



FACHLITERATUR  
EDITION ROMIOSINI  
ΕΠΙΣΤΗΜΗ



ICGL12 | 12<sup>th</sup> INTERNATIONAL CONFERENCE  
ON GREEK LINGUISTICS  
16 – 19 SEPTEMBER 2015  
FREIE UNIVERSITÄT BERLIN, CEMOG

# Proceedings of the ICGL12

vol. 1

The International Conference on Greek Linguistics is a biennial meeting on the study and analysis of Greek (Ancient, Medieval and Modern), placing particular emphasis on the later stages of the language.



**PROCEEDINGS OF THE ICGL12**  
**ΠΡΑΚΤΙΚΑ ΤΟΥ ICGL12**



**Thanasis Georgakopoulos, Theodossia-Soula Pavlidou, Miltos Pechlivanos,  
Artemis Alexiadou, Jannis Androutsopoulos, Alexis Kalokairinos,  
Stavros Skopeteas, Katerina Stathi (Eds.)**

**PROCEEDINGS OF THE 12<sup>TH</sup> INTERNATIONAL  
CONFERENCE ON GREEK LINGUISTICS**

**ΠΡΑΚΤΙΚΑ ΤΟΥ 12<sup>ΟΥ</sup> ΣΥΝΕΔΡΙΟΥ ΕΛΛΗΝΙΚΗΣ  
ΓΛΩΣΣΟΛΟΓΙΑΣ**

VOL. 1



© 2017 Edition Romiosini/CeMoG, Freie Universität Berlin. Alle Rechte vorbehalten.  
Vertrieb und Gesamtherstellung: Epubli ([www.epubli.de](http://www.epubli.de))  
Satz und Layout: Rea Papamichail / Center für Digitale Systeme, Freie Universität Berlin  
Gesetzt aus Minion Pro  
Umschlaggestaltung: Thanasis Georgiou, Yorgos Konstantinou  
Umschlagillustration: Yorgos Konstantinou

ISBN 978-3-946142-34-8  
Printed in Germany

Online-Bibliothek der Edition Romiosini:  
[www.edition-romiosini.de](http://www.edition-romiosini.de)

*Στη μνήμη του Gaberell Drachman (†10.9.2014)  
και της Αγγελικής Μαλικούτη-Drachman (†4.5.2015)  
για την τεράστια προσφορά τους στην ελληνική γλωσσολογία  
και την αγάπη τους για την ελληνική γλώσσα*





## ΣΗΜΕΙΩΜΑ ΕΚΔΟΤΩΝ

Το 12ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας (International Conference on Greek Linguistics/ICGL12) πραγματοποιήθηκε στο Κέντρο Νέου Ελληνισμού του Ελεύθερου Πανεπιστημίου του Βερολίνου (Centrum Modernes Griechenland, Freie Universität Berlin) στις 16-19 Σεπτεμβρίου 2015 με τη συμμετοχή περίπου τετρακοσίων συνέδρων απ' όλον τον κόσμο.

Την Επιστημονική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Δώρα Αλεξοπούλου, Γιάννης Ανδρουτσόπουλος, Αμαλία Αρβανίτη, Σταύρος Ασημακόπουλος, Αλεξάνδρα Γεωργακοπούλου, Κλεάνθης Γκρώμαν, Σαβίνα Ιατρίδου, Mark Janse, Brian Joseph, Αλέξης Καλοκαιρινός, Ναπολέον Κάτσος, Ευαγγελία Κορδώνη, Αμαλία Μόζερ, Ελένη Μπουτουλούση, Κική Νικηφορίδου, Αγγελική Ράλλη, Άννα Ρούσου, Αθηνά Σιούπη, Σταύρος Σκοπετέας, Κατερίνα Στάθη, Μελίτα Σταύρου, Αρχόντω Τερζή, Νίνα Τοπιντζή, Ιάνθη Τσιμπλή και Σταυρούλα Τσιπλάκου.

Την Οργανωτική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Αλέξης Καλοκαιρινός, Κώστας Κοσμάς, Θεοδοσία-Σούλα Παυλίδου και Μίλτος Πεχλιβάνος.

Οι δύο τόμοι των πρακτικών του συνεδρίου είναι προϊόν της εργασίας της Εκδοτικής Επιτροπής στην οποία συμμετείχαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Γιάννης Ανδρουτσόπουλος, Αλέξης Καλοκαιρινός, Σταύρος Σκοπετέας και Κατερίνα Στάθη.

Παρότι στο συνέδριο οι ανακοινώσεις είχαν ταξινομηθεί σύμφωνα με θεματικούς άξονες, τα κείμενα των ανακοινώσεων παρατίθενται σε αλφαβητική σειρά, σύμφωνα με το λατινικό αλφάβητο· εξαίρεση αποτελούν οι εναρκτήριες ομιλίες, οι οποίες βρίσκονται στην αρχή του πρώτου τόμου.

Η Οργανωτική Επιτροπή του ICGL12



## ΠΕΡΙΕΧΟΜΕΝΑ

Σημείωμα εκδοτών .....	7
Περιεχόμενα .....	9
Peter Mackridge: <i>Some literary representations of spoken Greek before nationalism(1750-1801)</i> .....	17
Μαρία Σηφιανού: <i>Η έννοια της ευγένειας στα Ελληνικά</i> .....	45
Σπυριδούλα Βαρλοκώστα: <i>Syntactic comprehension in aphasia and its relationship to working memory deficits</i> .....	75
Ευαγγελία Αχλάδη, Αγγελική Δούρη, Ευγενία Μαλικούτη & Χρυσάνθη Παρασχάκη-Μπαράν: <i>Γλωσσικά λάθη τουρκόφωνων μαθητών της Ελληνικής ως ξένης/δεύτερης γλώσσας: Ανάλυση και διδακτική αξιοποίηση</i> .....	109
Κατερίνα Αλεξανδρή: <i>Η μορφή και η σημασία της διαβάθμισης στα επίθετα που δηλώνουν χρώμα</i> .....	125
Eva Anastasi, Ageliki Logotheti, Stavri Panayiotou, Marilena Serafim & Charalambos Themistocleous: <i>A Study of Standard Modern Greek and Cypriot Greek Stop Consonants: Preliminary Findings</i> .....	141
Anna Anastassiadis-Symeonidis, Elisavet Kiourti & Maria Mitsiaki: <i>Inflectional Morphology at the service of Lexicography: ΚΟΜΟΛεξ, A Cypriot Morphological Dictionary</i> .....	157

Γεωργία Ανδρέου & Ματίνα Τασιούδη: <i>Η ανάπτυξη του λεξιλογίου σε παιδιά με Σύνδρομο Απνοιών στον Ύπνο</i> .....	175
Ανθούλα- Ελευθερία Ανδρεσάκη: <i>Γατρικές μεταφορές στον δημοσιογραφικό λόγο της κρίσης: Η οπτική γωνία των Γερμανών</i> .....	187
Μαρία Ανδριά: <i>Προσεγγίζοντας θέματα Διαγλωσσικής Επίδρασης μέσα από το πλαίσιο της Γνωσιακής Γλωσσολογίας: ένα παράδειγμα από την κατάκτηση της Ελληνικής ως L2</i> .....	199
Spyros Armotistis & Kakia Petinou: <i>Mastering word-initial syllable onsets by Cypriot Greek toddlers with and without early language delay</i> .....	215
Julia Bacskai-Atkari: <i>Ambiguity and the Internal Structure of Comparative Complements in Greek</i> .....	231
Costas Canakis: <i>Talking about same-sex parenthood in contemporary Greece: Dynamic categorization and indexicality</i> .....	243
Michael Chiou: <i>The pragmatics of future tense in Greek</i> .....	257
Maria Chondrogianni: <i>The Pragmatics of the Modern Greek Segmental Markers</i> .....	269
Katerina Christopoulou, George J. Xydopoulos & Anastasios Tsangalidis: <i>Grammatical gender and offensiveness in Modern Greek slang vocabulary</i> .....	291
Aggeliki Fotopoulou, Vasiliki Foufi, Tita Kyriacopoulou & Claude Martineau: <i>Extraction of complex text segments in Modern Greek</i> .....	307
Αγγελική Φωτοπούλου & Βούλα Γιούλη: <i>Από την «Έκφραση» στο «Πολύτροπο»: σχεδιασμός και οργάνωση ενός εννοιολογικού λεξικού</i> .....	327
Marianthi Georgalidou, Sofia Lampropoulou, Maria Gasouka, Apostolos Kostas & Xanthippi Foulidi: <i>“Learn grammar”: Sexist language and ideology in a corpus of Greek Public Documents</i> .....	341
Maria Giagkou, Giorgos Fragkakis, Dimitris Pappas & Harris Papageorgiou: <i>Feature extraction and analysis in Greek L2 texts in view of automatic labeling for proficiency levels</i> .....	357

Dionysis Goutsos, Georgia Fragaki, Irene Florou, Vasiliki Kakousi & Paraskevi Savvidou: <i>The Diachronic Corpus of Greek of the 20th century: Design and compilation</i> .....	369
Kleanthes K. Grohmann & Maria Kambanaros: <i>Bilectalism, Comparative Bilingualism, and the Gradience of Multilingualism: A View from Cyprus</i> .....	383
Günther S. Henrich: „Γεωγραφία νεωτερική“ στο Λίβιστρος και Ροδάμνη: μετατόπιση ονομάτων βαλτικών χωρών προς την Ανατολή;.....	397
Noriyo Hoozawa-Arkenau & Christos Karvounis: <i>Vergleichende Diglossie - Aspekte im Japanischen und Neugriechischen: Veritäten - Interferenz</i> .....	405
Μαρία Ιακώβου, Ηριάννα Βασιλειάδη-Λιναρδάκη, Φλώρα Βλάχου, Όλγα Δήμα, Μαρία Καββαδία, Τατιάνα Κατσίνα, Μαρίνα Κουτσομπού, Σοφία-Νεφέλη Κύτρου, Χριστίνα Κωστάκου, Φρόσω Παππά & Σταυριαλένα Περρέα: <i>ΣΕΠΙΜΕ2: Μια καινούρια πηγή αναφοράς για την Ελληνική ως Γ2</i> .....	419
Μαρία Ιακώβου & Θωμαΐς Ρουσουλιώτη: <i>Βασικές αρχές σχεδιασμού και ανάπτυξης του νέου μοντέλου αναλυτικών προγραμμάτων για τη διδασκαλία της Ελληνικής ως δεύτερης/ξένης γλώσσας</i> .....	433
Μαρία Καμηλάκη: «Μαζί μου ασχολείσαι, πόσο μαλάκας είσαι!»: Λέξεις-ταμπού και κοινωνιογλωσσικές ταυτότητες στο σύγχρονο ελληνόφωνο τραγούδι.....	449
Μαρία Καμηλάκη, Γεωργία Κατσούδα & Μαρία Βραχιονίδου: <i>Η εννοιολογική μεταφορά σε λέξεις-ταμπού της ΝΕΚ και των νεοελληνικών διαλέκτων</i> .....	465
Eleni Karantzola, Georgios Mikros & Anastassios Papaioannou: <i>Lexico-grammatical variation and stylometric profile of autograph texts in Early Modern Greek</i> .....	479
Sviatlana Karpava, Maria Kambanaros & Kleanthes K. Grohmann: <i>Narrative Abilities: MAINing Russian-Greek Bilingual Children in Cyprus</i> .....	493
Χρήστος Καρβούνης: <i>Γλωσσικός εξαρχαϊσμός και «ιδεολογική» νόρμα: Ζητήματα γλωσσικής διαχείρισης στη νέα ελληνική</i> .....	507

Demetra Katis & Kiki Nikiforidou: <i>Spatial prepositions in early child Greek: Implications for acquisition, polysemy and historical change</i> .....	525
Γεωργία Κατσούδα: <i>Το επίθημα -ούνα στη ΝΕΚ και στις νεοελληνικές διαλέκτους και ιδιώματα</i> .....	539
George Kotzoglou: <i>Sub-extraction from subjects in Greek: Its existence, its locus and an open issue</i> .....	555
Veranna Kyprioti: <i>Narrative, identity and age: the case of the bilingual in Greek and Turkish Muslim community of Rhodes, Greece</i> .....	571
Χριστίνα Λύκου: <i>Η Ελλάδα στην Ευρώπη της κρίσης: Αναπαραστάσεις στον ελληνικό δημοσιογραφικό λόγο</i> .....	583
Nikos Liosis: <i>Systems in disruption: Propontis Tsakonian</i> .....	599
Katerina Magdou, Sam Featherston: <i>Resumptive Pronouns can be more acceptable than gaps: Experimental evidence from Greek</i> .....	613
Maria Margarita Makri: <i>Opos identity comparatives in Greek: an experimental investigation</i> .....	629
2ος Τόμος	
Περιεχόμενα .....	651
Vasiliki Makri: <i>Gender assignment to Romance loans in Katoitaliótika: a case study of contact morphology</i> .....	659
Evgenia Malikouti: <i>Usage Labels of Turkish Loanwords in three Modern Greek Dictionaries</i> .....	675
Persephone Mamoukari & Penelope Kambakis-Vougiouklis: <i>Frequency and Effectiveness of Strategy Use in SILL questionnaire using an Innovative Electronic Application</i> .....	693

Georgia Maniati, Voula Gotsoulia & Stella Markantonatou: <i>Contrasting the Conceptual Lexicon of ILSP (CL-ILSP) with major lexicographic examples</i> .....	709
Γεώργιος Μαρκόπουλος & Αθανάσιος Καρασίμος: <i>Πολυεπίπεδη επισημείωση του Ελληνικού Σώματος Κειμένων Αφασικού Λόγου</i> .....	725
Πωλίνα Μεσηνιώτη, Κατερίνα Πούλιου & Χριστόφορος Σουγανίδης: <i>Μορφοσυντακτικά λάθη μαθητών Τάξεων Υποδοχής που διδάσκονται την Ελληνική ως Γ2</i> .....	741
Stamatia Michalopoulou: <i>Third Language Acquisition. The Pro-Drop-Parameter in the Interlanguage of Greek students of German</i> .....	759
Vicky Nanousi & Arhonto Terzi: <i>Non-canonical sentences in agrammatism: the case of Greek passives</i> .....	773
Καλομοίρα Νικολού, Μαρία Ξεφτέρη & Νίτσα Παραχεράκη: <i>Το φαινόμενο της σύνθεσης λέξεων στην κυκλαδοκρητική διαλεκτική ομάδα</i> .....	789
Ελένη Παπαδάμου & Δώρας Κ. Κυριαζής: <i>Μορφές διαβαθμιστικής αναδίπλωσης στην ελληνική και στις άλλες βαλκανικές γλώσσες</i> .....	807
Γεράσιμος Σοφοκλής Παπαδόπουλος: <i>Το δίπολο «Εμείς και οι Άλλοι» σε σχόλια αναγνωστών της Lifo σχετικά με τη Χρυσή Αυγή</i> .....	823
Ελένη Παπαδοπούλου: <i>Η συνδυαστικότητα υποκοριστικών επιθημάτων με β' συνθετικό το επίθημα -άκι στον διαλεκτικό λόγο</i> .....	839
Στέλιος Πιπερίδης, Πένυ Λαμπροπούλου & Μαρία Γαβριηλίδου: <i>clarin:el. Υποδομή τεκμηρίωσης, διαμοιρασμού και επεξεργασίας γλωσσικών δεδομένων</i> .....	851
Maria Pontiki: <i>Opinion Mining and Target Extraction in Greek Review Texts</i> .....	871
Anna Roussou: <i>The duality of mīpos</i> .....	885

Stathis Selimis & Demetra Katis: <i>Reference to static space in Greek: A cross-linguistic and developmental perspective of poster descriptions</i> .....	897
Evi Sifaki & George Tsoulas: <i>XP-V orders in Greek</i> .....	911
Konstantinos Sipitanos: <i>On desiderative constructions in Naousa dialect</i> .....	923
Eleni Staraki: <i>Future in Greek: A Degree Expression</i> .....	935
Χριστίνα Τακούδα & Ευανθία Παπαευθυμίου: <i>Συγκριτικές διδακτικές πρακτικές στη διδασκαλία της ελληνικής ως Γ2: από την κριτική παρατήρηση στην αναπλαισίωση</i> .....	945
Alexandros Tantos, Giorgos Chatzioannidis, Katerina Lykou, Meropi Papatheohari, Antonia Samara & Kostas Vlachos: <i>Corpus C58 and the interface between intra- and inter-sentential linguistic information</i> ....	961
Arhonto Terzi & Vina Tsakali: <i>The contribution of Greek SE in the development of locatives</i> .....	977
Paraskevi Thomou: <i>Conceptual and lexical aspects influencing metaphor realization in Modern Greek</i> .....	993
Nina Topintzi & Stuart Davis: <i>Features and Asymmetries of Edge Geminates</i> .....	1007
Liana Tronci: <i>At the lexicon-syntax interface Ancient Greek constructions with ἔχειν and psychological nouns</i> .....	1021
Βίλλυ Τσάκωνα: <i>«Δημοκρατία είναι 4 λύκοι και 1 πρόβατο να ψηφίζουν για φαγητό»:Αναλύοντας τα ανέκδοτα για τους/τις πολιτικούς στην οικονομική κρίση</i> .....	1035
Ειρήνη Τσαμαδοῦ- Jacobberger & Μαρία Ζέρβα: <i>Εκμάθηση ελληνικών στο Πανεπιστήμιο Στρασβούργου: κίνητρα και αναπαραστάσεις</i> ...	1051
Stavroula Tsiplakou & Spyros Armostis: <i>Do dialect variants (mis)behave? Evidence from the Cypriot Greek koine</i> .....	1065
Αγγελική Τσόκογλου & Σύλα Κλειδή: <i>Συζητώντας τις δομές σε -οντας</i> .....	1077



Αλεξιάννα Τσότσου:	
<i>Η μεθοδολογική προσέγγιση της εικόνας της Γερμανίας στις ελληνικές εφημερίδες .....</i>	1095
Anastasia Tzilinis:	
<i>Begründendes Handeln im neugriechischen Wissenschaftlichen Artikel: Die Situierung des eigenen Beitrags im Forschungszusammenhang.....</i>	1109
Κυριακούλα Τζωρτζάτου, Αργύρης Αρχάκης, Άννα Ιορδανίδου & Γιώργος Ι. Ευδόπουλος:	
<i>Στάσεις απέναντι στην ορθογραφία της Κοινής Νέας Ελληνικής: Ζητήματα ερευνητικού σχεδιασμού .....</i>	1123
Nicole Vassalou, Dimitris Papazachariou & Mark Janse:	
<i>The Vowel System of Mišótika Cappadocian .....</i>	1139
Marina Vassiliou, Angelos Georganas, Prokopis Prokopidis & Haris Papageorgiou:	
<i>Co-referring or not co-referring? Answer the question!.....</i>	1155
Jeroen Vis:	
<i>The acquisition of Ancient Greek vocabulary.....</i>	1171
Christos Vlachos:	
<i>Mod(aliti)es of lifting wh-questions.....</i>	1187
Ευαγγελία Βλάχου & Κατερίνα Φραντζή:	
<i>Μελέτη της χρήσης των ποσοδεικτών λίγο-λιγάκι σε κείμενα πολιτικού λόγου .....</i>	1201
Madeleine Voga:	
<i>Τι μας διδάσκουν τα ρήματα της ΝΕ σχετικά με την επεξεργασία της μορφολογίας.....</i>	1213
Werner Voigt:	
<i>«Σεληνάκι μου λαμπρό, φέγγε μου να περπατώ ...» oder: warum es in dem bekannten Lied nicht so, sondern eben φεγγαράκι heißt und ngr. φεγγάρι .....</i>	1227
Μαρία Βραχιονίδου:	
<i>Υποκοριστικά επιρρήματα σε νεοελληνικές διαλέκτους και ιδιώματα .....</i>	1241
Jeroen van de Weijer & Marina Tzakosta:	
<i>The Status of *Complex in Greek.....</i>	1259
Theodoros Xioufis:	
<i>The pattern of the metaphor within metonymy in the figurative language of romantic love in modern Greek.....</i>	1275

# LEXICO-GRAMMATICAL VARIATION AND STYLOMETRIC PROFILE OF AUTOGRAPH TEXTS IN EARLY MODERN GREEK

Eleni Karantzola<sup>1</sup>, Georgios Mikros<sup>2</sup> & Anastassios Papaioannou<sup>1</sup>

<sup>1</sup>University of the Aegean, <sup>2</sup>University of Athens  
karantzola@rhodes.aegean.gr, gmikros@isll.uoa.gr,  
papaioannou@rhodes.aegean.gr

## Περίληψη

Στο άρθρο διερευνάται το υφομετρικό προφίλ κειμένων της πρώιμης νέας ελληνικής (16ος-17ος αι.) από τη Ζάκυνθο, που παραδίδονται σε αυτόγραφα χειρόγραφα, τα οποία διασώζουν το σύνολο των γλωσσικών επιλογών των συντακτών τους χωρίς την παρεμβολή αντιγραφών. Στο σώμα κειμένων της έρευνας εφαρμόζεται η Ανάλυση Κυρίων Συνιστωσών. Η μέτρηση της συχνότητας εμφάνισης των 1.000 συχνότερων λέξεων αναδεικνύει τη θεματική και κειμενική συγγένεια των κειμένων, καθώς και τη σύνδεσή τους με τον συγγραφέα τους. Σε διαφορετικές συσχετίσεις οδηγεί η σύγκριση των κειμένων βάσει επιλεγμένων γλωσσικών χαρακτηριστικών που εμφανίζονται στα κείμενα σε συμπληρωματική κατανομή.

*Keywords: Early Modern Greek, 16<sup>th</sup> and 17<sup>th</sup> c., Zante, vernacular, autograph manuscripts, stylometric profile, computational stylometry*

## 1. Introduction

It is well known and not surprising that there are considerable linguistic differences among manuscripts of more or less vernacular Greek texts in Byzantine and Early Modern times. In this respect, the choice to work with *autograph* manuscripts of major or minor scholars of this period (e.g. Papaioannou 2016) is extremely important,

concerning the reconstruction of written and spoken registers used during the 16<sup>th</sup> and 17<sup>th</sup> centuries.

Given the linguistic diversity of surviving manuscripts of a text, this paper addresses the specific question of the margins/extent of an author's language differentiation in his own autographs of the same text, in relation to the context of text's production. Unfortunately, a very limited number of autographs have been identified and are accessible (i.e. published texts). The mostly referred case of Manouil Malaxos, whose paraphrased version in vernacular (1562/1563) of his unpublished work "Nomokanon" has been recorded in at least eleven autographs, should be considered as a rather exceptional case.

In our study we will explore a very interesting and quite unusual case of autograph manuscripts from the Ionian islands region, and more specifically from Zante. The half of them are related to a minor scholar, Anthimos Skourtas (beginning of 17<sup>th</sup> c.), copier and translator of hagiographical works, while the rest belong to other minor scholars from Zante, namely Theodoros Raftopoulos (first half of 16<sup>th</sup> c.), and Efthimios Tzimpletis and Dionyssios Mountenos (first half of 17<sup>th</sup> c.).

The quantitative methods used in the analysis of our corpus can provide significant insights in a number of interesting research questions, such as how "deep" are the proposed lexico-grammatical changes observed in the examined texts in terms of stylistic integrity. Among a variety of stylometric textual features (Mikros 2016), we will investigate words frequency, as well as specific (dialectal) linguistic features of the texts, in order to compare the stylometric profile of all the texts and explore the impact of the various lexico-grammatical alternations in the distribution of important stylometric parameters in our corpus. More specifically, we are interested in seeing whether texts of different content from the same author and copies of the very same text copied by different persons can be differentiated using stylometric quantitative profiles.

## **2. Corpus and methodology**

### *2.1. Corpus*

The corpus that we analyze consists primarily of four texts associated with the name of Anthimos Skourtas. The two of them (skourtas\_1\_1 and skourtas\_1\_2) are considered as autographs, although the differences in handwriting are not negligible – com-

mon practice during the byzantine and postbyzantine period. Concerning the text, it is an unpublished translation in “koine” –by Skourtas– of the Life of St Jean. The first autograph (skourtas\_1\_1) was written in Strophadon Monastery in Zante (ms 3, 1617, 215<sup>v</sup>-225<sup>r</sup>), where Skourtas spent the biggest part of his monasterial life. The second presumed autograph manuscript with the same text resides in the Greek National Library (ms 2328, 17<sup>th</sup> c., 213<sup>r</sup>-230<sup>v</sup>); it is assumed to be written in Kefallinia island, some years later, and compared with the first, it contains a large number of language differentiations in all linguistic levels.

Two more manuscripts, also from Strophadon Monastery, complete Skourtas’ corpus: (i) a copy of Skourtas’ translation of the Life of St John, by a less literate monk (skourtas\_1\_3) (ms 10, 1<sup>r</sup>-36<sup>v</sup>), and (ii) a second unpublished translation made by Skourtas, of the Life of St Spyridon –an autograph text as well–, some pages before the translation of the Life of St Jean (ms 3, 163<sup>r</sup>-211<sup>r</sup>). Table 1 illustrates the profile of the four texts of Skourtas’ corpus:

Text	Manuscript	Work	Aut. vs. Copy	Words
skourtas_1_1	ms 3 (Strophades, Zante)	Life of St Jean	autograph	6,436
skourtas_1_2	ms 2328 (National Library, Athens)	Life of St Jean	autograph (?)	6,208
skourtas_1_3	ms 10 (Strophades, Zante)	Life of St Jean	copy	6,409
skourtas_2	ms 3 (Strophades, Zante)	Life of St Spyridon	autograph	15,583

Table 1 | Profile of Skourtas’ texts

In a comparative perspective, four more autograph texts from the same place and from the same period, were included in the corpus (Table 2): (i) an unpublished translation/version of the hagiographical novel “Varlaam and Ioassaf” (ms 14, 1676, 17-148), elaborated by Dionyssios Mountenos, a monk known for his code copying activity in the Strophadon monastery in the second half of the 17<sup>th</sup> c., (ii) a chronographic text (Marc. Gr. II, 1513, 110<sup>r</sup>) of the notary Theodoros Raftopoulos concerning the earthquakes that happened in Zante (1513-1521), and two texts of Efthymios Tzimpletis, descendant of an illustrious family of Zante, which are included in the ms 99 of Kolyva

collection (Athens, first half of 17<sup>th</sup> c.), i.e. (iii) a sermon (1-111), and (iv) a translation of the Life of St Helen (196<sup>r</sup>-200<sup>v</sup>).

<b>Text</b>	<b>Manuscript</b>	<b>Work</b>
mountenos	ms 14 (Strophades, Zante)	Varlaam and Ioassaf
raftopoulos	Marc. Gr. II (Venice)	chronographic text
tzimpletis_1	ms 99 (Kolyva collection, Athens)	sermon
tzimpletis_2	ms 99 (Kolyva collection, Athens)	Life of St Helen

Table 2 | Other autograph texts from Zante

## 2.2. Methodology

In our research, computational stylistic methods will be used, in order to tackle the research hypotheses stated in section 1. Computational stylistics is a rapid-changing interdisciplinary field studying the quantitative aspects of texts and their correlation with author’s characteristics (e.g identity, gender, age, personality etc.). The basic stylometric assumption is that each writer possesses an idiosyncratic way of using his/her linguistic competence, a “human stylome” as has been referred by Van Halteren et al. (2005), and this can be traced using quantitative methods. Research in the last two decades have extended the use of stylometric methodology to various other fields of scientific inquiry including education (readability analysis of educational material (Pitler & Nenkova 2008)), neurocognitive pathology (stylometric evidence of neurocognitive pathology in the written linguistic production of patients (van Velzen, Nanetti & de Deyn 2014)) and literary history (e.g. clustering techniques for literary genre analysis, of topic modelling, intertextual analysis, and of other stylometric or computational approaches to explore literary texts (Schöch & Jannidis 2013)).

In this paper, we specifically employ the Delta method (Burrows 2002), one of the most robust distance measures for authorship attribution (Juola 2008; Stamatatos 2009). It has been shown to render very useful results in different text genres (Hoover 2004) and languages (Eder & Rybicki 2013). Delta is based on the frequency of the

$n$  most frequent words of a corpus. These frequencies are normalized and then converted to  $z$  scores, i.e. the extent to which the normalized word frequencies within a particular text are above or below average for the set of texts as a whole. In a later step the table of the  $z$  scores is converted into a table of “distances” between documents in which high values show that a document pair is dissimilar to each other, and low values show that a document pair is highly similar.

This table of “distances” can be further analyzed using various multivariate statistical and visualization methods. We employed Principal Component Analysis (PCA), a robust multivariate method which can restructure the table of text “distances” and reduce its dimensions in only 2 principal components or factors. These new variables/dimensions are linear combinations of the original ones and are uncorrelated with one another. They are also orthogonal to the original dimension space and capture as much of the original variance in the data as possible.

The first principal component is the one that captures the greatest variability (covariance) in our data and the second is the next orthogonal (uncorrelated) direction of greatest variability. All graphs produced using Delta as distance measure and PCA as statistical analysis can be interpreted using the Euclidean distance semantics, i.e. the further two texts the more stylistically dissimilar they are and vice versa. Moreover, texts grouped together under  $x$  axis (first principal component) or  $y$  axis (second principal component) share stylistic similarities under a more general stylistic dimension.

### **3. Stylometric analysis**

#### *3.1. Frequent words*

We first measured the frequency of occurrence of the 1,000 most frequent words in the Skourtas corpus (whole texts). Using the resulting term ~ document table we calculated a PCA (based on correlation matrix) and we plotted the first 2 dimensions. The PCA solution was satisfactory since both dimensions were able to explain nearly 90% of the total data variation (the first PC (PC1) explains the 69.5% of the data variation while the second PC (PC2) explains the 20.3%). Moreover, culling has been set to 50%, meaning that words that appear at least in the half of the texts of the corpus were considered in the analysis.

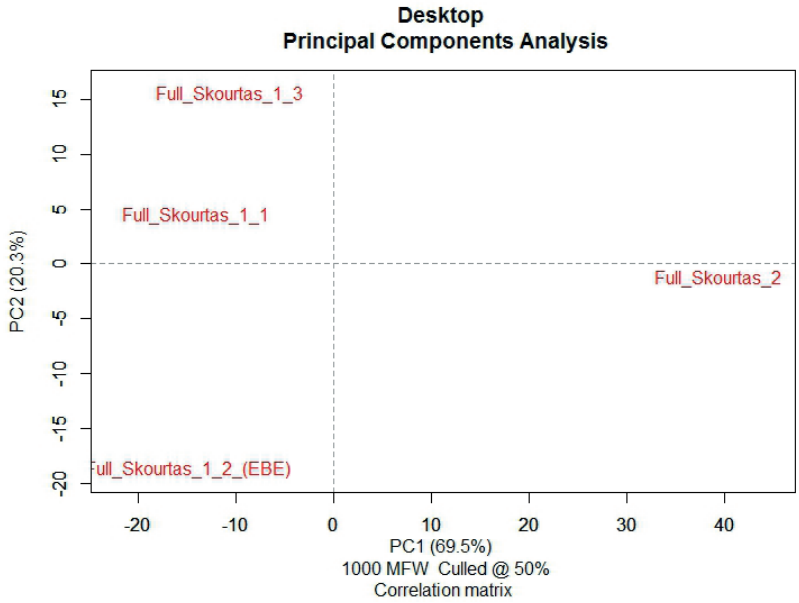


Figure 1 | Stylometric position of Skourtas' (full) texts

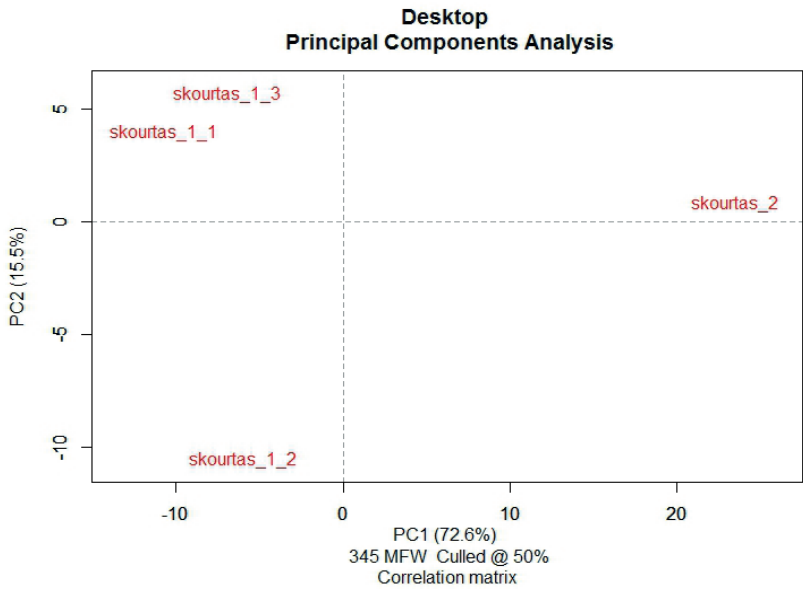


Figure 2 | Stylometric position of Skourtas' texts (samples)

In Figure 1, we can see the relative position of the four Skourtas' texts. The translation of "The Life of St. Spyridon" (Full\_Skourtas\_2) is located far right in the graph, clearly indicating that this text is completely different in terms of vocabulary usage. As far as the other three texts are concerned, we can see the following: The copy of the translation of the Life of St. John (Full\_Skourtas\_1\_3) is stylistically closer to the Zante version of the text than the alleged autograph version of the same text held by the National Library (Full\_Skourtas\_1\_2\_(EBE)).

In order to expand the scope of our research, we enlarged our corpus with four more autograph texts from Zante, described in section 2.1. Since our new texts are samples from bigger documents, we first had to check whether the PCA solution offered previously for the Skourtas' texts holds valid also using sample and not whole texts. Thus, we replicated the analysis using the same parameters and the results confirmed fully our initial conclusions, as Figure 2 suggests.

After establishing the robustness of our analysis in smaller text samples, we proceeded to the stylometric investigation of the new four texts from Zante (tzimpletis\_1, tzimpletis\_2, mountenos and raftopoulos) compared to Skourtas' texts. The exact length of the eight text samples is indicated in Table 3:

	<b>Text</b>	<b>Words</b>
	mountenos	460
	raftopoulos	662
	skourtas_1_1	689
	skourtas_1_2	691
	skourtas_1_3	700
	skourtas_2	507
	tzimpletis_1	561
	tzimpletis_2	607

*Table 3 | Length of samples*

We performed again a PCA using the 1,000 most frequent words (57 after 50% culling) based on the correlation matrix of the term~document table. The results of the PCA performed are displayed in Figure 3.



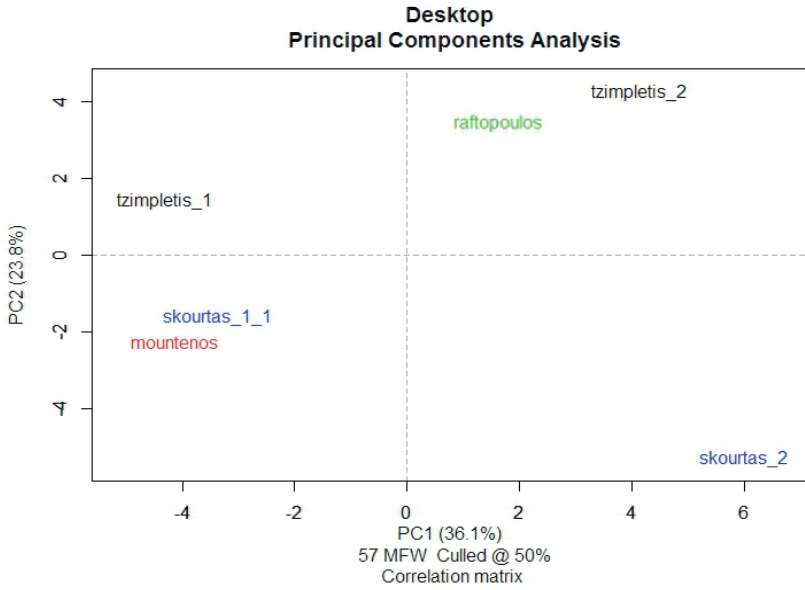


Figure 3 | Stylometric position of texts from Zante

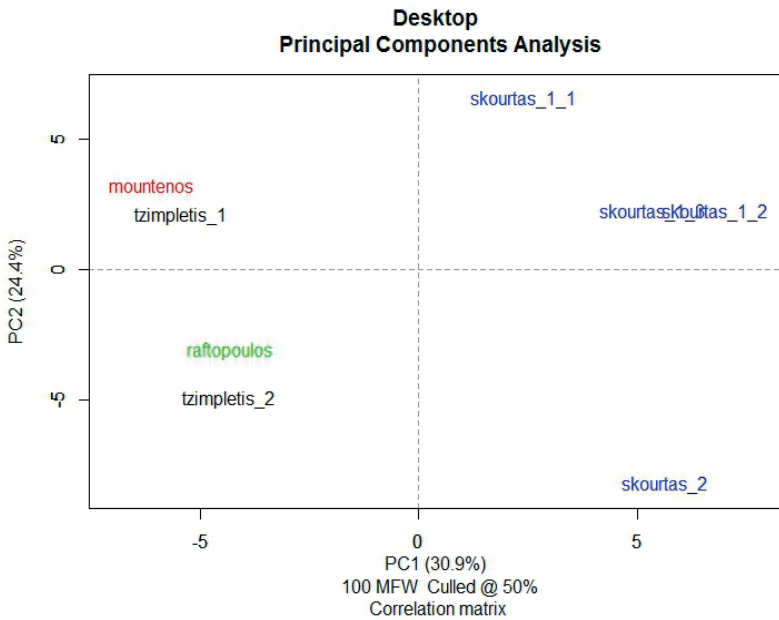


Figure 4 | Stylometric position of Skourtas' texts vs. other texts from Zante

Figure 3 clearly indicates that PCA using the most frequent words can reliably discriminate the authorial trace of the four authors in question. The horizontal dimension distinguishes further Skourtas' from Tzimpletis' texts while the vertical dimension splits different texts from the same authors (tzimpletis\_1 vs. tzimpletis\_2 and skourtas\_1\_1 vs. skourtas\_2).

A natural extension to our methodology was to test our stylometric method using now all text samples from Skourtas and compare them with the other authorgraphs from Zante. We performed PCA using the same parameters with the previous analyses (1,000 most frequent words culled at 50%). The results are displayed in Figure 4.

As we can see, the PCA analysis correctly puts all versions of Skourtas' texts at the right half of the plot, while it continues to differentiate Tzimpletis' texts in the horizontal dimension.

### 3.2. Selected linguistic features

Our approach up to this point consisted in using the most frequent words, in order to uncover underlying stylistic differentiations between the texts of our corpus and exploit them for comparison purposes along various (socio – stylo) linguistic dimensions such as authorship or genre. As already mentioned, the two autographs of Skourtas (skourtas\_1\_1 and skourtas\_1\_2) display a wide number of lexico-grammatical differences. For the purposes of the present research, we limit ourselves to point out four of them, which appear often enough and consequently affect the whole image of the texts:

- The ending of past tense *-τουv*<sup>1</sup> vs. *-τον*, e.g. *ήτουv/ήτον, εγράφουv/εγράφετον*
- The use of gerund vs. participle, e.g. *δίδονταc/δίδοντεc, σκοπόνταc/σκοπόντεc, ανοίγονταc/ανοίγουσα* etc.
- The spelling *-ο-* vs. *-ω-* of the gerunds, e.g. *δίδονταc/δίδωνταc*
- The form *έκαμνε* vs. *έκανε*

---

<sup>1</sup> The endings *-τουv* are characteristic of the dialects of the Ionian islands (cf. Karantzola & Papaioannou 2014).

In order to study the variation of the above mentioned feature in Skourtas' whole texts, we calculated their raw frequencies. Figure 5 displays their distribution of the selected features across Skourtas' texts.

It is evident that the National Library version (skourtas\_1\_2), the only one not written in Zante, displays a highly contrastive frequency pattern compared to the other three versions. This is illustrated even better if we calculate the normalized<sup>2</sup> frequencies of the features (see Figure 6).

Following to this, PCA on the basis of the abovementioned selected features resulted in a new stylometric position of Skourtas' texts. Although our initial stylometric analysis based on the most frequent words differentiated the four versions on the dimension of topic (Life of St. John vs. Life of St. Spyridon), our new analysis based on the selected linguistic features produced a different mapping compared to Figures 1 and 2:

In fact, the second presumed autograph of Skourtas' translation of the Life of St John (skourtas\_1\_2) stays apart, while the three other texts, his autograph (skourtas\_1\_1), the copy of it produced in Strophadon monastery (skourtas\_1\_3) and the autograph of Skourtas' translation of the Life of St Spyridon (skourtas\_2) look closer to each other.

#### 4. Conclusions

This paper examined the stylometric profile of texts written in (more or less vernacular) Early Modern Greek during the 16<sup>th</sup> and 17<sup>th</sup> centuries. Our corpus, mainly being consisted of autographs of five scholars from Zante –among which two autographs of the same text (skourtas\_1\_1 and skourtas\_1\_2) together with a copy of it (skourtas\_1\_3)–, allowed us to investigate whether texts of different content from the same author and copies of the very same text copied by different persons can be differentiated using stylometric quantitative profiles.

At first, the measurement of the frequency of occurrence of the 1,000 most frequent words (after culling) in Skourtas' corpus, clearly demonstrated that the main divide is in terms of text vocabulary (and theme). Since skourtas\_2 was the only sample from a completely different text (translation of the Life of St Spyridon), the result (Figure 1)

---

2 Normalized frequency take into account the different text sizes and adjust the features' frequencies so that can be compared across texts with unequal length. Since we had two antagonistic feature types we used a percentage ratio transformation i.e. Normalized Frequency of Feature i = (raw frequency of type 1 of feature i / raw frequency of type 2 of feature i) X 100.

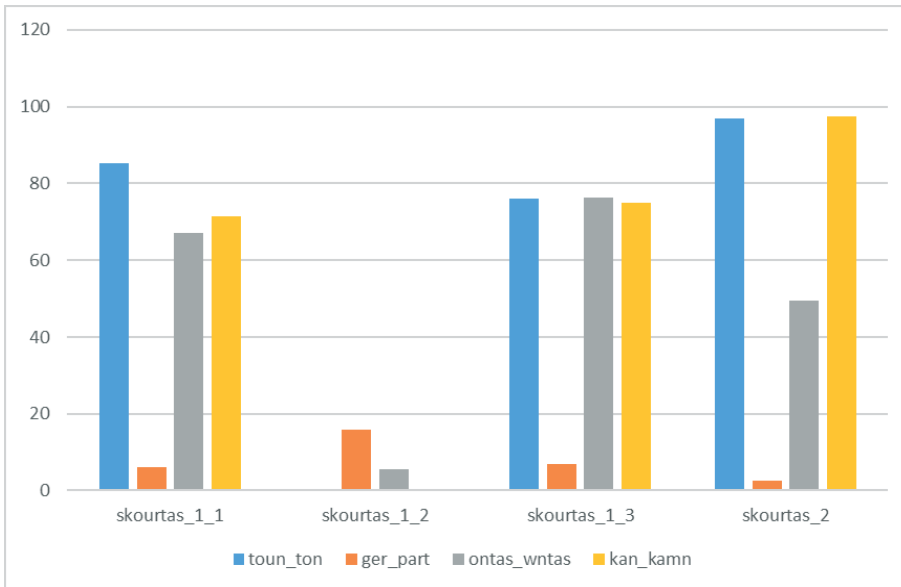


Figure 5 | Distribution of selected features across Skourtas' texts

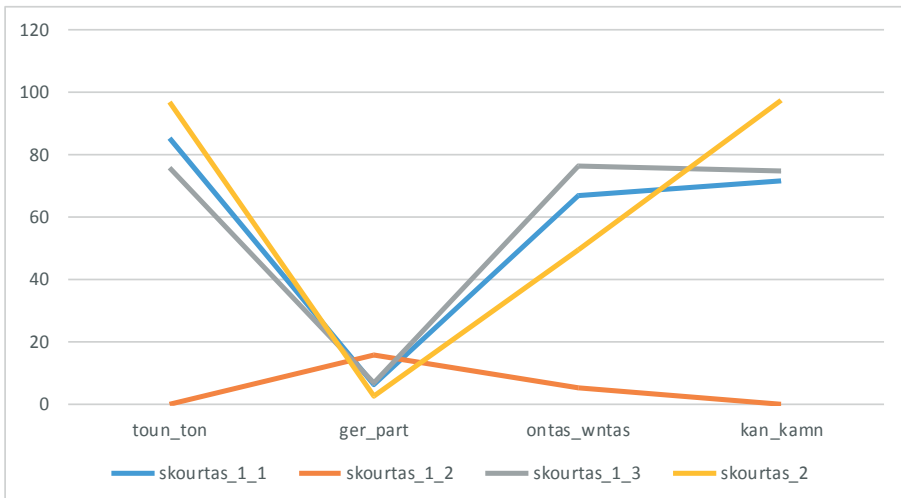


Figure 6 | Comparison of Skourtas' texts on the basis of selected features (normalized frequencies)

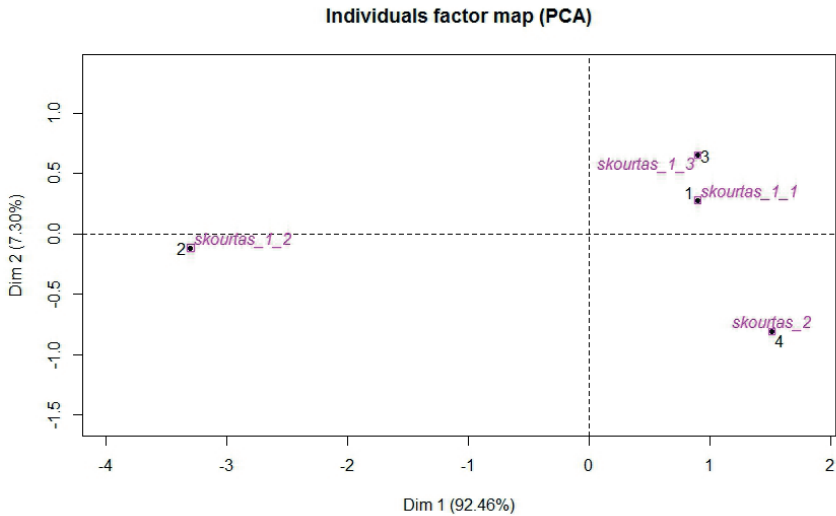


Figure 7 | Stylometric position II of Skourtas' (full) texts

is not surprising. The PCA also correctly put (Figure 3 and Figure 4) all the texts of an author, i.e. Skourtas or Tzimpletis, at the same half of the plot, which means that authorship is stylistically recognizable.

As far as our specific research question is concerned, the results seem not to be conclusive. According to the criterion of 1,000 most frequent words (Figure 1), as was applied in Skourtas' texts, the copy of a text (by a different person) is more similar to the original than a second autograph, of the same text, by the author. The stylometric analysis based on selected linguistic features of the texts indicates further (Figure 7) that in the case of Skourtas' autographs we have to assume either that the production of his second (presumed) autograph of his translation of the Life of St John is a completely "new" version, probably in the sense of *de-dialectisation* attested in some authors of that period like Ioannis Morezinos (Kakoulidi-Panou, Karantzola & Halvatzidaki 2007, π'), or that at the end the second manuscript is *not* an autograph of him.

It is certain that a thorough investigation of this topic is required; this work could be considered as a small contribution towards this direction.

## References

- Burrows, John F. 2002. “Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship.” *Literary and Linguistic Computing* 17 (3):267–87.
- Eder, Maciej, and Jan Rybicki. 2013. “Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution.” *Literary and Linguistic Computing* 28 (2):229–36. doi: 10.1093/llc/fqs036
- Hoover, David L. 2004. “Testing Burrow’s Delta.” *Literary and Linguistic Computing* 19 (4):453–75.
- Juola, Patrick. 2008. “Authorship Attribution.” *Foundations and Trends in Information Retrieval* 1 (3):233–334. doi: 10.1561/1500000005
- Kakoulidi-Panou, Eleni, Eleni Karantzola, and Maria Halvatzidaki, eds. 2007. *Ioannis Morezinos. Klini Solomontos. Istories thavmaton tis Panagias (1599)*. Irakleio: Etairia Kritikon Istorikon Meleton.
- Karantzola, Eleni, and Anastassios Papaioannou. 2014. “Sxeseis Eptanissiakon kai Kritikon Dialekton sta Anagennisiaka – Proima Neoellinika Keimena tou 17ou ai.: i Periptosi tis *Evgenas*.” In *Proceedings of the IX<sup>th</sup> Panionian Congress (Paxi island, 26-30/5/2010)*, edited by Aliko Nikiforou, B:127–47. Paxi: Etairia Paxinon Meleton.
- Karantzola, Eleni, and Nikolaos Lavidas. 2016. “Xarakteristika tis Kerkyraikis tou 16ou Aiona.” *Meletes gia tin Elliniki Glossa* 36:129–50.
- Mikros, George K. 2016. *Ypologistiki Yfometria*. URI: <http://hdl.handle.net/11419/4860>
- Papaioannou, Anastassios. 2016. “Peza Keimena Logion toy 16ou kai 17ou Aiona se Dimodi Elliniki Glossa, Vasei Aftografon Xeirografon Kodikon.” PhD diss., University of the Aegean.
- Pitler, Emily, and Ani Nenkova. 2008. “Revisiting Readability: a Unified Framework for Predicting Text Quality.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (Honolulu, Hawaii, 25-27/10/2008)*, 186–95. Association for Computational Linguistics.
- Schöch, Christof, and Fotis Jannidis. 2013. “Quantitative Text Analysis for Literary History.” Report on a DARIAH-DE Expert Workshop *DARIAH-DE Working Papers*. Göttingen: GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität.
- Stamatatos, Efstathios. 2009. “A survey of Modern Authorship Attribution Methods.” *Journal of the American Society for Information Science and*

*Technology* 60 (3):538–56. doi: 10.1002/asi.21001

van Velzen, Marjolein H., Luca Nanetti, and Peter Paul de Deyn. 2014. “Data Modeling in Corpus Linguistics: How Low May we Go?” *Cortex* 5:192–201. doi: <http://dx.doi.org/10.1016/j.cortex.2013.10.010>