



26 October 2015

Terttu Nevalainen
Professor of English Philology
Department of Modern Languages
P.O.Box 24 (Unioninkatu 40)
00014 University of Helsinki
Finland

To Whom It May Concern

Assessment of the DIACHRONIC CORPUS OF GREEK OF THE 20TH CENTURY

The Diachronic Corpus of Greek of the 20th Century is an important complement to the two existing major digital corpora that have been compiled for the study of the Greek language. Unlike these two present-day language resources, the new corpus compiled by Professor Goutsos and his team is unique in that it enables systematic diachronic research into the history of the Greek language over the last century.

The Diachronic Corpus of Greek of the 20th Century consists of over 3.3 million running words representing various text types that are made up of both spoken and written sources. The compilers' emphasis has been oriented towards speech and on a variety of speech-based, speech-purposed and speech-like genres, realized in text types such as public speeches, newsreels, film scripts and drama. The written sources sampled include fiction, news, academic writing, as well as popularized non-fiction. To the extent it has been possible, all genres have been sampled from the nine decades (1900–1980) decade by decade.

The corpus provides a valuable resource for tracing actual language use and diachronic changes in the Greek language in both spoken and written language domains. A number of questions have been raised by past research, e.g., regarding the realizations of societal diglossia, the distinctions between katharévoussa and demotic varieties of Greek, reflected in and cutting across genres. The new GreekCorpus20 allows empirical studies on a range of changing linguistic patterns over time and genre, and thus an informed assessment of the direction of language change as it is realized in speech and writing. It also enables a systematic approach to the standardization of Greek in the course of the 20th century – something that was not possible before this fine new resource became available.

Terttu Nevalainen

Terttu Nevalainen

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HUMANISTINEN TIEDEKUNTA
HUMANISTISKA FAKULTETEN
FACULTY OF ARTS

Nykykielten laitos, PL 24
(Unioninkatu 40)
00014 Helsingin yliopisto
<http://www.helsinki.fi/nykykielet/>

Department of Modern Languages, P.O.Box 24
(Unioninkatu 40)
FIN-00014 University of Helsinki, Finland
<http://www.helsinki.fi/modernlanguages/>



ΚΑΤΕΡΙΝΑ Θ. ΦΡΑΝΤΖΗ
ΑΝ. ΚΑΘΗΓΗΤΡΙΑ
ΤΜΗΜΑ ΜΕΣΟΓΕΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Ρόδος, 23/10/2015

Diachronic Corpus of the 20th Century (Greek Corpus 20)

Αξιολόγηση

Ο σχεδιασμός και η κατασκευή ενός Σώματος Κειμένων της γενικής γλώσσας, το οποίο δεν έχει φτιαχτεί για συγκεκριμένη εφαρμογή αλλά φιλοδοξεί να χρησιμοποιηθεί σε ποικιλία εφαρμογών, εμπλέκει πολλές παραμέτρους και κριτήρια απαραίτητα για την εξασφάλιση της πληρότητας και εγκυρότητας της εξαχθείσας πληροφορίας από τη χρήση του. Όταν ο σχεδιασμός και η κατασκευή αφορά σε Διαχρονικό Σώμα Κειμένων, οι παράμετροι που εμπλέκονται για τη σχεδίαση και κατασκευή του είναι σαφώς περισσότερες και κάνουν την αντιμετώπιση του θέματος πολυπλοκότερη: δυσκολία εύρεσης γλωσσικού υλικού, δυσκολία εξόρυξης του, όχι πλήρης κάλυψη όσο αφορά στα εργαλεία διαχείρισης της γλώσσας (π.χ. θέματα πολυτονικού και αυτόματης αναγνώρισης χαρακτήρων), όχι πάντα συνέχεια των κειμενικών ειδών στο χρόνο, θέματα πνευματικών δικαιωμάτων όσο αφορά στα δικαιώματα χρήσης του υλικού.

Το δυναμικό σήμερα *Diachronic Corpus of the 20th Century (Greek Corpus 20)* καλύπτει ένα ερευνητικό – και όχι μόνο κενό – με τη δυνατότητα εξυπηρέτησης ποικίλων ερευνητικών αναγκών που αφορούν στη μελέτη διαχρονικού γλωσσικού υλικού. Ήδη μετρά πάνω από 3,3 εκ. λέξεις ενώ το τελικό του μέγεθος θα ανέρχεται στις 20 εκ. λέξεις, 6 φορές μεγαλύτερο από το σημερινό του μέγεθος. Στο σημερινό του μέγεθος καλύπτει 9 είδη λόγου και πάνω από 15 κειμενικά είδη, τα οποία θα συμπληρωθούν με το μέγιστο δυνατό αριθμό και άλλων ειδών λόγου/κειμενικών ειδών. Η κατανομή του υλικού στο χρόνο δίνει τη δυνατότητα στις/στους ερευνήτριες/ητές για εύρος μελετών γλωσσολογικών θεμάτων διαχρονικής χρήσης της γλώσσας αλλά και θεμάτων της γλώσσας που αφορούν σε συγκεκριμένη χρονική περίοδο. Όπως και παρατηρείται από τις/τους συντελεστές σχεδίασης και κατασκευής του σώματος κειμένων, εκτός των κοινά αποδεκτών δυσκολιών όσο αφορά στη σχεδίαση και κατασκευή διαχρονικού σώματος κειμένων, οι δεκαετίες 1920, 1930 και 1950 παρουσίασαν ιδιαίτερες δυσκολίες που αφορούσαν στην εύρεση και συλλογή του υλικού σε σχέση με τις υπόλοιπες δεκαετίες, ακόμα και αυτές του 1900, 1910 και 1940.

Σημαντικό μέρος του *Diachronic Corpus of the 20th Century (Greek Corpus 20)* είναι η ύπαρξη υλικού πολύ κοντά στον καθαρά προφορικό λόγο. Ο προφορικός λόγος παρουσιάζει ιδιαίτερα προβλήματα στη συλλογή και επεξεργασία του αλλά είναι απαραίτητος σε ιδιαίτερα ενδιαφέρουσες εφαρμογές θεμάτων της Ανάλυσης Λόγου, της Κοινωνιογλωσσολογίας, της Γλωσσικής Αλλαγής κ.α. Προσωπικά γράμματα, δημόσιες



ομιλίες, θεατρικοί διάλογοι, διάλογοι από ταινίες και ειδήσεις προσφέρουν ποικιλία κειμενικών ειδών για έρευνα ενώ συνεντεύξεις είναι υπό επεξεργασία για την τελική ενσωμάτωσή τους στο σώμα κειμένων.

Όσο αφορά στο όλο πρόγραμμα για το *Diachronic Corpus of the 20th Century (Greek Corpus 20)*, το σώμα κειμένων συνοδεύεται από σχετικό ιστότοπο, βιβλιογραφία για τα διαχρονικά σώματα κειμένων, συλλογή ερευνητικών προγραμμάτων για τη σχεδίαση και κατασκευή διαχρονικών σωμάτων κειμένων και τα πρακτικά ημερίδας η οποία οργανώθηκε ειδικά για το θέμα αυτό. Ευρήματα σχετικά με την σχεδίαση και κατασκευή του αλλά και την εφαρμογή του έχουν ήδη δημοσιευτεί σε διεθνή συνέδρια (βλ. *Goutsos, D. & Fragaki, G. 2014. Recent language change in Greek: Design of the Diachronic Corpus of Greek of the 20th century. In Kotzoglou G. et al. (eds.) Selected Papers of the 11th International Conference on Greek Linguistics. Rhodes, University of the Aegean, 318-329* και *Fragaki, G. & Goutsos, D. 2014. Geek Diglossia in the 20th century: a historical corpus linguistics approach. Presented at the 12th International Conference on Greek Linguistics (ICGL12), Berlin*).

Το πρόγραμμα σχεδίασης και κατασκευής Διαχρονικού Σώματος Κειμένων της Ελληνικής για τον 20^ο αιώνα, *Diachronic Corpus of the 20th Century (Greek Corpus 20)*, είναι καταρχήν σημαντικό όσο αφορά το θέμα, το σκοπό του και τους στόχους του. Η ποικιλία των εφαρμογών θα το καταστήσουν σημαντικό εργαλείο στα χέρια των ερευνητριών/ών όχι μόνο της θεωρητικής και εφαρμοσμένης γλωσσολογίας αλλά και άλλων επιστημονικών πεδίων στα οποία η γλώσσα είναι αντικείμενο μελέτης (Κοινωνικές Επιστήμες, Πολιτικές Επιστήμες, Επιστήμες Μέσων Μαζικής Ενημέρωσης, Φιλολογία, κ.ά). Είναι σωστά σχεδιασμένο ώστε να περιέχει εύρος ειδών λόγου και κειμενικών ειδών. Είναι σωστά πλαισιωμένο με την απαραίτητη σχετική εργογραφία ενώ υπάρχει η απαραίτητη μέριμνα για τη διαδικτυακή παρουσίασή του ώστε να είναι εφικτή η χρήση του από τις ερευνήτριες/ές για τις/τους οποίες/ους θα αποτελέσει πολύ χρήσιμο εργαλείο.

ΚΑΤΕΡΙΝΑ Θ. ΦΡΑΝΤΖΗ

ΑΝ. ΚΑΘΗΓΗΤΡΙΑ
ΠΛΗΡΟΦΟΡΙΚΗ – ΕΠΕΞΕΡΓΑΣΙΑ ΣΩΜΑΤΩΝ ΚΕΙΜΕΝΩΝ
ΤΜΗΜΑ ΜΕΣΟΓΕΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΔΗΜΟΚΡΑΤΙΑΣ 1, ΡΟΔΟΣ, 85100
email: frantzi@rhodes.aegean.gr

KATERINA T. FRANTZI

ASSOCIATE PROFESSOR
INFORMATICS – CORPUS PROCESSING
DEPARTMENT OF MEDITERRANEAN STUDIES
FACULTY OF HUMANITIES
UNIVERSITY OF THE AEGEAN
1, DIMOKRATIAS AVE., RHODES, 85100, GREECE
email: frantzi@rhodes.aegean.gr