

Original Paper

An Application of Learner Corpus Analysis in Grammatical Comprehension—*Task Based Language Teaching and Corpus*

Linguistics

Katerina Florou¹

¹ National and Kapodistrian University of Athens, Greece

Received: May 19, 2025

Accepted: May 27, 2025

Online Published: June 20, 2025

doi:10.22158/fce.v6n1p1

URL: <http://dx.doi.org/10.22158/fce.v6n1p1>

Abstract

In this study, inspired by the corpus error analysis, its pedagogical application and the directions of Task Based Language Teaching, we compared the written production of two groups of students. The first was a group that applied the focusing on form phase by observing the past participle, in examples from native speakers and in grammar rules. The second was a group of students that applied the focusing on form phase by observing their own errors and comparing them with corresponding examples of a native speaker corpus. To investigate our hypothesis, we developed four learner corpora. The first two (one of each group of students) were produced during the task phase and the other two at the end of the lesson as a follow up activity. The results of the comparison showed that students' exposure to their own errors can improve the comprehension of a grammatical element (Note 1).

Keywords

Task Based Learning, Corpus Driven Learning, Focus on form, Italian as foreign language

1. Introduction

In recent years, Task-Based Language Teaching (TBLT) has emerged as a prominent approach within the field of language education, praised for its emphasis on using tasks to drive learning. According to Bula-Villalobos and Murillo-Miranda (2019), “Task-based Language Teaching alludes to a communicative approach which is grounded in the usage of tasks (real-life and pedagogical) as a central element for language instruction.” This definition underscores the centrality of tasks in facilitating meaningful language use and acquisition, aligning with the broader communicative approach that prioritizes interaction and practical application over rote memorization and isolated grammar drills.

One of the significant aspects of TBLT is its focus on form analysis to enhance language awareness among students. Willis (2016) argues that analyzing linguistic forms within tasks helps students explore language intricacies and develop a nuanced understanding of different aspects of the foreign language. This process involves engaging with both correct and incorrect forms, offering a dual pathway to linguistic competence. By examining correct forms, learners can internalize accurate language usage, while analyzing errors provides insights into common pitfalls and areas requiring further attention.

The use of learner corpora, particularly error-tagged corpora, is a powerful tool in this regard. Learner corpora consist of textual data produced by language learners, which can be systematically analyzed to identify patterns of errors and developmental stages in second language acquisition (SLA). According to Granger (2002), “Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose.” These corpora offer authentic insights into learners’ interlanguage—the evolving linguistic system that learners create as they progress toward full proficiency.

However, focusing on erroneous forms, while informative, carries inherent risks. As Ellis (2006) notes, teachers “should seek to address errors or gaps in the students’ L2 knowledge” during task performance. This corrective approach aims to scaffold learners’ development by providing timely feedback and targeted instruction. Yet, an overemphasis on errors might lead to a disproportionate focus on negative aspects of learner language, potentially affecting learners’ confidence and motivation. It raises the question of whether learners’ interlanguage, characterized by its developmental nature and variability, can be considered truly “authentic.” (Note 1)

The authenticity of learner language is a contentious issue in SLA research. While errors are an integral part of the learning process, representing learners’ current stage of development, they also reflect the ongoing struggle to master the target language. This dual nature of learner language—as both a genuine reflection of learner progress and a site of imperfection—poses a challenge for educators. Balancing the need to address errors with the goal of fostering a positive, confidence-building learning environment is crucial.

In this context, TBLT offers a robust framework for integrating form-focused instruction with communicative practice. By embedding language analysis in meaningful tasks, TBLT encourages learners to apply their knowledge to practical, real-world scenarios. This task-oriented approach not only promotes linguistic accuracy but also enhances learners’ overall communicative competence. The dynamic interplay between task performance and form analysis facilitates a more holistic understanding of the target language, supporting learners’ gradual and sustained language development.

In conclusion, TBLT represents a significant advancement in language pedagogy, emphasizing the use of tasks to create engaging and effective learning experiences. By incorporating form analysis and leveraging learner corpora, TBLT provides a comprehensive approach to language instruction that

addresses both the strengths and challenges of learner language. As educators continue to refine their practices, balancing error correction with the cultivation of a supportive learning environment will be key to unlocking the full potential of TBLT in fostering language proficiency.

2. Literature Review

What is the connection between TBLT and Corpus Linguistics? Firstly, TBLT emphasizes the importance of using authentic language samples, as highlighted by Ellis (2005) in his principles for TBLT: “Exposure to authentic language is significant; Language should be used for real purposes; tasks need to motivate students to use language; a focus on language should be established.” Later, Ellis (2009) integrated a focus on form with error and corrective feedback, suggesting that teachers should use corrective feedback strategies to address language form. Corrective feedback is linked to error analysis, considering that errors represent forms, and the providers of corrective feedback are often the learners themselves (classmates) (Su & Tian, 2016).

Conversely, Corpus Linguistics has long sought applications in language teaching. Flowerdew (2009) discusses the pedagogical use of Corpus Linguistics, distinguishing between corpus-based and corpus-driven teaching, and directly associates it with Data-Driven Learning (DDL). Boulton and Tyne (2013) similarly provide a critical overview of these approaches. According to Zaki (2020), “the field has witnessed a tremendous surge in corpus-based teaching approaches in second/foreign language classrooms, as teachers around the world realize the importance of language data in their classrooms and develop pedagogically sound means of incorporating corpora in their syllabi and classroom activities.” Numerous studies have explored the relevance of corpora to language teaching, such as Bernardini’s (2004) suggestions and the book by Granger, Hung, and Petch-Tyson (2002), which presents modern ideas beyond traditional concordances. More recently, research by Lacková (2022) and Aguilar and Luna have examined the application of Corpus Linguistics in teaching vocabulary and terminology.

Additionally, there are some studies that connect TBLT with the teaching of Italian as a foreign language. For example, the use of TBLT in a second-semester elementary Italian course (Means, 2020). This case study with its qualitative analysis of the students’ impressions demonstrates that the students believed the course was effective and enjoyable and met the overall goal of TBLT: What they learned in the classroom could be put to practical usage in the outside world. There are also studies that connect TBLT with technology: Baralt’s and Gomez (2017) paper which is a methodological guide for how to do TBLT via online video interactive tutorials, or Tomas’ and Reinders’ (2015) book with a collection of studies and experiment reports on how one can implicate the new in Task based teaching.

Despite the extensive research in both fields, there remains a noticeable gap in linking Corpus Linguistics with TBLT, even though both are fundamentally connected to language teaching.

3. Research Question

Bearing in mind that decades after Granger et al. (2002) defined learner corpora, and several years after Ellis (2006) proposed “focusing on form” during any phase of a TBL framework, the present study addresses the following question: How can learner corpora, whether error-annotated or not, provide opportunities to focus on form in an educational environment?

If corpus analysis and TBLT have developed parallel paths over the past 30 years, both aiming to enhance teaching practices, is it possible to effectively combine them in the classroom?

4. Methodology

In order to integrate TBLT with corpus-based learning and conduct a study based on learner corpora, we combined language courses with error analysis courses. Specifically, we used a course of the Department of Italian language and Literature. The experiment was conducted during a three-hour lecture of the “Error Analysis in the Italian Language” course.

4.1 The Learners

The students from the University Department of Italian Language and Literature are all adults, either Greeks or long-term residents of Greece, with Greek as their first language. Upon entering the department, their Italian proficiency is at a B1+ level. Throughout their studies, they take courses in theoretical and applied linguistics, Italian culture, and language courses, aiming to achieve a C2 proficiency level after four years. Their professional goal is to teach the Italian language. For this research, we selected 30 participants who were at a B2+ level, all of whom were in their third year of studies.

We divided the students into two groups randomly, and we asked them to work on their written production in Italian. Then a comparison of the written production of the two groups of students was conducted:

The first group (G1) aimed to apply the focusing on form phase by observing a particular form of the Italian language, the *past participle*, in examples from native speakers and in grammar rules.

The second group of students (G2) aimed to apply the focus on form phase by observing their own errors and comparing them with corresponding examples of a native speaker corpus so as to improve their performance in this form (e.g., past participle).

4.2 TBLT Framework and Corpus Development

Throughout the entire experiment, and for the purpose of investigating the hypothesis, we developed four learner corpora while following the TBLT framework. First, as a pre-task phase, the students watched the trailer of a classic American film, “Vacanze Romane”, dubbed in the Italian language.

During the task phase, the first two sub-corpora, one from each group of students, were produced. The students were required to deliver a text responding to the topic “Describe your experience of your last visit to Italy” with a minimum word limit of 300 words. The collection of these texts from the first group constitutes the G1 corpus of the task phase, and the collection of texts from the second group

constitutes the G2 corpus for the same phase. At the end of the course, as a follow-up activity, they had to produce another written text on the same topic and of the same length (see 3.4). The collection of these texts forms the G1 and G2 corpora of the follow-up activity.

A presentation of the texts produced during both phases by both groups is available here. As evidenced, not all students completed the follow-up activity; however, the word count of the two corpora remains comparable:

Table 1. The Corpus and the Sub-Corpora

	Task Phase		Follow up activity	
	G1	G2	G1	G2
n. texts	15	15	12	13
n. words	4317	4,509	4,589	4,726

The students who did not continue with the follow-up activity are those who did not complete the courses either for health reasons or due to workload

4.3.1 Focus on Form (First Group)

The first group of students began the first stage of the focus-on-form phase—after having written their texts—by presenting or reading them in class. In the second stage, students focused on the form of the past participle in three ways: they discussed in pairs the grammatical elements of a tense, e.g., *passato prossimo* and the use of the past participle in particular; they practised simple activities on the grammatical structure; and they listened to excerpts featuring the use of “*passato prossimo*” by native speakers.

4.3.2 Focus on Form (Second Group): “Error Annotation”

The students of the second group commenced the focus-on-form stage by creating a learner corpus, which involved compiling the texts produced by both groups with the assistance of the teacher. Following the creation of this corpus, the texts were randomly distributed among the students of the second group for manual error annotation, focusing exclusively on errors in verb usage.

Subsequently, as a group assignment, the students categorized the annotated errors based on their frequency. To facilitate this process, they utilized the software AntConc. In the final phase of their assignment, the students conducted a comparative analysis by searching for the same types of verb errors in a native speakers’ corpus. For this study, the Coris Codis corpus was used, allowing the students to identify and analyze the differences between the correct forms used by native speakers and the erroneous forms identified in their learner corpus.

4.3.2.1 Tag Set

The primary objective of this group is to develop a tag set that is straightforward and user-friendly for students to utilize during annotation tasks. The chosen tag set employs a single-level classification system that delineates the form of the error. Given that the part of speech is consistently the verb, annotators are provided with three principal categories: tense, person, and number. Additionally, two supplementary categories are included for instances where the erroneous form comprises two elements: auxiliary for incorrect grammatical or lexical choices of the auxiliary verb, and participle for incorrect grammatical or lexical choices of the participle. A residual category, another, is available for errors that do not fit easily into the aforementioned classifications.

Table 2. Tagset for Error Annotation

Tempo/Tense	T
Persona/Person	P
Numero/Number	N
Ausiliare/Auxiliary	AUS
Participio/Participle	PAR
Altro/Other	A

Observing the steps of the focus-on-form phase, it is evident that both groups began their respective activities immediately after the task phase. However, the nature of these activities differed between the groups. The first group transitioned into the focus-on-form phase following an individual activity, which emphasized personal engagement and reflection. In contrast, the second group entered the focus-on-form phase after engaging in a collaborative task that was more technical in nature, rather than purely pedagogical.

In summary, it can be stated that students in the second group were obligated to complete supplementary preparatory tasks prior to participating in the core activities of the focus-on-form phase. These preliminary tasks involved creating and annotating a learner corpus, which provided a technical foundation for their subsequent focus on grammatical accuracy and error analysis. This added layer of engagement may have influenced the nature and depth of their learning experience during the focus-on-form phase.

Table 3. The Stages of the Lesson in the Two Groups

Stages	First group (G1)	Second group (G2)
1 st	Present or read the texts in class	Collect texts and create a corpus with anonymized texts in group
2 nd	Discuss the grammatical elements of the tense in pairs	Select texts randomly and annotate errors on verbs manually

3 rd	Practice on activities on authentic texts	Categorize the errors by frequency using the Ant Conc in group
4 th	Listen to samples of natives on the same topic	Find analogies in NS Corpora using Coris Codis

4.4 Follow up Activity

During the follow-up activity, both groups wrote their assigned texts and submitted them via the e-class platform. At this stage, the teacher/researcher was responsible for creating two distinct follow-up activity corpora, one for each group (G1 and G2). Subsequently, the researcher undertook the task of annotating and counting the errors related to parts of speech, with a particular focus on the past participle of verbs.

In this phase, additional error annotation was required by the researcher to ensure comprehensive analysis. The error counts for each of the four corpora were meticulously recorded. These counts will be discussed with the students during the feedback phase to facilitate a deeper understanding of their linguistic errors and to guide further learning.

The specific error counts in each of the four corpora will serve as a basis for this discussion, enabling students to identify common mistakes and learn strategies for avoiding them in the future.

Table 4. The Number of Errors in the Two Learner Corpora

	<i>Task phase</i>		<i>Follow up activity</i>	
	G1	G2	G1	G2
n. errors in verbs	67	81	51	53
n. errors in the form	18	20	15	9

4.5 Discussion in Class/Feedback

As an optional phase of the framework, Willis (2016) adds the follow-up activity, the last step of which involves discussing “how they [the students] felt about the task and the task cycle and what they might like to do next or sometime later (and/or note this down in their diaries for their teacher to read later).” This particular step was incorporated into this study and was applied under the name of feedback. The detailed feedback process is designed to enhance the students’ grasp of grammatical structures and to improve their overall language proficiency. Therefore, both groups first discuss their errors in pairs and then as a group, focusing on the reduction in their errors and trying to point out all the positive steps of their involvement throughout the learning process.

During this analysis, students of G1 shared their experiences with students of G2, highlighting how helpful the practice activities and native speaker oral samples were. Similarly, students of G2 discussed the originality of error analysis, the use of an NLP tool (AntConc), and the experience of searching within a corpus.

At the end of this roundtable discussion, the TBLT framework followed in this experiment was revealed to the students (and future language teachers) as follows:

Table 5. Application of the TBLT Framework on the Study

<i>TBLT framework (Willis, 2016)</i>	<i>Steps of the present study</i>
Pre-Task	Watch a piece of film in the target language
Task Cycle	
Planning	Planning the writing of the narrative text
Report	Write the narrative text
Post task	Present the text in the class/Create a corpus with the texts
Focus on form	
Analysis	Discuss the grammar/Error annotation
Practice	Practice on grammar/Compare with native speakers' corpora
Follow up	Re-write the narrative
Feedback	

Such demonstration provokes more discussion and opinion exchange about the link between Corpus Error Analysis and TBLT.

5. The Results

According to the results, several observations can be made. The comparison results indicate that students' engagement with their own errors can substantially enhance their comprehension of specific grammatical elements. Insights from the feedback stage reveal that learners demonstrated increased confidence when rewriting on the same topic. This boost in confidence extended beyond their use of the *passato prossimo* (past perfect) to their overall writing skills.

While "focus on form" does not necessarily equate to a focus on grammatical form alone—as Long (2000) highlights that focus can also center on aspects such as meaning or linguistic forms—this study suggests that learners were inclined to refine their written expression toward greater accuracy.

Students in G2 reported a significant shift in their perception of error analysis, which, in turn, influenced their behavior as learners. They observed a notable decrease in the number of errors they made and, more importantly, felt actively involved in the learning process, particularly through the practice of annotating their own mistakes. (Note 2)

6. Conclusion

One of the aims of this study was to investigate the effectiveness of integrating corpus-based teaching, error analysis, and task-based learning. The results of the final error analysis confirm that this combination can be beneficial for students. For this specific group of learners, who are also future teachers, such studies could serve as valuable tools for their future professional practice. Furthermore, the feedback results suggest that, in addition to improving learners' accuracy, this approach enhances their overall awareness of the learning process. As Baralt and Gomez (2017) state, "the analysis is the time to explicitly bring learners' attention to forms that arise from their own production," underscoring the importance of this reflective practice in language learning. This research also highlighted that relying solely on quantitative results could lead to inconclusive findings. A qualitative research technique is essential for evaluating the impact of a new teaching model comprehensively.

Recognizing the potential effectiveness of this methodology, it can be applied to other educational contexts or levels, provided there is a more detailed description of the educational scenario or the use of a different tag set.

References

- Aguilar, C., & Luna, F. (2023). "TODAY IS ANOTHER DAY TO FIND YOU": USING CORPUS LINGUISTICS FOR LANGUAGE LEARNING. In *ICERI2023 Proceedings* (pp. 8042-8050). IATED. <https://doi.org/10.21125/iceri.2023.2055>
- Baralt, M., & Morcillo Gómez, J. (2017). *Task-based language teaching online: A guide for teachers*.
- Bernardini, S. (2004). Corpora in the classroom. *How to use corpora in language teaching*, 12, 15-36. <https://doi.org/10.1075/scl.12.05ber>
- Boulton, A., & Tyne, H. (2013). Corpus linguistics and data-driven learning: A critical overview. *Bulletin suisse de Linguistique appliquée*, 97, 97-118.
- Bula-Villalobos, O., & Murillo-Miranda, C. (2019). Task-based language teaching: Definition, characteristics, purpose and scope. *International Journal of English Literature and Social Sciences (IJELS)*, 4(6). <https://doi.org/10.22161/ijels.46.39>
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209-224. <https://doi.org/10.1016/j.system.2004.12.006>
- Ellis, R. (2006). The Methodology of Task-Based Teaching, *Asian EFL Journal*, Volume 8, Number 3 September 2006 Conference Proceedings: Task-based Learning in the Asian Context 39.

- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal: An electronic refereed journal for foreign and second language educators*, 1(1). <https://doi.org/10.5070/L2.V1I1.9054>
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International journal of corpus linguistics*, 14(3), 393-417. <https://doi.org/10.1075/ijcl.14.3.05flo>
- Granger, S. (2002). *A bird's-eye view of learner corpus research*. <https://doi.org/10.1075/llt.6.04gra>
- Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20(3). <https://doi.org/10.1558/cj.v20i3.465-480>
- Granger, S., Petch-Tyson, S., & Hung, J. (2002). Computer learner corpora, second language acquisition and foreign language teaching. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 1-257. <https://doi.org/10.1075/llt.6>
- Lacková M. (2022). Enlarging Students Vocabulary with the Help of Corpus-Driven Data. *Changes and Challenges of Increasing the Quality of Preparation of Prospective Teachers in European Countries*, 67.
- Long, M. H. (2000). Focus on form in task-based language teaching. *Language policy and pedagogy: Essays in honor of A. Ronald Walton*, 179-192. <https://doi.org/10.1075/z.96.11lon>
- Means, T. (2020). Task-Based Language Teaching of Italian. *Italica*, 97(3), 567-587. <https://doi.org/10.5406/23256672.97.3.06>
- Thomas, M., & Reinders, H. (Eds.). (2015). *Contemporary task-based language teaching in Asia*. Bloomsbury Publishing.
- Willis, J. (2016). *A flexible framework for task-based learning*. (Revised article originally: In Challenge and change in language teaching, Eds. J. Willis and D. Willis, Oxford: Heinemann Macmillan).
- Zaki, M. (2021). Corpus - based language teaching and learning: Applications and implications. *International Journal of Applied Linguistics*, 31(2), 169-172. <https://doi.org/10.1111/ijal.12316>

Notes

Note 1. A pilot experiment from this study was presented at the Corpus Linguistic Conference 2023 at Lancaster University

Note 2. The data that supports the findings of this study are available from the corresponding author, upon reasonable request.