The Word Order in Western Asia Corpus

metadata for the data set

Hellenic — Pontic Greek (Romeyka)

13 July 2025
(doc v1.1)

Overview of the data set

data set	`hell_ponticgreek_romeyka`
contributors	Laurentia Schreiber
published	22 August 2021
texts	5
tokens	501 analyzed (597 total)
notes	XLS/TSV contain entire source text

Metadata for individual texts

Text A

source	Schreiber (Forthcoming: 08_04072019M_1) (WAV ✘ PDF ✓)
location	Ogene/Karacam, Caykara, Trabzon district, Turkey (40.5928° N, 40.2190° E)
recorded	2019
text type	personal narrative
speaker	01 — male, age 60
tokens	198 analyzed (223 total)

Text B

source	Schreiber (Forthcoming: 04_01072019F_12) (WAV ✘ PDF ✓)
location	Camlibel, Caykara, Trabzon district, Turkey (40.6800° N, 40.2287° E)
recorded	2019
text type	stimulus-based narrative
speaker	02 — female, age 50
tokens	47 analyzed (56 total)

Text C

source	Schreiber (Forthcoming: 05_03072019M_3/4) (WAV ✘ PDF ✓)
location	Besköy, Sürmene, Trabzon district, Turkey (40.9113° N, 40.1186° E)
recorded	2019
text type	stimulus-based narrative
speaker	03 — male, age 40
tokens	52 analyzed (66 total)

Text D

source	Schreiber (Forthcoming: 04_01072019F_1) (WAV ✘ PDF ✓)
location	Camlibel, Caykara, Trabzon district, Turkey (40.6797° N, 40.2287° E)
recorded	2019
text type	personal narrative
speaker	02 — female, age 50
tokens	147 analyzed (178 total)

Text E

source	Schreiber (Forthcoming: 04_01072019F_1) (WAV ✘ PDF ✓)
location	Camlibel, Caykara, Trabzon district, Turkey (40.6800° N, 40.2287° E)
recorded	2019
text type	stimulus-based narrative
speaker	02 — female, age 50
tokens	57 analyzed (74 total)

Background to the data set

The data were collected in June/July 2019 around Caykara, Trabzon district, Turkey. They comprise data from two dialects: Romeyka of Of as spoken in Caykara (texts A, B, D, and E), and Romeyka of Sürmene (text C). Since Romeyka is currently undergoing language shift, language competence and the amount of Turkish contact influence vary between speakers. The data comprise both free narratives in the form of personal stories and descriptions of everyday work, and texts elicited by means of a story telling task prompted by picture cards. All texts have been fully transcribed, translated in Turkish and English and morphologically glossed as part of the Romeyka corpus (Schreiber Forthcoming). Note that in some of the texts (B, C, D), the sext segment numbering in the XLS/TSV differs slightly from that in the source text.

References

Schreiber, Laurentia. Forthcoming. *The Romeyka corpus*.