



Anemoskala: corpus and concordances for major Modern Greek poets

Anemoskala. Corpus and concordances for major Modern Greek poets, Centre for the Greek language (ed.), 2017. <http://www.greek-language.gr/digitalResources/literature/tools/concordance/index.html> (Last Accessed: 15.11.2017). Reviewed by Anna-Maria Sichani (University of Ioannina (Greece) - King's Digital Lab), amsichani (at) gmail.com.

Abstract

This paper reviews Anemoskala, an online resource offering a fully searchable textual corpus of major Modern Greek Poets (19th and 20th century), alongside with concordances and other scholarly components. Anemoskala is a digital resource originally designed to support teaching and research, and this overall goal is closely related to its scholarly identity, structure, and selection criteria for the corpus. The article assesses Anemoskala by paying attention to issues of editorial enrichment, data modelling and information architecture and interface design of the corpus' data. Although the project follows a development workflow of high scholarly quality and rigor, parts of this work are currently missing from the end user as the resource does not allow any further reuse of its content while also lacking of a robust documentation of the technical part of its implementation, from encoding decisions to technologies used. As the project is still under development, suggestions related to its technical infrastructure, documentation, sustainability, reuse scenarios and overall user experience are also discussed.

Introduction

1 Anemoskala (*Ανεμόσκαλα*)¹ offers a fully searchable textual corpus of major Modern Greek Poets of 19th and 20th century, alongside with processing tools (concordance) and scholarly resources (biographies, bibliographies etc.). The digital resource is initially developed as a project by the Centre for the Greek Language (“Κέντρο της Ελληνικής Γλώσσας”),² co-financed by the Greek state and the European Union (European Social Fund), through the Operational Programme “Educational and Lifelong Learning”, NSRF 2007-2013. Anemoskala is part of the Center’s broader digital ecosystem of educational resources called Psifides (“Ψηφίδες”),³ which includes text corpora, linguistic tools, multimedia resources, thematic essays etc. for teaching and research purposes of Modern Greek Literature.

2 Anemoskala has been developed as an online literary corpus with editorial qualities and advanced functionalities as well as one of the first digital scholarly initiatives in the field of Modern Greek Literary Studies, aiming to combine creatively and efficiently computational techniques with high quality scholarly standards. Although the field is familiar with previously developed literary corpora (mainly for Early Modern Greek texts in CD-ROM format)⁴ as well as with concordances in printed form, Anemoskala is the first freely available online literary corpus originally compiled as such alongside analytical tools. In addition, by creatively bringing together methodologies such as corpus creation, text encoding and data modelling, the project employs a challenging workflow, not without costs to its overall design, as I will discuss later on. Given the fragmented and slow adoption of digital technologies and tools in the field of Modern Greek Literary studies, Anemoskala stands not only as a reference point for its educational and scholarly contribution to the community but also as an effective touchstone for opening up the discussion on best practices, methodologies and tools around digital scholarship, Digital Humanities and Modern Greek Studies.

Selection and design

3 Anemoskala contains a fully searchable corpus of the poetic oeuvre of 16 major Modern Greek poets of the 19th and 20th century, including the two Greek Nobel Prize-winning poets Seferis and Elytis. The resource thus succeeds to present not a complete but nevertheless a useful panorama of the Modern Greek Poetry, from traditional to modern, covering also a long period of diglossia in Modern Greece. Although the

language of the corpus is the Modern Greek, the corpus includes poetic works written in Katharevousa and Demotic, two distinct language variations of Modern Greek.⁵ This language pluralism, while historically interesting, is often a headache for corpus design and its further processing and this is why the project team has thoroughly established and developed a set of normalizations.



Fig. 1: “Home Page” of Anemoscala.

4 The resource is originally conceived, designed and self-classified using a double scholarly identity, as a set of textual corpora (“Σώματα κειμένων”) and concordances (“Συμφραστικοί πίνακες”) (based on the corpora) of major Modern Greek poets (Fig. 1). Anemoscala is, thus, structured around 16 scholarly units, one for each poet, containing the edited and structured complete works (where available), the concordance tool and a number of search options, alongside extensive biographical and bibliographical material. Based on the self-classification and on the suggested typology of different types of text collections as found in Neuber & Henny-Krahmer 2017, Anemoscala is a (literary) corpus. Nevertheless, the fact that the Modern Greek community used to commonly refer to and mainly use Anemoscala as “concordances” alongside the blurred use of terms

such “digital edition” (“ψηφιακή έκδοση”) in order to alternatively describe (the compilation of) the resource (Akritidou 2016), manifest the need to further clarify concepts, methodologies and (translated) terminologies of different kinds of digital text collections in Modern Greek studies also in line with an international discussion on this field.

5 This multitude and heterogeneity of material offered might often blur the typology, the identity (*is it a literary corpus, a concordance or a digital edition?*), the development process and the uses of this text collection itself, but at the same time celebrates the double goal of the resource: to provide online access to scholarly edited text corpora of Modern Greek poetry as well to tools for further processing and analysis of the textual collection. According to Adolphs, the main aim of literary corpora is “the preservation of text resources and the development of integrated digital infrastructures to enable individual and collaborative research that is resource intensive”(Adolphs 2006, 32). In this line, Anemoskala aims to preserve the Modern Greek textual heritage in the digital realm by offering a scholarly curated digital corpus as well as further processing tools such as concordances in order “to facilitate the teaching and research of Modern Greek poetry and to enable the reading experience and pleasure of the poetic word”.⁶ Originally designed as a public-facing scholarly resource, its audience is also well-defined while broad in principle: Anemoskala seeks to address and support the educational and research needs of Modern Greek Literature teachers, students and researchers, in Greece and abroad, and at the same time stands as a reference point for everyone to read and enjoy an one of a kind collection of the Modern Greek poetic tradition.

6 The above-mentioned goals of the corpus are also well-represented in the selection process of its content. The selection criteria for poets to be included are in line with their position in the Modern Greek educational curriculum (from primary school to University level courses) as well as in the Modern Greek Literary history. The choice to stick to the (narrow) traditional canon, thus, corresponds to the educational goal of the resource and underlines the - somehow still strong - (all-male) canon formation variable also in the digital life of Modern Greek scholarship. As for now, users have access to full works by Anagnostakis, Valaoritis, Varnalis, Kalvos, Kariotakis, Palamas, Sachtouris, Seferis and Solomos and, due to copyright issues, limited access (i.e. contextual concordances) to poetic works by Sinopoulos, Ritsos, Sikelianos. Under development are currently the subcorpora including the poetic oeuvre of Eggonopoulos, Elytis and Embirikos. Even if the current size of the corpus is not available, the amount of the

general concordance entries can be easily estimated to 57,674 tokens, plus 3,153 if we include the paratextual elements⁷; these numbers will significantly increase with the addition of the remaining sub-corpora by Eggonopoulos, Elytis and Embirikos.

Editorial enrichment, data modelling and technical infrastructure

7 The poetic texts of the corpus were initially acquired from canonical printed editions (without keeping additional material such as commentary, notes) and then converted in electronic form through manual typing. At this stage, the texts are also carefully edited based on well-documented scholarly guidelines, established especially for this corpus; in other words, the digital corpus has not a print counterpart as such. These scholarly guidelines and the editorial enrichment of the texts are Anemoskala's key scholarly features and they are also in line with the available functionalities of the corpus. The research team put a lot of energy and scholarly rigor in order to establish a set of rules and guidelines for the normalisation of the language, the orthography, the accentuation and the punctuation for the vast collection of poetic texts of different authors and chronological periods (19th & 20th century). While the normalization, as a process per se, is a usual practice for printed editions or for large anthologies aiming to modernize the texts, here it mainly enables the further digital processing of the textual material (i.e. advanced & customized searches, concordances etc.). Alongside modernization, special editorial decisions have been made in a case-by-case basis for the restructuring of individual poetic works (where there is no canonical edition of 'Collected works' available).

8 Once scholarly edited, the text corpus is tokenized and enriched with annotations and mark-up, having (again) in mind the concordance function of the resource. A TEI-based schema with project-specific extensions is used as a data model for the corpus in order to adjust to the needs of a dynamically generated and searchable concordance that harvests textual data. In more detail, Anemoskala employs a three-fold TEI markup strategy for further indexing the textual corpus of each author, which is also the basic unit of the architecture: 1) structural markup (as of the hierarchy of the textual content, e.g. poet's oeuvre <text>, poems' collections and poems <div>, poems' titles <head>, verse <l>), 2) procedural markup (as of the poems' typographical layout, e.g. emphasized text <hi>), 3) semantic markup (paratextual features such as motti

<epigraph>, place <placeName>, person <name>, speakers <speaker>, translation <foreign>, authorial or editorial notes <note> etc.⁸

The screenshot shows the Varnalis concordance tool interface. At the top, there is a header for Κώστας Βάρναλης (1884-1974) with a portrait and icons for 'Για τη ζωή και το έργο του' and 'Το ποιητικό του έργο'. Below this is a search bar with 'word:1b72' and a 'Βρες' button. A section titled 'Επιλογές αναζήτησης' (Search options) includes a dropdown for 'Επιλογή συλλογών' (Collection selection) set to 9, a dropdown for 'Κατάλογοι λέξεων' (Word catalogs) set to 1, a dropdown for 'Ειδικά Κριτήρια' (Special criteria) set to 1, and a checkbox for 'Ακριβής αναζήτηση' (Exact search). Below this, the 'Λημματικός τύπος: "φως"' (Lemma type: "φως") is displayed, with a link to 'Επανάληψη αναζήτησης σε όλους τους ποιητές' (Repeat search for all poets). A list of results is shown, with the first two lines highlighted: 'σε κόσμο πεταχτό· | εγώ το φως τους να 'μαι, | αυτά να 'ναι το φως μου' and 'το φως τους να 'μαι, | αυτά να 'ναι το φως μου | στου αμίλητου αυτού κόσμου'.

Fig. 2: Results of an advanced search using the concordance tool for “φως” in all of Varnalis’ collections.

The screenshot shows the Varnalis concordance tool interface with a sidebar on the left containing a list of poets, including Κώστας Βάρναλης. The main area shows search results for 'φως'. A dropdown menu is open, showing 'Λέξεις που συμπεριλαμβάνονται στον ΣΠΛ' (Words included in the SPL) with options: 'Τίτλοι-υπότιτλοι συλλογών-ποιημάτων' (Titles-subtitles of collections-poems), 'Τίτλοι έργων άλλων συγγραφέων' (Titles of works of other authors), 'Αρχαίες ελληνικές και ξένες λέξεις' (Ancient Greek and foreign words), 'Κύρια ονόματα (ανθρωπωνύμια-τοπωνύμια)' (Main names (personalities-places)), and 'Μεταφράσεις' (Translations). The search results show 'φως (1)' and 'TO ΦΩΣ ΠΟΥ ΚΑΙΕΙ'.

Fig. 3: Results of an advanced search using the concordance tool for “φως” in all of Varnalis’ collections’ titles.

9 The concordance, as a core functionality of the resource based on the corpus, is then generated ‘on the fly’ and based on the original word-level indexing (tokenization),

the above-mentioned markup categories and the dynamic search facets provided. The TEI tagging, thus, corresponds to the available search options for the corpus: the resulting concordances and lists of words/tokens (lexical and/or semantic) are generated dynamically based on advanced customized searches to the text collection (by author/s or specific collection) (Fig. 2) or to specific textual sections (i.e. only in titles, in motti, or in specific paratextual features etc.) (Fig. 3) as well as searches for names and places, words in ancient Greek, etc.

10 Based on such a workflow⁹, the creation of the corpus with rich tagging of text properties offers to the user not only a scholarly edited text but also the option to perform advanced searches on the annotated corpus. From an operational point of view, designing and developing a TEI-encoding strategy in order to serve a specific processing task or tool (in this case, that of the concordance) alongside the non-availability of the TEI files, are decisions that significantly limit the functionality and (re)use of the project's outcomes from the outset. This is not a common decision neither for a text corpus enterprise nor for a TEI-based project, as the main virtues of these type of projects are the outcomes' flexibility and interoperability: a painstakingly marked-up text or a corpus can be further manipulated, visualized, processed in many ways, even more than its original creators ever could have imagined. This endeavour of mirroring the project's methodological decisions to a specific functionality with limited analytical power is manifested in Anemoskala's case also from the fact that the editorial and scholarly decisions can be found under the "Search guidelines" tab. These compromises are well-known to the project team; in Akritidou's words:

the rich encoding required for the (digital) edition of the texts, especially the editions of poetic texts alongside concordances, does not merely act as an addition to the traditional tasks of word lists creation, stylistic research et.c. but poses itself a series of interpretation issues that further reveal the non-neutrality of the digital (and print) editing decisions.

(Akritidou 2016, 665-6.)

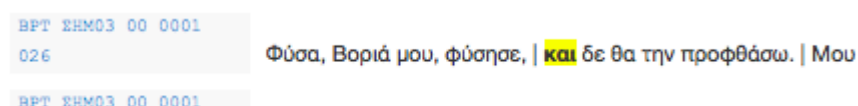


Fig. 4: Concordance results: reference abbreviations.

11 In terms of technical infrastructure and implementation, Anemoskala lacks of a clear documentation, as I will also elaborate later on, although it uses a number of specialized technologies to perform tasks such as the tokenization of the corpus (probably using regular expressions), the concordance functionality and the structuring of the corpora and the contextual data (probably using a database structure). This omission, I think, stands as a huge obstacle for the user in order to recognize e.g. the reference abbreviations (poet/collection/collection's section/poem/line) in the concordance results as "traces" of the encoding ([Fig. 4](#)) and the database structure or the alphabetical vertical axis in the concordance results as the alphabetical order over the token types. This might also explain why the average user sometimes struggles to understand the overall content design and the functionalities of the resource.

Content design, interface and navigation

12 Anemoskala succeeds in structuring this rich textual corpus containing the oeuvre of major Modern Greek poets alongside a vast amount of scholarly material and analytical tools by implementing a content design that further manifests and enhances its original scholarly aims. Having a double scholarly identity, as a textual corpus and as a concordance tool (based on the corpus), Anemoskala's multi-level content architecture and interface design are precisely founded on this double axis while enabling their multiple interconnections.

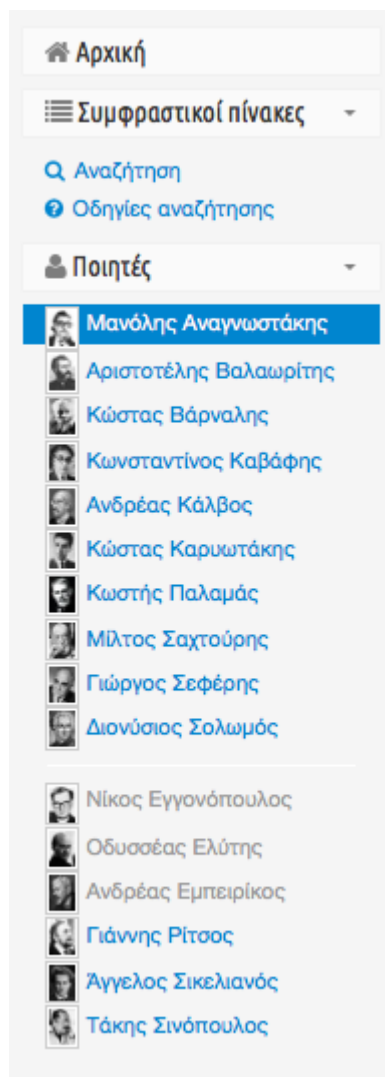


Fig. 5: The fixed left-side navigation bar.

13 The “Home page” (Αρχική) (see also [Fig. 1](#)) of the resource is structured into two distinct horizontal areas, corresponding to the main contents of the resource: the *top area*, with a general introduction note - instead of “About” -, image links to “Concordances” (Συμφραστικοί πίνακες λέξεων), that is to the general concordance of the corpus in its entirety, “biographical & bibliographical information / educational applications” (εργοβιογραφικές πληροφορίες/ διδακτικές εφαρμογές), currently under development, “Credits / Contributors” (Συστατικά στοιχεία), a faceted search bar (for the

corpus in its entirety) and the *bottom area*, with individual sub-corpora and/or related concordances using images of the poets as links.

14 The double character of the resource is also well-captured in the fixed left-side navigation bar, available throughout the resource, which contains, in addition to the “Home page” button (Αρχική), a first section under “Concordances”, with the sub-tabs “Search” (Αναζήτηση) (i.e. general concordance) and “Search guidelines” (Οδηγίες αναζήτησης) and a second section under “Poets”, with a list of all poets under separate active tabs (except from poets whose work is under copyright) (Fig. 5). Using these two sections in the navigation bar, the user can easily switch between the general concordance function and an individual poet’s corpus, or even navigate through a different poet’s corpus by making an adequate choice from the list of poets.

15 Not surprisingly, there are two main ways to navigate and use this resource, corresponding mainly to its double identity, as a textual corpus and as a concordance tool.

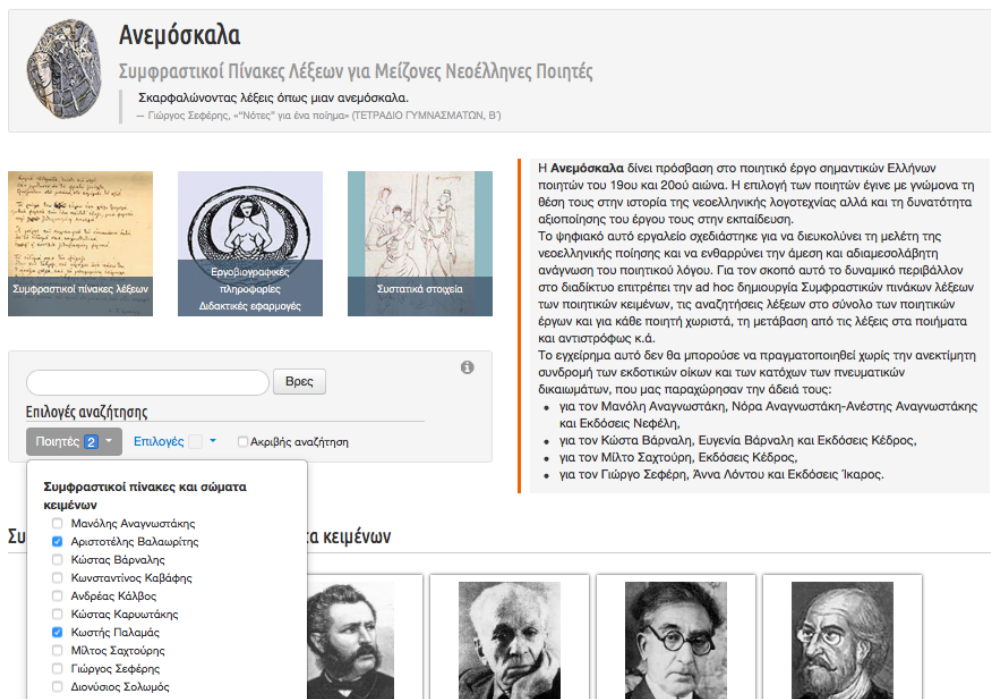


Fig. 6: Access to General Concordance function from the “Home Page” using the faceted search bar.

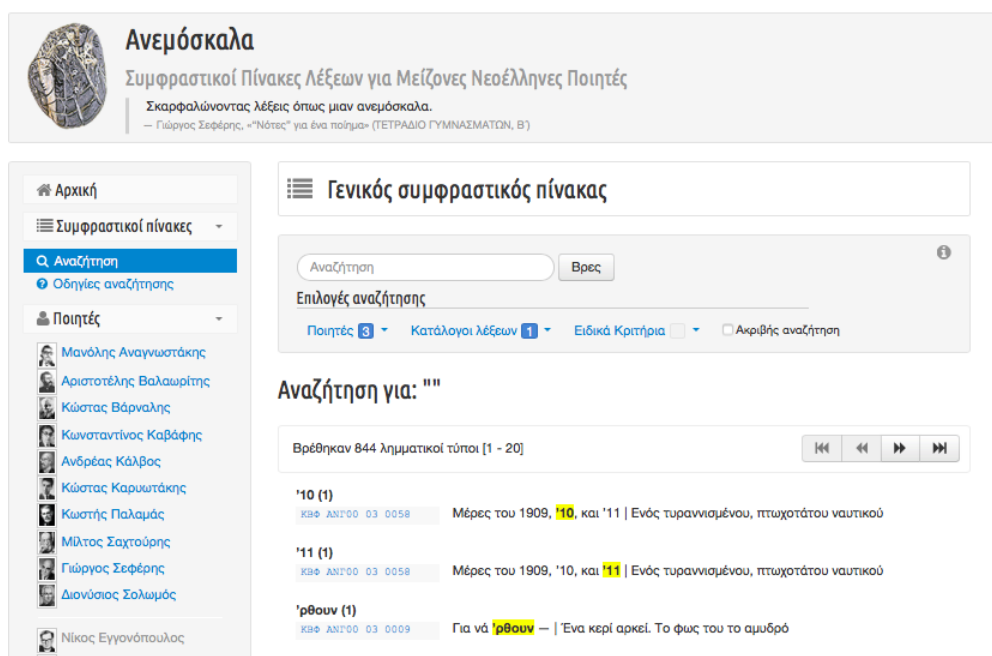



Fig. 7: Access to General Concordance via the “Search” tab in the left-side main navigation bar alongside the faceted search bar.

16 On the one hand, a user wishing to use the “Concordances/Search” function as such can perform advanced searches in the sub-corpora of more than one poets or even to the corpus in its entirety. This can be done either using the “Concordance” image link or the faceted search bar, both available in the “Home page” (Fig. 6) or the “Search” tab in the left-side main navigation bar alongside, the customized choices (Fig. 7).

17 On the other hand, a user wishing to access a poet’s digital corpus can choose and access the individual corpus using either the relevant image link on the “Home page” or the poet’s name from the list of poets in the main navigation bar.



Ανεμόσκαλα

Συμφραστικοί Πίνακες Λέξεων για Μείζονες Νεοέλληνες Ποιητές

Σκαρφαλώνοντας λέξεις όπως μian ανεμόσκαλα.

— Γιώργος Σεφέρης, «Νότες» για ένα ποίημα (ΤΕΤΡΑΔΙΟ ΓΥΜΝΑΣΜΑΤΩΝ, Β)

Αρχική


Συμφραστικοί πίνακες

Αναζήτηση

Οδηγίες αναζήτησης

Ποιητές

- Μανώλης Αναγνωστάκης
- Αριστοτέλης Βαλαωρίτης
- Κώστας Βάρναλης
- Κωνσταντίνος Καβάφης
- Ανδρέας Κάλβος
- Κώστας Καρυωτάκης
- Κωστής Παλαμάς
- Μίλτος Σαχτούρης
- Γιώργος Σεφέρης
- Διονύσιος Σολωμός**
- Νίκος Εγγονόπουλος
- Οδυσσεύς Ελύτης
- Ανδρέας Εμπειρίκος
- Γιάννης Ρίτσος
- Άγγελος Σικελιανός
- Τάκης Σινόπουλος



Διονύσιος Σολωμός (1798-1857)

Για τη ζωή και το έργο του

Το ελληνόγλωσσό έργο του

Αναζήτηση

Βρες

Επιλογές αναζήτησης

Επιλογή συλλογών ☐ Κατάλογοι λέξεων ☐ Ειδικά Κριτήρια ☐ ☐ Ακριβής αναζήτηση

Το έργο του Διονυσίου Σολωμού αποτελεί μείζον εκδοτικό πρόβλημα της νεοελληνικής φιλολογίας, ήδη από την έκδοση των σολωμικών *Ευρισκομένων* * στην Κέρκυρα το 1859 από τον Ιάκωβο Πολυλά, δύο χρόνια μετά τον θάνατο του ποιητή. Κατά τη διάρκεια της ζωής του, μικρό μέρος του υπό συνεχή επεξεργασία έργου του δημοσιεύτηκε: τα σολωμικά χειρόγραφα φυλάσσονται σήμερα στο Μουσείο Σολωμού και επιφανών Ζακυνθίων, στην Ακαδημία Αθηνών και στην Εθνική Βιβλιοθήκη, ενώ εκδόθηκαν σε πανομοιότυπη έκδοση αρχικά από τον Λίνο Πολίτη και πλέον στη σειρά *Διονυσίου Σολωμού Αυτόγραφα Έργα, ενότητες 1-15*.^[1]

Μια ψηφιακή έκδοση του σολωμικού έργου θα μπορούσε να λύσει το εκδοτικό πρόβλημα με την τεχνική του υπερκειμένου, που θα επέτρεπε την ταυτόχρονη εμφάνιση αυτογράφων, μεταγραφών και εκδοτικών παραλλαγών· κάτι τέτοιο, ωστόσο, υπερβαίνει τα όρια του συγκεκριμένου προγράμματος. Με αυτούς τους περιορισμούς, το μόνο ολοκληρωμένο ποιητικό σώμα του σολωμικού έργου σε ηλεκτρονική μορφή θα ήταν η ψηφιοποίηση της έντυπης έκδοσης των αυτογράφων. Καθώς, ωστόσο, το ηλεκτρονικό περιβάλλον της Ανεμόσκαλας δεν απευθύνεται μόνο στον ειδικευμένο ερευνητή αλλά και στους μαθητές και στο ευρύτερο κοινό, προτιμήθηκε η συγκρότηση ενός ψηφιακού corpus του ελληνόγλωσσου -και όχι αποκλειστικά ποιητικού- έργου του ποιητή, με βάση τη χρηστική έκδοση Πολυλά, όπως αναθεωρήθηκε από τον Λίνο Πολίτη:

Διονύσιου Σολωμού, *Άπαντα, τόμος πρώτος: Ποιήματα*, επιμέλεια-σημειώσεις Λίνου Πολίτη, Αθήνα: Ίκαρος, 1999 [1948].

Διονύσιου Σολωμού, *Άπαντα, τόμος δεύτερος: Πεζά και ιταλικά*, επιμέλεια-σημειώσεις Λίνου Πολίτη, Αθήνα: Ίκαρος, 1999 [1955].

Η ψηφιακή έκδοση του ελληνόγλωσσου σολωμικού corpus στο περιβάλλον γλωσσικής τεχνολογίας της Ανεμόσκαλας, σε μονοτονικό και με σύγχρονη ορθογραφία, ακολουθεί, επομένως, σε γενικές γραμμές την πρόταση κατάταξης του Λίνου Πολίτη, με κάποιες εσωτερικές διαφοροποιήσεις (όπως είναι η μεταφορά ποιημάτων μεταξύ των «συμβατικά πάντα συγκροτημένων» «συλλογών» σύμφωνα με νεότερες χρονολογήσεις ή ταυτίσεις, η ενσωμάτωση ποιημάτων από τις Σημειώσεις και το Επίμετρο της έκδοσης Πολίτη κλπ.).^[2] Δεχόμενοι τη διάρθρωση της έκδοσης Πολυλά-Πολίτη, ενσωματώνουμε και τις κυριότερες παραλλαγές από την ίδια έκδοση, καθώς στόχος είναι η αναζήτηση να περιλαμβάνει όσο το δυνατόν μεγαλύτερο μέρος του γλωσσικού υλικού του ελληνόγλωσσου

Fig. 8: The scholarly unit around Dionysios Solomos' oeuvre, Introductory-editorial note.

Ανεμόσκαλα
 Συμφραστικοί Πίνακες Λέξεων για Μείζονες Νεοέλληνες Ποιητές
 Σκαρφαλώνοντας λέξεις όπως μιν ανεμόσκαλα.
 — Γιώργος Σεφέρης, «Νότες» για ένα ποίημα» (ΤΕΤΡΑΔΙΟ ΓΥΜΝΑΣΜΑΤΩΝ, Β')

Διονύσιος Σολωμός (1798-1857)

Αναζήτηση

Επιλογές αναζήτησης
 Επιλογή συλλογών ☐ Κατάλογοι λέξεων ☐ Ειδικά Κριτήρια ☐ Ακριβής αναζήτηση ☐

[ΤΑ ΠΡΩΤΑ ΠΟΙΗΜΑΤΑ]

- [Εἰς κόρην ἣ οποῖα αναθρέφετο μέσα εἰς μοναστήρι]
- [Ἀνθαύλα]
- [Ἀνάμνησις]
- [Εἰς φίλον ψυχορραγούντα]
- [Τὸ ὄνειρό]
- [Πρὸς τὸν κύριον Λοδοβίκον Στράνη]
- [Πρὸς τὸν Λοδοβίκο Στράνη]
- [Κάκιωμα]
- [Ἰδοὺ που λαμπυρά...]
- [Ὡδή εἰς τὴ σελήνη]
- [Ἠ σκιὰ τοῦ Ὁμήρου]
- [Ἠ Ευρυκώμη]
- [Ὁ θάνατος τῆς ορφανῆς]
- [Ὁ θάνατος τοῦ βοσκού]
- [Ἠ ωικαύλα]

[ΤΑ ΠΡΩΤΑ ΠΟΙΗΜΑΤΑ]
 [ΥΜΝΟΣ Εἰς τὴν ΕΛΕΥΘΕΡΙΑΝ]
 [Εἰς τὸν ΘΑΝΑΤΟΝ τοῦ ΛΟΡΔ ΜΠΑΙΡΟΝ]
 [ΠΟΙΗΜΑΤΑ 1823–1833]
 [Ὁ ΛΑΜΠΡΟΣ]
 [Ὁ ΚΡΗΤΙΚΟΣ]
 [Οἱ ΕΛΕΥΘΕΡΟΙ ΠΟΛΙΟΡΚΗΜΕΝΟΙ]
 [Ὁ ΠΟΡΦΥΡΑΣ]
 [ἌΛΛΑ ΠΟΙΗΜΑΤΑ ΤΗΣ ΤΕΛΕΥΤΑΙΑΣ ΠΕΡΙΟΔΟΥ]
 [ΣΑΤΙΡΙΚΑ]
 [ΜΕΤΑΦΡΑΣΕΙΣ]
 [ΔΙΑΛΟΓΟΣ]
 [Ἠ ΓΥΝΑΙΚΑ ΤΗΣ ΖΑΚΥΘΟΥ]

Fig. 10: The scholarly unit around Dionysios Solomos' oeuvre, His Poetic oeuvre / Corpus.

18 Around each poet, a scholarly unit is organized as an individual web environment containing:

1. a central navigation bar with
 - the “poet’s ID” (poet’s name, date of birth & death, picture, copyright information) (Fig. 8),
 - a tab to “Life and work” (“Για τη ζωή και το έργο”), containing rich multimedia contextual material, using content (e.g. images, facsimiles, high-quality scholarly or editorial notes) also from external providers (e.g. Europeana, University Libraries, Public Domain) (Fig. 9), and
 - a tab to “His Poetic oeuvre” (“Το ποιητικό του έργο”), accessing his text corpus, structured as a hyperlinked table of contents with collections/ groups of poems (Fig. 10),
2. a faceted search bar, containing the search box and four (4) search categories (Collections, Word lists, Special criteria, Exact Match - in separate drop down multiple select checkbox lists) and
3. an introductory-editorial note, which is only available in the entry page of each scholarly unit.

19 The user can access, navigate and search through the corpus of a selected poet either by using the table of contents of his oeuvre or by performing a number of advanced searches using the faceted search bar. Furthermore, the user can navigate from the search results to the full text and vice versa, using the available reference abbreviations of the encoding, except from copyrighted works.

20 In terms of interface design, Anemoskala overall offers a simple and clear layout that further facilitates user experience and navigation through the vast content, functionalities and complex architecture of the resource. More specifically, the individual scholarly units for the 16 poets share the same basic structure, layout and functionalities. This design choice on the one hand contributes to a uniform visual identity and gives the sense of a 'seriality' to the textual corpus as a whole and on the other hand creates a pattern for the user's navigation. Specifically, for the Concordance function, either for an individual corpus or a group of subcorpora, the fixed faceted search bar with clear categories and the navigation guides (arrows, pagination, number of results) are well-designed; a "clear previous search" option will also facilitate the user navigation.

21 Furthermore, the vertical navigation bar along the left is a wise design decision due to the density of content (i.e. the list of 16 poets) as its purpose is chiefly to support the user's browsing. While the navigation bar across the entire resource assists the user to understand the resource's structure and to navigate throughout, the absence of this very navigation bar from the "Home Page" creates a disruption in the user flow. Instead of making information more easily accessible, this somehow unnecessary complexity actually makes it more difficult for end-users to get the information they are looking for. Just to give two examples of this design failure: the "Contributors" page is only accessible via the "Home Page"; the faceted search bar in the "Home Page" and the faceted search bar under "Search" in the navigation bar, both of them designed for the general concordance, have slightly different facets). In addition, as individual rich web environments are created around each scholarly unit/poet, a (local) 'Home button' per unit is also required. As the multiplicity of paths to enter the same piece of information or to perform the same task is one of the main characteristics of the resource, a breadcrumbs trail, a 'back' button and a generic 'search' button for the resource in its entirety would be welcoming additions in order to boost user experience.

22 The consistent use of graphical user interface elements throughout the site, especially input controls (image links with mouse hovers, buttons, text fields, checkbox

lists, icons and message boxes) and navigation components (arrows and pagination mainly for the concordance results), is of great assistance to the user in performing main navigation tasks with efficiency and satisfaction while keeping to an aesthetic awareness. All in all, the front-end navigation adequately reflects the back-end content architecture alongside a clear, functional and uniform design that succeeds in delivering a smooth user experience. As currently Anemoskala is still under development (without a clear completion plan available), features of responsive web design alongside a URL-structuring strategy will significantly improve the discoverability and the overall usability of the resource.

Access & Publication

23 However, in spite of the many advantages of a free readily available edited corpus of Modern Greek Poetry, since Anemoskala's conception there have been two open issues regarding the level of 'access' offered by the resource.

24 Firstly, due to copyright, there is a limited access to a number of works for which the project was not able to obtain permission for full online access (i.e. works by other poets of "Generation of the Thirties" and the first post-war generation). The status of intellectual rights and copyright of each poet is well-documented in the individual introductory-editorial note of each scholarly unit and further details regarding the publication terms are given there. That means, the user can still search the copyrighted subcorpus and visualise the search results only through the concordance function at line level but then she can't access the full text to which it belongs.

25 Secondly, while Anemoskala achieves to offer a rich textual corpus of Modern Greek Poetry, its publication policy does not allow any kind of download, export or further reuse of any part of the textual corpus (plain text, TEI-XML files, metadata, concordance results). According to the project's disclaimer, "*the copyright to the digital encoded version is held by the Centre for the Greek Language, in conjunction with the various holders of intellectual property, and redistribution is not allowed in order to protect the rights of the print edition publishers*".

26 By making a literary corpus available on 'read-only' or 'see-but-not-touch' mode, for browsing (where full text is not copyrighted) and/or for performing a specific and highly controlled (using the guided navigation in the faceted search) analytical task (concordances), the project becomes 'self-limited' as of the possible visualisations and

uses of the data. As Hockey reminds us in one of the first books on the development and the uses of electronic texts in the Humanities back in 2000:

the production of high-quality electronic texts and resources is very labour intensive. It is imperative to find ways of making these resources as broadly multipurpose and reusable as possible. This must be done in the context of different theoretical perspectives in scholarship and in the context of a broadening range of users.

(Hockey 2004, 165.)

27 Literary corpora, as Anemoskala, and in general text collections should not be seen as the end-product or an end in itself but as the starting point for sophisticated computational processing activities: for critical or genetic digital editions; for a number of large-scale comparative approaches and methods of distant reading through their tagging and availability of the metadata; for a vast spectrum of data representation and computational analysis techniques such as data visualizations, topic modelling, stylometry, text mining, lexical patterning, or even machine learning etc. As for now, Anemoskala's corpus is not used in any other processing framework besides the original concordance function, the results of which could be then consulted only to support traditional scholarly outputs (articles, monographs et.c.). On the contrary, such a rich corpus could be seen as the starting point of highly advanced computational analysis endeavours to explore, for example, the diachronic evolution of literary themes or linguistic patterns. Such computationally intensive approaches to the literary stand mainly as a hypothesis testing device for literary analysis and as markers of (inter)disciplinary methodological shifts: *"text analysis is to participate in literary critical endeavour in some manner beyond fact-checking... to assist the critic in the unfolding of interpretative possibilities... its purpose should be to generate further 'evidence', [which] is not definitive but suggestive of grander arguments and schemes"* (Ramsay, 2011, 10). After all, digital tools in textual studies enable us to move beyond 'screen essentialism' towards experimentation of epistemological and methodological boundaries.

Documentation and long-term maintenance



Fig. 11: The scholarly and editorial documentation under the “Search guidelines”.

28 As Anemuskala aims to offer an updated, modernized and edited version of texts of Modern Greek Poetry, the project team seeks to provide well-documented editorial decisions, explaining in high detail the normalization rules they followed and the scholarly enrichment of the texts. Based on these guidelines, a list of search scenarios and FAQ are also part of the scholarly documentation, helping the user to understand and navigate through the collection while indirectly manifesting the single processing use of the resource, through the concordance tool. As mentioned earlier in this review, the fact that all the above mentioned documentation can be found loosely structured under the “Search guidelines” tab (“Οδηγίες αναζήτησης”) further blurs the methodology, the development workflow and the overall goals of the resource (Fig. 11). Documentation and information about the editorial history and conditions of each particular subcorpus (aka metadata), are available through the introductory-editorial note on each scholarly unit.

29 On the other hand, a detailed technical documentation, including the infrastructure and the data architecture employed, is highly missing - the technical development of the resource has been undertaken by an external collaborator (Centre for Research and Technology Hellas - CERTH). This lack of any kind of documentation regarding the technical implementation is more clearly manifested when compared with

the available detailed and laborious philological documentation. In my eyes, this omission also manifests the still influential distinction between scholarly and technical work, especially in fields and communities of practice such as Modern Greek Studies with limited exposure and institutional support towards the field of digital scholarship. In general, the creation of a digital textual collection and resource demands and in the same time celebrates the creative, interdisciplinary and collaborative work of textual and literary scholars, computational linguists, digital humanists and computer engineers. All this combined knowledge and expertise should be well-documented, openly available for review and consultation for future similar attempts.

30 As for project documentation, the resource provides detailed description of the interdisciplinary project team, their roles and contribution under a dedicated image link (Contents” / Συστατικά στοιχεία) available only at the “Home page”. It should also be mentioned that although the content of the resource is regularly updated and adequately documented as such, including quality control provision, a citation mechanism is not yet provided as part of its scholarly documentation, significantly reducing its scholarly usability in the short and in the long term. Finally, given the fact that Anemoskala is a publicly funded project offering freely available textual resources that are widely used in various educational and research contexts, a robust sustainability plan for the long-term maintenance and preservation would be also highly expected. To this end, a real-world assessment of its actual scholarly use cases as well as a collaborations and partnerships plan with libraries, archives, publishers and other stakeholders in order to enrich the textual collection with cross-domain content and outputs could provide interesting insights for the future development plan.

Overall conclusions and recommendations

31 To conclude, Anemoskala is an invaluable digital scholarly resource for the educational and research community of Modern Greek Literary studies by providing a gateway for high-quality edited literary corpora, analytical tools, alongside impressive media-rich contextual material, in an accessible, overall easy for the user to navigate way. Besides its scholarly and philological virtues, Anemoskala stands also as a milestone for the digital scholarship in Modern Greek Humanities by promoting and supporting the adoption of Digital Humanities best practices, established standards, computational technologies, and collaborative workflows among the Greek research community.

32 As an ongoing project with expected progress in the next months, we would expect the resource not only to grow in content, textual data and analytical power but also to amend inconsistencies and gaps in the documentation and small design failures that will further enhance its usability and its contribution to the field. Anemoskala can also generate and benefit from vivid and creative discussions regarding the methodology, the development, the tools and the applications of digital corpora and text collections in various interdisciplinary contexts, both in national and international level. In order to play such as role, it is important to create and provide robust and transparent documentation of the rigorous editorial work, the technical infrastructure and implementation and the overall workflow (an English version of the documentation would be also of high importance).

33 As mentioned earlier in this review, the lack of user empowerment and engagement with the textual data is, in my view, one of the major shortcomings of Anemoskala. I claim that the browsing and search options should by no means be the end-goal of a textual collection development. On the contrary, the audience's creative and productive interaction with the collection can be enhanced by allowing users to read and/or download various textual formats and contents, to creatively reuse them, also through aggregation mechanisms (API); by offering additional advanced analytical and interactive tools; by enabling web 2.0 functionalities and environments (e.g. Wikis) for users to contribute, comment or collaboratively annotate the texts of the collection. By opening up Anemoskala to its users in a more interactive and creative way, the project team should have in mind that communication should not be one-way traffic on the resource - especially when digital-natives, such as pupils and students, are among the audience. In fact, a dialogue between the resource and the intended audience should be genuinely established: a "contact" page, social media plugins, alongside an active community space with use cases, news and discussions, will allow the project team to "meet and learn" from its users and their needs.

34 Finally, Anemoskala's digital text collection corpus can be placed as the starting point of an ecosystem of digital scholarship initiatives and projects that will build upon the existing edited corpus. This infrastructure logic, fundamental for the development and longevity of digital text corpora and resources, given the examples of Oxford Text Archive¹⁰ or the Women's Writers Project¹¹, is a sustainable operational model for producing digital textual resources of ongoing scholarly relevance and contributing to the advancement of digital scholarship in the field at large. Hopefully Anemoskala will

capitalise expertise, achievements, and experiences on digital text corpus development and will raise expectations not only for quality digital scholarly resources in the field but also for a Modern Greek digital literary studies agenda and beyond.

Notes

1. The name of the resource “Anemoskala” (rope ladder) is derived from a phrase of a poem of George Seferis, also included in the corpus: “climbing words as if up a rope ladder” (“Σκαρφαλώνοντας λέξεις όπως μίαν ανεμόσκαλα”).

2. <https://web.archive.org/web/20180217084726/http://greeklanguage.gr/>.

3. <https://web.archive.org/web/20180217084813/http://www.greek-language.gr/digitalResources/index.html>.

4. Philippides, Dia M. L. & Holton David (with technical assistance John L. Dawson), 2014. *Erotokritos. As the disk spins. CD-ROM (PC/MAC) [Ερωτόκριτος: Του δίσκου τα γυρίσματα]*. Athens, Ermis Publications.

5. Katharevousa, a conservative form of the Modern Greek language conceived in the early 19th century, was widely used in the public sector and literature, while Demotiki, the modern vernacular form of the Greek language, was used mainly in everyday life. The tension between the two linguistic varieties, also known as the ‘Greek language question’, was a highly controversial dispute in the 19th and 20th centuries about the official language of the Greek nation, also linked to national identity issues, and it was finally resolved in 1976 in favour of Demotic. See also Peter Mackridge, 1985, *The Modern Greek Language*, Oxford University Press.

6. “Introductory Note” <https://web.archive.org/web/20180217085234/http://www.greek-language.gr/digitalResources/literature/tools/concordance/index.html>.

7. Amounts estimated from the total list of tokens available under the full search function.

8. This additional information about the encoding has been supplied on private email communication with Maria Akritidou, part of the scientific team coordinating the project - I am more than grateful to Maria for being so helpful and generous in providing parts of information regarding several aspects of the resource’s development. As I will argue later on, more of this information should be made available on the resource’s documentation.

9. Especially for the case of the poetic work of George Seferis, the workflow is different. The original printed concordance was first converted to a database structure and then the work of Seferis was compiled around this service and used as a point of entry to the concordance and vice versa.

10. <https://web.archive.org/web/20180217085326/https://ota.ox.ac.uk/>.

11. <https://web.archive.org/web/20180217085638/http://www.wwp.northeastern.edu/>.

References

Adolphs, Svenja, 2006. *Introducing electronic text analysis. A practical guide for language and literary studies*, Routledge.

Akritidou, Maria 2015. "Modern Greek Studies in the digital era. The development of a digital corpus of Modern Greek Poetry at CGL using the TEI international standard" ["Οι νεοελληνικές σπουδές μπροστά στην ψηφιακή πραγματικότητα. Η ανάπτυξη ψηφιακού σώματος νεοελληνικής ποίησης από το ΚΕΓ με το διεθνές πρότυπο Text Encoding Initiative"], *Proceedings 5th European Congress of Modern Greek Studies of the European Society of Modern Greek Studies Thessaloniki, 2-5 October 2014 Continuities, Discontinuities, Ruptures in the Greek World (1204-2014): Economy, Society, History, Literature*, Vol.IV, K.Dimadis (ed), Athens: Society of Modern Greek Studies, p.663-691.

Hockey, Susan, 2004 [2000]. *Electronic Texts in the Humanities*, Oxford University Press.

Neuber, Frederike & Henny-Krahmer Urlike, 2017. "Editorial: Reviewing Digital Text Collections", *RIDE*, Issue 6: Text Collections, <https://web.archive.org/web/20180222141455/https://ride.i-d-e.de/issues/issue-6/editorial-reviewing-digital-text-collections/>.

Ramsay, Stephen, 2011. *Reading Machines: Toward and Algorithmic Criticism (Topics in the Digital Humanities)*, University of Illinois Press.

Factsheet

Resource reviewed	
Title	Anemoskala. Corpus and concordances for major Modern Greek poets
Editors	Centre for the Greek language
URI	http://www.greek-language.gr/digitalResources/literature/tools/concordance/index.html
Publication Date	2017
Date of last access	15.11.2017

Reviewer	
Surname	Sichani
First Name	Anna-Maria
Organization	University of Ioannina (Greece) - King's Digital Lab
Place	London
Email	amsichani (at) gmail.com

General Information		
Bibliographic description	Can the text collection be identified in terms similar to traditional bibliographic descriptions (title, responsible editors, institution, date(s) of publication, identifier/address)? (cf. Catalogue 1.1)	yes
Contributors	Are the contributors (editors, institutions, associates) of the project documented? (cf. Catalogue 1.3)	yes
Contacts	Is contact information given? (cf. Catalogue 1.4)	no
Aims		
Documentation	Is there a description of the aims and contents of the text collection? (cf. Catalogue 2.1)	yes
Purpose	What is the purpose of the text collection? (cf. Catalogue 2.2)	Research, Teaching, General purpose

Kind of research	What kind of research does the collection allow to conduct primarily? (cf. Catalogue 3.1.8)	Qualitative research
Self-classification	How does the text collection classify itself (e.g. in its title or documentation)? (cf. Catalogue 2.3)	Corpus
Field of research	To which field(s) of research does the text collection contribute? (cf. Catalogue 2.2)	Literary studies
Content		
Era	What era(s) do the texts belong to? (cf. Catalogue 2.5)	Modern
Language	What languages are the texts in? (cf. Catalogue 2.5)	Greek
Types of text	What kind of texts are in the collection? (cf. Catalogue 2.5)	Literary works
Additional information	What kind of information is published in addition to the texts? (cf. Catalogue 2.5)	Introduction, Context material, Bibliography
Composition		
Documentation	Are the principles and decisions regarding the design of the text collection, its composition and the selection of texts documented? (cf. Catalogue 3.1.1-3.1.3)	yes
Selection	What selection criteria have been chosen for the text collection? (cf. Catalogue 3.1)	Language, Genre
Size		
Texts/records	How large is the text collection in number of texts/ records? (cf. Catalogue 3.1.4)	unknown
Tokens	How large is the text collection in number of tokens? (cf. Catalogue 3.1.4)	< 100,000
Structure	Does the text collection have identifiable sub-collections or components? (cf. Catalogue 3.1.5)	yes
Data acquisition and integration		
Text recording	Does the text collection record or transcribe the textual data for the first time? (cf. Catalogue 3.1.6)	yes

Text integration	What kind of material has been taken over from other sources? (cf. Catalogue 3.1.6)	Full texts, Metadata
Quality assurance	Has the quality of the data (transcriptions, metadata, annotations, etc.) been checked? (cf. Catalogue 3.1.7)	yes
Typology	Considering aims and methods of the text collection, how would you classify it further? For definitions please consider the help-texts. (cf. Catalogue 3.1.8)	Corpus
Data Modelling		
Text treatment	How are the textual sources represented in the digital collection? (cf. Catalogue 3.2.1)	Normalized transcription, Orthographic transcription
Basic format	In which basic format are the texts encoded? (cf. Catalogue 3.2.4)	XML
Annotations		
Annotation type	With what information are the texts further enriched? (cf. Catalogue 3.2.2)	Semantic annotations, Structural information
Annotation integration	How are the annotations linked to the texts themselves? (cf. Catalogue 3.2.2)	Embedded
Metadata		
Metadata type	What kind of metadata are included in the text collection? (cf. Catalogue 3.2.3)	Descriptive, Structural
Metadata level	On which level are the metadata included? (cf. Catalogue 3.2.2)	Collection parts/ components, Individual texts
Data schemas and standards		
Schemas	What kind of data/metadata/annotation schemas are used for the text collection? (cf. Catalogue 3.2.4)	Customized standard schema
Standards	Which standards for text encoding, metadata and annotation are used in the text collection? (cf. Catalogue 3.2.4)	TEI
Provision		

Accessibility of the basic data	Is the textual data accessible in a source format (e.g. XML, TXT)? (cf. Catalogue 4.1)	no
Download	Can the entire raw data of the project be downloaded (as a whole)? (cf. Catalogue 4.2)	no
Technical interfaces	Are there technical interfaces which allow the reuse of the data of the text collection in other contexts? (cf. Catalogue 4.2)	none
Analytical data	Besides the textual data, does the project provide analytical data (e.g. statistics) to download or harvest? (cf. Catalogue 4.3)	no
Reuse	Can you use the data with other tools useful for this kind of content? (cf. Catalogue 4.4)	no
User Interface		
Interface provision	Does the text collection have a dedicated user interface designed for the collection at hand in which the texts of the collection are represented and/or in which the data is analyzable? (cf. Catalogue 5.1)	yes
User Interface questions		
Usability	From your point of view, is the interface of the text collection clearly arranged and easy to navigate so that the user can quickly identify the purpose, the content and the main access methods of the resource? (cf. Catalogue 5.3)	yes
Access modes		
Browsing	Does the project offer the possibility to browse the contents by simple browsing options or advanced structured access via indices (e.g. by author, year, genre)? (cf. Catalogue 5.4)	yes
Fulltext search	Does the project offer a fulltext search? (cf. Catalogue 5.4)	yes
Advanced search	Does the project offer an advanced search? (cf. Catalogue 5.4)	yes
Analysis		

Tools	Does the text collection integrate tools for analyses of the data? (cf. Catalogue 5.5)	yes
Customization	Can the user alter the interface in order to affect the outcomes of representation and analysis of the text collection (besides basic search functionalities), e.g. by applying his or her own queries or by choosing analysis parameters? (cf. Catalogue 5.5)	yes
Visualization	Does the text collection provide particular visualizations of the data? (cf. Catalogue 5.6)	no visualization
Personalization	Is there a personalisation mode that enables the users e.g. to create their own sub-collections of the existing text collection? (cf. Catalogue 5.7)	no
Preservation		
Documentation	Does the text collection provide sufficient documentation about the project in general as well as about the aims, contents and methods of the text collection? (cf. Catalogue 6.1)	no
Open Access	Is the text collection Open Access? (cf. Catalogue 6.2)	no
Rights		
Declared	Are the rights to (re)use the content declared? (cf. Catalogue 6.2)	yes
License	Under what license are the contents released? (cf. Catalogue 6.2)	No explicit license / all rights reserved
Persistent identification and addressing	Are there persistent identifiers and an addressing system for the text collection and/or parts/objects of it and which mechanism is used to that end? (cf. Catalogue 6.3)	none
Citation	Does the text collection supply citation guidelines? (cf. Catalogue 6.3)	no
Archiving of the data	Does the documentation include information about the long term sustainability of the basic data (archiving of the data)? (cf. Catalogue 6.4)	no
Institutional curation	Does the project provide information about institutional support for the curation and sustainability of the project? (cf. Catalogue 6.4)	no

Completion	Is the text collection completed? (cf. Catalogue 6.4)	no
Personnel		
Editors	M. Akritidou A. Gioti Chr. Daniel K. Dimopoulou F. Kolitsi A. Margoni P. Pantzarelak Aik. Tiktououlou Sp. Spyrououlos E. Floka	
Encoders	E. Anthououlou M. Akritidou A. Mavrououlos Our. Voskaki Aik. Makri	
Programmers	Information Technologies Institute - Centre of Research & Technology - Hellas	
Advisors	I. N. Kazazis	
Designers	T. Yiannou	
Administrators	T. Yiannou M. Akritidou A. Margoni St. Mpoteli	
Contributors	G. Alvanidou P. Andriououlou Our. Voskaki K. Dimopoulou P. Karatzoglou Sp. Kastanioti M. Koumalas P. Kypraiou A. Margoni A. Mavrououlos V. Mpatsidou St. Sporiki M. Akritidou V. Vasileiadis T. Yiannou A. Gioti Chr. Daniel K. Dimopoulou F. Kolitsi P. Pantzarelak Aik. Tiktououlou	