
Cypriot Greek Lexicography: An online lexical database

Charalambos Themistocleous, Marianna Katsoyannou, Spyros Armosti & Kyriaki Christodoulou

Keywords: *web-service, Cypriot Greek, dialectal lexicography, text to speech.*

Abstract


This article presents an online dictionary environment, with enhanced sorting and searching functionalities and a text to speech feature, for hearing the pronunciation of the words. The online dictionary environment has been developed as part of the ‘Syntychies’ research program. ‘Syntychies’ online environment is a pioneering web-service for Greek dialectal lexicography and it is the first of its kind for Cypriot Greek.

1. Introduction

‘Syntychies’ is a research project for the production of linguistic lexicographic resources undertaken at the University of Cyprus during 2006–2010. The research goal of the project focuses on the study of Cypriot Greek (henceforth CG) vocabulary and its written representation. An online web service on ‘Syntychies’ website (<http://lexcy.library.ucy.ac.cy/>) has been created in order to allow access to the ‘Syntychies’ lexical database. Three main principles guide the structure and the form of the website: the accessibility, the efficiency and the user friendliness. In the rest of this article the graphical user interface (henceforth GUI) will be presented, with a short description of the data shown on screen; a presentation of the sorting and searching capabilities follows and then the text to speech features are described.

2. Graphic al User Interface (GUI)

The Graphic al User Interface (GUI) is shown in Figure 1.


**Πανεπιστήμιο
Κύπρου**
ΑΡΧΙΚΗ ΣΧΕΔΙΑΣ ΣΥΝΤΥΣΣΕΣ ΕΡΕΥΝ. ΟΜΑΔΑ ΔΗΜΟΣΙΕΥΜΑΤΑ ΒΟΗΘΕΙΑ

ΛΕΞΙΛΟΓΙΚΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ ΤΗΣ ΚΥΠΡΙΑΚΗΣ ΔΙΑΛΕΚΤΟΥ

Σε ποια στήλη ψάχνω;
 Ποια λέξη ή τμήμα λέξης ψάχνω;
 Τι είδος λέξης ψάχνω;
 Βοήθεια

Πώς πληκτρολογούνται τα κυπριακά;

Λήμμα	Φωνή	Άλλες μορφές	Άλλες γραφές	Γραμμ. κατηγορία	Προφορά	Κλίση	Σχόλια
α				επιφώνημα	a		
α-				πρόθημα	a		
αάπη		αγάπη		ουσιαστικό, θηλυκό	αάρι		
αατημένος		αγατημένος		επίθετο, τριγενές	ααριμένος	-ος, -η, -ον	
αατητικός		αγατητικός		ουσιαστικό, διγενές	ααριτικός	-ος, -ή/-ιά	
αατώ		αγατώ		ρήμα	ααρό		
αβάδωτος		αβάωτος		επίθετο, τριγενές	ανάδωτος	-ος, -η, -ον	
αβάωτα				επίρρημα	ανάωτα		
αβάωτος		αβάωτος		επίθετο, τριγενές	ανάωτος	-ος, -η, -ον	

1 2 3 4 5 6 7 8 9 10 ...

Figure 1. The ‘Syntychies’ GUI.

The GUI of the web service has minimal design and constitutes the homepage of the ‘Syntychies’ website. The fields currently are in Greek but there are plans for extended language support. The cream colored search bar is predominant at the top of the screen; at the left and at the right of the search bar there searching tools—combo boxes—that allow the user to make sophisticated research on the database. A table follows with the data.

3. General Description of the fields

The columns of the data-table contain the following information:

- (1) *Lemma (Λήμμα)*. The lemma is the head of the record. A standard search executes a search on this column.
- (2) *Voice (Φωνή)*. This field contains buttons that allow—when pressed—the user to hear the pronunciation of the word in the Lemma column.
- (3) *Other forms (Άλλες Μορφές)*. For most lemmas there are variants—mostly geographic—and these are represented in this column.
- (4) *Other spellings (Άλλες Γραφές)*. Other spellings of the word types exist since there is no standardized orthography for CG and these are presented here.
- (5) *Grammatical Category (Γραμματική Κατηγορία)*. This column shows the grammatical category a word belongs to.
- (6) *Pronunciation (Προφορά)*. The pronunciation of each word in the International Phonetic Alphabet is reported for each word in this field.
- (7) *Inflection (Κλίση)*. The inflection categories provide a description of the word’s morphological paradigm.
- (8) *Comments (Σχόλια)*. The comments are shown in the final column and provide more information concerning selected lemmas.

The web-service allows the user to sort each table in ascending or in descending order, just by double clicking each column header.

4. Search Capabilities

The user can search for any given word just by typing the word in the search bar and pressing enter or clicking at the button labeled ‘Αναζήτηση’ (search). The build-in search capabilities of the web-service allow finer search such as:

- (1) *By using expressions:* The star key < * > is stands for one or more characters while the question mark < ? > represents one character; so the search results for *τος, are all the existing CG words ending in < τος >, while a search for ?τος will provide one result: έτος (year).
- (2) *Search in other fields:* Searching is also enabled for the fields [Other forms (Άλλες Μορφές)], [Other spellings (Άλλες Γραφές)] and [Pronunciation (Προφορά)] by making the appropriate selection in the left combo-box. This capability allows the user to find information that is not part of the main lemma. The user can even search for different pronunciations in IPA; this functionality is rear in e-dictionaries but it is of utmost importance for researchers, linguists, phoneticians and speech pathologists.
- (3) *Constraining the search:* The user may choose to reduce the search results to certain grammatical categories by making the appropriate selection in the right combo-box.

5. Text to Speech

The text to speech application allows the user to hear the pronunciation of the word. This application consists of an algorithm transcribing the word to IPA (Themistocleous 2011) and a text to speech system (Xydas and Kouroupetroglou 2001). The text to speech systems allows the users to hear the words. It resolves the ambiguity occurring with homographs (cf. Jurafski and Martin 2000: 791) and informs the CG native and non-native speakers about the pronunciations of a lemma.

6. Conclusions

There are no electronic lexical databases for CG; ‘Syntychies’ research program bridges this gap, with the online web-service and by the use of modern technologies. ‘Syntychies’ web-service aims to provide free online lexicographic recourses not only to academics, researchers and scholars but to anyone interested in CG.

References

- Jurafski, D. and J. Martin 2000.** *Speech and Language Processing*. New Jersey: Prentice Hall.
- Xydas, G. and G. Kouroupetroglou 2001.** ‘The DEMOSTHeNES Speech Composer.’ In *Proceedings of the 4th ISCA Tutorial and Workshop on Speech Synthesis, SSW4*, 167–172.
- Themistocleous, C. 2011.** ‘Computational Greek Phonology: IPAGreek.’ In *Proceedings of 10th International Conference of Greek Linguistics*.