Can We Teach in a Better Way the Usage of the Verb Complements after Analyzing our Students Written Productions?
A Study in the Lexical Bundles of a Learner Corpus of the Italian Language

Katerina Florou

National and Kapodistrian University of Athens

Abstract

Some Corpora using automatic or semi-automatic tools can provide a list of all repeated word combinations (2-gramms, 3-gramms ect). This method has been used in the last twenty years to analyze lexical bundles in different genres and registers (Paquot & Granger 2012: 135). The results of these analysis are a significant addition to both linguistic research and foreign language teaching.
For the present paper, an enriched form of a corpus was used, which is also a diachronic project. The UniCorpus with a total of approximately 80,000 words was created during the last 4 years by students of Italian Language studying in a Greek University, in the department of Italian Language and Literature. The written productions that have been selected for this analysis are of the narrative genre. The Corpus was analyzed focusing at the n-grams that form the most frequent verbs and verb forms. The purpose is to observe whether there is an overuse or an underuse in specific lexical bundles, which are directly linked to patterns of verb complements.
Analyzing the results quantitatively, we can establish a persistence of the students in repetitive patterns. A possible explanation for this observation is the confidence they feel when they use the same phrases (Nesselhauf 2005: 69) which are probably the ones they know better. Moreover, it becomes apparent that few, but repeated errors of the students are related to the mother tongue and its interference (Paquot & Granger 2012: 136) in learning Italian.
The analysis of this kind of n-grams reveals a technique that can distinguish linguistic patterns related to a part of speech (in our case the verbs) or more, starting from those that are common to the whole corpus and proceeding to those that distinguish the linguistic variety of one student from another (Aarts & Granger 2014: 140). The above as a conclusion is immediately usable in teaching in order to improve both the students' performance and the targeting or the evaluation of the course.

Keywords: *Language patterns, learner corpora, teaching Italian language*

1. Introduction
The analysis of lexical bundles, which are repetitive word combinations, has become a popular research method in the field of corpus linguistics. This paper aims to investigate the usefulness of this method for improving foreign language teaching, specifically the usage of verb complements in Italian. The analysis is based on a corpus of written productions by students of Italian at a Greek university. The corpus was enriched and diachronic, covering a period of four years. The purpose of the analysis is to observe whether students tend to overuse or underuse specific lexical bundles related to patterns of verb complements, and whether these patterns are related to the students' mother tongue.

2. Background
Research on language acquisition suggests that learners acquire language through exposure to language input and interaction. There have been numerous studies over the past two decades that have utilized automatic or semi-automatic tools to analyze repeated word combinations, also known as lexical bundles, in various genres and registers. This method has been shown to provide valuable insights into language usage and has implications for both linguistic research and foreign language teaching.
One prominent study in this area is the work of Biber et al. (1999), who analyzed the frequency and distribution of lexical bundles in a corpus of written and spoken English. They found that lexical

bundles are a key feature of language use, and that they vary according to genre and register. This work has been influential in shaping subsequent research in the field.

Another important study is the work of Hyland (2008), who analyzed the use of lexical bundles in academic writing. He found that certain bundles are particularly common in this genre, and that they play a key role in creating coherence and cohesion in academic texts. This work has been useful for both linguistic research and for teaching academic writing to non-native speakers.

In addition to analyzing the frequency and distribution of lexical bundles, researchers have also examined their function and meaning in different contexts. For example, Biber and Conrad (2009) analyzed the use of lexical bundles in conversation, and found that they serve a range of functions, including expressing agreement, disagreement, and uncertainty.

Overall, the use of automatic or semi-automatic tools to analyze lexical bundles has proven to be a valuable approach in language research. By identifying common word combinations, researchers can gain insights into the structure and function of language in different contexts, and this knowledge can be applied to improve language teaching and learning.

3. Methodology

The UniCorpus, which contains approximately 80,000 words, was analyzed using the AntConc software (Anthony, 2022). It is a Learner corpus, the data of which was collected as part of a course in the Department of Italian Language and Literature, with the participation of 120 students to date. Students level (according to CEFR) is rather advanced, C1. The Corpus is divided in 4 sub corpora according to the year of production. The focus of this particular study was on the most frequent verbs and verb forms in the corpus, and the n-grams that formed these verbs and verb forms. The corpus consisted of written narratives, and the analysis was performed to detect any overuse or underuse of specific lexical bundles. The aim was to identify patterns related to verb complement usage and to assess whether these patterns were related to the students' mother tongue. The results were analyzed quantitatively to establish the persistence of students in using repetitive patterns, and to identify the errors that were related to the interference of the mother tongue in learning Italian.

Among the most frequent verbs appearing in this corpus are auxiliaries and modals. Observing the clusters that are formed in all the sub corpora of all four years we see a repeating pattern in the verbs Volere (want) and Sapere (know), which in Italian language and also in Greek -the students' mother tongue- are used in two ways: either with an objective complement that follows the verb and may be a subordinate clause or a noun, or with an infinitive, which follows the verb and in no case is a preposition interposed (as in other cases of verbs). The research was focused on clusters of three or four word following the verbs of interest. The results of UniC as to these patterns are shown below:

Patterns	Number of appearances	Patterns	Number of appearances
Sapere +phrase	44	Volere +phrase	13
Sapere +noun	3	Volere +noun	1
Sapere +infinitive	14	Volere +infinitive	62
Total	61		76

Table 1: Number of appearances in all cases

Examples from the corpus in the patterns of both verbs:

...sapevo che esistono.. (I knew they exist)

...sapevo parlare un po'.. (I knew speaking a bit)

Sapevamo esattamente di cosa si tratta. (We knew exactly what it was about)

...ho saputo lavorare a distanza. (I learned to work online)

...sapiamo molto bene la sua sorella. (we know very well your sister)

...voglio analizzarla perché era un'esperienza brutta... (I want to analyze it because it was a very bad experience)

Lo voglio per me stessa... (I want it for myself)

Vogliamo sradicare dalla memoria... (we want to from our memory)

Volevo che i miei genitori [...] ammalassero. (we wanted that our parents got sick)

Ho voluto di andare a fare la spesa... (I wanted to go and make some shopping)

From the total appearances of those two verbs, the infinitive form was excluded. As it is shown at the examples, the verb types used by the learners in this learner corpus were mostly in past tenses (imperativo-imperative and passato prossimo-past perfect) and in present tense, which is logical given the genre of narration. However, in the above results (and in some of the examples) they are some mistaken clusters. The cases of errors are as shown in the table below:

Patterns	No of cases	Errors	Patterns	No of cases	Errors
Sapere +phrase	44	3	Volere +phrase	13	5
Sapere +noun	3	0	Volere +noun	1	0
Sapere +infinitive	14	1	Volere +infinitive	62	4
Total	61	4		76	9

Table 2: Number of appearances and errors

The total of errors compared to the total of appearances is, at first glance, not important. Nevertheless, there is need to examine exactly where these mistakes are located, because students seem to often use the patterns voglio/so+che+verb and voglio/so+infinitive, precisely, because they feel safe with something that they know how to use, and it is familiar to them from their mother tongue. However, the secondary verb in the first pattern is in some cases in the wrong tense or in the wrong declension and a preposition in the second pattern is inserted between voglio/so and infinitive, in a way that strongly resembles a pattern from the Greek language: θέλω να/ξέρω να.

Examples of errors of both cases:

...ho saputo a guidare ... (I learnt to drive)

Voglio di essere ingrata… (I want to be grateful)

…vuole che si interessa… (he wants that he is interested...)

4. Results and Discussion:

The analysis of the UniCorpus revealed that students tend to persistently use repetitive patterns of verb complement usage in their written productions. This persistence can be explained by the students' confidence in using familiar phrases (Nesselhauf, 2005). Additionally, the analysis revealed that few, but repeated errors made by the students were related to the influence of their mother tongue on learning Italian. These errors were identified by analyzing the n-grams related to verb complement usage.

The analysis of the UniCorpus also revealed that this method can be used to distinguish linguistic patterns related to the usage of a specific part of speech, such as verbs. This technique can also be used to identify individual linguistic variety among students (Aarts & Granger, 2014).

In order to use the results in teaching Italian as a foreign language we have to take into consideration that in this level of proficiency students' interlanguage is far from mother tongue interference (Granger, 2015), therefor they tend to rely on syllabus and textbooks. It is obvious, that syllabuses and textbooks can be improved if they follow the observations of such research. Moreover, examples like the ones that were registered in this study can assist students to identify their errors and improve their performance in Italian language. Clusters and n-grams of word retrieval tools are useful in the classroom activities and they can be used not only by teachers but also by students especially of this level. This type of learning based on their own data could be proved more stimulus.

5. Conclusion:

The analysis of lexical bundles related to verb complements in the UniCorpus has revealed the usefulness of this method for foreign language teaching. The persistent use of repetitive patterns among students can be addressed by focusing on the appropriate usage of verb complements in Italian. Additionally, the errors related to the interference of the mother tongue can be identified and addressed. The analysis of n-grams related to verb complements can also be used to target specific areas of the course for improvement. In the future a study on how these patterns can be taught through Learner Corpora can make a significant contribution to teaching.

References

[1] Aarts, J., & Granger, S. (2014). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In Learner English on computer, Routledge,132-141.

[2] Anthony, L. (2022). AntConc (Version 4.1.4) [Computer Software]. Tokyo, Japan: Waseda University.

[3] Biber, D. (1999). A register perspective on grammar and discourse: variability in the form and use of English complement clauses. Discourse studies, 1(2), 131-150.

[4] Biber, D., & Conrad, S. (2009). Register. Genre, and Style.

[5] Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, *1*(1), 7-24.

[6] Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. International journal of applied linguistics, 18(1), 41-62.

[7] Nesselhauf, N. (2005). Collocations in a Learner Corpus. Amsterdam, the Netherlands: John Benjamins.

[8] Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. Annual Review of Applied Linguistics, 32, 130–149