

# Diachronic Corpus of Greek of the 20th century: Design and implementation



Dionysis Goutsos, Georgia Fragaki, Irene Florou, Vasiliki Kakousi & Paraskevi Savvidou  
National and Kapodistrian University of Athens

## The project

This poster presents the research project "Diachronic Corpus of Greek of the 20th century", involving the compilation and analysis of a corpus of Greek of the 20th century (Greek Corpus 20). The project is funded in the frame of the action ARISTEIA I.

International advisory board:

- Claudia Claridge, University of Duisburg-Essen
- Mark Davies, Brigham Young University
- Hendrik De Smet, KU Leuven
- Susan M. Fitzmaurice, University of Sheffield
- Marianne Hundt, University of Zurich
- Christian Mair, University of Freiburg
- Terttu Nevalainen, University of Helsinki
- Fabio Tamburini, Università di Bologna
- Sean Wallis, University College London

## Issues of compilation

**Availability of data/text types:**

- spoken data > development of micro-corpora (cf. Culpeper & Kytö 2010):  
speech-like: private letters  
speech-based: Parliament speeches, newsreels  
speech-purposed: theatrical plays, film scripts
- written data > development of polytonic OCR for Greek 20<sup>th</sup> century texts

**Representativeness:**

- Based on the structure of CGT (1990-2010)
- Texts are selected on the basis of how representative they are of particular text types and *not* of their language variety. The question of diglossia will not constitute a sampling parameter but will be investigated in the stage of analysis, as an outcome of corpus compilation.

## Greek Corpus 20

Κωδικός κειμένου	Τύπος κειμένου	Χρονιά	Μήκος κειμένου (words)
1	Εφημερίδα	1900	100.000
2	Εφημερίδα	1910	120.000
3	Εφημερίδα	1920	150.000
4	Εφημερίδα	1930	180.000
5	Εφημερίδα	1940	200.000
6	Εφημερίδα	1950	220.000
7	Εφημερίδα	1960	250.000
8	Εφημερίδα	1970	280.000
9	Εφημερίδα	1980	300.000
10	Εφημερίδα	1990	320.000
11	Εφημερίδα	2000	340.000
12	Εφημερίδα	2010	360.000
13	Εφημερίδα	2015	380.000
14	Εφημερίδα	2020	400.000
15	Εφημερίδα	2025	420.000
16	Εφημερίδα	2030	440.000
17	Εφημερίδα	2035	460.000
18	Εφημερίδα	2040	480.000
19	Εφημερίδα	2045	500.000
20	Εφημερίδα	2050	520.000
21	Εφημερίδα	2055	540.000
22	Εφημερίδα	2060	560.000
23	Εφημερίδα	2065	580.000
24	Εφημερίδα	2070	600.000
25	Εφημερίδα	2075	620.000
26	Εφημερίδα	2080	640.000
27	Εφημερίδα	2085	660.000
28	Εφημερίδα	2090	680.000
29	Εφημερίδα	2095	700.000
30	Εφημερίδα	2100	720.000
31	Εφημερίδα	2105	740.000
32	Εφημερίδα	2110	760.000
33	Εφημερίδα	2115	780.000
34	Εφημερίδα	2120	800.000
35	Εφημερίδα	2125	820.000
36	Εφημερίδα	2130	840.000
37	Εφημερίδα	2135	860.000
38	Εφημερίδα	2140	880.000
39	Εφημερίδα	2145	900.000
40	Εφημερίδα	2150	920.000
41	Εφημερίδα	2155	940.000
42	Εφημερίδα	2160	960.000
43	Εφημερίδα	2165	980.000
44	Εφημερίδα	2170	1.000.000
45	Εφημερίδα	2175	1.020.000
46	Εφημερίδα	2180	1.040.000
47	Εφημερίδα	2185	1.060.000
48	Εφημερίδα	2190	1.080.000
49	Εφημερίδα	2195	1.100.000
50	Εφημερίδα	2200	1.120.000
51	Εφημερίδα	2205	1.140.000
52	Εφημερίδα	2210	1.160.000
53	Εφημερίδα	2215	1.180.000
54	Εφημερίδα	2220	1.200.000
55	Εφημερίδα	2225	1.220.000
56	Εφημερίδα	2230	1.240.000
57	Εφημερίδα	2235	1.260.000
58	Εφημερίδα	2240	1.280.000
59	Εφημερίδα	2245	1.300.000
60	Εφημερίδα	2250	1.320.000
61	Εφημερίδα	2255	1.340.000
62	Εφημερίδα	2260	1.360.000
63	Εφημερίδα	2265	1.380.000
64	Εφημερίδα	2270	1.400.000
65	Εφημερίδα	2275	1.420.000
66	Εφημερίδα	2280	1.440.000
67	Εφημερίδα	2285	1.460.000
68	Εφημερίδα	2290	1.480.000
69	Εφημερίδα	2295	1.500.000
70	Εφημερίδα	2300	1.520.000
71	Εφημερίδα	2305	1.540.000
72	Εφημερίδα	2310	1.560.000
73	Εφημερίδα	2315	1.580.000
74	Εφημερίδα	2320	1.600.000
75	Εφημερίδα	2325	1.620.000
76	Εφημερίδα	2330	1.640.000
77	Εφημερίδα	2335	1.660.000
78	Εφημερίδα	2340	1.680.000
79	Εφημερίδα	2345	1.700.000
80	Εφημερίδα	2350	1.720.000
81	Εφημερίδα	2355	1.740.000
82	Εφημερίδα	2360	1.760.000
83	Εφημερίδα	2365	1.780.000
84	Εφημερίδα	2370	1.800.000
85	Εφημερίδα	2375	1.820.000
86	Εφημερίδα	2380	1.840.000
87	Εφημερίδα	2385	1.860.000
88	Εφημερίδα	2390	1.880.000
89	Εφημερίδα	2395	1.900.000
90	Εφημερίδα	2400	1.920.000
91	Εφημερίδα	2405	1.940.000
92	Εφημερίδα	2410	1.960.000
93	Εφημερίδα	2415	1.980.000
94	Εφημερίδα	2420	2.000.000
95	Εφημερίδα	2425	2.020.000
96	Εφημερίδα	2430	2.040.000
97	Εφημερίδα	2435	2.060.000
98	Εφημερίδα	2440	2.080.000
99	Εφημερίδα	2445	2.100.000
100	Εφημερίδα	2450	2.120.000
101	Εφημερίδα	2455	2.140.000
102	Εφημερίδα	2460	2.160.000
103	Εφημερίδα	2465	2.180.000
104	Εφημερίδα	2470	2.200.000
105	Εφημερίδα	2475	2.220.000
106	Εφημερίδα	2480	2.240.000
107	Εφημερίδα	2485	2.260.000
108	Εφημερίδα	2490	2.280.000
109	Εφημερίδα	2495	2.300.000
110	Εφημερίδα	2500	2.320.000
111	Εφημερίδα	2505	2.340.000
112	Εφημερίδα	2510	2.360.000
113	Εφημερίδα	2515	2.380.000
114	Εφημερίδα	2520	2.400.000
115	Εφημερίδα	2525	2.420.000
116	Εφημερίδα	2530	2.440.000
117	Εφημερίδα	2535	2.460.000
118	Εφημερίδα	2540	2.480.000
119	Εφημερίδα	2545	2.500.000
120	Εφημερίδα	2550	2.520.000
121	Εφημερίδα	2555	2.540.000
122	Εφημερίδα	2560	2.560.000
123	Εφημερίδα	2565	2.580.000
124	Εφημερίδα	2570	2.600.000
125	Εφημερίδα	2575	2.620.000
126	Εφημερίδα	2580	2.640.000
127	Εφημερίδα	2585	2.660.000
128	Εφημερίδα	2590	2.680.000
129	Εφημερίδα	2595	2.700.000
130	Εφημερίδα	2600	2.720.000
131	Εφημερίδα	2605	2.740.000
132	Εφημερίδα	2610	2.760.000
133	Εφημερίδα	2615	2.780.000
134	Εφημερίδα	2620	2.800.000
135	Εφημερίδα	2625	2.820.000
136	Εφημερίδα	2630	2.840.000
137	Εφημερίδα	2635	2.860.000
138	Εφημερίδα	2640	2.880.000
139	Εφημερίδα	2645	2.900.000
140	Εφημερίδα	2650	2.920.000
141	Εφημερίδα	2655	2.940.000
142	Εφημερίδα	2660	2.960.000
143	Εφημερίδα	2665	2.980.000
144	Εφημερίδα	2670	3.000.000
145	Εφημερίδα	2675	3.020.000
146	Εφημερίδα	2680	3.040.000
147	Εφημερίδα	2685	3.060.000
148	Εφημερίδα	2690	3.080.000
149	Εφημερίδα	2695	3.100.000
150	Εφημερίδα	2700	3.120.000
151	Εφημερίδα	2705	3.140.000
152	Εφημερίδα	2710	3.160.000
153	Εφημερίδα	2715	3.180.000
154	Εφημερίδα	2720	3.200.000
155	Εφημερίδα	2725	3.220.000
156	Εφημερίδα	2730	3.240.000
157	Εφημερίδα	2735	3.260.000
158	Εφημερίδα	2740	3.280.000
159	Εφημερίδα	2745	3.300.000
160	Εφημερίδα	2750	3.320.000
161	Εφημερίδα	2755	3.340.000
162	Εφημερίδα	2760	3.360.000
163	Εφημερίδα	2765	3.380.000
164	Εφημερίδα	2770	3.400.000
165	Εφημερίδα	2775	3.420.000
166	Εφημερίδα	2780	3.440.000
167	Εφημερίδα	2785	3.460.000
168	Εφημερίδα	2790	3.480.000
169	Εφημερίδα	2795	3.500.000
170	Εφημερίδα	2800	3.520.000
171	Εφημερίδα	2805	3.540.000
172	Εφημερίδα	2810	3.560.000
173	Εφημερίδα	2815	3.580.000
174	Εφημερίδα	2820	3.600.000
175	Εφημερίδα	2825	3.620.000
176	Εφημερίδα	2830	3.640.000
177	Εφημερίδα	2835	3.660.000
178	Εφημερίδα	2840	3.680.000
179	Εφημερίδα	2845	3.700.000
180	Εφημερίδα	2850	3.720.000
181	Εφημερίδα	2855	3.740.000
182	Εφημερίδα	2860	3.760.000
183	Εφημερίδα	2865	3.780.000
184	Εφημερίδα	2870	3.800.000
185	Εφημερίδα	2875	3.820.000
186	Εφημερίδα	2880	3.840.000
187	Εφημερίδα	2885	3.860.000
188	Εφημερίδα	2890	3.880.000
189	Εφημερίδα	2895	3.900.000
190	Εφημερίδα	2900	3.920.000
191	Εφημερίδα	2905	3.940.000
192	Εφημερίδα	2910	3.960.000
193	Εφημερίδα	2915	3.980.000
194	Εφημερίδα	2920	4.000.000
195	Εφημερίδα	2925	4.020.000
196	Εφημερίδα	2930	4.040.000
197	Εφημερίδα	2935	4.060.000
198	Εφημερίδα	2940	4.080.000
199	Εφημερίδα	2945	4.100.000
200	Εφημερίδα	2950	4.120.000
201	Εφημερίδα	2955	4.140.000
202	Εφημερίδα	2960	4.160.000
203	Εφημερίδα	2965	4.180.000
204	Εφημερίδα	2970	4.200.000
205	Εφημερίδα	2975	4.220.000
206	Εφημερίδα	2980	4.240.000
207	Εφημερίδα	2985	4.260.000
208	Εφημερίδα	2990	4.280.000
209	Εφημερίδα	2995	4.300.000
210	Εφημερίδα	3000	4.320.000
211	Εφημερίδα	3005	4.340.000
212	Εφημερίδα	3010	4.360.000
213	Εφημερίδα	3015	4.380.000
214	Εφημερίδα	3020	4.400.000
215	Εφημερίδα	3025	4.420.000
216	Εφημερίδα	3030	4.440.000
217	Εφημερίδα	3035	4.460.000
218	Εφημερίδα	3040	4.480.000
219	Εφημερίδα	3045	4.500.000
220	Εφημερίδα	3050	4.520.0