

Original Paper

Informal Correspondence by Greek Learners of the Italian Language: A Study Based on Learner Corpora, Native Corpora and Textbooks

Katerina Florou^{1*}

¹ Department of Italian Language and Literature, National and Kapodistrian University of Athens, Athens, Greece

* Katerina Florou, Department of Italian Language and Literature, National and Kapodistrian University of Athens, Athens, Greece

Received: July 31, 2019

Accepted: August 13, 2019

Online Published: August 28, 2019

doi:10.22158/fet.v2n3p159

URL: <http://dx.doi.org/10.22158/fet.v2n3p159>

Abstract

The aim of this study is to compare various lexical structures between a learner corpus of students with Italian as a foreign language and a reference monolingual Italian corpus. More specifically, the first is a learner corpus (part of a wider learner corpus) comprised of Greek students studying Italian as a foreign language while the second is the CWIC reference corpus of native Italian speakers.

The research findings help us explain the role of didactic material in comprehending linguistic structures that are found in informal letters/emails and, moreover, they provide us valuable information regarding the use of the same lexical structures by native speakers.

Keywords

foreign language teaching, contrastive linguistics, linguistic structures

1. Introduction

During the last two decades Corpora have established their position in the field of Computational linguistics and Foreign Language Teaching and have formed an autonomous sector (Corpus Linguistics) (McEnery, 2012).

This specific research is an attempt to detect converging and diverging elements in written speech produced by Greek students learning the Italian language and by Italian native speakers; furthermore, more generalized results could be expanded and produced which could be exploited by various teaching applications.

In this paper Learner Corpora and Native Corpora have been used in order to show that there are systematic differences between the student's interlanguage and native Italian language concerning the frequency of words, phrases and structures, with some elements being overused and others underused (Granger, 1998).

Learner Corpora are large collections of oral or written texts produced by learners of a foreign language (Granger, 2004). Consequently, more and more studies now turn to corpora as a source of primary data. While corpus-based methodologies have increased in sophistication, the use of corpus data is also associated with a number of unresolved problems (Arppe, Gilquin, Zeschel, Hilpert, & Glynn, 2010). Learner Corpora can be studied regarding the learners' errors or they can be compared with other learner corpora or Native Corpora. Such researches have been carried out in order to study, for example, the frequency of the use of verbs (Ringdom, 1998), the use of connectors (Altemberg & Tapper, 1998), the vocabulary (Lenko-Szymanska, 2005), the use of prepositions (Diez-Bedmar & Casas-Pedrosa, 2006), the use of modal verbs (Aijmer, 2002).

The present paper, even though it draws upon all the previous research experience, differs because it compares the same genre of written speech in the Italian language and, most of all, because the results of such research can be used effectively in Foreign Language Teaching.

Before introducing the main analysis, it would be useful to clarify why this kind of speech/genre was chosen to be studied. Firstly, it is very common during language certification exams (of all foreign languages and not Italian in particular) for learners to be asked to write an informal letter or email, so it is something which learners practice thoroughly. It is also a kind of writing that is taught at the beginning of the course, even before teaching basic linguistic structures and, finally, it is among the first written communication tasks.

Moreover, it is a very convenient communication activity, since students see the immediate application -mostly professionals and students who are familiar with online communication- and it does not require special guidance, as there is always something more to be said or written to a friend without the use of formulated expressions. This satisfies the need for simplified communication skills, which is covered with the use of basic vocabulary. That need is eminently evident in the early stages of foreign language acquisition, when the student seeks linguistic realization through lexical structures that he/she has already obtained in his/her mother language and searches for the lexical items with which he/she feels more familiar in the target language and, at the same time, constitute its own basic vocabulary (Ιακώβου, Μαρκόπουλος, & Μικρός, 2003).

This particular choice of genre also serves the purpose of applying the research results in teaching or in the development of textbooks.

2. Research Hypothesis

The present study examines whether the structures of opening and closing a letter/e-mail as well as the connective structures within the letter/e-mail written by the learners of Italian language are identical to those that have been taught in class and/or those that native speakers use. We observed that in most Italian language textbooks aimed at Greek students a template letter is used which includes some ways of addressing the recipient and others to conclude the letter. The students of the Italian language do not always apply the lexical rules and there is a need to determine whether these rules actually describe patterns of native speakers. Furthermore, the informal e-mail is one of the first communicative tasks taught in class (due to its simplicity) and it is useful to learn some words or phrases that will be used as connectors. Some connectors are already known and others are taught to facilitate this communicative task. But students and teachers often wonder if memorizing all these words is really effective, or if it is more natural to use fewer (as they might imagine Italians native speakers do). In order to study the abovementioned claim, we used Learner and Native Corpora to achieve both speed in data processing and reliability in findings.

3. Comparison of the Two Corpora

The Corpora used and contributed to the formulation of the conclusions was the Learner Corpus named Sub-IFLG Corpus (Florou, 2009) and the Native Corpus of Italian language named CWIC-lettere (Miceli & Kennedy, 2005). The first is a Corpus which is constructed of written texts of adult students who attend a private Italian language institute and consists of 66 texts with a total of 8,342 words. It is a collection of letters, written by learners, which cover the same topic: "During vacation you are writing to a friend inviting him to join you". This is a sub-Corpus of a larger Learner Corpus, IFLG, of 20,000 words, which has been error-tagged by a semiautomatic tool developed in the University of Athens named "Episimiotis" (Koutsis, Markopoulos, & Mikros, 2007).

The second is a Native Corpus used as a reference Corpus, as a linguistic resource that is used to compare the language of students with that of native speakers. It has been compiled by Kennedy and Miceli in a University of Australia and it is a smaller part of CWIC. For the use of the above data special permission was granted by the corpus developers restricted to academic use. The CWIC-lettere consists of 40 letters which total amount of 8,648 tokens (Miceli & Kennedy, 2005). In order for the comparisons to be reliable, the two corpora should be equal in terms of size. For the accurate analysis and counting of the structures under study, a corpus analysis tool is necessary (Chambers, 2005) and, in this case, the one that we used was WordSmith (v. 6.0). This tool was used mainly for counting the frequency of occurrence of the searched linguistic features.

3.1 Structures of Address Forms

In order to facilitate the writing of a letter/email for students of Italian, textbooks for Italian as a foreign language suggest some specific structures for addressing the recipient of the letter/email. The use of these structures is not obligatory (teachers do not restrict students to these phrases), while native

speakers use some of those mentioned in the textbooks and also others not suggested to students. The percentages of structures of address forms in both Corpora are shown in the following Table 1:

Table 1. Structures of Address Forms

STRUCTURES	OF	Sub-IFLG	%	Lettere-CWIC	%
INTRODUCTION					
Caro		26	39,5	3	7,5
Cara		36	54,5	10	25
Cari		0	0	5	12,5
Carissimo		1	1,5	7	17,5
Carissima		3	4,5	8	20
Ciao		0	0	3	7,5
RECIPIENT'S NAME		0	0	4	10
Total		66	100	40	100

Note. Caro=Dear (m), cara=Dear(f), cari=Dear (pl), Carissimo=Dearest(m), carissima=Dearest(f), ciao=Hallo.

The most striking observation is that the letters/emails of native speakers have a greater variety of address forms. This seems quite natural since students do not write spontaneously, but take on a role and, therefore, they choose the most simple and popular form. The second area of concern is that letters with only the name of the recipient or a simple greeting are not used by Greek students. Should these forms be used, since they are not recommended in the textbooks, they are corrected by the teachers. The high frequency terms *Cara* and *Carissima*, both in Sub-IFLG and the CWIC-lettere, appear more often, because the chosen recipient (imaginary or real) is a woman and not a man; in addition, regarding the Learner Corpus, there are more female students than male students who attend Italian language courses. On the other hand, CWIC does not keep any personal data of the native speakers who offered the original material.

It should be noted that in the IFLG the use of *caro* and *cara* becomes excessive in relation to what natural speakers use, with a significant difference. In contrast, in CWIC the forms *cari* and *carissima* are widely used. The same forms seem to be avoided by Italian learners because they are rarely addressed in plural and a form like *carissima* in Greek language sounds overly affectionate.

3.2 Closing Structures

Regarding closing structures, textbooks suggest some standard phrases but in this case there is more variety. Nevertheless, it is possible to group these structures into sets with approximately the same meaning, as shown below Table 2:

Table 2. Closing Structures

STRUCTURES OF CLOSING	Sub-IFLG	%	Lettere-CWIC	%
Ti bacio/baci/bacioni/ baci e abbracci/ baci e saluti	40	60,6	13	32,5
Ti abbraccio/un abbraccio	8	12,1	10	25
Saluti/un saluto	8	12,1	1	2,5
A presto	1	1,5	3	7,5
Con affetto	2	3,1	1	2,5
Con simpatia	0	0	2	5
Ciao	0	0	4	10
Ti aspetto	2	3,1	0	0
SENDERS' NAME	4	6	4	10
-	1	1,5	2	5
Total	66	100	40	100

Note. Ti bacio=I kiss you/ baci=kisses/ bacioni=big kisses/ baci e abbracci=kisses and hugs/
baci e saluti=kisses and greetings, Ti abbraccio=I hug you/ un abbraccio=a hug, Saluti=greetings/ un saluto=a greeting, A presto=in a minute, Con affetto=With tenderness, Con simpatia=with affection, Ciao=bye, Ti aspetto=I expect you.

As far as closing a letter/email is concerned, there are more ways used by native speakers. The phrase *Ti aspetto* is not used in the Italian Corpus, but appears frequently in the IFLG since it is semantically associated with one of the topics of the letters (a topic regarding an invitation was given to the students). Without doubt, what stands out from the above table is an agreement on the most popular closings, but also a prevalence of phrases related to kissing in the Learner Corpus. In all likelihood, the learners are strongly influenced by the Greek language and the corresponding expressions used in similar cases. There is a particular overuse of phrases *Ti bacio*, *baci*, etc., or even *Ti abbraccio*, *un abbraccio* etc. and also *un saluto*, *saluti*. On the other hand, in IFLG there is a minimum appearance of *a presto* while there is no use of *con simpatia* or *ciao*, probably because they are not taught and they do not resemble anything that would be used in the Greek language.

3.3 Connective Structures

When the students of the intermediate level of Italian are taught the structure of a letter/email, they have the opportunity to see some types of connection; structures that are either singular words or verbal sets, some belong to the category of links and others are adverbs or even verbal types (e.g., gerund); for this specific research all previous fall under the label of connectors.

The table below lists all the connectors proposed by textbooks in the specific unit on how an email should be written, i.e., those that learners of Italian use, but those that are used by native speakers are also included.

Table 2. Connectors

CONNECTORS	Sub-IFLG	%	Lettere-CWIC	%
Quindi	3	8,3	10	29,4
Così	5	13,9	0	0
Dunque	2	5,5	2	5,9
Di conseguenza	1	2,8	1	2,9
Invece	1	2,8	3	8,8
Tuttavia	0	0	0	0
Comunque	2	5,5	10	29,4
Eppure	0	0	1	2,9
Al contrario	0	0	0	0
Inoltre	12	33,3	0	0
In più	0	0	1	2,9
D' altra parte	1	2,8	0	0
Concludendo	0	0	0	0
Riassumendo	0	0	0	0
In altri termini	0	0	0	0
In breve	0	0	0	0
Infatti	2	5,5	2	5,9
In seguito	0	0	2	5,9
Alla fine	5	13,9	2	5,9
Infine	2	5,5	0	0
Total	36	100	34	100

Note. Quindi=then, Così=so, Dunque=therefore, Di conseguenza=accordingly, Invece=instead, Tuttavia=however, Comunque=nevertheless, Eppure=and yet, Al contrario=conversely, Inoltre=also, In più=plus, D' altra parte=on the other hand, Concludendo=in conclusion, Riassumendo=summing up, In

altri termini=in other words, In breve=in short, Infatti=in fact, In seguito=after, Alla fine=eventually, Infine=finally.

The differences in the use of connectors for both Corpora appear to be quite large. Of a total of twenty connectors appearing in text books, only seven are used in both Corpora (*quindi, dunque, di conseguenza, invece, comunque, infatti, alla fine*). Six of all connectors, even though recommended by the text books, are not used by learners; even native speakers do not commonly use them (*tuttavia, al contrario, concludendo, riassumendo, in altri termini, in breve*).

Furthermore, it should be noted that *quindi, comunque, and invece* are predominantly used by Italians for connection (with percentages of 29.4%, 29.4% and 8.8% respectively). On the contrary, Greek students do not use those connectors and prefer others like *così* and *inoltre* (with percentages of 13.9% and 33.3% respectively), which have no appearance in CWIC. Essentially, a different choice of connector is used to express the same sense: e.g., *quindi* and *comunque* in CWIC are used to introduce a conclusion, while the Greek students prefer for the same role *così*. The same holds true with *inoltre* that Greek students use for adding some information, while native speakers achieve the same effect with *in seguito* and *in più*, which occur at a frequency of 5.9% and 2.9% respectively.

Finally, it is worth noting that *alla fine* and *infine* (with percentages of 13.9% and 5.5% respectively), considered necessary by textbooks and teachers to complete a written letter/ email, are used by Greek learners while Italians omit them.

4. Concluding Remarks

In this research the forms of address for opening and the forms for concluding a letter/email were studied, including the use of connectors in the studied texts.

One first conclusion that emerges is that no choice is given to the students, either by their books or their teachers, to use more simple forms of address, such as the recipient's name.

Another issue that arises is that learners are being restricted by structures presented by both the textbook and the teacher and, therefore, do not experience the freedom of expression that a native speaker has.

Finally, teachers and textbooks do not give the appropriate attention to the pragmatical meaning of the connectors used by native speakers. This results in inefficient communication that can reveal the non-nativeness of the speaker.

This study demonstrated that the use and exploration of Learner and Native Corpora is a fruitful methodological choice (Granger, 2003), and can be applied in the classroom, so that foreign language learners can become more aware of the disparities between their interlanguage and the language that native speakers produce.

One of the suggestions for further research would be to compare a Learner Corpus of a larger number of words with a Native Corpus of Greek language (consisting of letters/emails) for the detection of interference from the mother tongue.

In addition, IFLG Corpus could be expanded with formal letters/emails (from learners of a higher level) so that a similar comparison with texts of the corresponding level and genre of the CWIC could be performed. We can use this comparison in order to check whether the textbooks can provide adequate linguistic resources for effectively teaching these forms.

Publishers and authors of textbooks should use Corpora with authentic material for suggesting realistic language use.

This study and similar investigations can easily be applied to the full range of teaching and the process of language acquisition.

These results affect classroom teaching and motivate teachers to lead students to approach foreign language authentic texts.

Of most importance is that such investigations use authentic pragmatic contexts, such as the emails to which foreign language students will respond, and in which the same students will be asked to communicate with native speakers.

References

- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 55-76). John Benjamins Publishing Company, Amsterdam/Philadelphia. <https://doi.org/10.1075/llt.6.07aij>
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners'. In S. Granger (Ed.), *Learner English on Computer* (pp. 80-93). Longman, London and New York.
- Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. In *Corpora 5* (pp. 1-27). <https://doi.org/10.3366/cor.2010.0001>
- Chambers, A. (2005). Integrating Corpus Consultation in Language Studies. *Language Learning and Technology*, 9(2), 11-125.
- Diez Bedmar, M., & Casas Pedrosa, V. (2006). The use of prepositions by Spanish learners of English at University level: Main problems. In *7th Conference on Teaching and Language Corpora (TaLC) Proceedings* (pp. 42-44). Paris.
- Florou, K. (2006). The experience of creating a Learner Corpus: The Italian as Foreign language to Greek students (IFLG). In *7th Conference on Teaching and Language Corpora (TaLC) Proceedings* (pp. 174-176). Paris.

- Granger, S. (2004). Computer Learner Corpus Research: Current Status and Future Prospects. In M. Conn, & T. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123-145). Amsterdam and Atlanta: Rodopi. https://doi.org/10.1163/9789004333772_008
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-17). Longman, London and New York.
- Ιακώβου, Μ., Μαρκόπουλος, Γ., & Μικρός, Γ. (2003). Καθορισμός Βασικού Λεξιλογίου μέσω ΗΣΚ: Πρακτική εφαρμογή στη διδασκαλία της Ελληνικής Ξένης Γλώσσας (translation: Determination of Basic Vocabulary through Corpora: Application on teaching Greek as Foreign language). In *6^ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας*. Retrieved August 1, 2019, from <http://www.philology.uoc.gr/conferences/6thICGL/ebook/a/iakovou&Markopoulosµs.pdf>
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77-90.
- Koutsis, I., Markopoulos, G., & Mikros, G. (2007). Episimiotis. A multilingual tool for hierarchical annotation of texts. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of the Corpus Linguistics Conference CL2007, 27-30 July, 2007*. University of Birmingham, UK. Retrieved August 1, 2019, from http://ucrel.lancs.ac.uk/publications/CL2007/paper/243_Paper.pdf
- Lenko-Szymanska, A. (2005). The role of L1 influence and L2 instruction in the choice of rhetorical strategies by EFL learners. In B. Lewandowska-Tomaszczyk (Ed.), *PALC'2005. Practical Applications in Language Corpora*. Peter Lang.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Ringdom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on Computer* (pp. 41-52), Longman, London and New York.