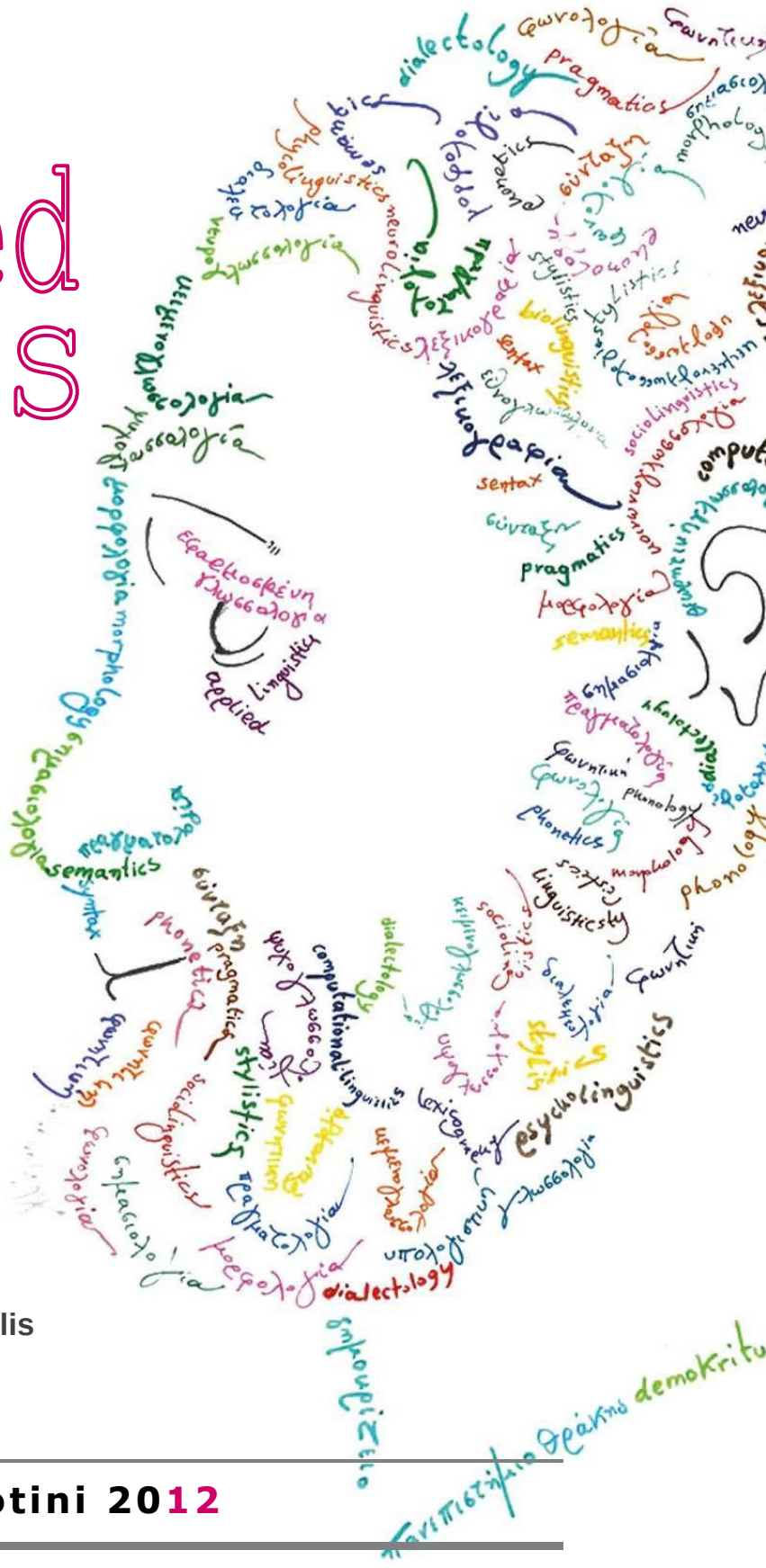
DEMOCRITUS UNIVERSITY of THRACE

selected papers of the 10th International Conference of Greek Linguistics



Edited by

Zoe Gavriilidou
Angeliki Efthymiou
Evangelia Thomadaki
Penelope Kambakis-Vougiouklis

Komotini 2012

ISBN 978-960-99486-7-8

Τυπογραφική επιμέλεια

Νίκος Μαθιουδάκης
Ελένη Παπαδοπούλου
Ελίνα Χατζηπαπά

Σχεδιασμός εξώφυλλου

Νίκος Μαθιουδάκης

MEASURING THE GREEK INFINITIVE THROUGH TIME

Marina Vassiliou
University of Athens
marina.vassiliou@gmail.com

ABSTRACT

The current work consists in a series of statistical analyses of Greek infinitival structures, aiming at illustrating the syntactic behaviour of the Greek infinitive through time. Spanning the period 5 BC – AD 16, the specific structures are drawn from texts of various authors and divergent topics, being representative of the four synchronies of the Greek language. The text corpus employed exceeds 5 million words in size, within which the infinitival occurrences approximate 102,000. To the best of our knowledge, measurements of such scale are presented for the first time, allowing a diachronic study of the infinitive use supported by statistical tests.

Keywords: Greek infinitive, Accusative case, statistical analysis, discriminant models

1. Introduction

The current paper, being part of an ongoing work, the object of which is the syntactic behaviour of the Greek infinitival structures through time, presents a set of statistical analyses of the evolution of the Greek infinitive. These analyses were based on measurements of infinitival occurrences over an extensive corpus of texts dated from 5 BC to AD 16. The specific Corpus of Infinitival Structures (named hereafter **CIS**) has more than 5 million words and contains approximately 102,000 infinitival occurrences, which were manually annotated and classified by the author on the basis of their syntactic function.

The paper has the following structure: Section 2 provides detailed information on how CIS was compiled and what its contents are. Section 3 describes the categories into which the infinitival occurrences were classified, while Sections 4 and 5 present a set of initial observations on the data, which allow a profiling of the synchronies in terms of the use of the infinitive. The subsequent section 6 reports on the statistical analyses performed, while Section 7 concludes the paper.

2. Data description and collection

The Corpus of Infinitival Structures comprises infinitival structures drawn from texts of various authors and divergent topics, such as history, chronography, rhetoric, philosophy, tragedy, comedy, poetry, novel, epic, mythology, ecclesiastical texts, hagiography etc.[1] These texts span a pretty extensive period, ranging from 5 BC to AD 16, and are representative of the four synchronies (or periods[2]) of the Greek language, namely (i) **Classical Greek** [**CL**] (5 – 4 BC), (ii) **Hellenistic Koine** [**KO**] (3 BC – AD 4), (iii) the **Early Byzantine** period [**EB**] (AD 5 – 10) and (iv) the **Late Byzantine** period [**LB**] (AD 11 – 16). It should be noted that special care was taken for the texts of the last three periods to be of the vernacular variety.

[1] The corpus of infinitival structures does not include text material from inscriptions and papyri, as these may well be considered as a discrete register reflecting the status of spoken Greek at the Koine period, which will be studied in conjunction to the current contents of the corpus at a later stage of this research.
[2] It should be noted that the terms "period" and "synchrony" are used here interchangeably.

In Z. Gavriilidou, A. Efthymiou, E. Thomadaki & P. Kambakis-Vougiouklis (eds), 2012,
Selected papers of the 10th ICGL, pp. 592-602. Komotini/Greece: Democritus University of Thrace.

The primary resource employed for compiling the corpus was the Thesaurus Linguae Graecae (TLG)[3]. In particular, all the texts of the first three periods originate from TLG, while only a few texts (didactic and medical texts, poems and novels) of the Late Byzantine period, not being available via TLG, were drawn from printed resources (for instance Legrand (1881 & 1896); Wagner (1874)).

The compilation of the corpus and the corresponding dataset involved a three-step process. First, the texts were exhaustively searched for infinitival forms of all tenses (Present, Future, Aorist and Present Perfect) and voices (Active and Non-Active). The search, which was based on the corresponding infinitival endings, was performed automatically in the case of texts available in electronic form (including naturally the ones originating from TLG). On the other hand, those texts available only in printed form were manually searched for detecting the given infinitives.

Subsequently, all retrieved infinitival occurrences were examined manually within their context so that they could be assigned to a category reflecting their syntactic function (see Section 3 for more details on the properties of these categories).

The final step of the processing involved measuring the infinitival occurrences on the basis of various criteria such as author, work, topic, syntactic category and synchrony. This dataset of measurements served then as the input of the statistical analyses.

Table 1 summarises the features of the CIS, which is divided into 4 sub-corpora, each corresponding to a period. All sub-corpora have roughly the same size, ranging from 1 to 1.5 million words, while the infinitival occurrences number approximately 102,000. The last column contains information on the topics covered by the texts of each sub-corpus.

Synchrony	Number of texts	Number of words	Infinitival occurrences	Topic
Classical Greek (**CL**)	431	1,004,194	37,828	Comedy, history, rhetoric, philosophy, tragedy
Hellenistic Koine (**KO**)	261	1,548,520	26,026	New Testament, Septuagint, apocalyptic texts, ecclesiastical texts, hagiography, theology, history, philosophy, mechanics
Early Byzantine (**EB**)	126	1,493,411	28,541	Chronography, philosophy, ecclesiastical texts, hagiography, history, poetry, theology
Late Byzantine (**LB**)	98	1,405,123	9,568	Poetry, chronography, novel, epic, hagiography, comedy, mythology, tragedy, medical texts, satirical texts
Totals	**916**	**5,451,248**	**101,963**	

Table 1 Profile of the Corpus of Infinitival Structures

3. Classification of infinitival structures

As mentioned before, the infinitival occurrences, which were retrieved, were classified into categories, reflecting their syntactic function. These categories were defined on the basis of two criteria, namely the type of syntactic structure and the case of the infinitival subject (Accusative vs. non-Accusative). The combination of these two criteria yielded fifteen (15) categories, being illustrated in Table 2.

Category name	Infinitival structure	Example
sc	**Subject control**	ἐγὼ τὴν μνείαν τοῦ θανάτου **ἀναγγεῖλαι** οὐ δύναμαι
oc	**Object control**	Οἱ Ἀθηναῖοι προεῖπον **τοῖς στρατηγοῖς** μὴ **ναυμαχεῖν**
aci	**Verb complement & Accusative subject** (Accusativus-cum-Infinitivo)	Ὁ δὲ βασιλεὺς […] ἠβουλήθη ἑτέραν **ἀγαγέσθαι νύμφην** τῷ Ἰωάννῃ
4a	**Personal/Impersonal structure**	ἡ δὲ ἡμετέρα ἀρχὴ χαλεπὴ δοκεῖ **εἶναι**
4b	**Personal/Impersonal structure & Accusative subject**	Αἰσχρόν ἐστι τὸν ἰσχυρὸν **Δαρεῖον** βασιλέα […] **δουλωθῆναι** δυστυχῶς ἀντάρτῃ Μακεδόνι

[3] http://www.tlg.uci.edu/

Category name	Infinitival structure	Example
5a	**Main verb of a subordinate** (normally temporal or final) **clause**	**ὁ Σίμων** καὶ τὴν ἱερὰν σύγκλητον καὶ τὸν δῆμον Ρωμαίων εἰς τοσοῦτο κατεπλήξατο, **ὡς** θεὸς **νομισθῆναι**
5b	**Main verb of a subordinate clause & Accusative subject**	καὶ συνέρχεται πάλιν ὄχλος, **ὥστε** μὴ **δύνασθαι αὐτοὺς** μηδὲ ἄρτον φαγεῖν
6a	**Simple infinitive** (a) with a mood (imperative or optative) function (b) in fixed expressions	* ὄψον δὲ μὴ πέμπετε μηδὲ **πιεῖν** * **ἁπλῶς δ' εἰπεῖν**, ἐναντία τοῖς νόμοις […] διατετέλεκε
6b	**Simple infinitive & Accusative subject**	δῴη ἡμῖν εὐφροσύνην καρδίας καὶ **γενέσθαι εἰρήνην**
7a	**Adverbial infinitive 1** (denoting reference or explanation) or **Complement of nominals**	* **τοῦτο** παρακελεύονται τοῖς δικασταῖς, μὴ **ἐθέλειν** ἀκούειν τῶν ἀπολογουμένων * **ἕτοιμος** ἔσομαι […] **πληρῶσαι** σοῦ τὴν αἴτησιν
7a1	**Adverbial infinitive 2** (final or resultative infinitive)	ἀπεστάλη προς τινα γέροντα **ἰάσασθαι** αὐτόν
7b	**Adverbial infinitive 1 & Accusative subject**	Ἔτυχε δὲ καὶ **τοῦτο** […] **ἐμπεσεῖν Λατῖνόν τινα**
7b1	**Adverbial infinitive 2 & Accusative subject**	ἐξελέξατο ἡμᾶς […], **εἶναι ἡμᾶς** ἁγίους καὶ ἀμώμους
8a	**Articulate infinitive**	μηδὲν μεριμνῶντες **τοῦ διαθρέψαι** αὐτά
8b	**Articulate infinitive & Accusative subject**	οὐκ εἶπε γὰρ ὁ Κύριος **τοῦ γενέσθαι σε** πετεινόν

Table 2 Categories of the infinitival structures studied

4. Profiling synchronies based on the distribution of the infinitival structures

An initial rough processing of the CIS lends support to the usual assumption regarding the distribution of the infinitive across the four synchronies. More specifically, the infinitival structures of all categories (as shown in Figure 1) exhibit a high frequency of occurrence in the Classical period, which progressively declines in the subsequent periods (with the exception of the Early Byzantine period, where a minor increase is observed over the previous period).

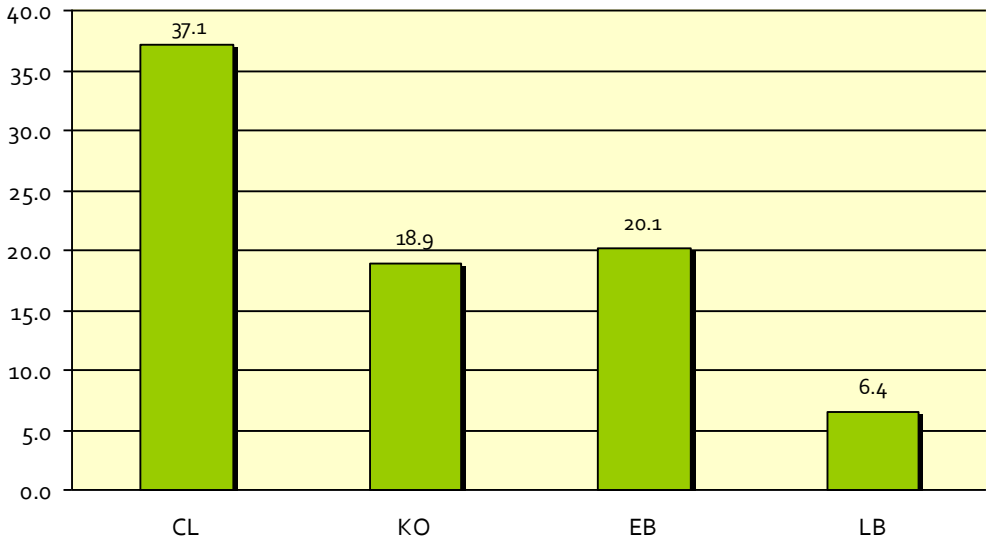


Figure 1 Frequency of occurrence of infinitival structures across the four synchronies *(values normalised per number of words)*

A more detailed picture is provided in Table 3, which illustrates the distribution of the infinitival categories in each individual period. The cases where the frequency of occurrence exceeds 5% are marked in bold.

Infinitival category	CL	KO	EB	LB
sc	**27,23%**	**27,86%**	**31,14%**	**53,63%**
oc	**10,00%**	**8,25%**	**10,22%**	**5,08%**
aci	**13,67%**	**7,62%**	**14,89%**	**7,35%**
4a	**19,23%**	**10,18%**	**8,32%**	**5,07%**
4b	**10,06%**	**7,53%**	**6,37%**	2,24%
5a	2,46%	1,81%	2,22%	1,71%
5b	1,57%	2,43%	4,26%	1,20%
6a	0,63%	0,95%	0,63%	0,87%
6b	0,53%	0,11%	0,09%	0,00%
7a	4,26%	3,09%	2,14%	1,87%
7a1	1,62%	**7,98%**	3,76%	2,93%
7b	1,49%	0,67%	0,88%	0,48%
7b1	0,21%	0,56%	0,23%	0,39%
8a	**5,24%**	**14,00%**	**8,88%**	**13,90%**
8b	1,78%	**6,95%**	**5,96%**	3,29%
Totals	**100%**	**100%**	**100%**	**100%**

Table 3 Distribution (in percentages) of categories of infinitival structures across synchronies

The data of Table 3 allow forming the profile of the four synchronies of Greek based on the distribution of the categories of the infinitival structures. More specifically, in **Classical Greek**, five types of structures are the prominent categories, namely subject (sc) and object (oc) control, aci and personal/impersonal (4a & 4b) structures, constituting more than 80% of all the infinitival occurrences. The articulate infinitive (8a) is also a frequent structure, with a frequency exceeding 5%.

The same five categories (sc, oc, aci and 4a & 4b) are still salient during **Koine**, approximating 62% of all the infinitival structures, even though, with the exception of subject control structures, a considerable decrease is observed in the frequency of each category. In parallel, circa 8% of the structures are final or resultative infinitives (7a1), while the use of the articulate infinitive (8a & 8b) is intensified over Classical Greek with a frequency exceeding 20%.

The total occurrences of subject and object control, aci and personal/impersonal structures increase over the residual categories during the **Early Byzantine** period, exhibiting a cumulative frequency of approximately 71%. It should be noted though that the frequency of the personal/impersonal structures decreases in comparison to the previous period, while the frequency of occurrence of oc and aci increases. Besides, the articulate infinitive (8a & 8b) remains a widely used structure (~ 15%), despite its decreased frequency during this period.

Finally, within the **Late Byzantine** period, subject control, aci and the articulate infinitive seem to be the enduring structures, representing approximately 75% of the total occurrences of infinitives.

5. Distributional behaviour of the infinitival structures across synchronies

With respect to the distribution of each individual infinitival category through time, it could be observed that most categories exhibit the same trend of gradual reduction in the number of occurrences, reflecting the gradual loss of the infinitive (cf. Jannaris, 1897:570-575; Joseph, 1983:38 & 50; Horrocks, 2010:93-94), despite any minor distribution divergences (see Figure 2).

In particular, the **subject control** structures seem to form the most enduring category for all periods, always representing more than 25% of the infinitival occurrences. It is noteworthy that in the Late Byzantine period the specific structures are the predominant category (their frequency rising to circa 54%).

Object control infinitives, **aci** and **personal/impersonal** (4a & 4b) structures have initially (Classical Greek) a high frequency (more than 10%), which is gradually reduced in Koine, reaching on average no more than 7% in the two last periods.

The use of the infinitive as the main verb of a subordinate clause (category **5**) or in replacement of a mood (category **6**) or with an adverbial function (category **7**) is generally fairly limited, although a minor increase is observed within the middle of the synchrony continuum, namely in Koine and in the Early Byzantine period. In all periods their frequency of occurrence reaches on average no more than 5%, this fact obviously justifying the early loss of such structures.

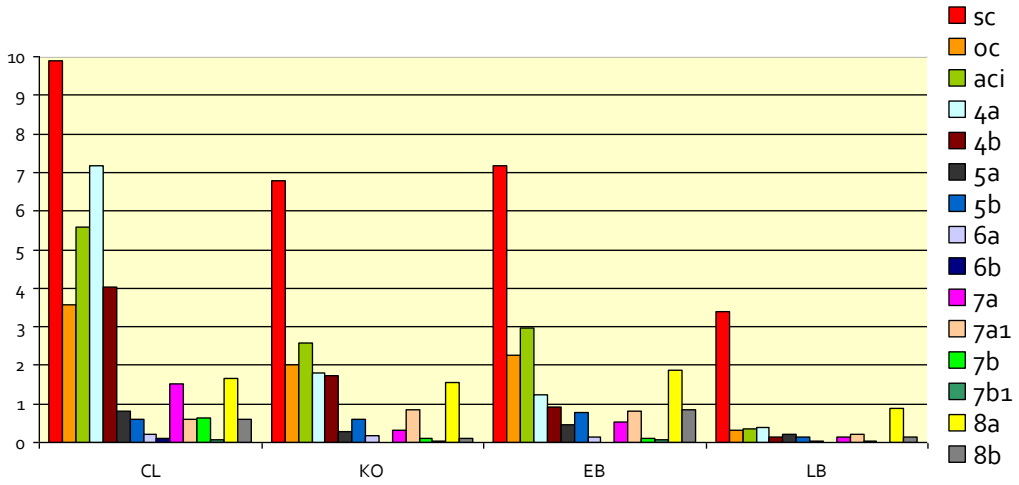


Figure 2 Frequency of occurrence of the 15 infinitival categories across the four synchronies (*values normalised per number of words)*

Finally, a non-typical behaviour is observed with respect to the articulate infinitive (category **8**), for which the majority of occurrences are observed in the middle of the synchrony continuum, instead of at the initial period (Classical Greek), as would be expected in accordance to the general trend. The high usage rate within Koine (> 10%) could be attributed to the fact that the articulate infinitive was frequently used instead of a subordinate clause (cf. Jannaris, 1897:577; Joseph, 1983:51; Horrocks, 2010:95). In a similar vein, this replacement pattern, which continues in the Late Byzantine period with this type of infinitive substituting not only subordinate clauses but also nouns and subjunctive verb forms, could explain the relatively high frequency of occurrence (~ 9%).

6. Analysing the data further

The aforementioned measurements served as a basis for obtaining a first picture of the distribution of the infinitival structures across the four synchronies. Additional experiments conducted were aimed at (i) examining the correlation between the distribution of infinitive and time (i.e. the four periods) and (ii) investigating whether the four periods constitute well-defined classes, which would be distinguishable in terms of the use of the infinitive.

The correlation between infinitive and time was studied via two types of statistical analyses, namely Levene's test (Levene 1960) and Analysis of variance – ANOVA (Fisher, 1925). In order to investigate the possibility of discriminating the four synchronies on the basis of the use of the infinitive, Discriminant Analysis (Klecka, 1980) was performed. For all statistical processing the SPSS (version 18) statistical package[4] was employed.

Before proceeding with the statistical analyses, it should be noted that in the following subsections the term "**infinitival variables**" refers to the measurements of the infinitival categories over the CIS and the term "**period variable**" denotes the synchrony.

[4] The SPSS statistical package was made available by the Computer Center of the University of Athens [URL: http://share.uoa.gr/public/Software/SPSS/SPSS18_for_Windows.iso].

6.1 Levene's test

Levene's test is a method for assessing the null hypothesis of the homogeneity of variance; in the current case the null hypothesis being that the variance of the infinitival variables remains stable across all levels of the period variable. If the output of Levene's test exceeds a certain significance value *p* (this being set normally to 0.05), then the null hypothesis is true. If, however, the resulting figures are below this threshold, then the null hypothesis is rejected and it can be concluded that the variance is significantly different across synchronies. Table 4 presents the *p* values obtained for each infinitival variable in the dataset. The *p* values exceeding the threshold of 0.05 are marked in bold.

Infinitival variable	Sig.
sc	0,001
oc	0,000
aci	0,000
4a	0,000
4b	0,000
5a	**0,926**
5b	0,000
6a	**0,066**
6b	0,000
7a	**0,067**
7a1	0,000
7b	0,000
7b1	0,010
8a	**0,201**
8b	0,000

Table 4 Levene's test significance (*p*) values obtained for each infinitival variable over the 4 periods

For almost all infinitival structures the variance is significantly different across the four periods, since the *p* value does not exceed the threshold of **0.05**. The only exceptions, where *p* is less than 0.05, are categories 5a (infinitive as the main verb of a subordinate clause), 6a (simple infinitive), 7a (adverbial infinitive 1) and 8a (articulate infinitive).

6.2 Analysis of Variance (ANOVA)

The analysis of variance was employed in order to investigate the effect of the period variable on the distribution of the infinitival structures. In this case, a significance threshold of **0.05** is also used, which, when being exceeded, indicates that the synchrony does not influence the values of the infinitival structures. The *p* values exceeding the threshold of 0.05 are marked in bold.

Infinitival variable	*F*	Sig.
sc	29,116	0,000
oc	26,610	0,000
aci	25,002	0,000
4a	63,091	0,000
4b	28,487	0,000
5a	3,438	0,018
5b	8,921	0,000
6a	1,570	**0,198**
6b	14,526	0,000
7a	48,943	0,000
7a1	14,054	0,000
7b	30,570	0,000
7b1	2,495	**0,061**
8a	7,370	0,000
8b	13,545	0,000

Table 5 ANOVA significance (*p*) values obtained for each infinitival variable over the 4 periods

The results confirm our intuition, since in all cases (except for two) $p < 0.05$, which signifies that the period factor substantially influences the distribution of the infinitival structures. The variables failing the test are 6a (simple infinitive) and 7b1 (adverbial infinitive 2 with an Accusative subject).

6.3 Discriminant Analysis

The third type of analysis that was performed was discriminant analysis and it was intended to investigate whether the texts of the corpus could be discriminated in terms of synchrony on the basis of the infinitival structures. In other words, the aim was to see whether (or which of) the infinitival variables could characterise the texts from which they were retrieved and separate them into classes, representative of the four periods.

6.3.1 Experimental phase 1

The experimental set-up involved two models, namely a **full** discriminant model and a **stepwise** one. In the full discriminant all variables enter the model from the start, while in the **stepwise** discriminant the variables are gradually introduced in the model. The value of the *F* parameter for entering/removing variables was set to **5.00/4.99** for the present experiment.

Moreover, in both models the ten-fold cross-validation methodology was followed in order to verify the validity of the results obtained. According to this method, the dataset was divided into ten equally-sized subsets, so that each subset contained the same amount of texts from each period. This allowed for ten runs of each discriminant analysis to be performed. In each run the nine subsets served as a labelled training set, while the residual subset was provided as an unlabelled test set[5], used for evaluating the accuracy of discrimination.

The rationale behind the division of the dataset into subsets was to check whether the information residing in 90% of the dataset, the classification of which is known, was sufficient for generating a model that could correctly classify the residual 10% of the dataset, the origin of which is not known. So, by repeating this process for each of the ten subsets it is possible to determine with an adequate degree of accuracy the average performance of the discriminant model.

The classification accuracy obtained is presented in Table 6, while Table 7 illustrates the variables that were used as discriminators in the stepwise model. The accuracy is not particularly high, ranging from 64% to ~ 69% for both models; it is though more than 2.5 times higher than the random rate of 25%, indicating the successful discrimination of groups. Besides, it is remarkable that the stepwise model employs only a limited set of variables, namely six (6), for assigning each text to a period.

	Full discriminant	**Stepwise discriminant** *[F = 5,00/4,99]*
Classification accuracy	**63,55%**	**68,47%**
Number of discriminant variables	**15**	**6**

Table 6 Accuracy obtained with the full and stepwise discriminant models (experimental phase 1)

Infinitival variable category	**Number of variables**	**Number of variables retained in the model** *[Stepwise discriminant]*	
sc	1		0
oc	1		0
aci	1	aci	1
4 (personal/impersonal)	2	4a	1
5 (infinitive as main verb)	2		0
6 (simple infinitive)	2	6b	1
7 (adverbial infinitive)	4	7a & 7a1	2
8 (articulate infinitive)	2	8b	1
Total	**15**		**6**

Table 7 Types of variables retained in the stepwise discriminant model (experimental phase 1)

[5] The terms "labelled" and "unlabelled" refer to known and unknown period membership respectively.

Table 8 presents the confusion matrix yielded for one ten-fold experiment of the stepwise discriminant, which indicates the amount of texts correctly identified as originating from a specific period. It can be observed that both ends of the synchronic continuum, that is Classical Greek (CL) and Late Byzantine (LB), constitute well-defined, discrete periods, since they exhibit a high classification accuracy (~ 92% of the CL texts were correctly characterised as originating from that synchrony and ~ 89% of the LB texts were also correctly characterised as originating from the corresponding period).

On the contrary, discriminating between the two periods in the middle of the synchronic continuum is not so straightforward. It is observed that 25% of the Koine texts are classified into the Early Byzantine period. In a similar vein, more than 30% of the Early Byzantine texts were predicted by the model as belonging to Koine.

		Predicted period membership according to the model			
		CL	**KO**	**EB**	**LB**
Actual period membership	**CL**	**91.7%**	8.3%	0.0%	0.0%
	KO	4.4%	**46.7%**	31.1%	17.8%
	EB	5.0%	25.0%	**62.5%**	7.5%
	LB	0.0%	5.0%	6.3%	**88.8%**

Table 8 Confusion matrix one ten-fold experiment of the stepwise discriminant model (experimental phase 1)

Figure 3 depicts the aforementioned picture by indicating the relative positions of patterns along the first two discriminant functions. The mapping of texts onto the pattern space shows that the CL and LB group centroids are far apart, while the KO and EB groups overlap and their centroids are very closely positioned in the space defined by the first two discriminant functions.

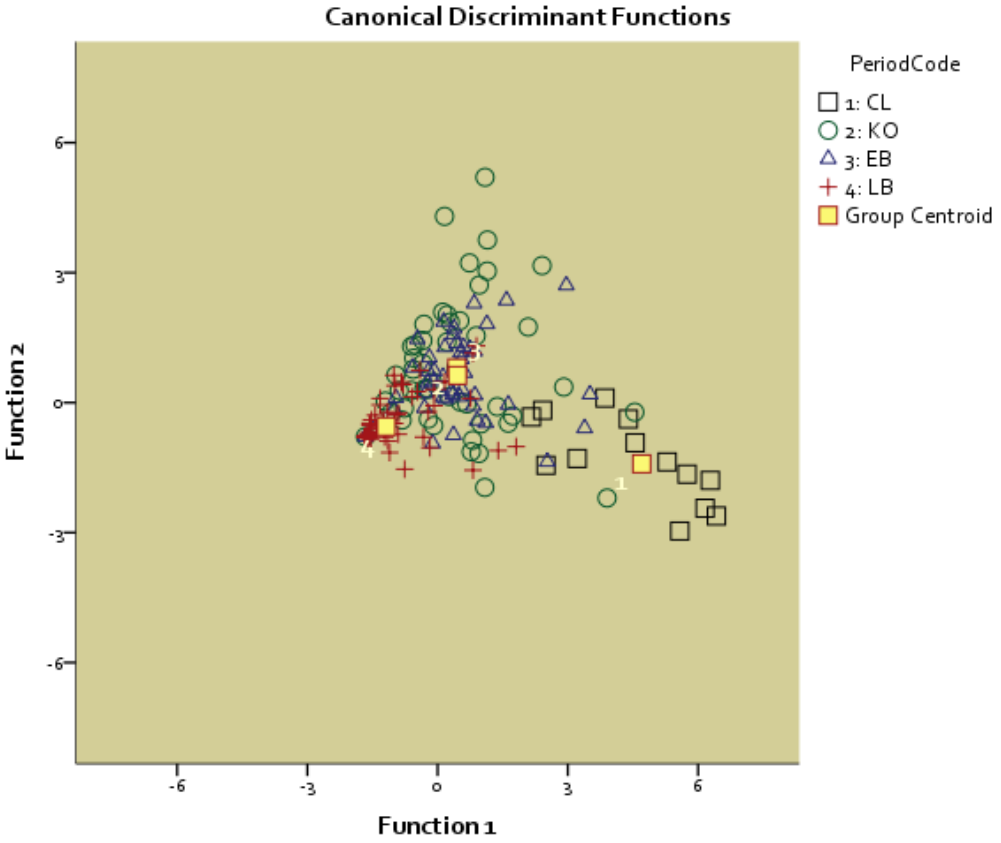


Figure 3 Mapping of the texts onto the space formed by the first two discriminant functions (experimental phase 1)

6.3.2 Experimental phase 2

As mentioned before, the boundaries between Koine and the Early Byzantine period are not so strict, as is evident from the low accuracy obtained when discriminating between texts for these two periods. This indicates that these two periods are very close and not easily distinguishable in terms of the use of the infinitive. Hence, a second experimental phase followed, which involved merging the Koine texts and the Early Byzantine ones into a single group, with the aim of investigating whether the classification accuracy of the model could be increased.

The experimental set-up was similar to the experimental phase 1 reported in Section 6.3.1. So, two models were employed, namely a **full** discriminant model and a **stepwise** one. The value of the *F* parameter for entering/removing variables was set again to **5.00/4.99**. Furthermore, the ten-fold cross-validation methodology was once more followed in order to verify the validity of the results obtained.

Table 9 presents the classification accuracy yielded in the second experimental phase. The accuracy substantially increases by ~ 17% and ~ 14% for the full and the stepwise model respectively, exceeding 80% in both cases. Moreover, in the case of both models the baseline rate of 33.3% for a 3-class discrimination experiment is comfortably exceeded, while the classification error is reduced by ~ 46% and ~44% respectively over the 4-class task.

	Full discriminant	**Stepwise discriminant** *[F = 5,00/4,99]*
Classification accuracy	**80,49%**	**82,31%**
Number of discriminant variables	**15**	**8**

Table 9 Accuracy obtained with the full and stepwise discriminant models (experimental phase 2)

Table 10 illustrates the variables that were used as discriminators in the stepwise model. The full discriminant model employs yet again fifteen (15) variables, whereas the variables of the stepwise model rise to eight (8), the two additional variables over the 4-class task being those related to the object control structures (oc) and the adverbial infinitive with an Accusative subject (7b).

Infinitival variable category	**Number of variables**	**Number of variables retained in the model** *[Stepwise discriminant]*	
sc	1		0
oc	1	**oc**	**1**
aci	1	Aci	1
4 (personal/impersonal)	2	4a	1
5 (infinitive as main verb)	2		0
6 (simple infinitive)	2	6b	1
7 (adverbial infinitive)	4	7a, **7b** & 7a1	3
8 (articulate infinitive)	2	8b	1
Total	**15**		**8**

Table 10 Types of variables retained in the stepwise discriminant model (experimental phase 2)

The confusion matrix (Table 11) generated for one ten-fold experiment of the stepwise discriminant model shows that the model is now able to predict more accurately the period membership of texts, since 80% of the KO & EB texts are correctly classified into this period. This indicates that the synchronies can be highly discriminated, when the Koine and Early Byzantine texts are grouped into one category.

		Predicted period membership according to the model		
		CL	**KO & EB**	**LB**
Actual period membership	**CL**	**91.7%**	8.3 %	0.0%
	KO & EB	5.9%	**80.0%**	14.1%
	EB	0.0%	10.1%	**89.9%**

Table 11 Confusion matrix for one ten-fold experiment of the stepwise discriminant model (experimental phase 2)

The high discrimination of the three groups also becomes evident when mapping texts onto the pattern space (cf. Figure 4), where minor overlaps are detectible, while the group centroids are positioned far from each other.

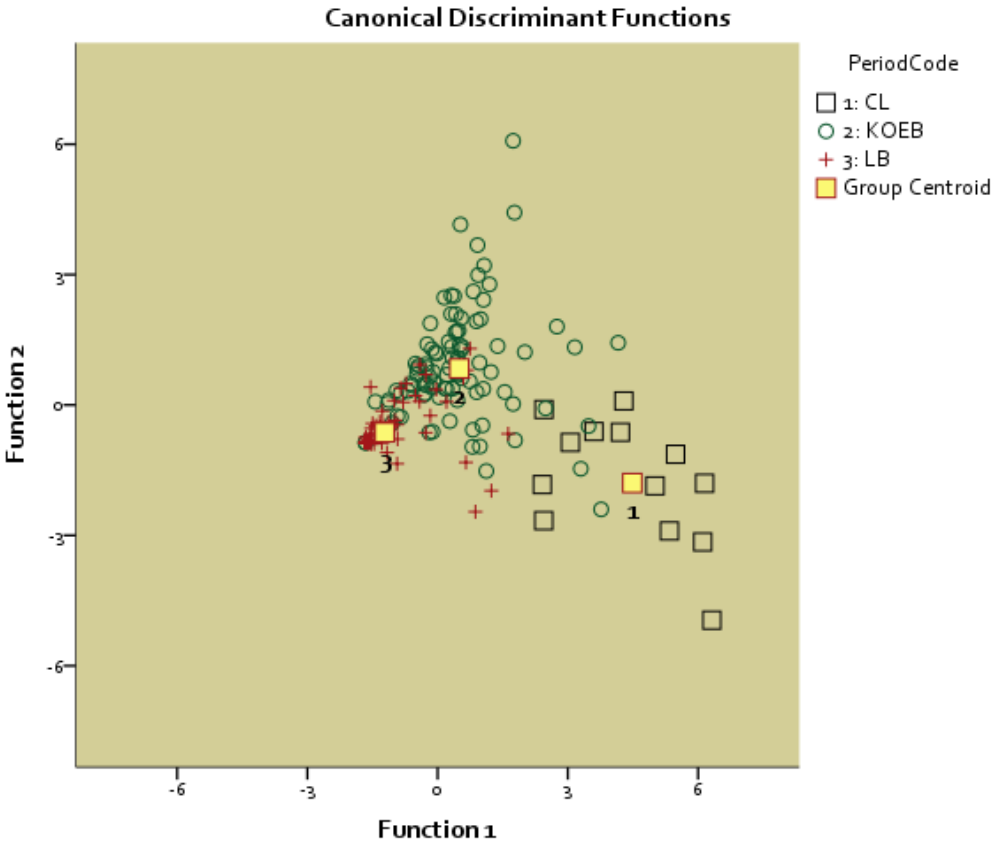


Figure 4 Mapping of the texts onto the space formed by the first two discriminant functions (experimental phase 2)

Finally, it is notable that the variables which serve as discriminating factors are either those whose frequency of occurrence systematically exhibits a declining trend through time, namely **4a** (personal/impersonal structures), **7a** and **7b** (adverbial infinitive 1 & complement of nominals with or without an Accusative subject), or those with substantial differences in frequency of occurrence across synchronies, that is **oc** (object control), **aci**, **6b** (simple infinitive with an Accusative subject), **7a1** (adverbial infinitive 2) and **8b** (articulate infinitive with an Accusative subject). The case of the infinitival subject also seems to play a significant role, since half of the discriminating variables correspond to structures with an Accusative subject.

7. Conclusions & future work

The present paper reported on a series of statistical analyses which were performed on an extensive corpus of infinitival structures. The aim was to draw some preliminary conclusions on the evolution of the Greek infinitive through time, supported by a statistical analysis.

The results obtained have shown that in all synchronies the categories exhibiting additively the highest frequency of occurrence are the subject and object control, aci and personal/impersonal structures together with the articulate infinitive. Besides, the subject control structures have been shown to form the most enduring category for all periods, always representing more than 25% of the infinitival occurrences.

Furthermore, the statistical analyses have asserted the correlation between time (i.e. the four synchronies) and the frequency of occurrence of infinitives and have highlighted a set of variables which can discriminate between synchronies in terms of the use of the infinitive.

It has already been mentioned that the present paper is part of an ongoing work, the object of which is the diachronic study of the Greek infinitival structures. Further experiments, either not reported here due to space restrictions or planned for the immediate future, include the employment of clustering techniques, discriminant analyses with variations based on text size or different values of the F parameter and the contrastive measurement of the frequency of occurrence of the final infinitive vs. the infinitive as complement of verbs of 'saying'.

Acknowledgments

The author wishes to acknowledge the valuable assistance of Dr. G. Tambouratzis and his comments on earlier drafts of this article. Of course all remaining errors are the author's responsibility.

References

Fisher, Ronald A. 1925. *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.

Horrocks, Geoffrey C. 2010. *Greek: A History of the Language and Its Speakers*. 2nd ed. Chichester/Malden, MA: Wiley-Blackwell

Jannaris, Antonius N. 1897. *An historical Greek grammar, chiefly of the Attic dialect as written and spoken from classical antiquity down to the present time, founded upon the ancient texts, inscriptions, papyri and present popular Greek*. Macmillan.

Joseph, Brian D. 1983. *The synchrony and diachrony of the Balkan infinitive*. Cambridge: Cambridge University Press

Klecka, William R. 1980. *Discriminant Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-019, Beverly Hills and London: Sage Publications.

Legrand, Émile. 1881. Bibliographie Hellénique (XVe- XVIe siècles). Tome deuxième. Paris.

Legrand, Émile. 1896. Bibliographie Hellénique (XVe- XVIe siècles). Tome huitième. Paris.

Levene, H. 1960. "Robust tests for equality of variances." In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann. Stanford, CA: Stanford University Press, pp. 278–292.

Wagner, Wilhelm. 1874. Carmina Graeca Medii Aevi. Lipsiae.