanding	1	20	3	27.3
Motivation boost	0	0	2	18.2
Self-assessment	1	20	1	9.1
Disadvantages of the Greek Grammar Checker				
Anxiety to the user due to numerous labelling	1	100	2	40
Inability to track all mistakes	0	0	3	60

Specifically, some of the participating students' comments were:

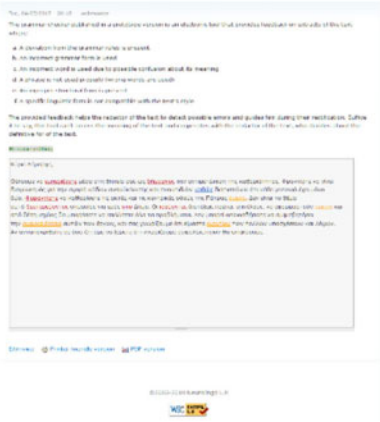
> "*Finally, the language lesson is being modernized" (Mary)*
> "*I am not afraid to write a text" (Helen)*
> "*I understand my mistakes" (Tom)*
> "*Dissimilar with the teacher's correction" (Jane)*
> "*Self-evaluation at last" (John)*

5.3 Rule generation

We posted all templates of the grammar checker in a friendly environment. The user types or copies the text at the lexical editor and sees the analysis of his/her text. The text analysis includes the morphological label of words and a note of text errors. Grammar rules are grouped into the following categories: error, info, and warning. The 'error' category describes errors that deviate from the templates; the 'info' category includes rules concerning the style; the 'warning' category



Sample of a participating student's text inputted into the Greek Grammar Checker



Sample of a participating student's text checked by the Greek Grammar Checker

Grammar rule description and correction suggestion by the Greek Grammar Checker to a participating student's text

Fig. 2 The process of grammar checking to a participating student's text through the Greek Grammar Checker

includes rules concerning information about the meaning of the words with conceptual confusion.

This formalism has the potential to utilize the morphological–stylistic features of the word forms described in an electronic morphological dictionary. By means of this formalism, it is possible to identify multi-word terms and in an automated way to extract the phrase—word form with the wrong information.

It is worth noting that in all cases, each error will be considered separately, and a set of corrective actions will be defined to address it.

The parser—the computational formalism—resulting from the creation of rules based on the templates mentioned is responsible for checking the correct syntax of orders. During this phase, it will read the tokens one by one giving all the morphological characteristics. The result of this morphological analysis is the identification of the various errors (syntactic and semantic) and the recording of the production tree that reflects the part of the structural analysis that is problematic. This tree will be the raw material for the next phase that is the semantic analysis before the intermediate code.

5.4 Software choice and design

The grammar (software) used in the above template implementation system follows the rationale of Unification Grammars. Context-free grammars are actually problematic in their use, primarily because their rules are necessarily redundant in order to cover all editorial phenomena (verbs with object, verb optionality, etc.), resulting in many syntactically incorrect. For this reason, lexical and grammar categories of context-free grammars are enriched with additional features that are relevant to each category. In other words, each of the above rules is accompanied by a series of limitations on the feature constraints of one or more of the categories that take part in the rule. In practice, checking is carried out if a restriction is met by the unification process. Two feature structures (i.e. a set of characteristic–value pairs) can be merged into a single structure if their common attributes have the same values. Grammars that use the formalism of unification are called Unification Grammars. Unification Grammars are a powerful and efficient representation of linguistic information and can describe phenomena much more complex than context-free grammars (Chomsky, 1965). The structure resulting from this procedure contains the synthesis of the information contained in both original structures together. The basic idea of formalism of a Unification Grammar is simple. The sentence analyser—the correct sentence separation including the abbreviations—contains technical representations of the semantic characteristics (e.g. case, etc.) of each component of a sentence (Savranidis, 1998).

In this analysis, the parser—the computational formalism—clarifies the morphological ambiguity (Gakis *et al.*, 2013; Orphanos and Christodoulakis, 1999). The description of the mistakes is done through Unification Grammars' formalisms that allow context sensitive constructs to be defined, in other words, considering their contextual environment (McCord, 1987; Pollard and Sag, 1987; Bratko, 1991).

In order to construct the Greek Grammar Checker, we used the 'Mnemosyne' environment, a complete state of the art NLP system used for information retrieval and information extraction in free text. This software has been developed in Java using a parallel and/or distributed architecture. The 'Kanon' formalism is used to describe complex syntactic structures, obeys the Unification Grammars, and belongs to the level of context-sensitive grammars (Chomsky, 1965). Unification Grammars have a powerful and efficient representation of linguistic information and describe much more complex phenomena than the context-free grammars.

The Mnemosyne environment constitutes a complete NLP system that incorporates advanced linguistic resources and computational tools aiming at the automatic extraction of structured information from unstructured electronic documents. It is mainly used for automatic processing of free-text documents. It ensures: processing big volumes of information, high precision in the recognition of named entities and events, and possibility of addition of new sources of information with low cost.

The advantage of Mnemosyne is that it incorporates linguistic information data. It has already been used in environments with large quantity items with very good results on the size of the input data, the processing speed, and the output precision. The language of the text was the Greek, but 'Mnemosyne' can handle all European languages.

In http://www.neurolingo.gr/el/online_tools/ggc, we posted all templates of grammar checker in a friendly environment (Fig. 1). User types or copies the text at the lexical editor and sees the analysis of his text (Gakis *et al.*, 2016).

6 Discussion and Conclusion

The identification of mistakes confirms the existence of many elements attributed to the particular features of the discourse, and, in particular, to its improbable

character and to the intense interference observed between the speaker and the listener. The consequence of this is ineffective and sometimes incomprehensible written speech. And this is, of course, natural, given that in the daily social reality, in which the user acquires the language, oral expression prevails. In young ages, according to Ochs (Ochs and Schieffelin, 1984), children have not progressed significantly in the acquisition of a planned discourse. Some users, therefore, seem to write a great deal what they would say. They produce, in other words, written speech in the way and terms that would produce oral speech. Using Labov's terminology (Labov, 1972), these 'errors' might have been considered as markers of aspects of the sociolinguistic and perhaps the psycholinguistic situation of the users, as markers of the dominant position of oral speech in their social encounters and the dominance of its traits in their linguistic development. As markers, namely, of the kind of speech in which they are the most and the deepest exposed, which they have mainly conquered, and which plays a very important role in their lives. In conclusion, these mistaken uses, if highlighted and assessed, help us to better understand not only the function of the various parts of the language system but also the way the mechanisms of linguistic analysis work in language acquisition and generally realization.

The goal of the current study was to define Grammar Checker templates based on native speakers' (secondary students) most common mistakes as well as to describe its possible utilization into the classroom (Ho, 2005).

As regards the first research question, the results of the current study show that most mistakes can be categorized in certain categories, which can then constitute templates based on the formalism 'Mnemosyne'. The Mnemosyne environment constitutes a complete NLP system that incorporates advanced linguistic resources and computational tools aiming at the automatic extraction of structured information from unstructured electronic documents. It is mainly used for automatic processing of free-text documents. It ensures processing of large volumes of information, high precision in the recognition of named entities and events, and possibility of addition of new sources of information with low cost (Gakis *et al.*, 2015). Mnemosyne is closer to the way of thinking of human and the natural writing process (Daiute, 1985).

Regarding the second research question, i.e. beliefs of the participants for the utilization of the Greek Grammar Checker in language teaching, the participant students focused mainly to the modernization the specific tool brings to the teaching of mother tongue (e.g. not focus on drills and text reviews by hand), while they believe that its main disadvantage is the fact that it can't track all mistakes. These results seem to be not in line with previous research relevant to the Greek Grammar Checker's possible utilization by educators, which revealed the elements of the systematic and collaborative character of the reviewing process (Kokkinos *et al.*, 2018). This is a rather interesting comparison, as it shows the different elements on which different groups of native speakers with different needs in the frame of a classroom focus. This finding underlines the multiple potentials that the specific tool can offer to different groups of users inside and outside the classroom. On the other hand, the element of its inability to track all mistakes definitely reveals that there is a need for more elaborate versions of the Grammar Checker, which would be more advantageous to students and their linguistic needs.

Overall, the high level of the originality of this work should be noted, since it is the first study not only describing a specific methodology of development of a Grammar Checker for Modern Greek but also utilizing data collected from native speakers in an authentic context (school). As noted previously, data collection through authentic texts is a crucial element underlined by numerous relevant studies (Granger *et al.*, 2002; Hunston, 2002). Additionally, the fact that a group of secondary students contributes for the first time to the development of grammar checker templates and have the chance to think critically on its possible utilization in language teaching is of great importance as relevant studies have been focusing until now on evaluation of Grammar Checkers on articles/determiners, prepositions, and collocations (Leech, 1998; Han *et al.*, 2006; De Felice and Pulman, 2008).

7 Limitations of the Study

There are limitations to the study. The most important of them is the limited sample size, which does not

offer the possibility for more advanced statistical analyses and more robust and generalizable results. However, it should be mentioned that such an undertaking was not possible in the current phase due to lack of funds, which limited significantly the scope of the current study. As shown in the participants' responses, future research should focus on new, more advanced versions of the Greek Grammar Checker and its broader utilization in the classroom (in relation to a control group as well) in order for more firm results for its contribution to students' linguistic competence to be drawn.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Programme 'Competitiveness, Entrepreneurship and Innovation', under the call 'RESEARCH–CREATE–INNOVATE' (project code:T1EDK-05094).

References

Arppe, A. (2000). Developing a grammar checker for Swedish. In Nordgård, T. (ed.), *Nodalida '99 Proceedings from the 12th Nordiske datalingvistikkdager, Department of Linguistics, University of Trondheim*, pp. 13–27.

Bernth, A. (2000). EasyEnglish: Grammar checking for non-native speakers. In *Proceedings of the Third International Workshop on Controlled Language Applications*, Seattle, WA, pp. 33-42.

Bick, E. and Didriksen, T. (2015). CG-3 - beyond classical constraint grammar. In Megyesi, B. (ed.), *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa 2015)*, Linköping University Electronic Press, Linköpings universitet, pp. 31-9.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, **17**: 102–18.

Bourjaili, R. (2014). Teachers' perceptions of the individual case studies' literacy performance and their use of computer tools. In *11th International Conference on Cognition and Exploratory Learning in Digital Age, (CELDA 2014)*, Faculty of Education, La Trobe University, Victoria, Australia, pp. 79–86.

Bratko, I. (1991). *Prolog Programming for Artificial Intelligence*, 2nd edn. Wokingham: Addison Wesley.

Bustamante, F. and León, F. (1996). GramCheck: a grammar and style checker. In *Proceedings of the 16th Conference on Computational Linguistics, COLING*, Copenhagen, 5–9 August 1996, pp. 175–81.

Cavaleri, M. R. and Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, **10**(1): A223–36.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Daiute, C. (1985). *Writing and Computers*. Reading, MA: Addison-Wesley.

De Felice, R. and Pulman, S. G. (2008). A classifier-based approach to preposition and determiner error correction in L2 English. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, pp. 169–76.

Ehsan, N. and Faili, H. (2010). Towards grammar checker development for Persian language. In *6th IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLPKE' 10)*, Beijing, China, pp. 150–7.

Ferris, D., Liu, H., Sinha, A., and Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, **22**: 307–29.

Gakis, P., Panagiotakopoulos, Ch., Sgarbas, K., and Tsalidis, C. (2012). Design and implementation of an electronic lexicon for Modern Greek. *Literary and Linguistic Computing*, **27**(2): 155–69.

Gakis, P., Panagiotakopoulos, Ch., Sgarbas, K., and Tsalidis, Ch. (2013). Analysis of lexical ambiguity in Modern Greek using a computational lexicon. *Literary and Linguistic Computing*, **27**(2): 20–38.

Gakis, P., Panagiotakopoulos, Ch., Sgarbas, K., and Tsalidis, Ch. (2016). Design and construction of the Greek Grammar. *Literary and Linguistic Computing*, **32**(3): 554–76.

Gakis, P., Panagiotakopoulos, Ch., Sgarbas, K., Tsalidis, Ch., and Verykios, V. (2015). Construction of a Modern Greek grammar checker through Mnemosyne formalism. *17th International Conference, Speech and Computer (SPECOM 2015)*, Athens, 20–24 September 2015, pp. 170–7.

Granger, S., Dagneaux E., and Meunier, F. (2002). *The International Corpus of Learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Han, N., Chodorow, M., and Leacock, C. (2006). Detecting errors in English articles usage by non-native speakers. *Natural Language Engineering*, **12**(2): 115–29.

Ho, C. (2005). *Empowering English Teachers to Grapple with Errors in Grammar*. Singapore: Nanyang University Press.

Holton, D., Mackridge, P., and Philippaki-Warburton, E. (2012). *Modern Greek Grammar*. Athens: Patakis.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

John, P. and Woll, N. (2018). Using grammar checkers in the ESL classroom: The adequacy of automatic corrective feedback. In Taalas, P., Jalkanen, J., Bradley, L., and Thouësny, S. (eds), *Future-proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018*. Jyväskylä, Finland: Research-publishing.net, pp. 118–23.

Iordanidou, A. (1999). *The Guidebook of Modern Greek*. Athens: Patakis.

Iordanidou, A. (2004). *The Guidebook of Modern Greek*, 2nd edn. Athens: Patakis.

Iordanidou, A. (2013). *Is It Error or Not. Behold the Question*. Athens: Metaixmio.

Kokkinos, T., Gakis, P., and Iordanidou, A. (2018). Exploring the role of Information and Communications Technology (ICT) in language teaching: the case of the Greek Grammar Checker. In *Proceedings of the International Conference on Educational Research "Confronting Contemporary Educational Challenges through Research"*, Patras, University of Patras, 30 June–2 July 2017, pp. 264–73.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Leech, G. (1998). Learner corpora: what they are and what can be done with them. In *Learner English on Computer*. London: Longman.

McCord, M. (1987). Natural language processing in prolog. In *Knowledge Systems and Prolog*. Reading, MA: Addison-Wesley, p. 291.

Mountifield, H. (2006). 'The information commons: a student-centred environment for IT and information literacy development'. In Martin, A., and Madigan, D. (eds), *Digital Literacies for Learning*. London: Facet Publishing, pp. 172–81.

Nava, E. and Heshaam, F. (2013). Grammatical and context-sensitive error correction using a statistical machine translation framework. *Software: Practice and Experience*, **3**(2): 187–206.

Ochs, E. and Schieffelin, B. (1984). Language acquisition and socialization: three developmental stories and their implications. In Shweder, R., and Levine R. (eds), *Culture and Theory: Essays on Mind, Self, and Emotion*. Cambridge: Cambridge University Press, pp. 276–320.

Orphanos, G. and Christodoulakis, D. (1999). Part-of-speech disambiguation and unknown word guessing with decision trees. In *Proceedings of EACL*, Bergen, Norway.

Papanastasiou, G. (2008). *Modern Greek Spelling*. Athens: Triantafyllides Foundation.

Park, J. C., Palmer, M., and Washburn, G. (1997). An English grammar checker as a writing aid for students of English as a second language. In *Proceedings of Conference on Applied Natural Language Process*, New Brunswick, NJ.

Petrounias, E. (2013). *Modern Greek Grammar and Comparative Analysis*. Athens: Zitis Publications.

Pollard, C. and Sag, A. (1987). *Information-Based Syntax and Semantics*, Vol. **1**. Stanford University Ventura Hall, Stanford, CA: CSLI Publications.

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, **26**: 81–114.

Savranidis, X. (1998). *Teaching Language with the Support of Computer*. Sentence Analyzer. http://www.etpe.gr/custom/pdf/etpe264.pdf.

Setatos, M. (1991). The language mistakes and their handling. *Filologos*, **63**: 17–39.

Notes

1 The ability of a computer program to understand human language as it is spoken.
2 H for Head, L for left, C for Context, and R for Right.
3 Style 2 contains the letters of the alphabet which emit the final -n.
4 We calculate competence in the grade they have in the Modern Greek language course.