Corpus of Modern Greek and the Greek Language Question

Maxim Kisilier
• 2011-2015: RAS program “Corpus Linguistics”: 14 corpora including Albanian and Modern Greek (http://web-corpora.net)

• Followed the model of the Eastern Armenian National Corpus (http://www.eanc.net)
Corpus of Modern Greek

- [http://web-corpora.net/GreekCorpus/search/](http://web-corpora.net/GreekCorpus/search/)
- Size: 35.7 million tokens
- Collection of written texts mostly from contemporary newspapers
- No disambiguation: each word is annotated with all possible analyses
Corpus of Modern Greek

- [http://web-corpora.net/GreekCorpus/search/](http://web-corpora.net/GreekCorpus/search/)
- Size: 35.7 million tokens
- Collection of written texts mostly from contemporary newspapers
- No disambiguation: each word is annotated with all possible analyses
- Occasional mistakes in the grammatical tags as annotation and compilation of the dictionary were automatic
1. Η Γυφτοπούλα  Αλέξανδρος Παπαδιαμάντης  1884

'Iδού, λάβετε.
Οἱ δύο μικροὶ ἐπαισμὸν τὴν ἐσπέραν ἐκείνην νὰ λιθοβολῆσι. Τὴν ἐπαύριον δῶμος ἤρχισαν ἕκ νέον τὰ αὐτά. Ιδοὺς η Ἀἴμα ὅτι ἀπέτυχε τὸ ἀκρως ἀφελεῖς ἐκείνο μέσον, ἐσκέφθη νὰ μεταβῇ εἰς τὸν οἶκον τοῦ ἄγροτον, καὶ νὰ παρακαλέσῃ τὴν μητέρα τῶν δύο χιλιετῶν νὰ περιορίσῃ ὅλους τὰς κακὰς ἔξεις τῶν. Τῷ ὁντὶ μετὰ δύο ἡμέρας μετέβη εἰς τὴν μικρὰν ἐπαυλίν. Η γυνὴ τοῦ χωρικοῦ ἤτο ἄνικανος ὅλως νὰ παιδεύσῃ τοὺς δύο μικροὺς διαβόλους. Οἱ μεγάλοι σώματος γυνῆ, ἔχουσα δύο ὑπερθέρεις μαστοὺς κρεμαμένους ἐξω τῆς τραχιλάς, καὶ ποτὲ δὲν κατώρθωσε νὰ τοὺς συμμαζέψῃ εἰς τὸ περιστήριον τῆς.
Corpus of Modern Greek

- No access to the full text
- Created by linguists
- Created for linguists
- Created for non-professional users
- User-friendly interface
- Advanced options
The Greek Language Question

• Diglossia or Polyglossia (Browning 1983)
• Terms *Katharevousa* and *Demotic Greek*
  appeared in 1796 and 1818, but
✓ in AG: literary dialects and spoken dialects
✓ in Middle Ages: historiography, hagiography, literature in vernacular/dialectal literature, dialects
The Greek Language Question

- Primary meaning of the terms *katharevousa* and *demotic Greek* differs from the modern one
- *Katharevousa*
  - Adamantios Korais
  - Yakovakis Neroulos
  - Emmanouil Roidis
  - Georgios Viziinos
  - Georgios Hadzidakis
  - Andreas Embirikos (prose)
The Greek Language Question

- Primary meaning of the terms *katharevousa* and *demotic Greek* differs from the modern one.
- *Katharevousa*
- *Demotic Greek*

- Yannis Psychariss
- Stratis Myrivilis
- Antonis Samorakis
- Nikos Kazantzakis
- Zirana Zatelli
- Rea Galanaki
The Greek Language Question

- Primary meaning of the terms *katharevousa* and *demotic Greek* differs from the modern one.
- *Katharevousa*
- *Demotic Greek*
- Poetry

- Yannis Psycharis
- Stratis Myrivilis
- Antonis Samorakis
- Nikos Kazantzakis
- Zirana Zatelli
- Rea Galanaki
Alexandros Papadiamandis

- 1851-1911
Alexandros Papadiamandis

• 1851-1911
• Many archaic features:
  • nominal morphology - τὴν διήγησθαι
Alexandros Papadiamandis

- 1851-1911
- Many archaic features:
  - nominal morphology - τὴν διήγησιν vs. τὴν διήγησην
Alexandros Papadiamandis

- 1851-1911
- Many archaic features:
  - nominal morphology - τὴν διήγησιν vs. τὴν διήγηση
  - verb morphology - ἐσηκώθη vs. σηκώθηκα, ἦσαν vs. ήταν
  - vocabulary - ὁδώρ vs. νερό, οὖς vs. φίδι
Alexandros Papadiamandis

BUT

• Lexical borrowings from:
  ➢ Turkish – ταμάμ (< tamam)
  ➢ Albanian - πουύπα (< çupë)
  ➢ Slavic – βάλτος (< Ƚламо)
• Features of vernacular: πλιά vs. πιά
Alexandros Papadiamandis

**BUT**

- Lexical borrowings from:
  - Turkish – ταμάμ (< tamam)
  - Albanian - τσούπα (< çupë)
  - Slavic – βάλτος (< длато)
- Features of vernacular: πλιά vs. πιά
Alexandros Papadiamandis

**BUT**

- Lexical borrowings from:
  - Turkish – ταμάμ (< tamam)
  - Albanian - τσούπα (< çupë)
  - Slavic – βάλτος (< ɓlamo)
- Features of vernacular: πλιά vs. πιά
- Dialectal features: γρουνίζω vs. γνωρίζω
Conclusions

• Tags *Katharevousa* and *Demotic Greek* for texts or/and authors is not a good solution
• *Katharevousa* or *Demotic Greek*? Who should decide?
Conclusions

- Tags *Katharevousa* and *Demotic Greek* for texts or/and authors is not a good solution
- *Katharevousa* or *Demotic Greek*?

Who should decide?

Corpus

User
Conclusions

• Choice between polytonic and monotonic

• Advanced search options (choice of flexions or inflexion types if possible)
Thank you for your attention!