

Greek Learner Corpus: Design and Implementation

Alexandros Tantos, Despina Papadopoulou,
Katerina Alexandri, Ifigenia Dosi, Georgia Fotiadou,
Katerina Pouliou & Paraskevi Savvidou

21ST INTERNATIONAL SYMPOSIUM
OF THEORETICAL & APPLIED LINGUISTICS

April 5-7, 2013¹

Outline

- Learner Corpora in Research and Teaching Foreign Languages
- Presentation of the Greek Learner Corpus (GLC)
- Demographics of the participants
- Preliminary Data Error Analysis

Learner Corpora and SLA - TFL

- The construction and exploitation of more LC would be very fruitful for both language teaching purposes and SLA (Díaz-Negrillo & Fernández-Domínguez, 2006)
- Learner Corpora → provide Authentic Language Data (Granger, 1998; Leech, 1998; Pravec, 2002)
- Errors → reflect the students' proficiency level and reveal aspects of their interlanguage

Learner Corpora & SLA

LC provide useful data in SLA research for:

- the testing of theoretical hypotheses in SLA
- the development of the learners' interlanguage
(for Greek see Dimitrakopoulou et al., 2004, 2006;
Ματθαίουδάκη et al., 2010; Παπαδοπούλου, 2005; Tsimpli et al., 2003)

Learner Corpora & Teaching FL

- Learner Corpora are of use in teaching foreign languages:
 - a) construction of teaching material
 - b) design of syllabi
 - c) design of (placement) tests
 - d) development of teaching techniques – strategies

(Granger et al. 2002: 122)

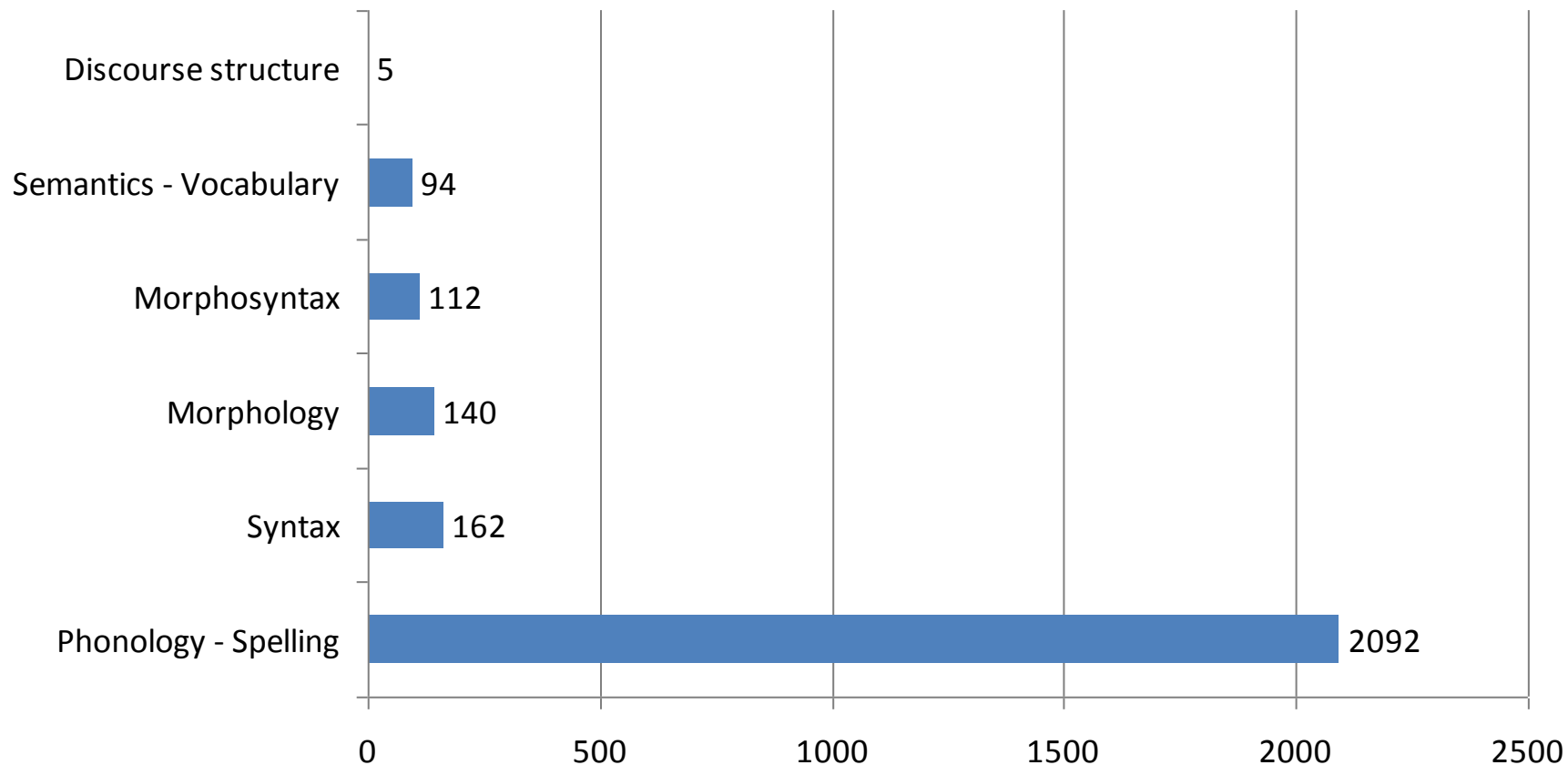
Previous LC for L2 Greek

- **Tzimokas (2010)**: 65,000 words and 291 texts → The first systematic attempt to codify a representative variety of *adult* learners' errors in Greek as a second/foreign language from an impressive number of L1 groups
- Data come from:
 - adult productions
 - a variety of genres
- It is based on a very detailed error annotation scheme and a customary editing and validation tool with in-line annotated files in an XML output format

LC in research for L2 Greek

- Studies for the main errors occurring in repatriated and foreign students' interlanguage of L2 Greek ('Integration of repatriated and foreign students in Greek Secondary Education', 2006-07/ 2007-08, Αναστασιάδη et al., 2007; 2008)
- Developmental data from written productions
- Only *systematic* errors have been analyzed

Error analysis



(Αναστασιάδη et al., 2008)

The Greek Learner Corpus (GLC)

- The first learner corpus assembled from written productions of learners in primary and secondary education
- AIMS of GLC:
 - a) to provide a more user-oriented (researchers, teachers & students) error & POS annotation schemes
 - b) to employ standardized means for the development and exploitation of language resources

GLC

- The GLC consists of pupils' written productions within the placement tests of proficiency in L2 Greek (generated in the framework of the **Action: “Educational & Linguistic Support for Reception Classes”**, sub-action: **Compiling Greek Learner Corpus**)
- It is being continuously expanded → For this presentation we focus on data from ~ 500 written productions (~ 33,000 words)
- Rich inventory of metadata, e.g. age, L1, age of onset, years of residence in Greece, years in the Greek educational system, parents' ability in Greek

The Annotation Methodology

- Informative (to the extend that annotation provided contains all necessary information)
 - Manageable (information included is not exhaustive in order to be easily manipulated by annotators)
 - Reusable (the annotation system is general enough –in order to be used in other languages as well
 - Flexible (accessibility for changes on the annotated texts)
 - Consistent (manual of Annotation system with examples)
- (Granger, 2003)

→ **Specific choices for the error annotation scheme of GLC**

Methodological steps towards an annotation scheme of GLC

- Error compilation and analysis circle follows the main methodological principles adopted in FRIDA (1998):
 - construction of error annotation scheme
 - compilation of error categories based on text samples, linguistic considerations and research data from L2 Greek
 - error annotation within UAM annotation tool
 - extraction of error data for statistical analyses
 - concordance-based tables of most frequent errors

(Granger, 2003)

Stand- Off Annotation and GLC

- Stand off Annotation, i.e. separating original data from annotation data in separate files
- It allows for:
 - a) extensibility of the resource in ANY kind of annotation
 - b) multiple error annotation efforts
 - c) merging of annotations and their qualitative & quantitative comparison

GLC – The Data

Data come from 3 writing tasks

a) task-related variables

- prepared
- cross-sectional

b) language-related variables

- genre: narration

Activity 1: common among Test I, Test II and Test III

Άσκηση 1

Οι εικόνες δείχνουν μια ιστορία.
Κοίταξε τις εικόνες και γράψε την ιστορία.

Μια φορά κι έναν καιρό

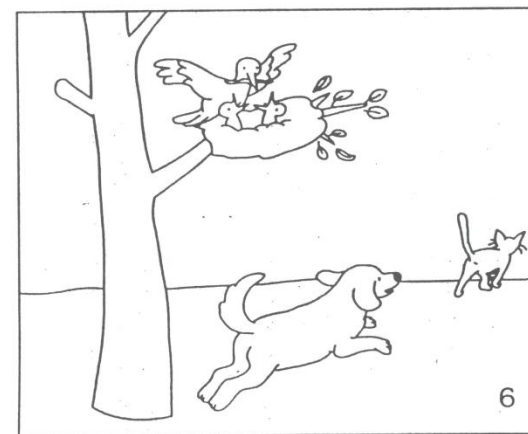
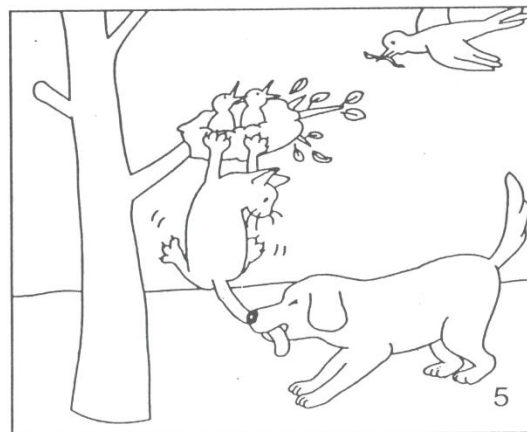
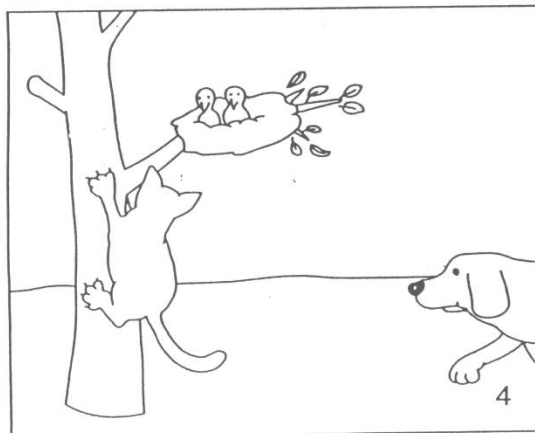
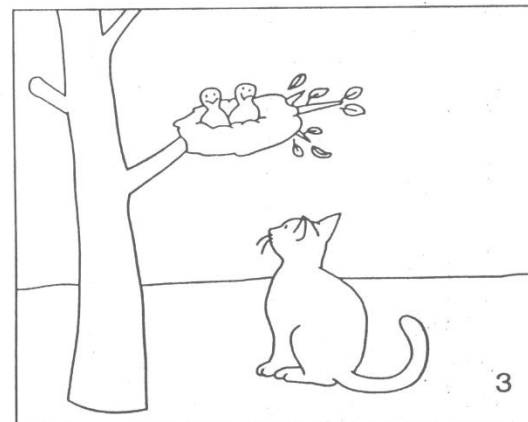
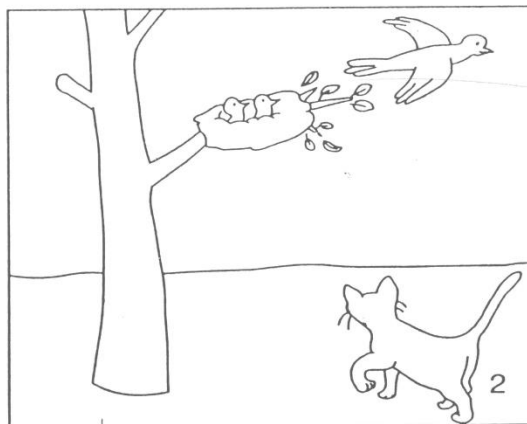
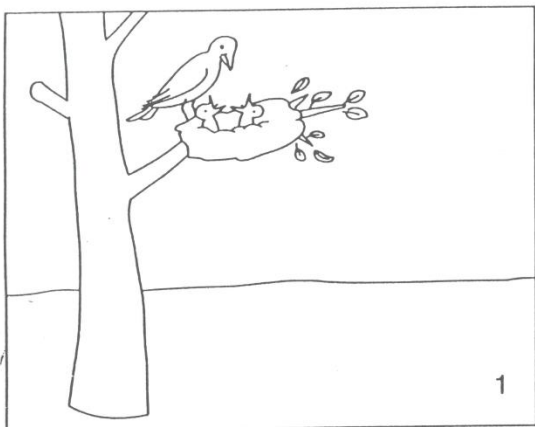
.....

.....

.....

.....

The Cat Story



Χθες ήταν Κυριακή και πήγες μια εκδρομή με τους γονείς σου και με φίλους σου. Γράψε στο ημερολόγιό σου πώς πέρασες. Οι ερωτήσεις θα σε βοηθήσουν.

Activity 2: Test II

- Με ποιους πήγατε εκδρομή;
- Πού πήγατε;
- Τι κάνατε στην εκδρομή;
- Τι θυμάσαι περισσότερο από την εκδρομή; Έγινε κάτι που σε έκανε να γελάσεις, να στενοχωρηθείς ή να φοβηθείς;
- Πώς τελείωσε αυτή η εκδρομή;



Activity 2: Test III

Θεσσαλονίκη 12 / 2 / 20...

..... χερα σου.

Ελπίζω να είσαι καλά.

Τις τελευταίες μέρες έχει πολύ καλό καιρό εδώ στη Θεσσαλονίκη. Έχει ήλιο και αρκετή ζέστη και ευχαριστείμαι να κάνω βόλτα στην παραλία της πόλης. Δυστυχώς όμως έχουμε ακόμα σχολείο και πρέπει να διαβάζω αρκετά κάθε μέρα και έτσι δεν έχω πολύ χρόνο να χαίνομαι.

Σε δυο εβδομάδες όμως θα πάμε εκδρομή στις Πρέσπες με το σχολείο. Δεν έχω πάει εκεί και θέλω πολύ να δω αυτές τις λίμνες.

Γράψε μου τα νέα σου όσο πιο σύντομα μπορείς. Έχετε πάει εσείς καμιά εκδρομή; Πώς τα πάρα σας;

Πολλά καιρετίσματα στους δικούς σου

Γιώργος

Πριν από δύο μέρες πήρες το παραπάνω γράμμα από έναν φίλο σου που μένει στη Θεσσαλονίκη. Γράψε και εσύ ένα γράμμα για να του απαντήσεις, λέγοντας:

- Αν έχεις πάει εκδρομή και πού πήγες.
- Πώς ήταν εκεί;
- Τι σου άρεσε στην εκδρομή; Γιατί;
- Τι δεν σου άρεσε; Γιατί;

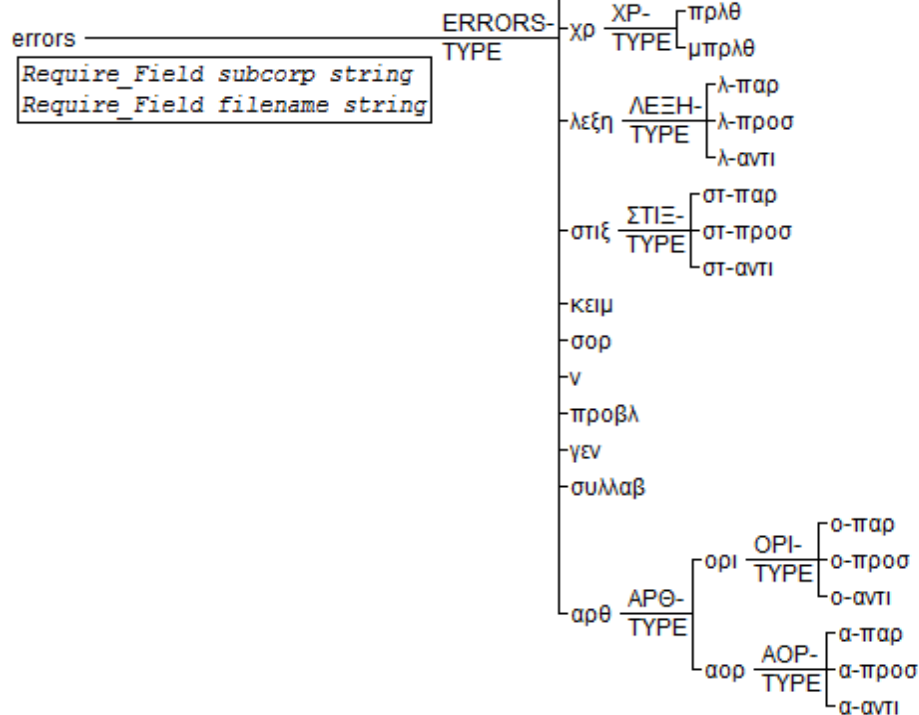
The Error Annotation Scheme of GLC: choices, characteristics

- It includes (a) **linguistic errors**, e.g. determiners, tense, aspect, etc. and (b) **error categories**, e.g. omission, substitution, addition

(Dulay, Burt and Krashen 1982)

- Our error annotation scheme aims at thoroughly describing the pupils' errors avoiding to provide a theoretical interpretation of the errors
- corrected output is not included as part of its annotation

The Error Annotation Scheme



| | | |
|-------|-------|--|
| _ΟΡΘ | -ΛΕΞ | φολιά/ ευδομάδα/ φαγίτο/ κε ανεβικιε/ πέζαμε/ προί |
| | -ΓΡΑΜ | παρη/ δοσι/ ταηση/ στω εκδρομι/ κιταξι |
| | -ΣΥΓΧ | τιλες/ παραπολι/ τινα/ μαδεν/ σαφίνο μετα/ τιλες |
| | -ΚΕΦ | ...ήταν Δύο πουλάκια.../ ...έναν καιρό Ήταν μια.../... με φωνάζει Η μάνα.../ ...στην θαλασσά Βορειθίκα. .. /την τραβixε απο την ουρα της Εκεινι την ωρα.... τα Φαγίτο/ |
| | -ΠΕΖ | στην Βόρια αμερίκη/ σο μεσολογγι/ γιώργο/ αλβανια/ τότε επέστφεψε και η μαινό των μικνφών με τροφη στο στόμα. ο σκυλος κυνήγησε τη γάτα ενω η μάνα..... |
| | -ΚΩΔ | δύο |
| _ΓΡΑΦ | -ΛΕΞ | πηλί (αντί για πουλί)/ δεντο (αντί για δέντρο) / εγρομι (αντί για εκδρομή)/ πολί (αντί για πουλί)/ μοχθρις (αντί για μοχθηρήs) |
| | -ΓΡΑΜ | έκατζε/ ...ήθελε να τους φα.../ πηγαμ |
| _ΤΟΝ | -ΠΡΟΣ | βρεί/ ναί/ τήν/ τρείς / χθές/ πίγάμε |
| | -ΠΑΡ | σκυλος/ φαει/ δαγοσε/ εκει / ξεκινησαμε/ένα |
| | -ΑΝΤΙ | επέσα/ γονόις/ ερχέτε / αύτοι/ ούρα |

| | | |
|-------|-------|--|
| _ΣΥΜΦ | -ΑΡΙ | ...ηταν ωραια το φαγιτο.../ στο μορα/ η λιμνες/ αυτοί η εγδρόμη / τα Φαγιτο |
| | -ΓΕΝ | η πουλί/ στους πουλιά/ το ουρά/ ενας περιστέρι |
| | -ΠΡΟΣ | τον (αντί για «με»)/ εγώ τα περνα/ η μανα επεστρεψα |
| | -ΠΤΩ | ενας σκιλο |
| _ΓΕΝ | | μια δέντρα/ στον δερντο / τα παιδιά της (αντί για του) / ένα ντήσκο |
| _ΛΕΞΗ | -ΠΑΡ | |
| | -ΠΡΟΣ | ...πέζαμε μετά αγόρια κινηγητό (*πέζαμε) ποδοσφερο βολεη |
| | -ΑΝΤΙ | ...αλλά αυτή τη στιγμή ήρθε ένας σκύλος (αντί για «εκείνη») |
| _ΣΤΙΞ | -ΠΑΡ | ...πέζαμε μετά αγόρια κινηγητό ποδοσφερο βολεη... την τραβixε απο την ουρα της Εκεινι την ωρα..../ καθίσαμε όλοι μαζί κάτω από τα δέντρα για να φάμε Στη συνέχεια συνεχίσαμε το παιχνίδι |
| | ΠΡΟΣ | ειδαμαι ολιτιν,ακροπολι/ ο μπαμπάς μου |
| | -ΑΝΤΙ | ...κατο απο την φωλλια, Μετα... (αντί τελείας) |
| _ΚΕΙΜ | | ...έρχετε η γάτα και και φέυχι.../ ... Και πιο μετα εινταν πανο στο δενρο... |
| _ΣΟΡ | | ...Το πουλί για να τους φωρη φαγητο και εμφανιστηκαι |

| | | |
|------|--------|--|
| _ΟΨΗ | -ΣΥΝ | ...άρχισαμε να παίξουμε.../ ...τελειώσαμε να παίξουμε... |
| | -ΜΣΥΝ | ...Εκεί πήγαμε και πέζαμε.../ ...ένας σκύλος την αρπαζέ.../ ...ηταν πηλί ορεα ηχε δεντρα λούλουδια διαβαζαμε παραμιθια.../ ...Επεζαμε με τους φηλους μου γελάσαμε.../ και εγω εχω πολα μαθήματα και να διαβαζω |
| _ΧΡ | -ΠΡΛΘ | |
| | -ΜΠΡΛΘ | ...εχώ Πάη (αντί για πήγα).../ ...όταν πιγα εκδρομι ηταν σουπερ μου αρεσει (αντί για άρεσε).../ ...Η λιμνες, στιν εκδρομι δεν μου αρεσουν (αντί για άρεσαν) τα μαγαζια δεν μου αρεσουν γιατι ητανε... |

The Error Annotation Scheme:

Examples

1) arxísame na *péksume

started.**PERF.1PL** SUBJ play.**PERF.1PL**

Error tag: **_ASP_PERF**

2) *éнас peristéri

MASC.SG.NOM pigeon.**NEUT.SG.NOM**

Error tag: **(a) _AGR_GEN (b) _GEN**

3) *mía déndra

FEM.SG.NOM tree.**NEUT.PL.NOM**

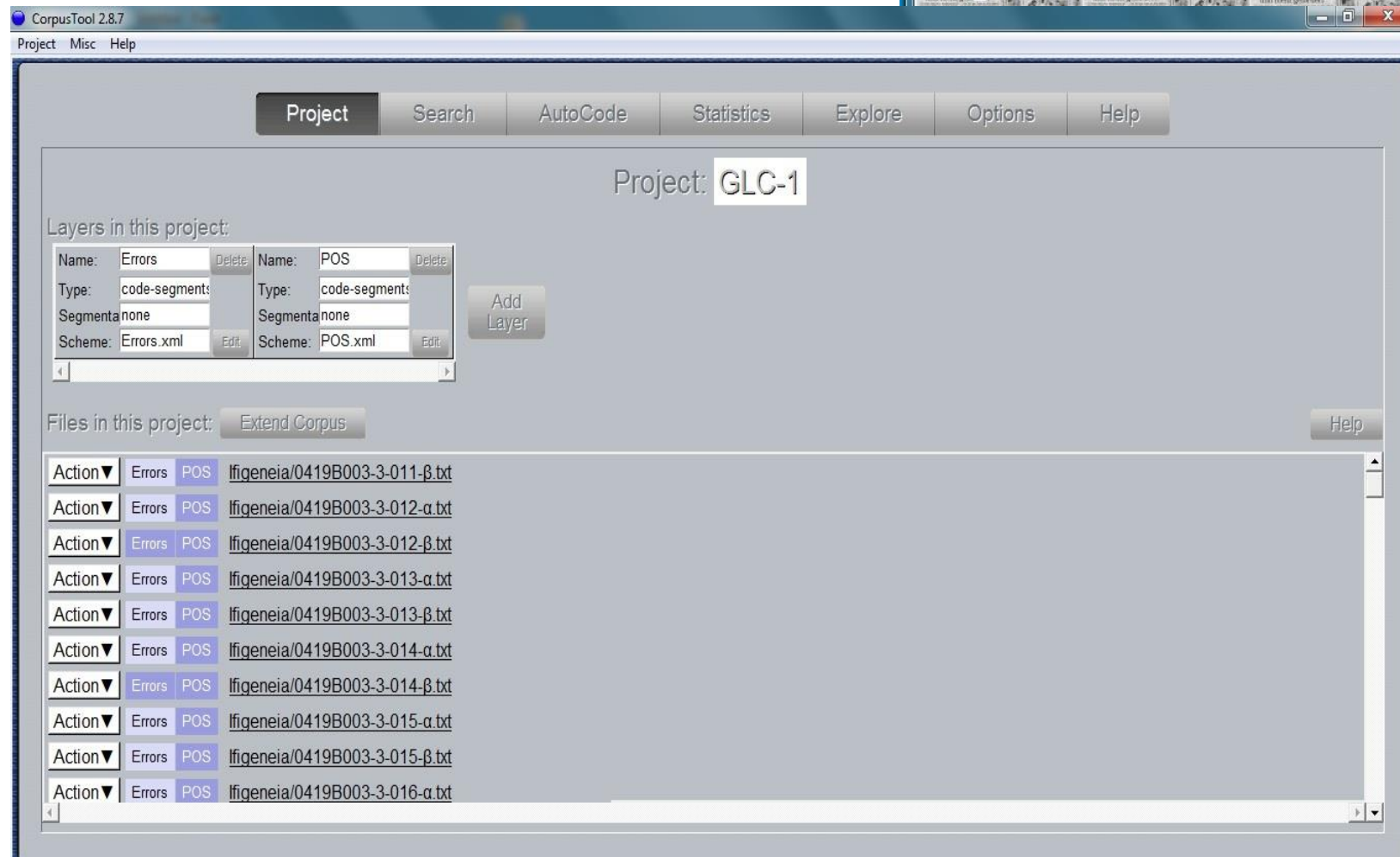
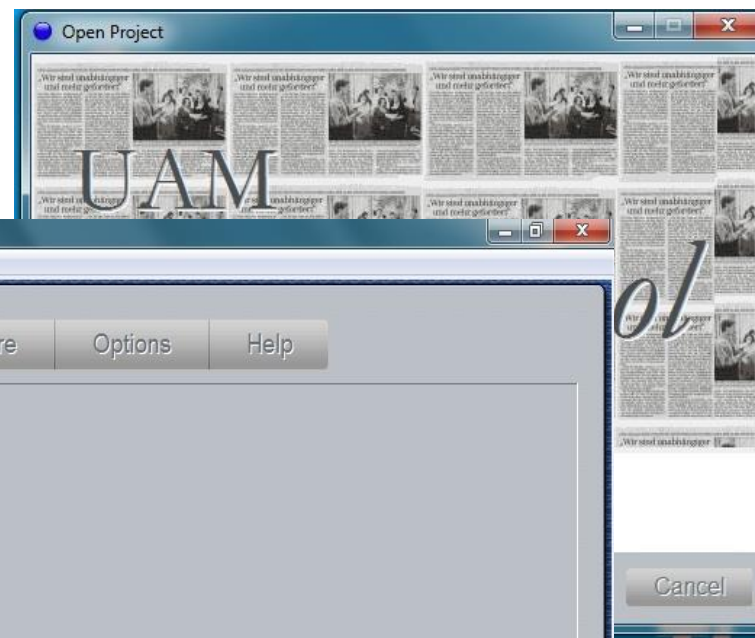
Error tag: **(a) _GEN, (b) _AGR_GEN/NUM**

POS

- GLC is partly annotated with respect to parts of speech (POS)
- POS tagging on a learner corpus is an enterprise of an entirely different nature than on a standardized corpus
- POS annotation provides useful information for the learners' developmental stages and language use as well as the accurate error analysis

UAM Corpus Tool

(O'Donnell, 2008)



Errors-Original

The screenshot displays the CorpusTool 2.8.7 interface. The main window shows a project named 'KERKURA/0523A010-2-002-β.txt'. The 'Layers in this project' section lists two layers: 'Errors' and 'POS'. The 'Files in this project' section lists several files, including 'KERKURA/0523A008-3-006-α.txt', 'KERKURA/0523A008-3-006-β.txt', 'KERKURA/0523A008-3-007-α.txt', 'KERKURA/0523A008-3-007-β.txt', 'KERKURA/0523A008-3-008-α.txt', 'KERKURA/0523A008-3-008-β.txt', 'KERKURA/0523A010-2-001-α.txt', 'KERKURA/0523A010-2-001-β.txt', 'KERKURA/0523A010-2-002-α.txt', and 'KERKURA/0523A010-2-002-β.txt'.

Two windows are open for error analysis for the file 'KERKURA/0523A010-2-002-β.txt'. The top window shows the original text with green annotations: 'Αγαπημένο μου ημερολόγιο πήγαμε με την ταξή' (underlined), 'στος βολος στο κυματογράφο ειδαμε μια' (underlined), 'τένεια την τένεια που ειδαμε οχί δεν είχε τίποτα' (underlined), 'να γελάσεις οχί στεναχώρηθικά ούτε φοβηθικά' (underlined), and 'τελειωσε πολύ ωραία ήταν καιώραια τενεια.' (underlined). The bottom window shows the same text with the annotations removed. The bottom window also has a 'Comment' field and a 'Save Text' button.

The Windows taskbar at the bottom shows the Start button and several application icons, including File Explorer, Media Player, and various office applications.

POS Scheme

CorpusTool 2.8.7

Project Misc Help

Project Search Auto

Layers in this project:

| Name: | Delete | Name: | Delete |
|---------------------|--------|---------------------|--------|
| Errors | | POS | |
| Type: code-segments | | Type: code-segments | |
| Segmenta: none | | Segmenta: none | |
| Scheme: Errors.xml | Edit | Scheme: POS.xml | Edit |

Add Layer

Files in this project: Extend Corpus

| Action | Errors | POS | File |
|--------|--------|-----|------------------------------|
| Action | Errors | POS | KERKURA/0523A008-3-006-α.txt |
| Action | Errors | POS | KERKURA/0523A008-3-006-β.txt |
| Action | Errors | POS | KERKURA/0523A008-3-007-α.txt |
| Action | Errors | POS | KERKURA/0523A008-3-007-β.txt |
| Action | Errors | POS | KERKURA/0523A008-3-008-α.txt |
| Action | Errors | POS | KERKURA/0523A008-3-008-β.txt |
| Action | Errors | POS | KERKURA/0523A010-2-001-α.txt |
| Action | Errors | POS | KERKURA/0523A010-2-001-β.txt |
| Action | Errors | POS | KERKURA/0523A010-2-002-α.txt |
| Action | Errors | POS | KERKURA/0523A010-2-002-β.txt |

Scheme: POS.xml

Start Feature: grammatic-categories Depth: 4 Zoom %: 50 Options

grammatic-categories

```

graph TD
    grammatic-categories[GRAMMATIC-CATEGORIES-TYPE] --> OYI[OYI-TYPE]
    grammatic-categories --> OYD[OYD-TYPE]
    grammatic-categories --> OYA[OYA-TYPE]
    grammatic-categories --> OYB[OYB-TYPE]
    OYI --> OYI_OYI[OYI-OYI-TYPE]
    OYI --> OYI_OYD[OYI-OYD-TYPE]
    OYI --> OYI_OYA[OYI-OYA-TYPE]
    OYI --> OYI_OYB[OYI-OYB-TYPE]
    OYD --> OYD_OYI[OYD-OYI-TYPE]
    OYD --> OYD_OYD[OYD-OYD-TYPE]
    OYD --> OYD_OYA[OYD-OYA-TYPE]
    OYD --> OYD_OYB[OYD-OYB-TYPE]
    OYA --> OYA_OYI[OYA-OYI-TYPE]
    OYA --> OYA_OYD[OYA-OYD-TYPE]
    OYA --> OYA_OYA[OYA-OYA-TYPE]
    OYA --> OYA_OYB[OYA-OYB-TYPE]
    OYB --> OYB_OYI[OYB-OYI-TYPE]
    OYB --> OYB_OYD[OYB-OYD-TYPE]
    OYB --> OYB_OYA[OYB-OYA-TYPE]
    OYB --> OYB_OYB[OYB-OYB-TYPE]
    
```

Help

EN

3:12 PM
2/4/2013



POS - Errors - Original

The screenshot displays the CorpusTool 2.8.7 interface with three overlapping windows. The background window shows the 'POS analysis for KERKURA/0523A010-2-002-β.txt' with a list of layers on the left and a text area containing Greek text with green brackets indicating POS tags. The middle window, titled 'Errors analysis for KERKURA/0523A010-2-002-β.txt', shows the same Greek text with green brackets and a list of errors below it. The foreground window, titled 'KERKURA/0523A010-2-002-β.txt', shows the Greek text and a list of errors with a 'Del' button and a text input field.

CorpusTool 2.8.7

Project Misc Help

Layers in this project:

| Name | Type | Segmenta | Scheme |
|--------|---------------|----------|------------|
| Errors | code-segments | none | Errors.xml |

Files in this project:

| Action | En |
|----------|----|
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |
| Action ▼ | En |

POS analysis for KERKURA/0523A010-2-002-β.txt

Αγαπημένο μου ημερολόγιο πήγαμε με την
στος βολος στο κυματογράφο ειδαμε μια
τένεια την τένεια που ειδαμε οχι δεν είχε τίποτα
να γελάσεις οχι στεναχώρηθείκα ούτε φοβηθείκα
τελειωσε πολύ ωραία ήταν καιώραια τενεια.

Errors analysis for KERKURA/0523A010-2-002-β.txt

Αγαπημένο μου ημερολόγιο πήγαμε με την ταξή
στος βολος στο κυματογράφο ειδαμε μια
τένεια την τένεια που ειδαμε οχι δεν είχε τίποτα
να γελάσεις οχι στεναχώρηθείκα ούτε φοβηθείκα
τελειωσε πολύ ωραία ήταν καιώραια τενεια.

KERKURA/0523A010-2-002-β.txt

Αγαπημένο μου ημερολόγιο πήγαμε με την ταξή
στος βολος στο κυματογράφο ειδαμε μια
τένεια την τένεια που ειδαμε οχι δεν είχε τίποτα
να γελάσεις οχι στεναχώρηθείκα ούτε φοβηθείκα
τελειωσε πολύ ωραία ήταν καιώραια τενεια.

Del errors + : b i u 14 A

Add

Save Text

Help

PARTICIPANTS - PRELIMINARY DATA

GLC - The participants

| Regions | Grade | Schools (N=25) | Participants (N=329) |
|-------------------|-----------|----------------|----------------------|
| Attica | Primary | 4 | 21 |
| | Secondary | 0 | 0 |
| Western Macedonia | Primary | 1 | 10 |
| | Secondary | 0 | 0 |
| Epirus | Primary | 2 | 14 |
| | Secondary | 2 | 27 |
| Thessaly | Primary | 4 | 45 |
| | Secondary | 2 | 24 |
| Ionian Islands | Primary | 1 | 6 |
| | Secondary | 0 | 0 |
| Central Macedonia | Primary | 3 | 98 |
| | Secondary | 2 | 53 |
| Crete | Primary | 4 | 31 |
| | Secondary | 0 | 0 |

GLC

The linguistic profile of the participants

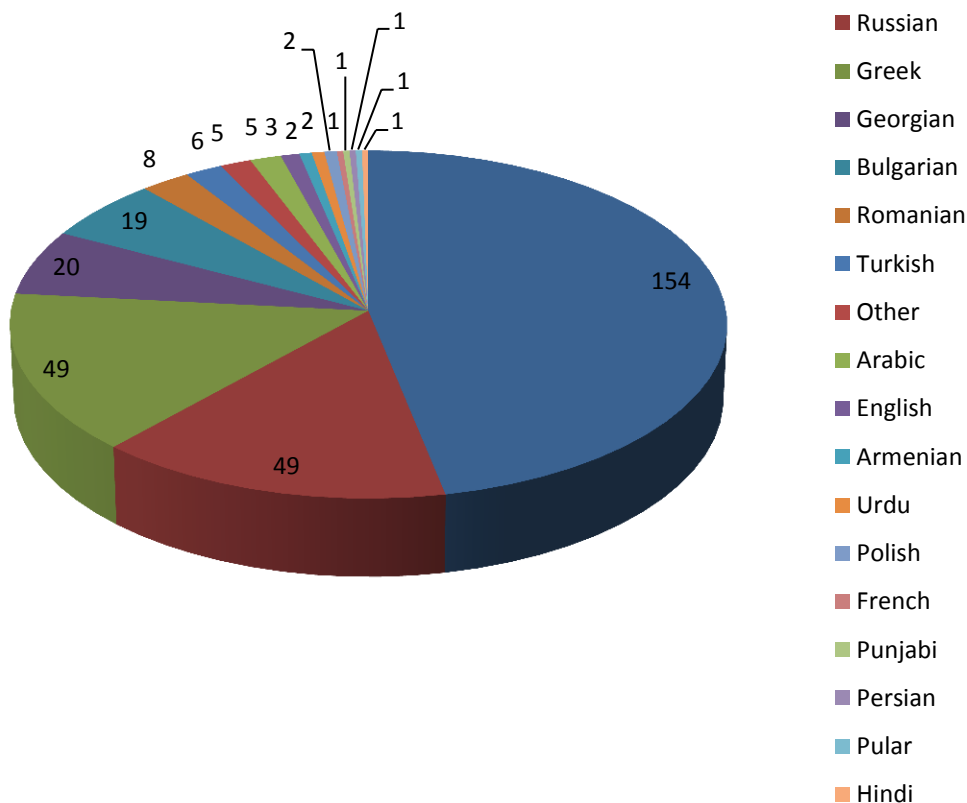
| Gender | | Age Groups | | | |
|----------------------------|-----------------------------|------------------------------|-------------------------------|--------------------------------|------------------------------|
| Male (N= 211, 64.1%) | Female (N=118, 35.9%) | 6-8 yrs (N=111, 33.7%) | 8-12 yrs (N=118, 35.9%) | 12-15 yrs (N= 73, 22.2%) | 15-19 yrs (N=27, 8.2%) |

| Years in Greece | | | | |
|-----------------------------|---------------------------|-------------------------------|-----------------------------|---------------------------------|
| 0-3 yrs (N=66, 20.1%) | 3-6 yrs (N=56, 17%) | 6-10 yrs (N=122, 37.1%) | >10 yrs (N=44, 13.4%) | Unspecified (N=41, 12.4%) |

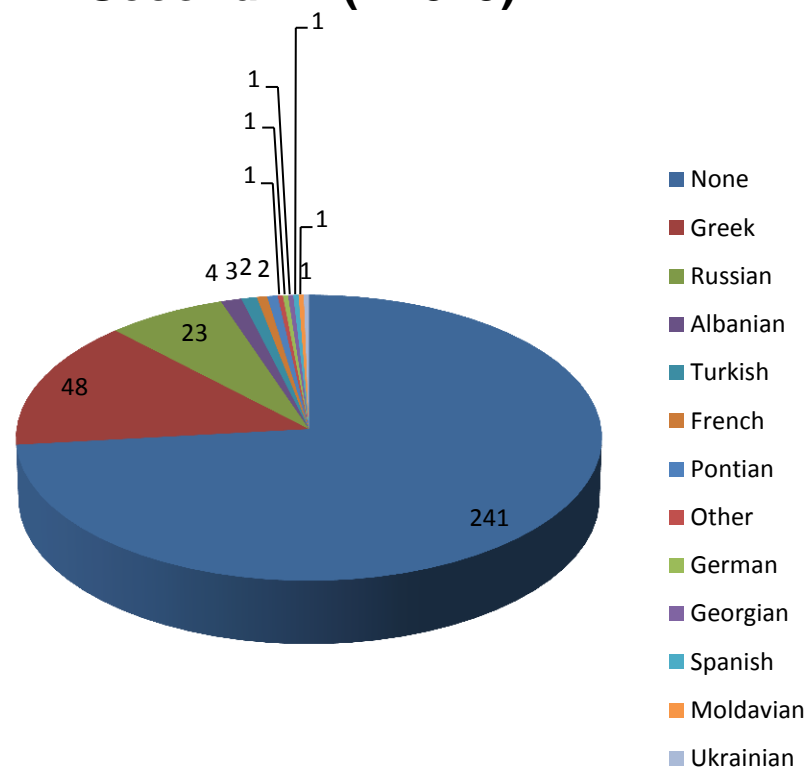
| Years in the Greek Educational System | | | | |
|---------------------------------------|-----------------------------|----------------------------|----------------------------|--------------------------------|
| 0-1 yrs (N=29, 8.9%) | 1-3 yrs (N=74, 22.6%) | 3-6 yrs (N=152, 46%) | >6 yrs (N=48, 14.6%) | Unspecified (N=26, 7.9%) |

L1 & second L1 of the participants

L1 (N=329)

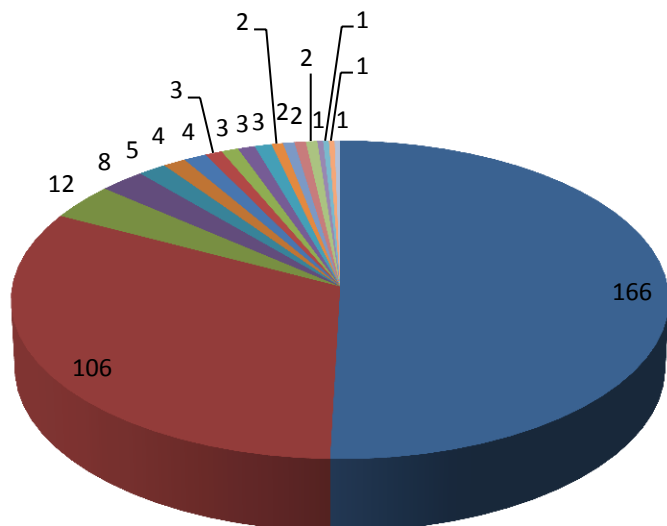


Second L1 (N=329)



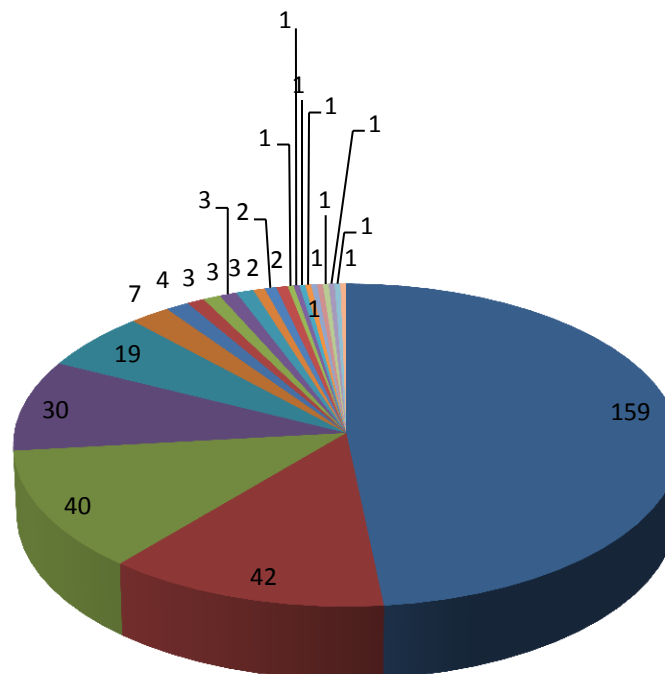
Country of Birth – Country of Origin of the participants

Country of Birth



Greece
 Albania
 Bulgaria
 Georgia
 Romania
 Germany
 Russia
 Pakistan
 UK
 US
 Unknown
 Cyprus
 Guinea
 Moldova
 Poland
 Algeria
 Iran
 Sudan
 Ukraine

Country of Origin



Albania
 Russia
 Georgia
 Greece
 Bulgaria
 Romania
 Unknown
 UK
 Syria
 Armania
 Pakistan
 Turkey
 Guinea
 Moldova
 Ukraine
 Sudan
 Algeria
 Azerbaijan
 Cyprus
 Germany
 India
 Iran
 35

The linguistic profile -Test 1

| Gender | | Age Groups | | |
|---------------------------------------|-----------------------------|------------------------------|-------------------------------|-----------------------------|
| Male (N= 41, 58.6%) | Female (N=29, 41.4%) | 6-8 yrs (N=65, 92.8%) | 8-10 yrs (N=2, 2.9%) | 10-14 yrs (N=3, 4.3%) |
| Years in Greece | | | | |
| 0-3 yrs (N=11, 15.7%) | 3-6 yrs (N=7, 10%) | 6-10 yrs (N=47, 67.2%) | Unspecified (N=5, 7.1%) | |
| Years in the Greek Educational System | | | | |
| 0-1 yrs (N=7, 10%) | 1-3 yrs (N=30, 42.9%) | 3-6 yrs (N=32, 45.7%) | Unspecified (N=1, 1.4%) | |
| Placement | | | | |
| Preliminary stage (N=5, 7.1%) | A1 (N=31, 44.3%) | | A2 (N=34, 48.6%) | |

The linguistic profile -Test 2

| Gender | | Age Groups | | |
|---------------------------------------|-----------------------------|------------------------------|---------------------------------|-------------------------------|
| Male (N= 58, 55.8%) | Female (N=46, 44.2%) | 6-8 yrs (N=46, 44.3%) | 8-10 yrs (N=54, 51.9%) | 10-12yrs (N=4, 3.8%) |
| Years in Greece | | | | |
| 0-3 yrs (N=10, 9.6%) | 3-6 yrs (N=20, 19.2%) | 6-10 yrs (N=63, 60.6%) | Unspecified (N=11, 10.6%) | |
| Years in the Greek Educational System | | | | |
| 0-1 yrs (N=1, 0.9%) | 1-3 yrs (N=14, 13.5%) | 3-6 yrs (N=81, 77.9%) | 6-8 yrs (N=2, 1.9%) | Unspecified (N=6, 5.8%) |
| Placement | | | | |
| A1 (N=4, 3.9%) | A2 (N=75, 72.1%) | | B1 (N=25, 24%) | |

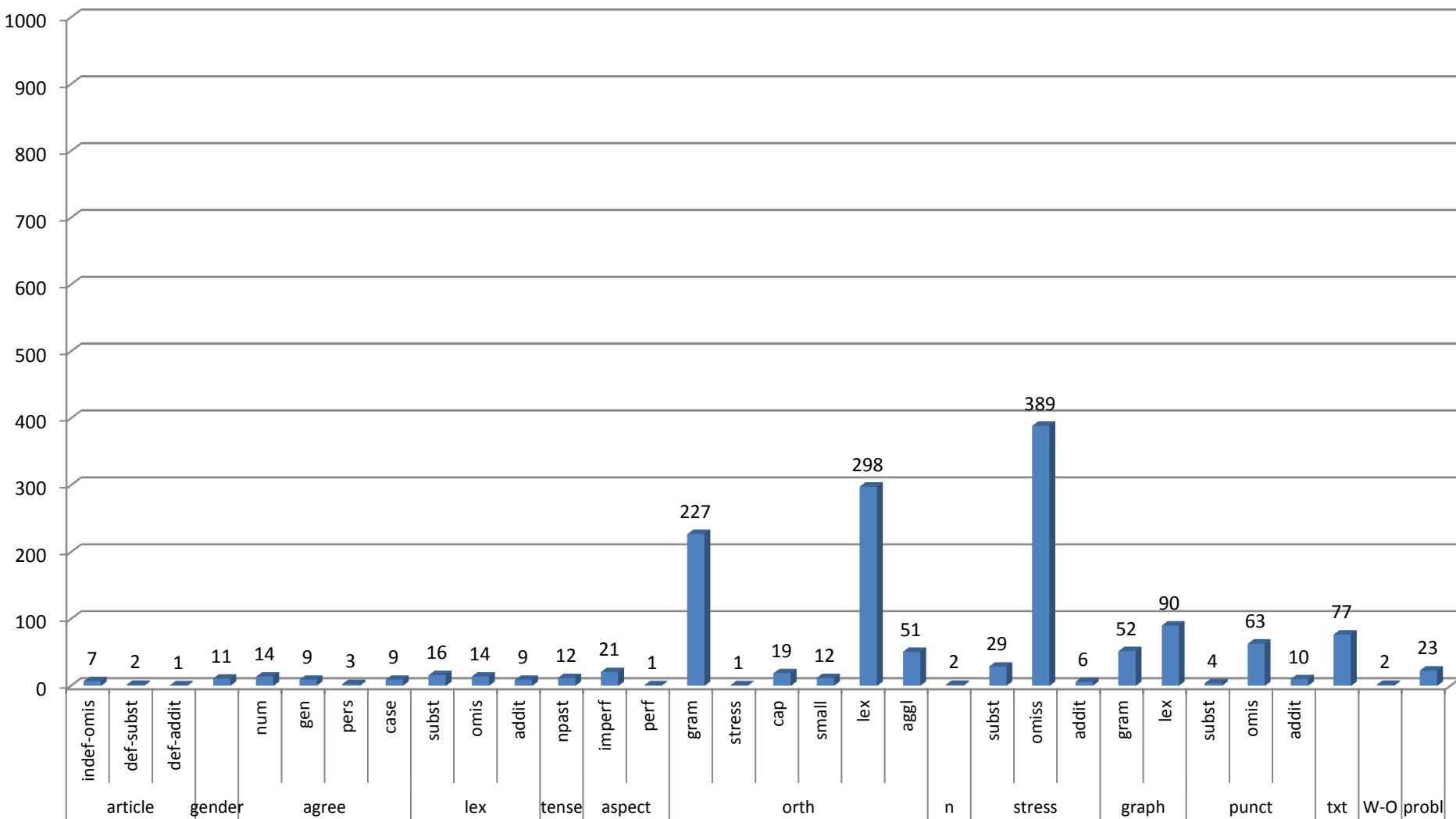
The linguistic profile -Test 3 (Primary)

| Gender | | Age Groups | | |
|---------------------------------------|-----------------------------|-------------------------------|--------------------------------|---------------------------------|
| Male (N= 32, 62.7%) | Female (N=19, 37.3%) | 10-12 yrs (N=45, 88.2%) | 12-14yrs (N=6, 11.8%) | |
| Years in Greece | | | | |
| 0-3 yrs (N=6, 11.8%) | 3-6 yrs (N=9, 17.6%) | 6-10 yrs (N=7, 13.7%) | 10-12 yrs (N=13, 25.5%) | Unspecified (N=16, 31.4%) |
| Years in the Greek Educational System | | | | |
| 1-3 yrs (N=9, 17.6%) | 3-6 yrs (N=18, 35.3%) | 6-8 yrs (N=16, 31.4%) | Unspecified (N=8, 15.7%) | |
| Placement | | | | |
| A2 (N=6, 11.8%) | B1 (N=41, 80.4%) | B2 (N=4, 7.8%) | | |

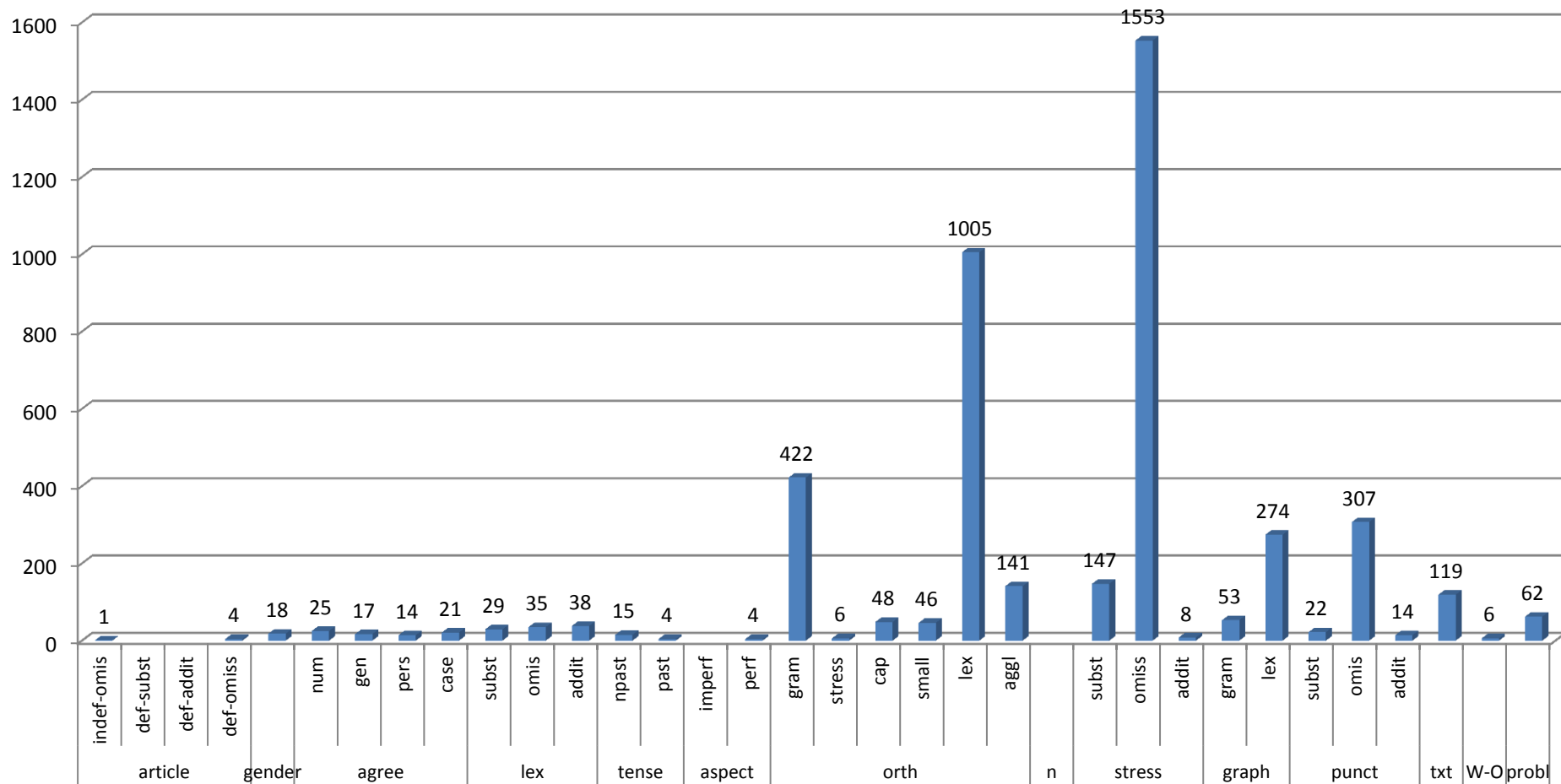
The linguistic profile - Test 3 (Secondary)

| Gender | | Age Groups | | | |
|---------------------------------------|-----------------------------|-----------------------------|-------------------------------|-------------------------------|---------------------------------|
| Male (N= 80, 77%) | Female (N=24, 23%) | 12-14 yrs (N=53, 51%) | 14-16 yrs (N=43, 41.3%) | 16-19yrs (N=8, 7.7%) | |
| Years in Greece | | | | | |
| 0-3 yrs (N=39, 37.5%) | 3-6 yrs (N=20, 19.2%) | 6-10 yrs (N=6, 5.8%) | >10 yrs (N=30, 28.8%) | Unspecified (N=9, 8.7%) | |
| Years in the Greek Educational System | | | | | |
| 0-1 yrs (N=21, 20.2%) | 1-3 yrs (N=21, 20.2%) | 3-6 yrs (N=21, 20.2%) | 6-8 yrs (N=14, 13.5%) | >8 yrs (N=16, 15.4%) | Unspecified (N=11, 10.5%) |
| Placement | | | | | |
| below A1 (N=2, 1.9%) | A2 (N=4, 3.8%) | B1 (N=86, 82.7%) | | B2 (N=12, 11.6%) | |

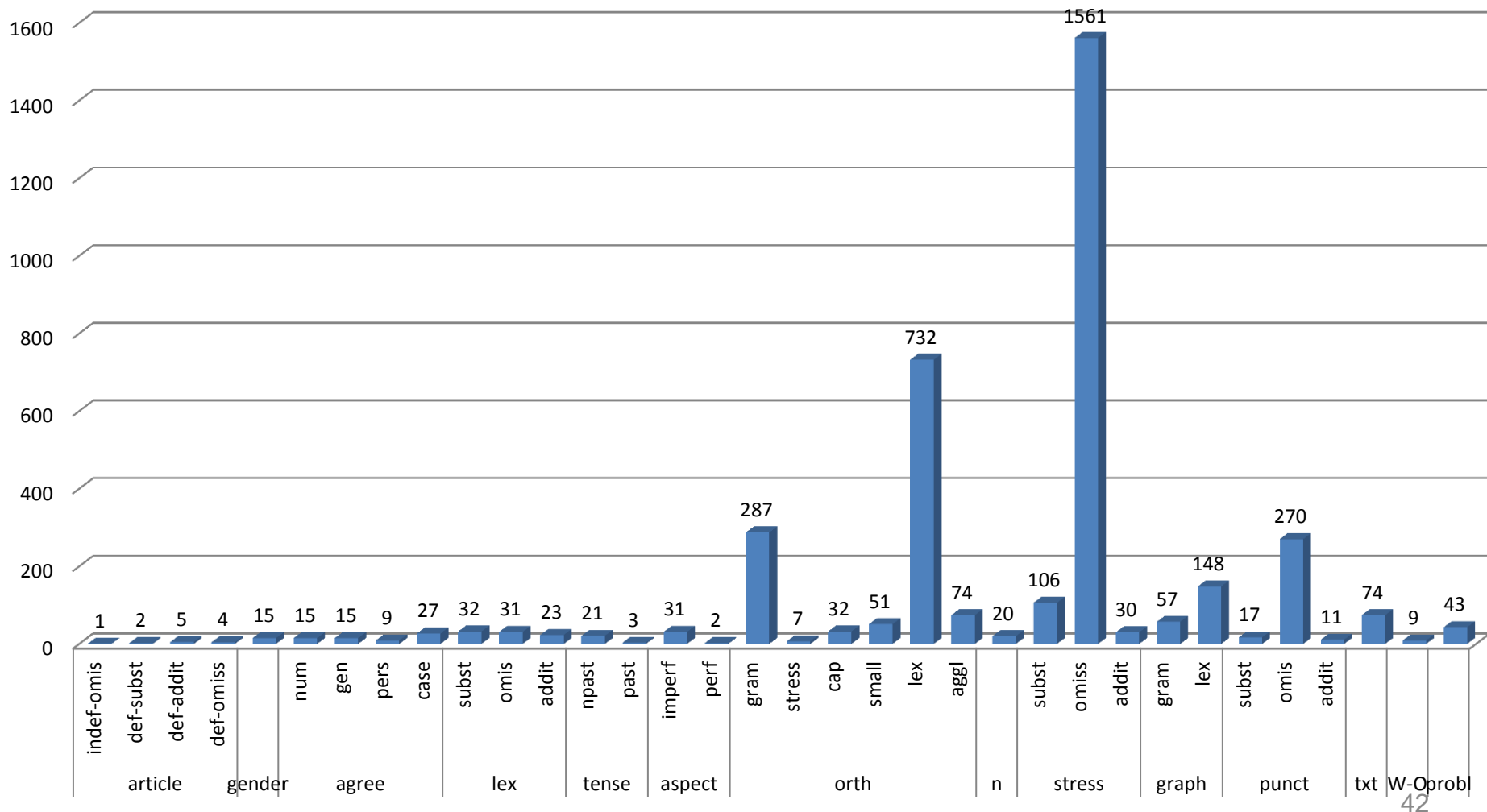
Test 1-Distribution of errors



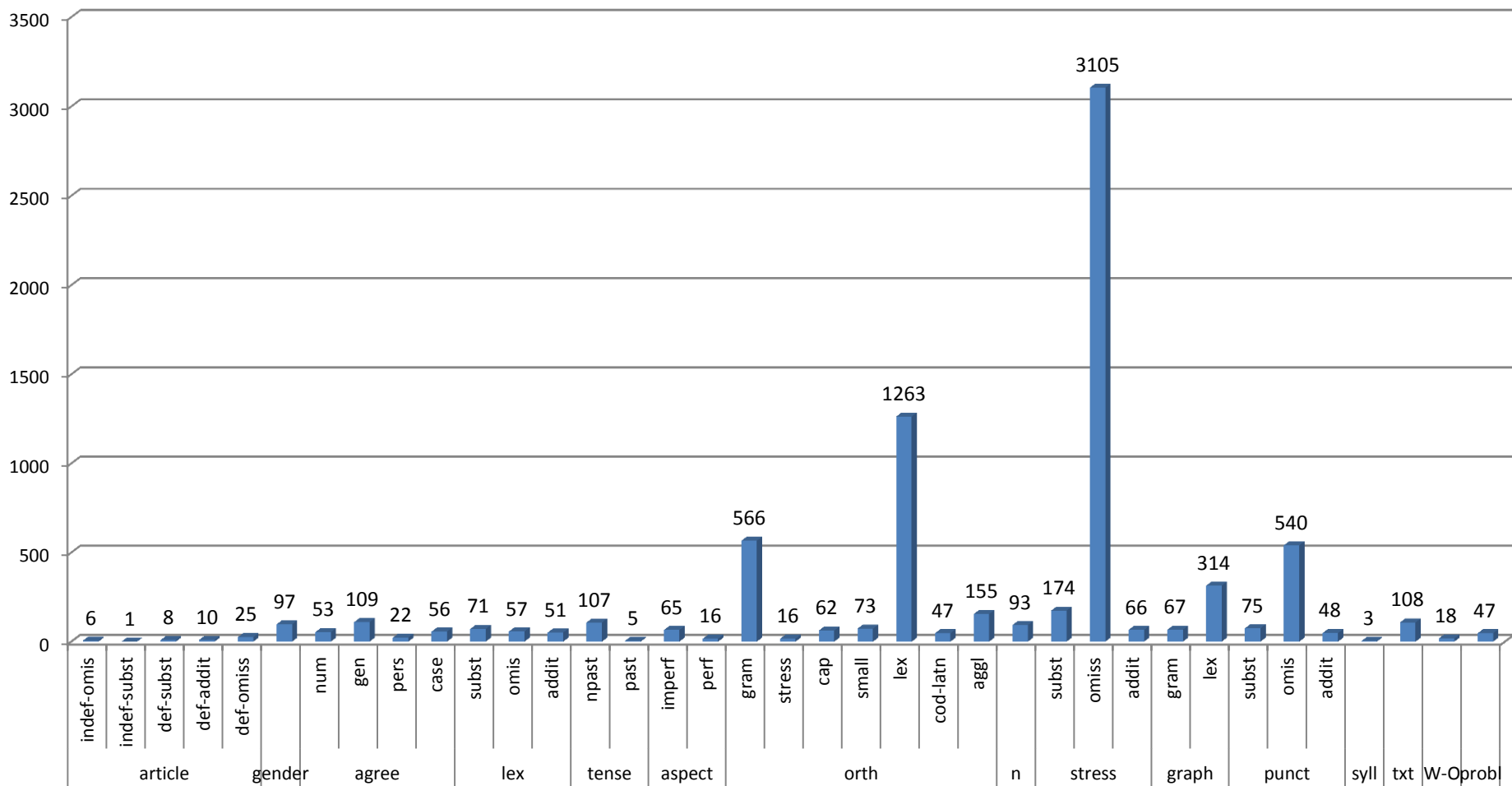
Test 2-Distribution of errors



Test 3 (primary)-Distribution of errors

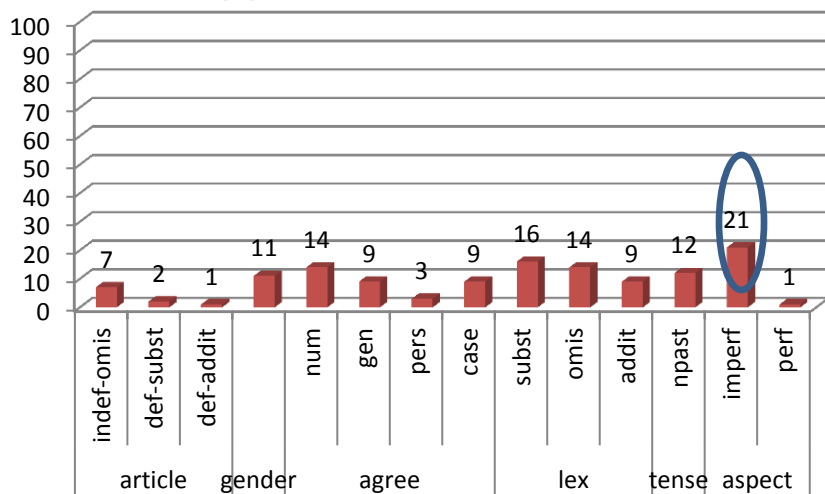


Test 3 (secondary)-Distribution of errors

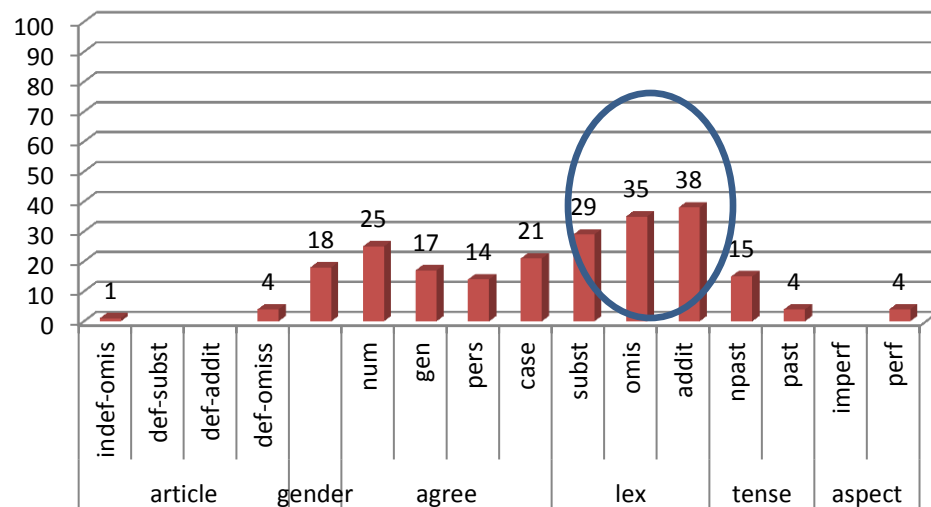


Morpho-syntactic errors

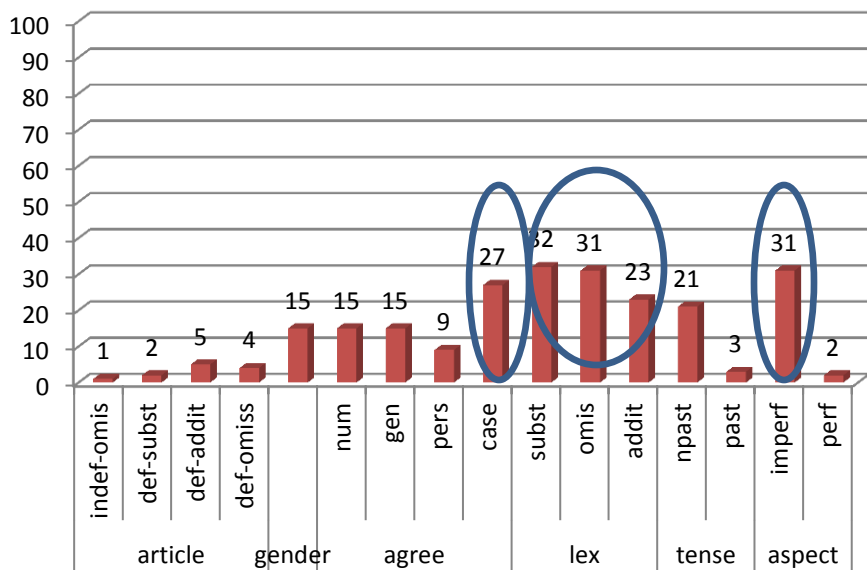
Test 1



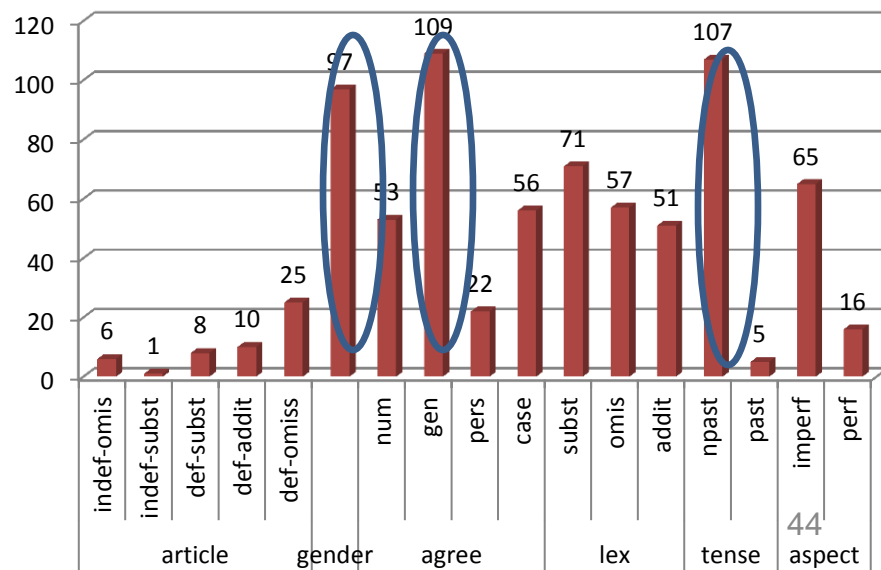
Test 2



Test 3 (primary)



Test 3 (secondary)



Future Development of GLC

- Continuous growth of GLC / updating: new pupils' productions
- Constant update and revision of the error annotation scheme: addition of linguistic errors such as, clitics, prepositions, voice morphology, complementizers, conjunctions etc.
- Further error analyses with respect to the metadata variables
- Completion of POS annotation stage

References

- Αναστασιάδη-Συμεωνίδη, Α., Ε. Βλέτση, Μ. Μητσιάκη, Β. Μποζονέλος, Β. Χούμα. 2008. Τα γλωσσικά λάθη των μαθητών της ελληνικής ως δεύτερης γλώσσας και ο ρόλος της Γ1 στις πολυπολιτισμικές τάξεις του Γυμνασίου. In International Conference "European year of intercultural dialogue: discussing with languages-cultures". School of French Language and Literature, A.U.TH, 597-612 (Available at: <http://www.frl.auth.gr/sites/congres/Interventions/GR/anastassiadis-symeonidis.pdf>)
- Αναστασιάδη Α., Α. Ζουραβλιόβα, Μ. Μητσιάκη & Γ. Φωτιάδου. 2007. Τα φωνολογικά λάθη στη διαγλώσσα των σλαβόφωνων μαθητών της ελληνικής ως δεύτερης/ξένης γλώσσας. Διεθνές συνέδριο της ΕΕΕΓ, Θεσσαλονίκη. <http://www.enl.auth.gr/gala/14th/Papers/Greek%20papers/Anastasiadi-Symeonidi&Zouravliova&Mitsiaki&Fotiadou.pdf>
- Díaz-Negrillo, A. & J. Fernández-Domínguez. 2006. Error tagging systems for learner corpora. *RESLA* 19, 83-102.
- Dimitrakopoulou, M., S. Kalaitzidou, A. Roussou & I. M. Tsimpli. 2004. Clitics and Determiners in the L2 Greek Grammar. Proceedings of the 6th International Conference on Greek Linguistics. <http://www.philology.uoc.gr/conferences/6thICGL>
- Dimitrakopoulou, M., Fotiadou, G., Roussou, A., Tsimpli, I.-M. (2006) «*Features and Agree Relations in L2 Greek*» in A. Belletti, E. Bennati, C. Chesi, E. Di Domenico and I. Ferrari (eds) *Language Acquisition and Development*, Proceedings of GALA 2005, Cambridge Scholars Press, 167-172
- Dimitrakopoulou, M., Fotiadou, G., Roussou, A., Tsimpli, I.-M. (2007) «*Features and Agree Relations in L2 Greek*», Proceedings of the 7th International Conference on Greek Linguistics (ICGL7), University of York. <http://83.212.19.218/icgl7/Tsimpli-et-al.pdf>
- Dulay, H., M. Burt & S. Krashen. 1982. *Language Two*. New York: OUP.
- Granger, S. 1998. The computer learner corpus: A versatile new source of data for SLA research. In S. Granger, ed., *Learner English on computer*, 3-18. London: Addison Wesley Longman.
- Granger, S., Dagneaux, E. & Meunier, F. (eds.) 2002. *The International Corpus of Learner English*. Handbook and CD-ROM, Universitaires de Louvain, Louvain-la-Neuve. Available from <http://www.i6doc.com>

- Granger, S. 2003. The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37(3), 538-545. Longman.
- Leech, G. 1998. Learner corpora: What they are and what can be done with them. In S. Granger, ed., *Learner English on computer*, xiv-xx. London: Addison Wesley.
- Ματθαιουδάκη, Μ., Ι. Κίτσου & Δ. Τζιμώκας. 2011. Η χρήση της ρηματικής όψης στη νέα ελληνική ως δεύτερη/ξένη γλώσσα: πορίσματα εμπειρικής έρευνας από τις εξετάσεις πιστοποίησης ελληνομάθειας, στο Μελέτες για την Ελληνική Γλώσσα. Πρακτικά της 31ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας. 16-18 Απριλίου 2010. Φιλοσοφική Σχολή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη: ΙΝΣ, 317-328
- O'Donnell, M. 2008. The UAM CorpusTool: Software for corpus annotation and exploration. In Bretones Callejas, Carmen M. et al. (Eds.) *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*, (pp. 1433-1447) Almería: Universidad de Almería.
- Παπαδοπούλου, Δ. 2005. Η παραγωγή της ρηματικής όψης από σπουδαστές της ελληνικής ως δεύτερης/ξένης γλώσσας». *Εφαρμοσμένη Γλωσσολογία* 21, 39-54.
- Pravec, N. A. 2002. Survey of learner corpora. *ICAME Journal* 26: 81-114.
- Tsimpli, I.M., Roussou, A., Dimitrakopoulou M., & S. Kalaitzidou. 2003. The production of clitics and articles by Slavic speakers of Greek. *Proceedings of the 13th International Conference on Applied Linguistics*. Thessaloniki.
- Tzimokas D. 2010. Ηλεκτρονικό σώμα κειμένων (ΗΣΚ) εκμάθησης της νέας ελληνικής ως δεύτερης γλώσσας: προς ένα ερευνητικό και διδακτικό εργαλείο. In *Proceedings of 30th Annual Meeting of the Department of Linguistics*. Thessaloniki. 602-616.

Thank you!!!