Evaluating Specifications for Controlled Greek

Marina Vassiliou
Institute for Language & Speech Processing
Epidavrou & Artemidos 6, Maroussi, 15125 Athens, Greece
mvas@ilsp.gr

Stella Markantonatou
Institute for Language & Speech Processing
Epidavrou & Artemidos 6, Maroussi, 15125 Athens, Greece
marks@ilsp.gr

Yanis Maistros
Department of Electrical & Computer Engineering, National Technical University of Athens, Athens, Greece
maistros@softlab.ece.ntua.gr

Vangelis Karkaletsis
Software and Knowledge Engineering Laboratory
Institute of Informatics and Telecommunications, N.C.S.R. "Demokritos"
vangelis@iit.demokritos.gr

Abstract

In this paper we report on the set of controlled language specifications defined for Modern Greek and the development of the respective style checker. We will focus on the effectiveness and suitability of these specifications by assessing the performance of a commercial machine translation system over controlled texts and will comment on the evaluation results. For our experiments we have used the SYSTRAN MT system (English-into-Greek language pair). We will show that an improvement in translation is feasible, when a text compliant with controlled language specifications enters a MT system. Finally, we will propose a third parameter for setting CL specifications.

1 Introduction

In this paper we outline the first, to the best of our knowledge, attempt to define specifications for controlled Modern Greek and develop the relevant authoring tool[1], aiming at the production of texts, which make a good input to MT systems (Huijsen 1998).

Our main concern here, though, is to test the validity of the aforementioned specifications by evaluating the translations of controlled texts. To this end, SYSTRAN has been used as a testing MT platform.

The linguistic and formatting specifications have been defined on the basis of the following design principles:

1. **Language level**: Reduction in ambiguity and redundancy together with effective terminology management
2. **Formatting level**: Controlled text layout, given that the text layout reflects textual structuring
3. **Implementation**: Use a development platform compatible to most current applications – Create a functional and user-friendly tool

The evaluation procedure involved the cross-checking of the translations of two types of text: (a) "raw" texts and (b) the same texts modified to conform with the controlled language specifications. The effect of the CL specifications on the translations was then measured.

2 Linguistic Specifications

Basically, we have constructed a machine-oriented controlled language (cf. Huijsen 1998). A checker was implemented and the majority of the specifications and instructions have been interpreted for machine use. However, some of the instructions are not exactly machine-oriented; they are recom-

[1] This tool is the result of the research project "SCHEMATOPIISI: Integrated environment for the development and exploitation of Greek controlled sub-languages" (ΕΠΕΤΙΙ/98), which was funded by the Greek General Secretariat of Research & Technology.

mendations, which allow the user a certain degree of freedom when processing his/her text.

The specifications we have used here comprise both the machine-oriented instructions and the human-oriented ones.

2.1 Language level

By defining the language level specifications, we aimed at eliminating ambiguity and redundancy (both lexical and structural) at the morphological, lexical and clause level.

At the morphological level, our primary effort has been to constrain *'politipia'*, a phenomenon particular to Modern Greek, which is manifested mainly by means of a variety of (mostly inflectional) endings. Thus, it is very often the case that for the same nominal/verbal root and the same grammatical properties (number, tense, person etc) more than one word forms are available (1) - (2).

These different endings often correspond to stylistic differences, which are inappropriate in a controlled language framework. Therefore, we have reduced the number of acceptable endings to one. Furthermore, we have excluded words, which reflect a certain speech style (3).

(1) **–ουν** and **–ουνε** [προσφέρ**ουν** & προσφέρ**ουνε** (= they offer)]
(2) **–ης** and **–εως** [πόλ**ης** & πόλ**εως** (= of the city)]
(3) ένεκα (= because of)

At the lexical level, our main concern has been to control the use of ambiguous words and phrases. In particular, we have set constraints on several parts of speech, which assume a variety of functions. For example, conjunctions introducing several semantic types of subordinate clauses have been excluded (4). The same holds for prepositions, which display a multitude of meanings or reflect a certain speech style (5). In most cases an alternative word or phrase is offered. Moreover, we have tried to restrict the use of pronouns (6) - the most characteristic case is the relative pronoun "**που**" (= who, which), which lacks any features for number, gender, or case - as well as of adverbs (7), whose morphological variance sometimes leads to semantic confusion. Finally, words not suitable for a controlled language environment (e.g. interjections) have been forbidden.

(4) **Αφού** δείξετε πάνω στο εικονίδιο και διπλοπατήσετε, το εικονίδιο ανοίγει. *(forbidden)*
a. **Όταν** δείξετε πάνω στο εικονίδιο και διπλοπατήσετε, το εικονίδιο ανοίγει.
[=***When*** *you point to the icon and double-click, the icon opens.*]
b. **Εάν** δείξετε πάνω στο εικονίδιο και διπλοπατήσετε, το εικονίδιο ανοίγει. [= ***If*** *you point to the icon and double-click, the icon opens.*]
(5)
a. Για την ονομασία των αρχείων να χρησιμοποιείτε λατινικούς χαρακτήρες **δίχως** κενά. *(forbidden)*
b. Για την ονομασία των αρχείων να χρησιμοποιείτε λατινικούς χαρακτήρες **χωρίς** κενά. [accepted alternative]
[= *For naming files, you should use Latin characters* **without** *blanks*.]
(6)
a. Αυτοί οι διακομιστές ανήκουν σε εξειδικευμένες υπηρεσίες, **που** ονομάζονται Φορείς Παροχής Υπηρεσιών Internet. *(forbidden)*
b. Αυτοί οι διακομιστές ανήκουν σε εξειδικευμένες υπηρεσίες, **οι οποίες** ονομάζονται Φορείς Παροχής Υπηρεσιών Internet.
[= *These servers belong to specialised services, which are called Internet Service Providers.*]
(7) ακριβ**ά** (= expensively) & ακριβ**ώς** (=precisely) <ακριβός [=expensive]

The linguistic specifications also support an effective management of terminology (Table 1). As a case study we have taken the thematic domain of computer goods and have built an extensive database of approximately 3.600 multilingual terms (one- or multi-word terms as well as acronyms). The respective constraints concern the way terms appear in the text. Moreover, we have used a checking mechanism, which crucially depends on the various fields of this database, in order to achieve successful term detection and recognition.

At the clause level, our aim was to decrease structural complexity by forbidding specific configurations such as indeclinable participial structures exhibiting a variety of possible meanings (8), iterative phrase sequences such as Genitive nouns (9) or Prepositional Phrases, varied word or con-

stituent ordering (10) or continuous embedding. The number of the available punctuation marks is also limited.

The participle *'πατώντας'* in (8) could admit either a temporal reading (= when you click) or a conditional one (=if you click). In this case, the end user is prompted to avoid the participial construction and use the relevant subordinate clause instead. Likewise in (9a) the succession of Genitive nouns gives rise to complexity, so the end user is required to reduce their number. In a similar vein, in (10) the version in (b) is indicated as the correct clause sequence, in an attempt to constrain variance in constituent ordering.

(8) **Πατώντας** στο κουμπί *Νέα διεύθυνση*, μπορείτε να δημιουργήσετε μία νέα καταχώρηση στο βιβλίο διευθύνσεων. *(forbidden)*

a. **Όταν πατήσετε** στο κουμπί *Νέα διεύθυνση*, μπορείτε να δημιουργήσετε μια νέα καταχώρηση στο βιβλίο διευθύνσεων. [= ***When you click*** *New Address, you can make a new entry in the Address Book.*]

b. **Εάν πατήσετε** στο κουμπί *Νέα διεύθυνση*, μπορείτε να δημιουργήσετε μια νέα καταχώρηση στο βιβλίο διευθύνσεων. [= ***If you click*** *New Address, you can make a new entry in the Address Book.*]

(9)

a. Το μενού Μορφοποίηση (Format) περιέχει εντολές και επιλογές για τη βελτίωση **της εικόνας των περιεχομένων του παραθύρου**. *(forbidden)*
[= *The Format menu contains commands and choices for the improvement* ***of the view of the contents of the window***.]

b. Το μενού Μορφοποίηση (Format) περιέχει εντολές και επιλογές για να βελτιώσετε **την εικόνα των περιεχομένων του παραθύρου**. [= *The Format menu contains commands and choices, so that you can improve* ***the view of the contents of the window***.]

(10)

a. Πατήστε στο κουμπί *Αποθήκευση*, για να αποθηκεύσετε το έγγραφο. *(forbidden)*
[=*Click Save, in order to save the document.*]

b. Για να αποθηκεύσετε το έγγραφο, πατήστε στο κουμπί *Αποθήκευση*. [= *In order to save the document, click Save.*]

3 Formatting Specifications

At the formatting level we have tried to establish a standard correspondence between textual structuring and the text layout. Our objective is to avoid ambiguity and vagueness not only with respect to language, but also with respect to text formatting. Therefore, the various kinds of text (titles, headers, captions, normal text, warning text etc) must be easily discernible. This has been achieved with a formatting DTD, in which differentiating textual parametres such as font, font size, line spacing etc are defined. These parameters render each kind of text easily recognisable from one another.

4 Implementation

At the implementation level we followed two paths: the first one gives a Word output (Petasis et al 2002), whereas the other approach provides an XML – HTML output, which is browsable by any Web Browser. In this way, the opportunity is given to the user to employ the core system in various environments (Markantonatou et al. 2002). Both implementations make use of the same inventories of "forbidden" words / phrases and terminological databases and sets of surface rules, which constrain the occurrences of certain structures in the period. However, they differ in the underlying technology for text processing. including the morphological one.

4.1 Word-based implementation (Demokritos)

Technical writers are able to call the controlled language checker through their word processor (MS Word is used in the current implementation). This allows users to check the format and language of their documents in a similar way as a spelling/syntax checker (Petasis et al. 2002). The technical document is first converted into an XML format in order to be processed by the checker (Fig. 1). The checker outputs the identified errors in a format “understandable” by the word-processor in order to let users view their errors. The checker checks both text language (correct application of controlled language grammar and vocabulary) and text format (e.g. line spacing, fonts style and size). The XML text is first processed using linguistic resources (restricted terminology, vocabulary, grammar) and tools (tokeniser, sentence splitter, part of speech tagger, case tagger,

morphological analyser, lexical analyser) in order to apply the language checker. Language checking involves lookup of a terminological database (termbase) and a database of forbidden words as well as checking for paragraph and sentence size, number of sentence clauses, correct appearance of terms, application of syntax restrictions, etc. The text is also checked using a format DTD (Document Type Definition) in order to locate possible errors in format.

The linguistic resources and tools used have been developed using Ellogon, a text engineering platform developed by NCSR "Demokritos" (Petasis et al. 2002). *Ellogon* was used not only as the development platform for the checker, but also as a means for embedding it under Microsoft Word.

4.2 Web-based implementation (NTUA)

The Web-based version of the authoring tool (Fig. 2) is running on a server to which the end user is connected. Users may submit to the tool their texts, which are validated by invoking the linguistic engine, a software system resident to the server. This engine triggers a client application, which produces the final output after checking is accomplished. Any XML annotated document can serve as input to this version.

The end user invokes the Web version of the Authoring Tool, supplies the system with his/her document and selects the group(s) of checkings she/he wants the system to execute. The input text is first processed by the underlying linguistic engine, which performs the following distinct tasks:

(a) Normalisation: sentence splitting and tokenisation, performed by the normaliser
(b) Part-of-Speech tagging and Grammatical Annotation: performed by the "Lexifanis" PoS Tagger (Kotsanis et al. 1985)
(c) Lemmatisation and Case Disambiguation: carried out by "QuickLem" (Kotsanis et al. 1987)

The obtained linguistic information is added to the existing XML structure in the form of PAROLE conformant tags.

The "QuickLem" lemmatiser consults a database of inflectional endings and a limited set of contextual rules (Maistros et al. 2001) Contextual rules are used to resolve case ambiguity. None of the aforementioned tools makes use of a morphological lexicon. This is advantageous, because the overall application relies on "light" tools and a restricted amount of linguistic resources.

5 Experimental Methodology

5.1 Selection and processing of corpora

For the purposes of the evaluation[2] of the impact that controlled language specifications may have on the machine translation process and output, we have collected a series of corpora consisting of texts, which were extracted from various technical documents and manuals in the Informatics thematic domain (*Corpus I*). The texts comprising *Corpus I* were originally written in Greek and they did not observe any controlled language or sublanguage specifications.

It should be mentioned that it was pretty difficult to find and process technical documents originally written in Greek, as the common practice followed in the Greek market is the translation of documents into the Greek language. We have decided, however, to exclude those kinds of texts from our experiments, aiming at obtaining representative results.

As a first step the texts of *Corpus I* were translated using SYSTRAN (*Corpus I_ trans*). Then the original texts of *Corpus I* were checked and manually corrected in accordance with the Controlled Greek specifications[3] mentioned above (*Corpus II*). The final step was the automatic translation of these checked texts (*Corpus II_trans*).

In a nutshell, we have created and processed 4 types of corpora:

i. **Corpus I**: the initial texts collected (in Greek)
ii. **Corpus I_trans**: the translation of the initial texts (in English)
iii. **Corpus II**: the corrected version of the initial texts (in Greek)
iv. **Corpus II_trans**: the translation of the corrected texts (in English)

[2] See Bourgeoys (2002) for a different approach in evaluation. She has examined a set of "raw" texts and a different set of technical documents, which were then cross-checked against their translations.
[3] We did not perform any checking with respect to text formatting, as this parameter does not have any effect on the translation output.

5.2 Testing

As a second step, we compared the aforementioned corpora on four (4) parallel windows. We tried to detect the differences in the translation output resulting from the editing of the non-CL original texts. More specifically, our aim was to find whether the changes/corrections at the morphological, lexical and syntactic level had a positive effect and rendered substantially improved translations.

For each change of the text we checked the SYSTRAN output and classified it as **correct** (= improved), **unchanged** (= invariable) or **wrong** (= non-improved). Then, for each class of constraints (e.g. lexical constraints, word order constraints etc) the number of total changes was counted as well as the number of total correct, unchanged and wrong SYSTRAN outputs. The relevant percentages were calculated as follows:

(total correct SYSTRAN output/total changes) x 100
(total unchanged SYSTRAN output/total changes) x 100
(total wrong SYSTRAN output/total changes) x 100

Before analysing the results of the testing phase, we will make a short reference to the basic characteristics and functions of the SYSTRAN MT platform.

6 Assessment of SYSTRAN

6.1 SYSTRAN basic features

SYSTRAN is a commercial MT system of the 'direct translation' type (Hutchins 1999b). In principle, the system is based on direct translations supported by bilingual lexica. It performs a word-to-word substitution and a number of word-reordering rules in the target language.

However, SYSTRAN is more sophisticated. Texts of the source language are transformed into abstract representations of 'meaning'. These representations are language-independent and are intended to be unambiguous and to provide the basis for the generation of texts into one or more target languages (Hutchins 1999b).

Like many other machine translation systems, SYSTRAN can be divided into (a) source language analysis, (b) bilingual transfer and (c) target language synthesis.

SYSTRAN performs morphological analysis, homograph analysis included, and it attempts to define clause boundaries. It performs a shallow syntactic parsing, by identifying the immediate constituents of sentences, i.e. subject and predicate, and by establishing basic syntactic relationships e.g. between a verb and its complement(s). In the bilingual transfer procedural phase the meanings of words are retrieved from the various dictionary files. Finally, in the target language synthesis phase, morphology-checking and word-rearrangement routines are performed (SYSTRAN, Internal report 1993).

At this point it should be mentioned that SYSTRAN is a MT system tuned to the sub-language(s) of the EU administrative texts, on which the terminological resources of the system mainly draw. Of course, it is possible to feed SYSTRAN with domain-specific terminology and, consequently, improve its output.

7 Evaluation

The evaluation procedure, which follows the methodology described in section 5.2, concerns solely the specifications defined at the language level, i.e. morphological, lexical and syntactic ones, since the formatting specifications do not in any way affect the translation process and the subsequent outcome.

The chart in Figure 3 depicts the results obtained, indicating for each category the percentage of cases, where the translation output has been improved or not, as well as remained the same.

The translation output has been characterised as 'improved' on the basis of the following criteria:

(a) accurate translation of words following the "disambiguation process"
(b) precise translation of terms
(c) correct identification of syntactic relations
(d) correct identification of anaphora relations

7.1 Language level specifications

Morphological level (Morphological variety): Our effort to constrain morphological variety and the corresponding stylistic differences reflected

was not proven fruitful, since SYSTRAN seems to have a good command of the rich morphological system of Modern Greek. As a consequence, SYSTRAN hardly ever fails to provide the same translation for morphologically different, but semantically similar words, apart from a few exceptions (e.g. αν και → even if vs. αν κι[4] → if also).

Lexical level (General vocabulary): The best translation results are detected after the "*disambiguation process*", that is after the specified forbidden words have been removed or the alternative recommended non-ambiguous words or phrases have been substituted for the ambiguous ones. In (11) - (13) the (a) sentences contain the forbidden words in bold and the (b) sentences the preferred alternatives, also in bold.

(11)

a. αντίστοιχα προγράμματα γενικής φύσης, **που** περιέχονται
   SYSTRAN output: *corresponding programs of general nature* ***that is*** *contained*
b. αντίστοιχα προγράμματα γενικής φύσης, **τα οποία** περιέχονται
   SYSTRAN output: *corresponding programs of general nature* ***which are*** *contained*

(12)

a. στο **επάνω** μέρος της οθόνης
   SYSTRAN output: *in the* ***on*** *part of screen*
b. στο **πάνω** μέρος της οθόνης
   SYSTRAN output: *in the* ***above*** *part of screen*

(13)

a. **Αφού** ολοκληρώσετε το μήνυμά σας
   SYSTRAN output: ***After*** *you complete your message*
b. **Όταν** ολοκληρώσετε το μήνυμά σας
   SYSTRAN output: ***When*** *you complete your message*

Lexical level (Terminology): The results regarding the translation of domain-specific terms are rather disappointing. Predictably, SYSTRAN fails nearly in every case to render the appropriate equivalent term in English. Characteristic examples of this inefficiency are phrases such as "*maternal card*" or "*left winger key of mouse*", which are given as the English equivalents of the terms "*μητρική κάρτα*" (= *mother board*) and "*αριστερό πλήκτρο του ποντικιού*" (= *left mouse button*) respectively.

Syntactic level (Clauses & complex constructions): A pretty **improved** translation output is noticeable, when certain complex clausal structures are corrected in accordance to the CL specifications.

More specifically, it has been observed that SYSTRAN can handle short clauses and sentences in a more effective way.

Likewise, a better translation output (14) may be the result of a systematic and invariable word ordering.

(14)

a. οι ιδιότητες αυτές (NP + Dem) [= *this attributes*]
b. αυτές οι ιδιότητες (Dem + NP) [= *these attributes*]

The same holds with respect to sequences of multiple Genitive nouns, when these are simplified into sequences of the type "Genitive noun + Prepositional Phrase".

The translation output is equally improved, when the indeclinable Present Participles, which are inherently ambiguous, are expanded into a corresponding subordinate clause or are replaced by a declinable verb form with a corresponding rearrangement of the whole clause (15).

(15)

a. **Πιέζοντας** υπερβολικά, μπορεί να σπάσετε την κάρτα. [= *Pressing excessively, they can you break the card*]
b. Προσοχή! Μην **πιέσετε** υπερβολικά και σπάσετε την κάρτα. [= *Attention! Do not press excessively and break the card*]

8 Conclusions

The evaluation procedure described in the previous section is an indication that a substantial improvement in translations of texts conforming to controlled language constraints is attainable.

A possible feasible solution, in order to get a substantially improved output, would be to provide SYSTRAN (or, in general, any machine translation system) with information about domain-specific terminology.

[4] "και" and "κι" constitute different morphological types of the same word "και" (= and).

None the less, a more viable and effective solution would be to adjust the controlled language specifications to the MT commercial system meant to be used in a given application. This one-way tuning to the idiosyncrasies of a MT system, functioning as a third parameter for defining controlled language specifications - apart from the human and the machine factors - could lead to a substantial improvement of its performance. We are led, thus, to a three-way distinction of CLs:

(a) Human-oriented CLs
(b) Machine-oriented CLs and
(c) MT-oriented CLs

References

Bourgeoys J. 2002: *Perspectives et Limites de l' Utilisation de l' Anglais Simplifie pour la Traduction Automatique*. M.A Thesis (Universite Laval)

Huijsen W.O. 1998: "Controlled Languages – An Introduction". In *Proceedings of the Second International Workshop on Controlled Language Applications*, 21-22 May 1998, Carnegie Mellon University, Pennsylvania, USA

Hutchins J. 1999b: "The development and use of machine translation systems and computer-based translation tools". *International Conference on Machine Translation & Computer Language Information Processing*, Chen Zhaoxiong (ed.) 26-28 June 1999, Beijing, China (pp. 1-16) [http://ourworld.compuserve.com/homepages/WJHutchins/ Beijing.htm]

Kotsanis Y. & Y. Maistros 1985: "Lexifanis" - A Lexical Analyzer of Modern Greek. In *Proceedings of the 2nd Conference of the European Chapter of the ACL*, Geneva, pp. 154-158

Kotsanis Y. & Y. Maistros & A. Zavras 1987: "Quicklem" - A Software System for Greek Word-Class Determination, Literary and Linguistic Computing, Vol. 2, No 4, pp. 242-244, Oxford University Press

Linguistic Description of SYSTRAN (Internal Report), Luxembourg, April 1993

Maistros Y., M. Vassiliou & S. Markantonatou 2001: QUICKLEM: How you find lemma without a morphological lexicon. In *Proceedings of the 5th International Conference on Greek Linguistics*, Université Paris 5, Sorbonne, pp. 71-74

Markantonatou S., V. Karkaletsis & Y. Maistros 2002: An Authoring Tool for Controlled Modern Greek. To be included in *Proceedings of the 2nd Hellenic Conference on Artificial Intelligence* (*SETN-2002*), Thessaloniki, 11-12 April 2002

Petasis G., V. Karkaletsis, G. Paliouras & C. D. Spyropoulos 2002: "Ellogon: A Text Engineering Platform". *In Proceedings of the 3rd Language Resources and Evaluation Conference* (*LREC 2002*), May 27 – 31, 2002, pp. 72 – 78

APPENDIX

Head Lemma	Head	Index	Greek Term	English Term	Acronym	Head of a multi-word term	Head Tagset	Definition	Thematic Domain	Source
Γλώσσα *(language)*	γλώσσα *(language)*	24	γλώσσα μηχανής	machine language	------	+	NoCmFeXxXx	Η απλούστερη και παλαιότερη μορφή γλώσσας, στην οποία οι εντολές του προγράμματος έχουν τη μορφή μίας ακολουθίας δυαδικών ψηφίων (bits με τιμές 0 ή 1). Τα ηλεκτρονικά κυκλώματα της Κεντρικής Μονάδας Επεξεργασίας ενός υπολογιστή είναι σε θέση να εκτελέσουν άμεσα ένα πρόγραμμα εντολών σε γλώσσα μηχανής. *(The oldest and simplest language, whose commands have the form of bit sequences. CPU's electronic circuits can directly execute a commend program in machine language)*	Γλώσσες Προγραμματισμού *(Programming languages)*	Εφαρμογές Πληροφορικής – Υπολογιστών *(Informatics)*
γλώσσα *(language)*	γλώσσα *(language)*	24	γλώσσα μηχανής	machine code	------	+	NoCmFeXxXx	Η απλούστερη και παλαιότερη μορφή γλώσσας, στην οποία οι εντολές του προγράμματος έχουν τη μορφή μίας ακολουθίας δυαδικών ψηφίων (bits με τιμές 0 ή 1). Τα ηλεκτρονικά κυκλώματα της Κεντρικής Μονάδας Επεξεργασίας ενός υπολογιστή είναι σε θέση να εκτελέσουν άμεσα ένα πρόγραμμα εντολών σε γλώσσα μηχανής. *(The oldest and simplest language, whose commands have the form of bit sequences. CPU's electronic circuits can directly execute a commend program in machine language)*	Γλώσσες Προγραμματισμού *(Programming languages)*	Εφαρμογές Πληροφορικής – Υπολογιστών *(Informatics)*
γλώσσα *(language)*	γλώσσα *(language)*		Γλώσσα Χαρακτηρισμού Υπερκειμένου	Hyper-Text Markup Language	HTML	+	NoCmFeXxXx	Γλώσσα περιγραφής και δημιουργίας ιστοσελίδων. Πρόκειται για ένα σύνολο κανόνων για τη σύνταξη ιστοσελίδων, οι οποίες περιέχουν κείμενο, εικόνες, συνδέσμους, αρχεία ήχου, video κ.λ.π. Οι ιστοσελίδες μπορούν να τοποθετηθούν σε ένα διακομιστή ιστού, ώστε να είναι διαθέσιμες και σε άλλους χρήστες μέσω του διαδικτύου. *(Language for annotating and creating web pages. It is a set of rules for designing web pages, which include text, images, links, audio files, videos etc. Web pages may be stored in a web server, so that other users can access them via the Internet).*	Διαδίκτυο *(Internet*	Εφαρμογές Λογισμικού *(Software applications)*

Table 1. Extract from the term database

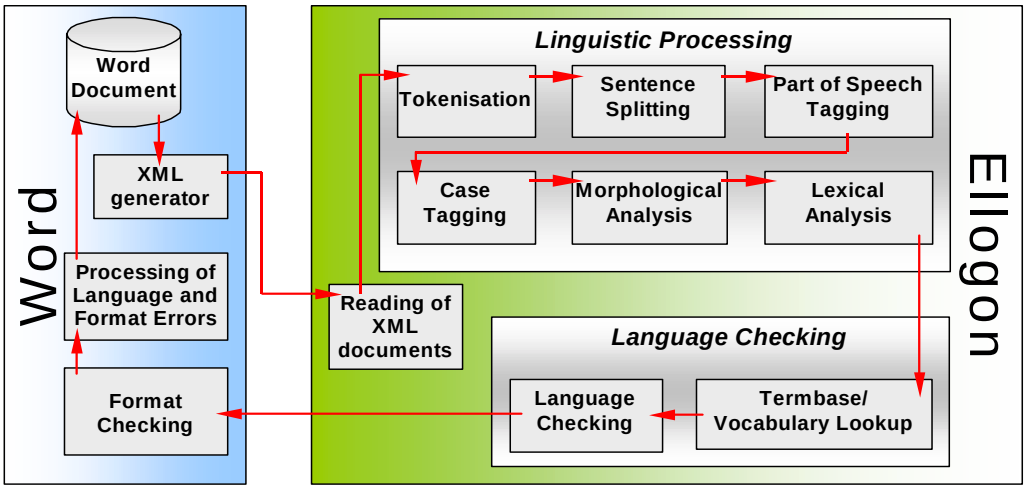

Figure 1. Architecture of the word-processor based controlled language checker

Figure 2. Presentation of the results of the WEB-based authoring tool

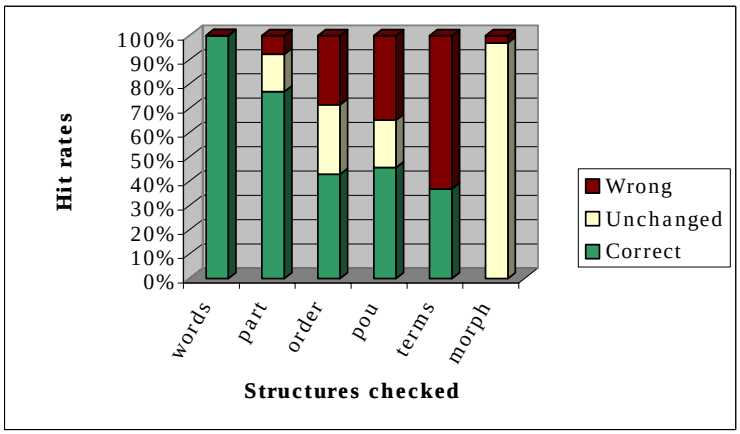

Figure 3. Presentation of preliminary results