Parallel-Learner Corpora or another way to confirm teachers’ intuitions

Katerina Florou

National and Kapodistrian University of Athens, Greece
Introduction

• The aim of the paper is to investigate learner and translator errors through learner translator corpora.
• The errors analyzed in this presentation are related to the narrative genre.
• The necessity of conducting this research is to verify or reject teachers’ intuition i.e. *although students learn certain structures early enough, even in an advanced level of the foreign language, or during their translation training, they tend to misuse them, due to their mother tongue interference.*
Teachers’ assumptions

• The advanced learners (or translator trainees) should know how to handle narrative speech since:
  ▫ They can use the time conjunctions which are easy to be matched to the ones in their mother tongue
  ▫ They have been taught in the early stages of learning the required tenses
Teachers’ intuition

• Advanced language learners (or translator trainees) commit mistakes during narration in the foreign language.

• Probably the mistakes are due to the transfer of similar forms or elements of their mother tongue.
Learner Translator Corpus

- **Place**: National and Kapodistrian University of Athens (Greece)
  - Department of Italian Language and Literature
  - Department of English Language and Literature
- **Participants**: Students
  - Of the first semester in Italian language
  - Of the first semester in English Language
- **Source text**: Essays written in Greek of two types of speech
  - Narrative speech
  - Academic speech
## Original Texts and Translations

<table>
<thead>
<tr>
<th></th>
<th>Original texts in Greek</th>
<th>Translations in Italian</th>
<th>Translations in English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative speech</td>
<td>77 texts</td>
<td>40 texts</td>
<td>37 texts</td>
</tr>
<tr>
<td></td>
<td>130,684 tokens</td>
<td>76,846 tokens</td>
<td>73,558 tokens</td>
</tr>
<tr>
<td>Academic speech</td>
<td>78 texts</td>
<td>41 texts</td>
<td>37 texts</td>
</tr>
<tr>
<td></td>
<td>131,769 tokens</td>
<td>75,394 tokens</td>
<td>69,975 tokens</td>
</tr>
</tbody>
</table>
Texts’ metadata

• Topics: two obligatory topics to write and translate in Italian or English
  ▫ “Describe the best and the worst moment of your life”
  ▫ “Which are the biggest social problems of our times”

• Time: No limit of time

• Size: Approximately 2000 words

• Tools: No suggested tools
Students’ metadata

- Age
- Sex
- Origin
- Knowledge of foreign languages
- Level in foreign languages knowledge
Corpus *Italogreco*

Contains:
- The narrative part of the whole corpus
- The translations in Italian language only

Analyses the errors of:
- The learning process
- The translation process

That’s why the typology of errors could be characterized as mixed
Processing the Italogreco

• **Frequency lists:**
  ▫ Revealed elements such as the Auxiliaries (“essere” and “avere”) and Time Conjunctions as most frequent in this genre

• **Concordancer:**
  ▫ Aligned source and translation texts at sentence level
  ▫ Having annotated the errors we isolate the elements in question

• **Error Annotator:**
  ▫ The tool used called “Episimiotis” was enriched with error translation tags
Linguistic Features under discussion (1)

- **Time conjunctions:**
  - Teaching time conjunctions is considered easy and because of their usefulness they are taught in early stages.
  - For Greek learners is extremely helpful matching time conjunctions in Italian with the ones in Greek as to their meaning.
  - In addition time conjunctions’ teaching is related with a number of rules, that learners have to apply systematically, within a number of examples and activities.
Time conjunctions in *Italogreco*

- Most of the frequent time conjunctions in the Italian language are used in the Corpus, such as:
  - Quando (when) 263 entries, appena (just) 16 entries, mentre (while) 53 entries, finchè (until) 2 entries, prima (before) 68 entries

- The most striking errors were those on the conjunctions considered easy by teachers and learners: “mentre” and “appena”
## Time Conjunctions

<table>
<thead>
<tr>
<th>CONJUNCTION</th>
<th>ENTRIES</th>
<th>NUMBER OF ERRORS</th>
<th>NUMBER OF INTERFERENCE ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentre</td>
<td>53</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Appena</td>
<td>16</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Comments on “mentre”

- Even though the errors of “mentre” are only a small amount (11%), almost all of them are annotated as interference errors. It is easy to verify the above looking at the (word to word) translation:
  - *mentre hanno creduto che lo abbia superato*..
    (Christodoulou_topic1_it.txt)
  - While *they have believed* that *he have overcome it*
  - Ενώ έχουν πιστέψει ότι το έχει ξεπεράσει...
- While the rule clearly says that in a context of precedent the conjunction “mentre” introduces an “imperfetto” (past continuous), and never a different tense. The correct form would be:
  - Mentre *credevano che lo avesse superato*...
Comments on “appena”

The 19% of “appena” entries is erroneous, but it is clear that are interference mistakes. An example:

- Il mio viaggio di conoscenza appena era cominciato. (dimitropoulou_topic1_it.txt)
- My trip of knowledge just was beginning
- το ταξίδι της γνώσης μου μόλις είχε αρχίσει.

The rule regarding the position of “appena” determines that when this conjunction is complementary to a composed tense, its collocation is between the auxiliary and the past participle. The correct form would be:

- Il mio viaggio di conoscenza era appena cominciato
Linguistic Features under discussion (2)

- Auxiliary “avere”
  - “Passato Prossimo” is the most used tense for the past therefore for narration.
  - Two auxiliary verbs are used to form this tense: “essere” (to be) and “avere” (to have).
  - “Avere” is used in most cases (all transitive and some intransitive verbs).
  - The majority of the Italian language course books contains a list of regular and irregular verbs (regarding their participle) that need either “avere” or “essere” to form “Passato Prossimo”.
  - Greek learners have to memorize and practice thoroughly, not only the list of the verbs, but also the rules on forming this particular tense, in a very early stage of learning the Italian language.
## Auxiliary “avere”

<table>
<thead>
<tr>
<th>ENTRIES</th>
<th>NUMBER OF ERRORS</th>
<th>ERRORS OF CHOICE IN AUX</th>
<th>SPELLING ERRORS OF PARTICIPLE</th>
<th>ERRORS OF CHOICE IN TENSE</th>
<th>OTHER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1607</td>
<td>64</td>
<td>33</td>
<td>25</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Comments on auxiliary “avere”

- It is obvious that in such a big amount as 1607 entries, only 4% is erroneous, but according to teacher “no mistake is forgiven in such an easy linguistic phenomenon!”
- Half of the errors are as a result of erroneous choice of auxiliary i.e. instead of using “essere” they use “avere”.
Examples of errors

• In learners’ errors there is a direct transfer of the verb “έχω” (to have) as an auxiliary, because of its similar use in Greek language, like in the examples bellow:
  ▫ *Ha rimasto* in carcere 10 anni. (Tanouzi_topic1_it.txt)
  ▫ *He has remained* in prison for 10 years.
  ▫ Έχει μείνει στη φυλακή 10 χρόνια.

• The correct form is:
  ▫ *É rimasto* in carcere [per] dieci anni.
  ▫ E la porta ha apperto e *ha entrato* mio papa...
    (Malama_topic1_it.txt)
  ▫ The door has opened and my dad *has entered*..
  ▫ Η πόρτα έχει ανοίξει και έχει μπει ο μπαμπάς μου...

• The correct form is:
  ▫ E la porta è stata aperta ed *è entrato* mio papà...
Conclusions

• We focused on two serious obstacles regarding the narration in translation, suspecting that this specific area needs more study.

• The quantitative description of the errors observed, revealed that two frequent time conjunctions (“mentre”, “appena”) are not always used correctly.

• Moreover, the auxiliary “avere” is used erroneously in cases where “essere” should be preferred.

• It is shown that advanced learners or translator trainees tend to transfer their mother tongue structures to Italian translations and tend to deviate from simple rules, that have already practice extensively.

• We assumed that the nature 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’. But as shown the interference (or influence from the source language) is common in both groups of participants.
Future Research

• Enlargement of the corpus by using more texts of the same genre to verify the presented results.
• Enlargement of the corpus by using an other genre of the same or other translated languages (an English learner translator corpus is already compiled).
• Extension of the corpus in other types of comparison e.g. learner translator corpus vs translation corpus, or learner translator corpus vs native speakers corpus.
• Application of the results in the development of FLT course books.
Thank you
Grazie
Ευχαριστώ