

# Learner Translator Corpus: Italogreco or Another Way to Confirm Teachers' Intuitions

Katerina Florou<sup>1</sup>

<sup>1</sup> National and Kapodistrian University of Athens, Greece

Correspondence: Katerina Florou, Department of Italian Language and Literature, Amyklon 18, 15231 Athens, Greece. E-mail: kathyflorou@ill.uoa.gr

Received: July 31, 2019

Accepted: August 22, 2019

Online Published: September 19, 2019

doi:10.5539/jel.v8n5p75

URL: <https://doi.org/10.5539/jel.v8n5p75>

## Abstract

This paper describes a project developed within an ongoing study at the University of Athens. In our previous studies we analyzed the errors of Greek learners of Italian language, using Learner Corpora evidence and we retrieved useful information about their interlanguage and its interaction to the language learning process. In this study we assume interlanguage is a tool that can help us to distinguish the language errors from the translation errors and identify the most frequent errors and the causes. For this reason, we focus on specific linguistic areas and we examine a learner translator Corpus. The results are interestingly different from the ones that a language teacher would expect.

**Keywords:** learner translator corpora, corpus analysis, foreign language acquisition, interlanguage analysis, linguistic features

## 1. Introduction

This project aims to provide a detailed description of an Italian-Greek learner translator Corpus (Italogreco) and to investigate the function of specific linguistic aspects of narrative speech (auxiliary verbs and time conjunctions). Furthermore, since our corpus includes both source texts in learners' mother tongue and translations, we will study how the developed interlanguage can affect the "translationese" (Kutuzov & Kuniloskaya, 2014) and the learning process at the same time.

In order to analyze the abovementioned research aims we developed the Italogreco Corpus, a parallel corpus of translations that trainees produced, using their own written productions in Greek as a source text. The learners' mother tongue is Greek; some of the students are bilinguals and one of their languages is Greek along with Russian, Albanian, and German. The corpus includes translations into Italian, because the most important goal of this project is to investigate the choices made by a Greek translator in Italian and how it is affected by his/her first language (Rawoens, 2010; Altenberg & Granger, 2007). The research aims put forward in this study can be investigated using empirical methods and more specifically using corpus analysis techniques. The implications of such study may be effective for the teaching process and also for the translation procedure.

## 2. Current Learner Translator Corpora in Language Research

The most notable work is the research on parallel corpora made by Altenberg and colleagues in several papers (Aijmer & Altenberg, 2000; Altenberg, 2007; Altenberg, 2010). However, an experience in Learner Translator Corpora it can be shown by two large projects. The first one is the multilingual MeLLANGE LTC which was developed by Castagnoli and colleagues (2006) meant to provide error tagged texts and linguistic annotation metadata. Floren and Sanz (2008) introduce ENTRAD, another Learner Translator Corpus based on translation in English by Spanish or French speaking trainees. Recently, two other Translator Corpora were built, the first by Castagnoli (2010), the Multiple Italian Student Translation Corpus (MISTiC) which was compiled for specific research purposes, and the Russian Learner Translator Corpus (Kutuzov & Kunilovskaya, 2014) which is the third learner translator corpus available online (after ENTRAD and MeLLANGE LTC). It is a large corpus, which provides sentence-level aligned multiple concordances and searchable metadata.

## 3. Corpus Design

In order to analyze our previous mentioned research aims we had to develop a learner translator corpus. The translations used were collected from students of the Department of Italian Language and Literature and the

Department of English Language and Literature at the National and Kapodistrian University of Athens in Greece. All translation texts were produced as part of the course “Italian Language I” during the Spring Semester 2017. Since our students were not experienced in translation tasks (they were all freshmen in the University), they were instructed to write first the essays in their mother-tongue (Greek) and then translate them in the L2 of their choice (Italian or English). The topics of the two essays were selected in such a way so that they can provide the maximum level of genre discrimination. The first one was titled “Describe the best and the worst moment of your life”, a more relaxed and personal topic belonging to the narrative genre. The second one was titled “Which are the biggest social problems of our times” and it is a more formal topic approaching the academic genre.

The corpus compiled used Greek as a source language and contains 265,281 tokens, while the learner translator corpus contains 293,298 tokens. The latter contains two target language subcorpora with nearly equal size (English sub corpus with 143,348 tokens, and the Italian sub corpus with 149,850 tokens). Furthermore, in order to homogenize the genre, we split in half the corpus, dividing the texts of narration from those of academic speech. The following table (

Table 1) presents the tokens contained in each of the sub corpora:

Table 1. Corpus size (number of tokens) across genres and languages.

	Original texts in Greek	Translations in Italian	Translations in English
Narrative speech	130,684	76,846	73,558
Academic speech	131,769	75,394	69,975

The students’ personal information was registered in a questionnaire that remained anonymous. Below is a list of the information that was collected from each respondent: age, sex, origin, foreign languages competence and the level achieved in them. In this paper we examine the sub corpus of the Italian translations of the narrative genre.

#### 4. Research Hypotheses

Our study will focus in the narrative part of our corpus and it will examine the grammatical errors observed in the translations obtained from our learners.

Moreover, we will try to develop an error ontology which will help us categorize the errors observed in the translation process as byproducts of the language learning or the translation procedure. For this reason, we had to develop a mixed error typology with labels covering both procedures.

The corpus processing included the generation of two frequency lists of the Greek sub corpus (narrative genre) and of the learner translator corpus of Italian (in the same topic). The inspection of the frequency lists revealed that the auxiliary verb “avere” and the time conjunctions are among the most frequent linguistic elements in the corpus and should be considered important in the language learning.

Another processing task that performed in the corpus was the concordancing. The concordances produced were manually aligned between source and translation texts at sentence level. Furthermore, we annotated all the grammatical errors found in the produced concordances and we isolated the linguistic features described above. Some issues worth of discussing are the following:

- a. The frequency lists show a significant use of the time conjunctions, in particular “quando” (when), “mentre” (while) and “appena” (as soon as). The errors observed in the concordances with “mentre” and “appena” seem to be categorized more as learners’ errors than translators’ based on similar previous research in the Greek ~ Italian pair (Florou, 2009).
- b. The prevalent tense used in the translators’ corpus is “passato prossimo” (an almost equivalent tense to present perfect) which is compiled by an auxiliary verb and a participle. We can notice a number of errors in verbs in “passato prossimo” that are related to the choice of the auxiliary (that might be either “essere” or “avere”).

The above-mentioned linguistic elements are important not only due to their frequency, but also to their significance in teaching them in Greek learners of Italian language, since frequently these errors in the written production of Italian can be associated with mother tongue transfer.

#### 5. Results

The two linguistic features chosen to be analyzed produced some results worthy to discuss as follows.

### 5.1 Time Conjunctions

Narration is a skill that students of the foreign or second language, as mentioned above, ought to achieve rather early. Furthermore, for Greek students, Italian time conjunctions are considered easy due to a direct matching of these connectors to Greek. This should facilitate the process of learning as well as the work of the translator. During the error annotation thought, we observed that two of those conjunctions created translation and grammatical inaccuracies. A quantitative description of the errors observed can be found in

Table 2.

Table 2. Error frequency of “mentre” and “appena” in our corpus.

CONJUNCTION	ENTRIES	NUMBER OF ERRORS	NUMBER OF INTERFERENCE ERRORS
<b>Mentre</b>	53	6 (11%)	5 (9%)
<b>Appena</b>	16	3 (19%)	2 (12.5%)

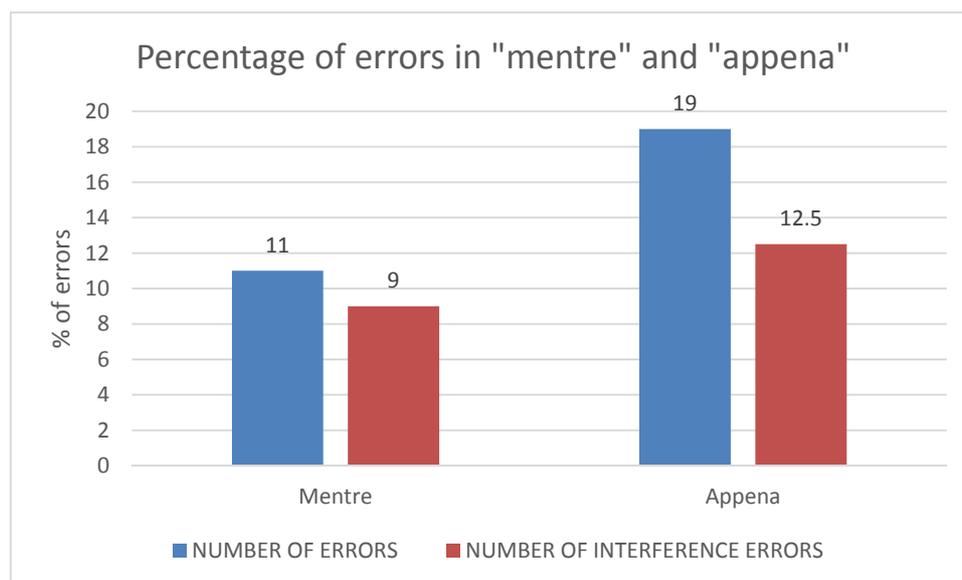


Figure 1. Related to Table 2

According to the table and the grapheme above, the erroneous phrases are 12.5% for “appena” and only 9.4% for “mentre”, and almost all of them are transfer errors. Some relevant examples follow:

Il mio viaggio di conoscenza era *appena* è cominciato. (dimitropoulou\_topic1\_it.txt)

Μόλις είχε αρχίσει το ταξίδι της γνώσης μου. (dimitropoulou\_topic1\_gr.txt)

(My knowledge trip *just* was beginning)

CORRECT FORM: Il mioviaggio di conoscenza era *appena* cominciato.

It is obvious that, even though it seems simple to apply the rule that says “appena”, among other conjunctions, should be placed between auxiliary verb and past participle when used in past tense (Note 1), learners are committing errors similar to this caused by mother tongue influence. In fact, the translation of the previous example is a “word to word” procedure and that is why “appena” is placed before the auxiliary. This is not considered a translation error; on the contrary, it is a language error as it is shown when parallelized to the equivalent phrase in the Greek corpus.

*Mentre* hanno creduto che lo abbia superato. (Christodoulou\_topic1\_it.txt)

Ενώ πίστευαν ότι το είχε ξεπεράσει... (Christodoulou\_topic1\_gr.txt)

(While they have believed that he has overcome it...)

CORRECT FORM: *mentre credevano* che lo avesse superato

The above-mentioned example shows that the rule for the conjunction “mentre” is not followed, although it is a non-exceptional rule (Note 2) that says that in a context of past “mentre” is always followed by past continuous (“imperfetto”); in this case the correct form of the verb is “credevano”. When we look at the equivalent text in the Greek Corpus, we notice that the tense that it is been used in Greek is of the same significance as the erroneous tense in the Italian translation (“passato prossimo”)

According to language teachers’ beliefs this kind of errors are not expected in this level of language knowledge. Nevertheless, trainees of translation are expected to commit errors more related to translation phenomena (such as the use of incorrect word, omission, disambiguation, idioms, etc.) Translator Learner Corpora provide us with increased explanatory power to understand the real causes of the errors.

### 5.2 Auxiliary AVERE

The second most sticking type of error in this corpus and most significant due to its role in the narration is the auxiliary “avere” (have). Indeed, auxiliary verbs in Greek language do not fully correspond to Italian. Furthermore, this topic is usually stressed in L2 books focusing on grammar rules. Greek students tend to transfer spontaneously the use of auxiliary from their mother tongue. They tend to connect the use of the Past tense in Greek with the most frequent auxiliary verb in Italian Passato Prossimo since they want to simplify the rule and create semantically equivalent phrases with the least effort.

In the table below (

Table 3) we present all the errors of “avere” as an auxiliary in the composed tenses and their break down in different error types:

Table 3. Error frequency of “avere” in our corpus.

ENTRIES	NUMBER OF ERRORS	ERRORS OF CHOICE IN AUX	SPELLING ERRORS OF PARTICIPLE	ERRORS OF CHOICE IN TENSE	OTHER
1607	67	33 (50 %)	25 (37%)	4 (6%)	2 (3%)

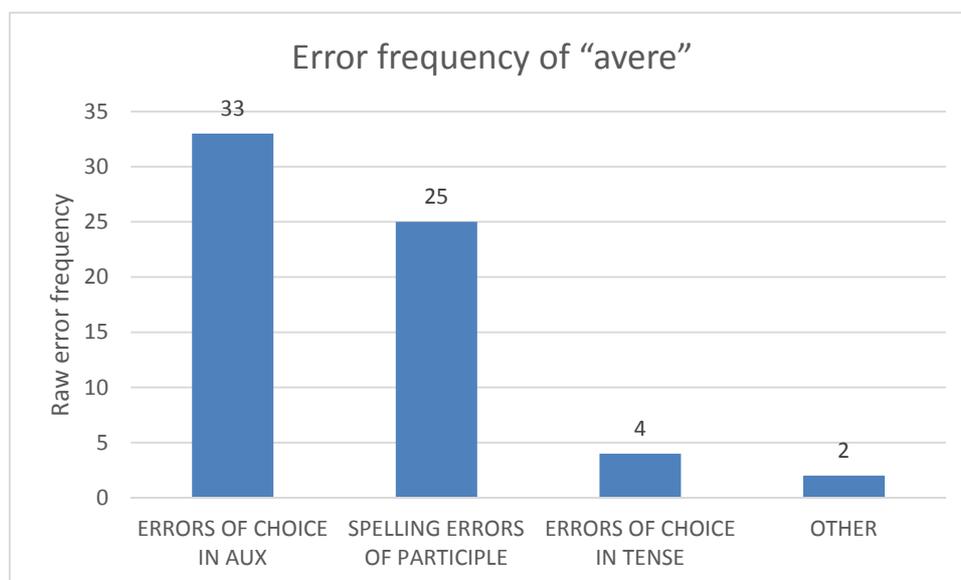


Figure 2. Related to Table 3

It is obvious that the problem is focused on the choice of the right auxiliary. In these 32 cases, students/translators should have chosen “essere” but they preferred “avere” for two reasons:

- 1) They connect “passato prossimo” (which in Greek matches to Past Tense) to this type of combination

(Avere+Participle) only, as we can notice at the following examples:

- a. Ed 'e venuto! *Ha seduto* serio, ma non triste... (Malama\_topic1\_it.txt)  
Και ήρθε! *Κάθισε* σοβαρός αλλά όχι θλιμμένος...(Malama\_topic1\_gr.txt)  
(And he came. *He sat* serious but not sad...)  
CORRECT FORM: Ed 'e venuto! *'E seduto* serio, ma non triste, ...
  - b. E dopo poco tempo, *ha diventato* guida in un forno... (Katsou\_topic1\_it.txt)  
και μετά από καιρό [γύρισε] και *έγινε* οδηγός σε φούρνο...(Katsou\_topic1\_gr.txt)  
(and after a while, *he became* driver for a bakery...)  
CORRECT FORM: e dopo poco tempo, *è diventato* guida in un forno...
- 2) They transfer the auxiliary «έχω» (have) from the Greek composed tenses, as an “avere” which has the same meaning. This is shown in the following examples:
- a. *Ha rimasto* in carcere 10 anni. (Tanouzi\_topic1\_it.txt)  
*Έχει μείνει* στη φυλακή 10 χρόνια.(Tanouzi\_topic1\_gr.txt)  
(*He has remained* in prison for 10 years)  
CORRECT FORM: *'Erimasto* in carcere 10 anni.
  - b. E la porta *ha aperto e ha entrato* mio papa... (Malama\_topic1\_it.txt)  
Η πόρτα *έχει ανοίξει* και *έχει μπει* ο μπαμπάς μου...(Malama\_topic1\_gr.txt)  
(The door *has opened* and my dad *has entered*...)  
CORRECT FORM: E la porta *è aperta e è entrato* mio papa...

In both cases, interlanguage is present as a factor of interference. This may be justified for beginners in language learning process, but as a phenomenon is rather unexpected for advanced learners of the foreign language, or even more for translator trainees. As shown also in the case of the conjunctions (see 5.1), the exploitation of Translator Learner Corpora is providing to researcher the necessary evidence for reliable conclusions.

## 6. Conclusions

Learner Translator Corpora can offer significant insights into the L2 learning processes. In this paper we presented the creation of a new linguistic resource that focuses in the parallel study of a number of frequent grammatical elements in Greek and Italian. The structure of the compiled corpus enables us to explore the nature of the errors observed, i.e., whether they belong to the learners' errors or the translators' and which is the role of the interlanguage.

The quantitative description of the errors observed revealed that two frequent time conjunctions (“mentre”, “appena”) are not used correctly, i.e., they are not placed correctly inside the phrase. Moreover, the auxiliary “avere” is used erroneously in cases where “essere” should be preferred.

In this way, we have shown that teachers' belief about the elements that an advanced learner should have assumed is not always a certainty. In the case of this study the elements under consideration are found erroneous not from a translative point of view, as expected, but due to mother tongue interference and interlanguage influence.

Future research will address whether it is useful and fruitful for students to be engaged in this kind of activity. It is shown that some results on Learner Corpora are applicable directly on the language acquisition process (Granger & Tyson, 1996) In our case translation trainees could use the Italogreco Corpus for creating language activities in which they should correct (their own, or their colleagues') mistakes.

In addition, this study could be expanded in Learner or Translator Corpora compiled of the same genre texts, from learners of a higher level in the same language or of the same level in another language (e.g., French) so that a comparison could be performed. In this case, we can see whether similar forms (e.g., “passé composé” and “passato prossimo”) are having the same impact to Greek students.

## Acknowledgments

I would like to thank all students of the fourth and sixth semester of the Department of Italian language and literature and the department of English language and literature of the Faculty of Philosophy, at the University of Athens. Without their help and their willingness this project could not be possible. Also special thanks to my

supervisor professor George Mikros for all practical and spiritual support.

## References

- Ajimer, K., & Altenberg, B. (2000). The English Swedish Parallel Corpus: A resource of contrastive research and translation studies. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 15–33). Amsterdam and Atlanta: Rodopi.
- Altenberg, B. (2002). Advanced Swedish learners' use of causative make: A contrastive background study. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (vol. 6, pp. 37–56). Amsterdam & Philadelphia: John Benjamins.
- Altenberg, B. (2002a). Using bilingual corpus evidence in learner corpus research. In S. Granger, J. Hung, S. Petch-Tyson & J. H. Hulstijn (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 37–54), Amsterdam & Philadelphia: John Benjamins. <https://doi.org/10.1075/llt.6.06alt>
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173–194. <https://doi.org/10.1093/applin/22.2.173>
- Castagnoli, S. (2010). *Variation and regularities in translation: insights from multiple translation corpora*. UCCTS 2010-Using Corpora in Contrastive and Translation Studies
- Castagnoli, S., Kunz, K., Kübler, N., & Volanschi, A. (2006). *Designing a learner translator corpus for training purposes* (pp. 1–19). Proceedings of Teaching and Language Corpora Conference TaLC 2006, Université Paris VII.
- Florén, C., & Sanz, R. (2008). The application of a parallel corpus (English-spanish) to the teaching of translation (entradproject). In M. Muñoz Calvo, C. Buesa-Gómez & M. A. Ruiz-Moneva (Eds.), *New Trends in Translation and Cultural Identity* (pp. 433–443). Cambridge Scholars Publishing.
- Florou, K. (2013) Comparing Error tagged Learner Corpora and learners' variables. *Modern Education Review* (ISSN 2155-7993).
- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15, 19–29. <https://doi.org/10.1111/j.1467-971X.1996.tb00089.x>
- Kutuzov, A., & Kunilovskaya, M. (2014). Russian Learner Translator Corpus: Design, Research Potential and Applications. International Conference on Text, Speech, and Dialogue, 315–323. [https://doi.org/10.1007/978-3-319-10816-2\\_39](https://doi.org/10.1007/978-3-319-10816-2_39)
- Rawoens, G. (2010). Multilingual corpora in cross-linguistic research: focus on the compilation of a Dutch-Swedish parallel corpus. In S. Bolasco, I. Chiari & L. Giuliano (Eds.), *JADT 2010: statistical analysis of textual data: proceedings of the 10th international conference* (1287–1294). Retrieved from <https://www.semanticscholar.org/paper/Multilingual-corpora-in-cross-linguistic-research%3A-Rawoens/5b301f1cbeedae2c1ec092aa65ff44db6682a01b>

## Notes

Note 1. Appena: Con valore temporale, da pochissimo; è sempre riferito a un v. di forma composta: a) con la forma completa del verbo si colloca tra l'ausiliare e il part.pass.: è a partito; (<http://dizionari.corriere.it/cgi-bin/sabcol/trova>).

Note 2. Attenzione!: in un contest passato la congiunzione mentre introduce sempre un imperfetto e mai un altro tempo:

([https://www.pul.it/cattedra/upload\\_files/16350/18%20Confronto%20fra%20passato%20prossimo%20e%20imperfetto.pdf](https://www.pul.it/cattedra/upload_files/16350/18%20Confronto%20fra%20passato%20prossimo%20e%20imperfetto.pdf)).

## Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).