

## **The verbs of the Italian language in informal letters:**

### **Comparing two Corpora**

**Katerina Florou**

*University of Epirus*

#### **Abstract**

The purpose of this paper is to prove the application of Contrastive Interlanguage Analysis on researching cases, which deal with the use of the Italian language by Greek students. By comparing a learner corpus with a native speakers' corpus we come to conclusions about the overuse or the underuse of the most frequently used verbs in informal letters. Furthermore, there can be additional observations on the differences in the use of tenses and person. Finally, by using facts we can give our estimation on the role of the foreign language textbooks in the Italian language classroom.

**Keywords:** Contrastive analysis, Italian as FL, learner corpus

#### *Abstracto*

El propósito de este estudio es probar la aplicación del Análisis Interlingüístico de Contrastes en casos de investigación que traten con el uso que los estudiantes griegos le dan al idioma italiano. Al comparar el corpus de un aprendiz con uno de un hablante nativo, llegamos a varias conclusiones sobre el uso exagerado y limitado de los verbos más comunes en cartas informales. Además, pudimos observar diferencias en el uso y el tiempo de los verbos y las personas. Finalmente, al utilizar hechos pudimos estimar el rol de los textos del aprendizaje de idiomas extranjeros en el aula italiana.

*Palabras claves: análisis de contrastes; el italiano como idioma extranjero (FL); corpus de un aprendiz*

**Katerina Florou** is a laboratory-collaborator at the University of Epirus, in Igoumenitsa, in the faculty of Foreign Languages Applied to Administration and Commerce. She has recently concluded her PhD Thesis at the University of Athens at the department of the Italian Language and Literature in the domain of Applied Linguistics. Her research interests are expanding from the field of Sociolinguistics to the field of Corpus Linguistics.

### **Introduction**

By error analysis, even if this is supported by the use of a computer, one can locate, count down, and observe the errors during the learning process and the assimilation of the foreign language, but in order for someone to get close to the conclusions concerning the cause of the errors, it is safer to make comparisons. Contrastive Interlanguage Analysis can provide us information on the overuse or underuse of structures by foreign language students and this is valuable, especially for the application of the results during the foreign language teaching.

### ***Previous Researches***

The bibliography, especially that of the recent years, has a lot of research on the Contrastive Interlanguage Analysis to show so far, even though it is a rather new scientific domain. As far as the first of the two branches is concerned (comparison of interlanguage with native speakers language) we can, for example, mention researches, like the research on the frequency of verbs in native speakers corpora and in sub-corpora of English language students with different mother tongue (Ringdom, 1993, pp. 41-52); the research on the different use of the adverbial connectors in an English native speakers corpus and in a corpus of Swedish students of the same language (Altenberg & Tapper, 1998, pp. 80-93); the research on the noun clauses in an English native speakers corpus and several sub-corpora of English language students with French, Chinese and Spanish as their mother tongue (Biber & Reppen,

1998, pp.145-158) ; the research on the different use of the auxiliary verbs in an English native speakers corpus and in a learner corpus of English language students of Swedish, French and German origin (Aijmer, 2002, pp. 55-76); the research on the use of adjectives by comparing an English native speakers corpus with a learner corpus of German students of the same language (Lorenz, 1999).

Moreover, there are other interesting researches like those of Diez - Bedmar and Casas - Pedrosa (2006, pp. 42-44) on the differences on the use of prepositions which are under study in an English native speakers corpus and in a learner corpus of students of English literature, in which the errors are marked; the research of Szymanska (2006, pp. 87-88), on the different use of the first person after comparing an American native speakers corpus and a learner corpus of Polish students of the English language.

### ***The Compared Corpora***

In a similar way to the above researches, there is going to be here a comparison between two corpora, but in this case, the language under study is the Italian language and of the two compared corpora, the one is an Italian native speaker corpus, for which permission has been granted to be used for strictly researching purposes, and the other is a learner corpus, which was constructed initially by the researcher for a study on error analysis and it consists of texts produced by Greek students of the Italian language (Florou, 2006). Its total size comes up to 20,000 words and the genres included are of two kinds: friendly letters and free written productions, both of them of intermediate level. The IFLG (Italian as Foreign Language for Greek students) corpus is an open corpus and is continually enriched with different genres and of different level.

From this IFLG, a sub-corpus was chosen with subject the informal letters of the Italian language students and its size was 8,598 words (66 letters). The size was considered sufficient because the facts that were chosen to be studied are of great frequency. It is true that for the vocabulary observations a wide range of samples is needed, while for the grammatical ones, even a corpus of 1,000 words is enough (Conrad, 1993, p. 3). In this case, there is going to be an observation of grammatical nature on well-known verbs of common usage and mainly closely connected to their syntactical function (e.g. *verbi modali*).

The choice of the letters was made based on the only criterion of their size, so as to be quite comparative with that of the native speakers' letters.

From the native speakers' corpora, which were mentioned before, the CWIC was chosen, because it is the only one that includes a sub-corpus of the same genre like the one to be compared, that are the friendly letters.

The CWIC is a written speech corpus made by Micelli, T. and Kennedy, C. at the Griffith University in Australia, in order to be used in class for the teaching of the Italian language (Micelli & Kennedy, 1999, p. 2). The texts included are of two categories: 1) Those written by non-professional writers and are divided in the below categories: a) personal letters, b) business letter writing, c) personal e-mails, d) business e-mails, e) e-mails on commercial catalogues, f) letters to business specialists, to magazine columns, and 2) those written by professional writers and are divided in these categories: a) specialists' responses, b) articles from daily columns in magazines, c) film reviews (Miceli & Kennedy, 2001, p. 78).

In order to make the comparison we took from the CWIC a sub-corpus of friendly letters (40) of total 8,648 words, so as to be compared with the IFLG. The

size was chosen so as to be proportional to the sub-corpus, and the genre, so as the comparison to be as reliable as possible.

In spite of all these, we should mention a difference: the IFLG letters are described as controlled written production, because they follow specific instructions that are given under certain circumstances, which the students have to impersonate; while the CWIC letters do not have a specific theme, but they are native speakers' free productions which were selected at random.

It is necessary to mention the above difference, because it will be taken into account during the data process.

### ***The genre***

In this point we should explain the reason why this specific genre is studied. The writing of an informal letter or an e-mail, is a very common requirement in exams for every language certificates (not only in the Italian language) so it is something that the students practice a lot. It is, also, a kind of writing that is taught in the first lessons, even before the teaching of basic grammatical structures is complete, and among the first written communicative tasks.

Furthermore, it is a convenient communicative task, because students can see its direct application - especially professionals and students who are familiar with Internet communication - and it does not require special instruction, as there is always something to be said or rather to be written to a friend without using complicated expressions. In that way, even the need for a simplified ability of communication is fulfilled, which is covered by using basic vocabulary (Iakovou, Markopoulos & Mikros, 2003). This need is mainly obvious in the first stages of the foreign language learning process, when the students asking for the achievement of language functions

similar to the language structures that they possess in their mother tongue, resort to the communicative use of the vocabulary elements of the target language with which they feel more comfortable and they consist the students' own basic vocabulary (Iakovou, Markopoulos & Mikros, 2003).

For these reasons, the informal letter was chosen to be studied and the aim of applying the research results in the classroom or in the textbooks writing is achieved as well, so as, beyond acquiring a certificate, the students of the Italian language to be, in fact, able to communicate with new friends-native speakers through e-mail.

### *The research*

#### *Verb frequency*

Initially, we decided to make a list of the verbal types that appear in both corpora up to five times and then make the lemmatization of the verbal types in order to make their observation easier.

But before describing the procedure, a table is presented; in which four of the most common used textbooks appear, in order to find out the time location when the friendly letter appears in comparison with the appearance of the grammatical phenomena, and especially in comparison with the verb. Two of the following textbooks have been used when teaching students whose written productions are part of the IFLG (Progetto and Allegro 1); and the first one represents those textbooks which are called traditional textbooks, while the fourth one is a book based mainly on authentic material.

**Table I:****Teaching sequence of the verbs and communicative abilities in the textbooks**

	<b>Qui Italia</b>	<b>Progetto1</b>	<b>Allegro1</b>	<b>Espresso</b>
1	ESSERE	ESSERE	STARE	ESSERE
2	ESSERCI	AVERE	ESSERE	AVERE
3	AVERE	CHIAMARSI	ABITARE	CHIAMARSI
4	V -ARE	V -ARE	ANDARE	V -ARE
5	V -ERE	V -ERE	ESSERCI	STARE
6	V -IRE	V -IRE	CHIAMARSI	FARE
7	FARE	ANDARE	PRENDERE	V -ERE
8	STARE	VENIRE	AVERE	VOLERE
9	ANDARE	DARE	FARE	PREFERIRE
10	VENIRE	SAPERE	FINIRE	V -IRE
11	DOVERE	STARE	POTERE	ANDARE
12	POTERE	USCIRE	DOVERE	GIOCARE
13	USCIRE	FARE	APRIRE	LEGGERE
14	VOLERE	GIOCARE	USCIRE	USCIRE
15	<b>"PAS.PROS."</b>	POTERE	VENIRE	PIACERE
16	<b>LETTERA</b>	VOLERE	SAPERE	ODIARE
17	"FUTURO"	DOVERE	PIACERE	ESSERCI
18	"IMPERFETTO"	<b>LETTERA</b>	<b>"PAS,PROS."</b>	POTERE
19	"TRAPASSATO"	<b>"PAS.PROS."</b>	<b>LETTERA</b>	VENIRE
20	PIACERE		"FUTURO"	<b>LETTERA</b>
21	"CONDIZIONALE"		"IMPERFETTO"	DOVERE
22	"IMPERATIVO"		"TRAPASSATO"	SAPERE
23	"PAS.REMOTO"			<b>"PAS.PROS."</b>

On the table we can notice many differences in the teaching sequence of the verbs, the tenses, and the letter communicative task. Some remarkable differences are that two out of the four books consider that *Passato Prossimo* (Present Perfect) is necessary to be taught before teaching letter writing. Moreover, three out of the four textbooks consider mentioning all the irregular verbs of the present necessary, while in the fourth we notice a “delay” in teaching the verbs “*dovere*” (must) and “*sapere*” (know). It is, thus, proved that some textbooks writers consider writing a letter more complicated, while others think that it is simply enough for someone who understands the present tense and its use.

The tool used to continue with the searching was the Wordsmith tool, which firstly provided us the word frequency lists, and then the verbal types lists. When the lemmatization was done, we used the Concordancer for the “*essere*” (be) and the “*avere*” (have) types (in both Corpora), so as to separate their literal use from that of the auxiliary. The results that arose are presented in the following table.

**Table II:**

**Frequency list of all verbs in the two corpora after lemmatization**

	IFLG				CWIC		
	VERBO	FRE	%		VERBO	FRE	%
1	ESSERE	349	418		ESSERE	306	3,67
2	AVERE	197	2,4		AVERE	181	2,17
3	DOVERE	45	0,5		FARE	48	0,57
4	SENTIRE	34	0,4		STARE	33	0,38
5	POTERE	67	0,8		SAPERE	29	0,34
6	ASPETTARE	26	0,3		ANDARE	25	0,3
7	FARE	51	0,6		MANDARE	10	0,12



8	SPERARE	24	0,3		VEDERE	10	0,12
9	ABBANDONARE	20	0,2		AUGURARE	10	0,12
10	ESSERCI	20	0,2		SPERARE	27	0,31
11	SAPERE	35	0,4		ESSERCI	6	0,07
12	STUDIARE	20	0,2		VOLERE	11	0,13
13	STARE	42	0,5		ARRIVARE	5	0,06
14	VENIRE	32	0,4		PARLARE	5	0,06
15	VOLERE	38	0,5		LAVORARE	5	0,06
16	ANDARE	24	0,3		PENSARE	11	0,13
17	CONTINUARE	10	0,1		TROVARE	11	0,13
18	SCRIVERE	22	0,3		DISPIACERE	5	0,06
19	PENSARE	33	0,4		PIACERE	5	0,06
20	RESTARE	8	0,1		PIOVERE	5	0,06
21	SCIARE	8	0,1		POTERE	10	0,12
22	CAPIRE	7	0,1		RIUSCIRE	5	0,06
23	TROVARE	24	0,3		SCRIVERE	5	0,06
24	CONSIGLIARE	6	0,1				
25	PRENDERE	6	0,1				
26	PROSEGUIRE	6	0,1				
27	CREDERE	18	0,2				
28	INVITARE	10	0,1				
29	DIRE	14	0,2				
30	OFFRIRE	5	0,1				
31	PASSARE	5	0,1				

Comparing these data from the two frequency lists, we can make some observations. First of all, we notice that the verbs in the IFLG that appear up to 5 times are more (31) than the verbs in the CWIC (23), and that the common verbs in the two columns are 13, but they differ in frequency. But it must be pointed that some

verbs owe their use to the letter topic, as far as the IFLG is concerned, and some others, like “*dispiacere*” (be sorry) and “*piacere*” (like), do not appear in the same learner corpus.

It is also noticeable that the modal verbs (modali) “*dovere*” (must), “*volere*” (want), “*potere*” (can) are widely used in the IFLG and as a result much more verbs in the learner corpus are in the infinitive form, because after modal verbs, in case of one-subject construction, an infinitive follows.

### ***Tenses frequency***

In a second phase of the research when comparing tenses, as they result from the initial table, and before making the lemmatization, new differences and similarities arise.

This time, in order to ascertain the importance of differences - wherever these arose- another automatic tool was used; the log-likelihood (LL), which automatically counts the importance of difference and the more the value (LL value) is that arises as a result, the more the two corpora differ in the use of the specific lexeme<sup>1</sup>. It should be noted here that the LL values which are higher than 3.8 verify the importance of the difference (Rayson & Garside, 2000, pp. 2-3).

---

<sup>1</sup> The log-likelihood is used both in speech parts frequency and in semantic elements frequency.

**Table III:****Frequency lists of tenses appearance in the two corpora**

	<b>IFLG</b>	<b>%</b>	<b>CWIC</b>	<b>%</b>	<b>LL</b>
PRESENTE	701	58,12	424	55,2	70,53
PASSATO PROSSIMO	146	12,1	216	28,12	13,22
IMPERFETTO	26	2,15	2	0,26	24,55
CONG. PRESENTE	44	3,64	9	1,17	25,38
CONG. PASSATO	6	0,49	4	0,52	0,41
FUTURO	34	2,81	11	1,43	12,46
FUTURO COMPOSTO	2	0,16	2	0,26	0,00
TRAPASSATO	2	0,16	14	1,82	10,05
CONDIZIONALE	17	1,4	5	0,65	6,99
INFINITO	228	18,9	81	10,54	73,70

As we can see from the above table, in both corpora the present tense is dominant, although for the IFLG an excessive use is recorded (LL: 70.53). *Passato Prossimo* (Present Perfect) (LL: 13.22) is used more by the Italians, and *Trapassato* (Past Perfect) as well (LL: 10.05) in double size. Perhaps here, we could mention a Greek students' difficulty in using the auxiliary and that is the reason why they avoid compound tenses.

As far as *Imperfetto* (Imperfect) is concerned (LL: 24.550), it is used excessively by Greek students (even though it is not about letters with a narrative topic), as well as *Futuro* (Future) (LL: 12.46), because it is also widely used in the Greek language. It is well known that Italian native speakers tend to use present tense to express an action in the near future and not verbs with morphemes that declare future. This is a habit that Greek students do not seem to have assimilated.

It is a surprise that *Presente Congiuntivo* (Subjunctive) (LL: 25.38) and *Condizionale Presente* (Conditional) -especially “*vorrei*” (I would like)- although they are tenses closely taught but in final lessons, are excessively used by Greek students. The cause can be located on the insistence of their assimilation during the learning process. Despite all these, for *Condizionale* we could express our reservation, due to the fact that the value of the importance of difference is small (LL: 6.99).

Finally, the infinitive form is particularly frequent in IFLG (LL: 73.70), as it has already been reported, due to its syntactical place.

### ***Person frequency***

A third attempt was the study of the person that is used in friendly letters and in the contrastive table, which arose, and there are some data worth commenting with regards to it.

### **Table IV:**

#### **Frequency lists of person appearance in the two corpora**

<b>PERSONA</b>	<b>IFLG</b>	<b>%</b>	<b>CWIC</b>	<b>%</b>	<b>LL</b>
1a SINGOLARE	484	40,13	278	36,2	57,59
2a SINGOLARE	213	17,66	39	5,07	133,19
3a SINGOLARE	211	17,49	254	33,07	3,74
1a PLURALE	14	1,16	48	6,25	19,52
2a PLURALE			15	1,95	20,71
3a PLURALE	56	4,64	53	6,9	0,10
INFINITO	228	18,9	81	10,54	73,70

Starting with the 1<sup>st</sup> person, it is obvious it is more used, especially in the learner corpus (LL: 57.59), probably because verbs like “*credere*” (believe), or “*pensare*” (think) encourage such use. Besides, letter topics have to do with making/rejecting an invitation or asking/giving advice and there is always a sender addressing to a receiver. This observation is also confirmed by the extensive use of the 2<sup>nd</sup> person in IFLG with a remarkable value in the importance of difference (LL: 133.19). But in the CWIC, the sender often speaks on behalf of their partner, and that is the reason why more verbs in the 1<sup>st</sup> person plural are used (LL: 19.52). They also address to more than one person sometimes, that is why the 2<sup>nd</sup> person plural appears (LL: 20.71), which does not happen in the IFLG.

It is remarkable that the 3<sup>rd</sup> person singular appears twice as much in the CWIC than in IFLG, which actually happens because of the extensive use of the 3<sup>rd</sup> person singular of the “*avere*” in the present tense, because it is mainly used as an auxiliary in present perfect and not in the present tense. However, from the value of the log-likelihood the difference, which was mentioned before, does not seem important enough. Nevertheless, we should notice the use of the specific verb in Presente and in Passato Prossimo in both corpora.

It is also confirmed in another way that in CWIC Passato Prossimo is widely used.

### ***Application***

Results like the ones mentioned before apply to the whole teaching process and to the language acquiring process.

First of all, the above results affect the classroom teaching and they give the motive to the teacher to lead the students closer to the foreign language through authentic texts.

Moreover, that kind of results make the textbooks better, so as during their writing process to avoid recommendations towards teachers and students that have nothing to do with the reality of the foreign language.

The most important of all is that this kind of researches contributes to the real-life communication situations, which for the present time are not letters, but e-mails, through which foreign language learners will communicate in a better way with the native speakers of the same language.

### 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). Amsterdam: John Benjamins Publishing Company.
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners written English. In S. Granger, (Ed.), *Learner English on computer* (pp. 80-93). New York: Longman.
- Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger, (Ed.), *Learner English on computer* (pp. 145-158). New York: Longman.
- Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, 27, 1-18.
- Diez-Bedmar, M., & Casas-Pedrosa, A. (2006, July 1-4). The use of prepositions by

- Spanish learners of English at University level: Main problems. *TaLC Proceedings*, Paris, pp. 42-44.
- Florou, K. (2006, July 1-4). The experience of creating a Learner Corpus: The Italian as foreign language to Greek students (IFLG). *TaLC Proceedings, Paris Universite*, Paris, pp. 174-176
- Iakovou, M., Markopoulos, G., & Mikros, G. (2003). Determination of the basic vocabulary through corpora for the teaching of Greek as a FL. *Proceedings of 2<sup>nd</sup> International Conference of the teaching of Greek as a FL, 19-21 September 2002, Athens*.
- Lenko-Szymanska, A. (2006, July 1-4). Self-mention in argumentative writing. *TaLC 2006 Proceedings*, Paris, pp. 87-88.
- Lorenz, G., R. (1999). Adjective intensification - learners versus native speakers: A corpus study of argumentative writing. *Language and Computers: Studies in Practical Linguistics 27*, Amsterdam: Rodopi.
- Miceli, T., & Kennedy, C. (1997). *A Computer-based resource for developing Italian writing skills*. Final Report For A 1997 National Teaching Development Grant, pp. 1-15.
- Miceli, T., & Kennedy, C. (2001). An evaluation if intermediate students approaches to corpus investigation. *Language Learning and Teaching*, 5(3), 77-90
- Rayson, P., & Gardise, R. (2000, October 1-8). Comparing corpora using frequency profiling. Proceedings of the workshop on “Comparing Corpora”, held in conjunction with the 38<sup>th</sup> annual meeting of the Association for Computational Linguistics (ACL2000) Hong Kong, pp. 1-6.
- Ringdom, H. (1998) Vocabulary frequencies in advanced learner English: a cross-

linguistic approach. In S. Granger, S. (Ed.). *Learner English on computer* (pp. 41-52). New York: Longman.

**Received: March 16, 2008**

**Published: June, 2008**