selected papers

of

the 10th International Conference of Greek Linguistics

Edited by

Zoe Gavriilidou
Angeliki Efthymiou
Evangelia Thomadaki
Penelope Kambakis-Vougiouklis

Komotini 2012

EXPLORING THE PRODUCTIVITY OF *-PIÓ* IN MODERN GREEK:
A CORPUS BASED STUDY

Angeliki Efthymiou
Democritus University
of Thrace, Greece
aefthym@eled.duth.gr

Georgia Fragaki
Aristotle University
of Thessaloniki, Greece
efraga@phil.uoa.gr

Angelos Markos
Democritus University
of Thrace, Greece
amarkos@eled.duth.gr

ABSTRACT

This is a corpus-based study of the morphological productivity of the Greek verb-forming element -pió*, treated here as an affixoid, gradually acquiring a suffixal status. Its productivity is measured in a corpus of 4,143,583 words (a part of the Corpus of Greek Texts) in relation to that of eight Greek verb-forming suffixes studied in Efthymiou et al. (2012). The results suggest that* -pió*'s productivity is similar to that of the moderately productive suffixes* -évo *and* -áro*. It is also found to be more productive in written registers, while it presents its highest productivity in opinion articles and its lowest one in literature.*

Keywords: corpora, frequency, Modern Greek, morphological productivity, *-pió*, text types, verb-forming elements

1. Introduction

This paper aims at exploring the productivity of the Modern Greek verb-forming element *-pió* in a subcorpus of the Corpus of Greek Texts (CGT), a reference corpus of Greek (see Goutsos 2010). We follow similar research on other languages, investigating frequency and morphological productivity with a corpus-based methodology (e.g. Baayen & Lieber 1991, Baayen 1992, Plag et al. 1999, Gaeta & Ricca 2003). This is part of a larger project studying the productivity of verb-forming elements, based on a large corpus of Modern Greek. Although *-pió*'s morphological status is debatable (i.e. whether it is a suffix or not), in the present study we treat this element as an affixoid moving towards acquiring a suffixal status, an element which is in a process of grammaticalization, becoming a suffix, according to relevant observations in the literature (see Anastassiadis-Symeonidis 1986, Giannoulopoulou 2000). On the basis of this, the morphological productivity of *-pió* is measured here in relation to that of Greek verb-forming suffixes *-áro, -éno, -évo, -íno, -ízo, -(i)ázo*, *-jázo* and *-óno*, studied in another paper (Efthymiou et al. 2012). This study can offer useful evidence on how productive *-pió* is in relation to the productivity of well-established suffixes. Here, a comprehensive productivity profile of *-pió* is attempted to be drawn by the examination of its productivity across written and spoken texts, as well as its productivity patterns across the five largest text types in the corpus (i.e. literature, news, popularized non-fiction, academic and opinion articles).

In summary, this paper aims at answering the following research questions:

(a) How productive is *-pió* in relation to the productivity scores of the eight verb-forming suffixes mentioned above?

(b) Is *-pió* equally productive across spoken and written registers?

(c) Is *-pió* more productive in certain text types?

The paper is organized as follows. The following section presents the main properties of the element under investigation, while section 3 is dedicated to a brief overview of the relevant literature

on the notion of productivity. In section 4 we explain how the data of this study was selected and outline the methodology used for data analysis. Section 5 describes the results of the research and section 6 discusses and summarizes the findings, along with drawing the implications of the study.

2. The Modern Greek verb-forming element *-pió:* morphological properties and status

In this section we outline the properties of the Modern Greek verb-forming element *-pió*, based on Anastassiadis-Symeonidis' (1986), Giannoulopoulou's (2000), Mela-Athanasopoulou's (2007) and Efthymiou's (2011a) analysis, and discuss its morphological status. For the semantic description of the derived verbs we will use the labels and glosses found in Plag (1999), namely causative/resultative 'cause to become x/turn into x', ornative 'make x go to/in/on something', locative 'make something go to/in/on x', inchoative 'become x', performative 'do x/perform x', similative 'act or be like x', instrumental 'use x' and stative 'be x' (see also Lieber 2004).

2.1 The structural and semantic properties of *-pió*

Historically, the bound Modern Greek morpheme *-pió* developed from the Ancient Greek verb *poiô* 'to make/do' (cf. Triandafyllidis Dictionary 1998). According to the literature, the verb-forming element *-pió* attaches both to nominal and adjectival bases. Interestingly, the majority of its adjectival bases are relational adjectives in *-ikós*. All *-pió* formations are transitive verbs and their meanings can be described as resultative, locative and ornative, as seen in example 1:

(1)   a.   *aplopió* 'simplify'   *aplós* 'simple'
          b.   *elaçistopió* 'minimize'   *eláçistos* 'minimal'
          c.   *stereopió* 'solidify'   *stereós* 'solid'
          d.   *γramatikopió* 'grammaticalize'   *γramatikós* 'grammatical'
          e.   *prosopopió* 'personify'   *prósopo* 'person'
          f.   *periθoriopió* 'marginalize'   *periθório* 'margin'
          g.   *omaδopió* 'divide into groups'   *omáδa* 'group'
          h.   *morfopió* 'to form'   *morfí* 'form'

According to Anastassiadis-Symeonidis (1986), *-pió* is extremely productive in forming neologisms with resultative meaning in Modern Greek and thus it enters in competition with other verb forming processes, like the verbal suffixes *-évo, -ízo,* and *-óno*. For example, the neologism *elino-pió* 'to turn into a Greek' has begun to replace the older synonymous parasynthetic verb *eks-elin-ízo* (*Élinas* 'Greek'). More specifically, Anastassiadis-Symeonidis (1986) argues that formation with *-pió* is more productive than parasynthetic verb formation for the following reasons:

    a) *-pió* is less sensitive to phonological, morphological and lexical restrictions than Modern Greek verb-forming suffixes,

    b) *-pió* formations are more transparent and predictable in meaning than parasynthetic formations (for the semantic description of the Modern Greek parasynthetic formations see Tsakou 2010 and Efthymiou 2011b),

    c) *-pió* is more likely to be phonotactically signalled because of the presence of the linking vowel *-o-*. This argument accords with Hay (2000), who argues that consonant-initial suffixes are more productive than vowel-initial ones, because they are more likely to provide phonotactic boundary signals.

    In addition, Anastassiadis-Symeonidi's claim about the predictability of *-pió* is in line with Efthymiou (2011a, 2001b), who argues that the meanings of *-pió* verbs are much more restricted than the meanings of Modern Greek parasynthetic or suffixed verbs. As already mentioned, *-pió* verbs are restricted to resultative, ornative and locative meanings, whereas suffixed or parasynthetic verbs can also express other meanings, such as similative, performative or privative meanings (e.g. *piθicízo* 'to imitate ape's behaviour', *apo-cefal-iz-o* 'to decapitate', *xorévo* 'to dance'). Finally, this observation also accords with Mela-Athanasopoulou (2007), who observes that inchoative meanings are only possible with the passive voice of *-pió* verbs.

2.2 The morphological status of *-pió*

According to several studies (e.g. Anastassiadis-Symeonidis 1986, Giannoulopoulou 2000, Ralli 2005, Dimela 2010), *-pió* is an affixoid, i.e. an element that shares properties both with derivation and compounding.[1] An affixoid is neither a stem nor an affix but is an element that is in process of being grammaticalized, gradually losing its lexical status as a stem and behaving more or less like an affix. The basic arguments found in the literature for treating *-pió* as an affixoid are the following:

a) Its homonymous free form has lost its independent status in Modern Greek. The use of the free form *pió* is restricted to fixed or archaic expressions.

b) There exist a few verbs in which *-pió* is a compound element and combines with prefixes (e.g. *ek-pió* 'to sell up', *para-pió* 'to counterfeit'). Nevertheless, all these forms are [+learned] verbs, which originate from Ancient Greek.

c) The Modern Greek bound element *-pió* expresses more meanings than the homonymous Ancient Greek free form *poiô* 'to do/make'.

d) A considerable number of Modern Greek *-pió* verbs, which are mostly used in academic terminology, are loan translations of English (or French) suffixed verbs (e.g. *γramatikopió* 'to grammaticalize').

e) *-pió* is not phonologically reduced or fused with its base. Furthermore it is phonotactically signalled because of the presence of the linking vowel *-o-*, which is found only in compounding.

Based on the above mentioned phonological, semantic and structural properties, *-pió* can be considered an affixoid, moving towards acquiring a suffixal status but still displaying no phonological attrition (see Anastassiadis-Symeonidis 1986 and Giannoulopoulou 2000 for details). In other terms, if compounding and derivation are thought of as the two poles of a morphological continuum, *-pió* is placed closer to suffixation and can be regarded as a suffix-like element.[2]

3. Morphological productivity

Morphological productivity is a central issue in the field of derivational morphology, since its measurement can help us describe the current and/or future involvement of a word formation process in the coinage of new words (cf. Bauer 1983: 18, Plag 2003: 44). The relevant literature has studied morphological productivity following different approaches, mainly qualitative or quantitative. This variety of approaches is justified by the different aspects of morphological productivity itself. According to Corbin (1987), productivity can be divided into two distinct phenomena: availability (translation of the French term "disponibilité", referring to how available a morphological process is to produce new words,) and profitability (translation of the French term "rentabilité", referring to how exploitable a morphological process is in order to create new words). In this sense, availability is a qualitative notion, since a morphological process is either available or not (Bauer 2001: 205), whereas profitability is a quantitative notion, since it reflects the extent to which a morphological process may be exploited to create new forms (cf. Plag 2006).

Qualitative approaches usually deal with various kinds of restrictions (i.e. structural, pragmatic, psycholinguistic etc.; see e.g. Plag 1999, Rainer 2005) that take part in word formation, while quantitative approaches deal with frequency and probability. As frequently observed, there is an inverse correlation between the number of restrictions and the productivity of a word formation process: the more restrictions apply, the fewer words will be derived (Booij 1977, Lieber 2010). This interrelation of qualitative and quantitative aspects of productivity, which is also observed by Plag (1999: 22), suggests that both quantitative and qualitative approaches have to be taken into consideration when talking about productivity (cf. e.g. Aronoff & Fudeman 2011, Plag 2006).

In this paper we explore the productivity of *-pió* mainly in quantitative terms. In particular, we follow Baayen and his collaborators (e.g. Baayen 1993, Baayen & Lieber 1991), who have proposed quantitative measures of productivity, defined with respect to the frequency of a given word-formation process. This frequency is measured in terms of "type frequency", i.e. the number of different words that occur with the morphological category of *-pió* (e.g. *nomimopió* 'to legitimate', *xrisimopió* 'to use'

---

[1] Other terms, which are frequently used by linguists in order to account for these borderline cases, are semi-affixes, affix-like elements, lexical suffixes or confixes (see, among others, Anastassiadis-Symeonidis 1986, Giannoulopoulou 2000, Ralli 2005, Mela-Athanasopoulou 2007).
[2] On the morphological continuum of compounding and derivation see Dalton-Puffer & Plag (2001), Ralli (2005) and Dimela & Melissaropoulou (2009), among others.

and *δimosiopió* 'to publish' are three types of the morphological category *-pió*), and "token frequency", i.e. the actual number of occurrences within the morphological category of *-pió* in a given text (e.g. *δimosiopió* 'publish$_{\text{1st.pers.sing.pres.}}$' and *δimosiopiís* 'publish$_{\text{2nd.pers.sing.pres.}}$' count as two occurrences, i.e. tokens, of the morphological category *-pió*).

Baayen and his collaborators (e.g. Baayen 1992, 1993, 2001, cf. also Baayen & Lieber 1991, Baayen & Renouf 1996, Plag et al. 1999) have elaborated a number of distinct and complementary corpus-based statistical measures of productivity, which all rely on the availability of large electronic text corpora. These measures can be operationalized by the concept of vocabulary growth, i.e. how frequently new word types that are formed by a morphological process are encountered when an increasing amount of text is sampled (Baayen 2001). This study relies on two well-known measures of morphological productivity proposed in this probabilistic framework.

One of Baayen's measures is termed "realized productivity". It is given by the number of word types *V* (or the vocabulary size) of a morphological category *C*, in a corpus of *N* tokens:

$$V = V(C, N)$$

This measure is also known as "type frequency" or "extent of use" (cf. Baayen 2008a, Plag 2006). According to Bauer (2001), realized productivity is restricted to past achievement (cf. Baayen 2008a) and cannot provide information as to whether a certain morphological process is available or not.

Another measure proposed by Baayen, known as "potential productivity", relies on the notion of "hapax legomenon". Hapax legomena (or hapaxes) are the types that occur only once in a corpus. The rate at which a vocabulary grows can be captured by the proportion of hapax legomena to the total number of tokens *N (C)* of all words with the morphological category *C*. Thus, potential productivity is given by:

$$P = V(1, C, N) / N(C)$$

According to Baayen and Lieber (1991), this ratio estimates the rate at which new types of a given morphological category are to be expected, given that the size of the sample of relevant observed types equals *N (C)*. This type of productivity, also called "productivity in the narrow sense" (cf. Plag 1999), serves to estimate the rate at which a morphological category enriches the vocabulary. The rate at which new types appear in the corpus can be visualized via a vocabulary growth curve. This curve reports vocabulary size (number of types, *V*) as a function of sample size (number of tokens, *N*) and *P* can be also seen as the slope of the tangent to this curve at *N* (Baayen 2001: 49-50).

However, potential productivity has been subjected to criticism (Baayen 1993, Bauer 2001) for ignoring type frequency and due to the fact that, since *P* is itself a function of *N*, its value depends on the size of the sample of the corpus. It is therefore problematic to compare directly a small subcorpus with a large subcorpus without distortion, due to the substantial differences in the overall sizes of the subcorpora and the substantial differences in *V* and *P* (Plag et al. 1999, Baayen 1992, 2008b). Fortunately, there are two ways to overcome this problem (Baayen 2008c: 272-274). The first is to compare the number of types across texts for the same text sizes. For larger texts, a random sample of the same size as the smallest text in the comparison has to be selected. The second way is to resort to models of vocabulary growth, especially developed for this purpose (see Baayen 2001 for an overview of these models).

4. Data and methodology

The data used for the measurement of productivity of *-pió* comes from the Corpus of Greek Texts (CGT), a synchronic and monolingual corpus of Modern Greek, including approximately 28 million words from a variety of spoken and written text types. (For more details on the corpus, see Goutsos 2010). For the aims of this study a subcorpus of the CGT (henceforth CGT4) was created by randomly selecting 4,155,036 words. The measurement of productivity in a corpus smaller than the CGT was considered as more manageable, since CGT is not tagged (e.g. morphologically or for parts of speech) and, consequently, the extraction of the verbs formed by *-pió* could only be done by manual pre-processing. CGT4 was stratified by text type, both spoken (conversation, interviews, news, public speech) and written (information items, law and administration texts, literature, anecdotes, popularized non-fiction texts, news, opinion articles, academic texts, private texts, such as private letters, electronic texts, diary, ephemera and procedural texts etc). All text types found in the CGT are represented in the CGT4, in the same proportion in which they occur in the former. It is also worth noting that all texts contained in the CGT4 are in Standard Modern Greek. Table 1 below presents the number of words and the percentage of the text types in CGT4.

Text type	Number of words	Percentage
Conversation	17,867	0.43
Interview	98,059	2.36
News (spoken)	44,874	1.08
Public speech	296,254	7.13
Spoken (Total)	*457,054*	*11.0*
Information items	16,620	0.40
Law and administration	186,145	4.48
Literature	382,263	9.20
Private	13,711	0.33
Popularized non-fiction	1,289,724	31.04
News	677,272	16.30
Other	55,262	1.33
Academic	633,643	15.25
Opinion articles	443,342	10.67
Written (Total)	*3,697,982*	*89.0*

Table 1 Size of text types in the CGT4

As can be seen in Table 1, the percentage of spoken data (11% of the corpus) is much lower than the percentage of written data (89% of the corpus). In addition, the percentages of the specific text types differ largely: for example, public speeches occupy approximately two thirds of the spoken corpus, while conversation hardly occupies one half percent out of the total 11% of the spoken corpus.

First, a frequency list of all words in the corpus was made using the wordlist function of Wordsmith Tools 5.0 (Scott 2008). The list was sorted in reverse alphabetical order (so that tokens with the same ending were clustered together) and all verbs formed by -*pió* were then manually identified. It is important to note that we did not search for the participles of -*pió* verbs, excluding thus from the list the (usually rare) periphrastic types of the present perfect and past perfect in the passive voice (*íme* 'to be' + past participle). This choice was also theoretically driven, since the inclusion of these periphrastic types in the tense system of Greek is controversial.[3]

After the compilation of the list, items with a problematic status were removed, following a standard procedure in the literature (see e.g. Plag 1999: 28-29, Plag et al. 1999: 214, Fradin et al. 2008: 38ff.). In particular, all verbs displaying derivational inner cycles, i.e. derived by prefixation (e.g. *apo-staθero-pió* 'to destabilize') and composition (e.g. *proto-xrisimo-pió* 'to use for the first time') were excluded. Items which fall into the category of "base-less derivatives", as Gaeta and Ricca (2003: 71) call them, were also removed from the lists. For instance, the verb *posostiko-pió* 'to quantify' was excluded, since its base **posostiko-* does not exist as such in Greek; this is rather a merging of *posotikós* and *posostó*. In addition, items "semantically opaque but formally analyzable" in terms of Plag et al. (1999: 214), such as *ikano-pió* 'to satisfy', which cannot be semantically related to the adjective *ikanós* 'capable', or *tropo-pió* 'to modify', which cannot be semantically related to the noun *trópos* 'means', were also excluded from the final list. Finally, a few verbs in which -*pió* is a compound element, combining with prefixes, such as *ek-pió* 'to sell up', *para-pió* 'to counterfeit' and *meta-pió* 'to remake/to alter', were not included. It is interesting to note that the majority of the types removed from the list were verbs derived by prefixation.

The application of these criteria reduced the number of tokens under study. Details for the number of tokens in the raw data, i.e. the total number of tokens in the corpus before clearing the data, are given in Table 2.

Suffixes	Raw data	Data after clearing		
	Tokens	Tokens	Types	Hapax legomena
-pió	6,195	5,130	132	40

Table 2 Frequencies in raw data and after clearing

---

[3] See, for example, Moser (1994: 140ff.), where it is supported that these structures in Standard Modern Greek are mainly stative predicates, something which means that they are not principally used for tense distinctions.

The tokens selected for the study were manually lemmatized and the number of hapax legomena was counted. The number of types, as well as the number of hapaxes found in the data included in the study, is also shown in Table 2. The selected tokens and types of *-pió*, as well as its hapaxes, were searched in the corpus as a whole, in the written and spoken registers and in each individual text type included in the CGT4. This search can offer valuable information about the relation of the productivity of *-pió* to written and spoken registers and text types.

The substantial differences in the overall size of the subcorpora (see Table 1) prompted us to resort to a family of parametric statistical methods, called Large-Number-of-Rare-Events (LNRE) models (for an overview see Baayen 2001). LNRE models can be used to quantify the relative productivity of two or more morphological processes by looking at their vocabulary growth rate as sample size increases. An LNRE model attempts to estimate the expected number of types (the vocabulary size) both at smaller sample sizes (interpolation) and at larger sample sizes (extrapolation), based on the counts of low frequency types in the corpus (the frequency spectrum). The technique of extrapolation produces the expected values of types for arbitrary values of tokens, larger than the empirical number of tokens. Currently, three major models are available: Generalized Inverse Gauss-Poisson (GIGP; Baayen 2001), finite Zipf-Mandelbrot and Zipf-Mandelbrot (fZM and ZM; Baroni & Evert 2006). These models are implemented in the package zipfR (Baroni & Evert 2006), a tool for lexical statistics in the R language, which is used in this study.

In order to address research question (a), regarding the productivity of *-pió* as compared to that of the eight Greek suffixes, an appropriate LNRE model is first computed and a corresponding growth curve is obtained for the suffix with the highest number of tokens (*-ízo*). Then, this model is used to extrapolate the growth curves of the other suffixes up to the size of *-ízo*. The appropriateness of each model for the data is assessed by a multivariate goodness-of-fit test (Baayen 2001: 118-122). To further assess model fit, the observed frequency spectrum of each suffix and *-pió* is compared with the fitted model predictions. The growth curves are graphed for 100 equally sized intervals and 95% confidence intervals are plotted around the curves. A 95% confidence interval for *V* gives the range of values that *V* is most likely to have when calculated for new corpora of the same design and size.

The difference in productivity of *-pió* across spoken and written registers (research question b), is explored in a similar manner to (a). The growth curves of *-pió* are plotted for 40 equally sized intervals up to the number of tokens sampled for it in the written subcorpus, the largest subcorpus of the CGT4. The growth rates *P* of the vocabulary are calculated for both the largest and the smallest sample size.

The investigation of differences in suffix productivity across different text types (research question c) is only relevant for the five largest text types in the CGT4 (literature, news, popularized non-fiction, academic and opinion articles). For the remaining text types, there are cases where the number of different tokens and types are not sufficient to fit an LNRE model or the model fit is not acceptable. In this context, an appropriate LNRE model is computed for each text type in the interval [1, 15316].

5. Results

The morphological productivity of *-pió* has been measured in relation to that of the eight verb-forming suffixes (*-áro, -éno, -évo, -íno, -ízo, -(i)ázo, -jázo* and *-óno*) studied in Efthymiou et al. (2012). The vocabulary growth curves of *-pió* and the eight suffixes in CGT4 are presented in Figure 1. The figure illustrates how vocabulary size, i.e. the number of types, shown on the vertical axis, increases as one reads through the tokens of the corpus, plotted on the horizontal axis. The number of types plotted corresponds to the expected vocabulary size $E[V(N)]$, i.e. the number of different types one may expect to count on average for a great many different orderings of the text fragments in a given corpus. A finite Zipf Mandelbrot model was fitted for the majority of suffixes, with the exceptions of *-ízo* and *-évo*, where a Generalized Inverse Gauss-Poisson (GIGP) model performed much better.
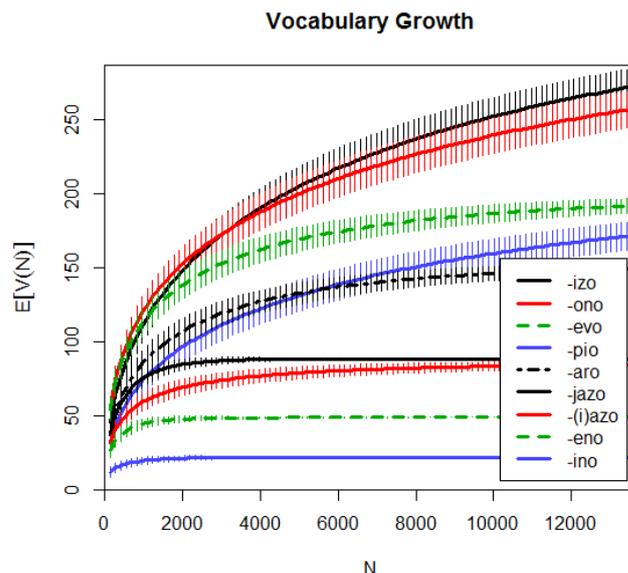
Figure 1 The expected number of types E[V(N)] for *-pió* and the eight suffixes in the CGT4 as a function of the size in tokens *N* of the suffix *-ízo* (with 95% confidence intervals)

The growth curves are plotted for 100 equally sized intervals between 1 and 15316, which stands for the number of tokens sampled for *-ízo* in the corpus. Figure 1 also plots 95% confidence intervals around the vocabulary growth curves. By means of the resulting vocabulary growth curves, we can easily compare the productivity of *-pió* and the eight suffixes for a range of different values of corpus sizes *N* (cf. Plag et al. 1999 for a similar approach to some English affixes). Two curves can be regarded as significantly different, if one is outside the confidence interval of the other.

According to our analysis in Efthymiou et al. (2012), there are three main sets of suffixes: (a) the very productive suffixes (*-ízo* and *-óno*), (b) the moderately productive suffixes (*-évo* and *-áro*) and (c) the least productive or unproductive suffixes (*-jázo, -(i)ázo, -éno* and *-íno*). The vocabulary curve of *-pió* is similar to that of moderately productive suffixes. In particular, *-pió* is less productive than *-ízo*, *-óno* and *-évo* and more productive than *-jázo*, *-(i)ázo*, *-éno* and *-íno*. The confidence intervals of *-pió* and *-áro* considerably overlap; thus it is difficult to decide which of the two is the most productive.

The results for *-pió* can be considered reliable, since the overall model fit was found to be satisfactory, when further assessed by comparing the observed frequency spectra of *-pió* with the fitted model predictions. A frequency spectrum is the number of types per frequency class, i.e. how many types occur once, twice and so on (cf. Baayen 2001). Figure 2 demonstrates the observed frequency spectrum and the fitted predictions for *-pió*. As the plot indicates, there are evidently some problems, especially at two, six and nine occurrences. These discrepancies can be attributed to possible violations of the randomness assumption of the models or to insufficient sample size. However, the overall fit appears to be satisfactory.

Frequency Spectrum: -pio
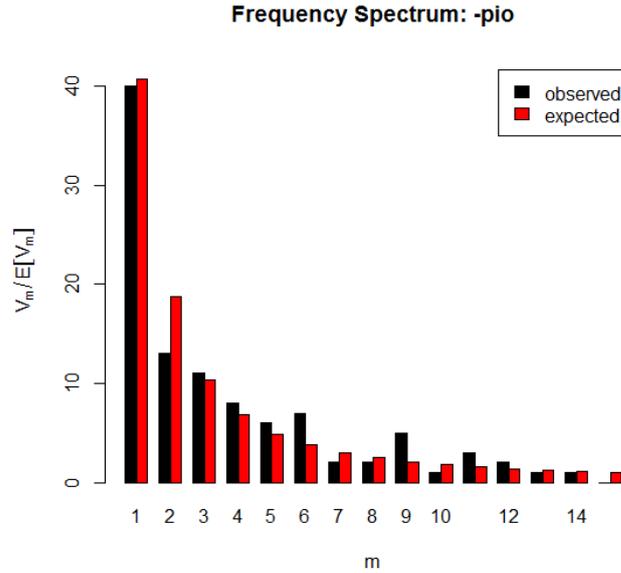
Figure 2 Frequency spectrum for *-pió* alongside predictions of the finite Zipf-Mandelbrot (fZM) model ($\chi^2(3) = 1.44$, p = 0.742). The y-axis shows the observed and expected number of types V that occur exactly m times. The plot shows the 15 types with the lowest frequencies.

5.1 Productivity of *-pió* across spoken and written registers

In order to estimate the productivity of *-pió* across the spoken and written registers of the CGT4, we have created the growth curves for 40 equally-sized intervals between 1 and the number of tokens sampled for the suffix in the written subcorpus. According to Figure 3, *-pió* is more productive in written than in spoken registers.

Vocabulary growth: -pio
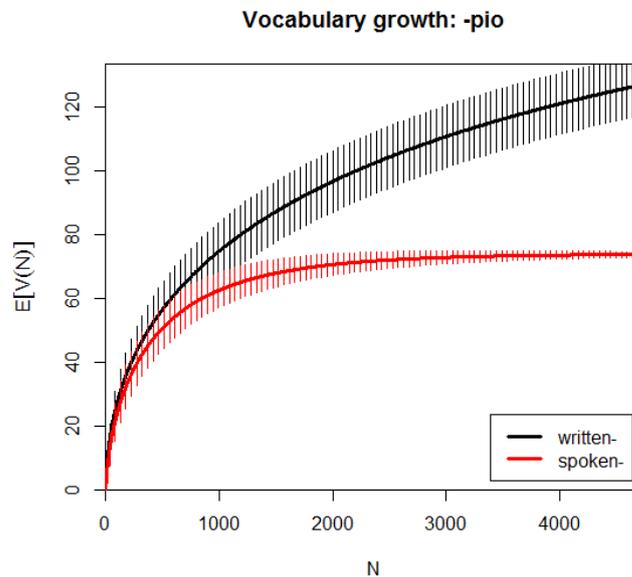
Figure 3 The expected number of types E[V(N)] for *-pió* (calculated by means of the finite Zipf-Mandelbrot (fZM) model) in the written and spoken subcorpora of the CGT4 as a function of the size in tokens *N* sampled for *-pió* in the written subcorpus (with 95% confidence intervals)

Table 3 shows the potential productivity of *-pió* based on the largest (i.e. written) and the smallest (i.e. spoken) sample size.

Suffix	N(Written)		N(Spoken)	
	Written	Spoken	Written	Spoken
-pio	0.0075	0.0001	0.0559	0.0474

Table 3 Potential productivity of *-pió* across written and spoken registers at different values of *N*

The results are similar with the ones derived from the growth curves: *-pió* is more productive in written than spoken registers, since it presents higher scores in the column "Written" for both points of reference (compare columns "*N*(Written)" and "*N*(Spoken)").

5.2 Productivity of *-pió* across text types

In this section we present a comparison of the contribution of *-pió* to the growth of the vocabulary of the largest written subcorpora in the CGT4, namely literature, news, popularized non-fiction, academic and opinion articles. Figure 3 visualises the growth curves for the five text types. The curves are plotted for 100 equally sized intervals between 1 and 15316, the number of tokens sampled for *-ízo* in the corpus. *-pió* tends to yield more types in opinion articles, academic texts and popularized non-fiction, while it is least productive in news and literature. It is also notable that the curve for opinion articles is almost always above the curves for the other text types, which shows the greater potential of *-pió* to form new words in opinion articles.
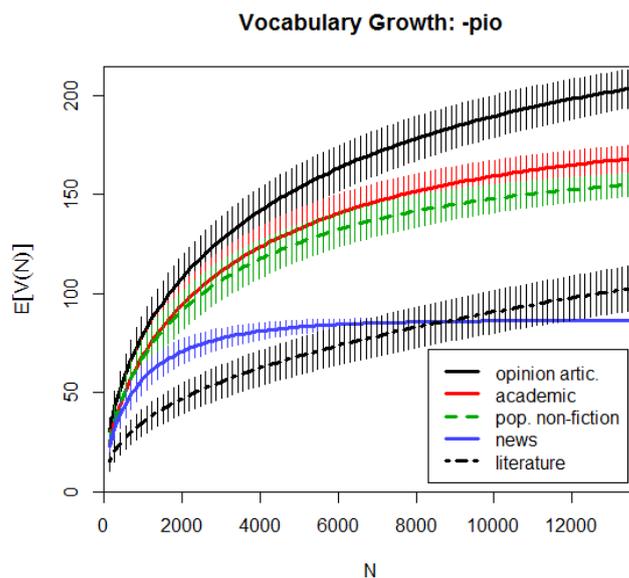
Figure 4 The expected number of types E[V(N)] for *-pió* in each text type as a function of the size in tokens *N* of the suffix *-ízo* (with 95% confidence intervals)

Table 4 shows the potential and realized productivity of the five text types in the interval [1, 15316] in ascending order.

Text types	Potential	Realized
Literature	0.0067	73.67
News	0.0051	79.58
Popularized non-fiction	0.0100	128.92
Academic texts	0.0110	134.06
Opinion articles	0.0133	157.75

Table 4 Realized and potential productivity of *-pió* by text type for *N* = 15316

The results are similar with the ones derived from the growth curves: *-pió* is more productive in opinion articles and less productive in literature. Table 4 also confirms the pattern observed in Figure 4, that *-pió* is more productive in the set of text types including opinion articles, academic texts and popularized non-fiction texts than in the set consisting of news and literature.

6. Discussion and conclusions

As shown in the previous section, the affixoid -*pió* has proven to be a moderately productive verb-forming element, which can be considered to belong to the same set of verb-forming elements as the suffixes -*évo* and -*áro*. More specifically, -*pió* is found to be less productive than -*ízo* and -*óno*, which are the most productive Modern Greek suffixes, and more productive than -*jázo*, -*(i)ázo*, -*éno* and -*íno*. In the set of verb-forming elements with moderate productivity -*évo* stands at some remove from the other members, while the relation between -*pió* and -*áro* is complicated. As observed in section 5 (see Figure 1), the confidence intervals of -*pió* and -*áro* considerably overlap and thus it is difficult to decide which is the most productive. Both verb-forming elements present a rather high number of hapaxes in the data (approximately one third of their types are hapax legomena; see Table 2 for -*pió* and Efthymiou et al. 2012 for -*áro*), something which suggests that the probability of finding new words formed by these elements is very high. Apart from their similarities regarding productivity scores and number of hapaxes, -*pió* and -*áro* are both relatively new verb-forming elements in the history of Greek, which seem to enter in competition with older Modern Greek verb-forming suffixes, such as -*évo*, -*ízo* and -*óno* (see e.g. Anastassiadis-Symeonidis 1986; cf. Efthymiou et al. 2012). They are also both related to the introduction of loan words in Greek, following, however, a different path: -*áro* shows preference for non-native bases (e.g. *tsekáro* 'to check', base: *tsek* 'check'), while -*pió* is frequently used in verbs which are loan translations of English (or French) suffixed verbs. This common characteristic of -*pió* and -*áro* can provide an explanation of their relatively high productivity, taking into account that other suffixes like -*ízo*, which were used in order to accommodate foreign loan words, are also very productive (see Anastassiadis-Symeonidis 1994, Ralli 2011).

Moreover, the attempt to refine the findings on the productivity patterns of -*pió* by looking at its productivity across written and spoken registers, on the one hand, and text types, on the other, has also yielded significant results. As was observed, -*pió* is more productive in written than spoken registers, a finding that can be related to the [+learned] character of the verbs it forms. It is worth noting here that the preference of -*pió* for written registers is not divergent, but is rather common; as our study of Greek verb-forming suffixes has indicated (Efthymiou et al. 2012), only two out of eight suffixes were found to be more productive in spoken than written registers. This result seems to confirm the observation found in the literature that written language is lexically richer than spoken language (Plag et al. 1999; cf. Biber et al. 1999: 53).

Although -*pió* and -*áro* are observed to have similar productivity scores, they show mirror image preferences as regards their productivity in spoken and written texts: -*pió* is more productive in written than spoken registers, whereas -*áro* appears to be more productive in spoken texts. As observed in Efthymiou et al. (2012), the high productivity of -*áro* in spoken texts can be related to the [-learned] or [+/-learned] character of its derivatives, which generally belong to everyday vocabulary, used in spoken interaction (e.g. *frenáro* 'to brake'). In contrast, -*pió* usually forms verbs with a [+learned] character, which are mostly expected in formal or written texts. The observation that two verb-forming elements with similar productivity scores can show opposite preferences in terms of registers suggests that the study of productivity across spoken and written registers can be very useful in drawing a refined picture of verb-forming elements.

Turning to the productivity pattern of -*pió* across text types, it should be noted that the search was only possible for the five largest text types (i.e. literature, popularized non-fiction texts, opinion articles, news and academic texts), because of the relatively small size of the other text types. As observed in section 5 (see Figure 4), it was found that -*pió* tends to yield more types in opinion articles, academic texts and popularized non-fiction, while it is least productive in news and literature. The study of the particular verbs formed by -*pió* in each text type shows that it is more productive in text types containing a large amount of terminology. For instance, opinion articles, in which -*pió* is most productive, contain many terms related to economy, politics or other social issues such as *refstopió* 'to liquidate', *iðiotikopió* 'to privatize', *pangozmiopió* 'to globalize', *anotatopiúme* 'to become a University, for Polytechnics'. Similarly, academic texts present a high number of academic terms such as *kanonikopió* 'to normalize', *lektikopió* 'to verbalize', *fisiolojikopió* 'to naturalize', *tiçeopió* 'to make something accidental'. Popularized non-fiction texts have a wide range of terminology related to various aspects of life (e.g. *eθnikopió* 'to nationalize', *astikopió* 'to urbanize', *aðiavroxopió* 'to make something waterproof'), as well as evaluative vocabulary such as *iroopió* 'to turn someone into a hero' and *periθoriopió* 'to marginalize'. Finally, news and literature abound in everyday vocabulary such as *xrisimopió* 'to use', *praγmatopió* 'to realize'. It is worth noting that the majority of the verbs mentioned above are loan translations of English or French suffixed verbs.

Furthermore, the combination of quantitative evidence with qualitative observations (e.g. semantic, syntactic or other properties of verb-forming elements) seems to explain the productivity rankings of the items studied, an assumption also found in the literature (see, for example, Plag 1999). The example of *-pió*, *-évo* and *-áro* is characteristic: although *-pió* is used with only three meanings (instead of seven for *-évo* and *-áro*), it appears that its fewer phonological and morphological constraints, in comparison with *-évo* and *-áro* (see section 2 and Efthymiou et al. 2012), balance out its smaller number of meanings so that it shows similar productivity scores with them.

Finally, we believe that the findings concerning the productivity of *-pió* may offer evidence about its morphological status. In particular, the relatively high productivity of *-pió* could be considered as supporting the view that it has become a suffix in Greek. This suggestion could form a research hypothesis to be explored in a future research e.g. by measuring *-pió*'s productivity in relation to that of other affixoids.

References

Anastassiadis-Symeonidis, Anna. 1986. "On the nature and productivity of the forming element *-pió*." *Studies in Greek Linguistics 1985*:49-70. [In Greek]

Anastassiadis-Symeonidis, Anna. 1994. *Neological Borrowing in Modern Greek*. Thessaloniki: Filosofiki Scholi [In Greek].

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: The MIT Press.

Aronoff, Mark, and Kirsten Fudeman. 2011. *What is Morphology?* Malden/Oxford/Chichester: Wiley-Blackwell.

Baayen, Harald. 1992. "A quantitative approach to morphological productivity." In Geert Booij, and Jaap van Marle (eds), *Yearbook of Morphology 1991*, 109-149. Dordrecht: Kluwer.

Baayen, Harald. 1993. "On frequency, transparency and productivity." In Geert Booij, and Jaap Van Marle (eds), *Yearbook of Morphology 1992*, 181-208. Dordrecht: Kluwer.

Baayen, R. Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.

Baayen, Harald. 2008a. "Corpus linguistics in morphology: Morphological productivity." In Anke Lüdeling, and Merja Kytö (eds), *Corpus Linguistics. An International Handbook*, 899-919. Berlin: Mouton de Gruyter.

Baayen, R. Harald. 2008b. languageR: Data Sets and Functions with Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. R Package Version 0.95.

Baayen, Harald. 2008c. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Baayen, Harald, and Rochelle Lieber. 1991. "Productivity and English derivation: A corpus-based study." *Linguistics* 29:801-843.

Baayen, Harald, and Antoinette Renouf. 1996. "Chronicling 'The Times': Productive innovations in an English newspaper." *Language* 72: 69-96.

Baroni, Marco, and Stefan Evert. 2006. The zipfR package for lexical statistics: A tutorial introduction. http://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf.

Bauer, Laurie. 1983. *English Word-Formation.* Cambridge: Cambridge University Press.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Booij, Geert. 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*. De Ridder: Lisse.

Corbin, Danielle. 1987. *Morphologie dérivationelle et structuration du lexique*. 2 Vols. Tübingen: Niemeyer.

Dalton-Puffer, Christiane, and Ingo Plag. 2001. "Categorywise, some compound-type morphemes seem to be rather suffix-like: on the status of *-ful, -type*, and *-wise* in Present Day English." *Folia Linguistica* 34(3-4):225-244.

Dimela, Eleonora. 2010. Prefixation in Modern Greek Dialects. Synchronic and diachronic approach. PhD Thesis, University of Patras. [In Greek]

Dimela, Eleonora, and Dimitra Melissaropoulou. 2009. "On prefix-like adverbs in Modern Greek." *Patras Working papers in Linguistics*. Volume 1. Accessed December 10, 2011. http://xantho.lis.upatras.gr/pasithee/index.php/pwpl/article/view/17/14

Efthymiou, Angeliki. 2011a. "The semantics of verb forming suffixes in Modern Greek." In *Proceedings of the 19th International Symposium of Theoretical and Applied Linguistics, 3-5 April 2009*. School of English, Aristotle University of Thessaloniki, 175-184.

Efthymiou, Angeliki. 2011b. "Are there any principles governing the order and combination of affixes in Modern Greek parasynthetic verbs?" Paper presented at the 3rd Vienna Workshop on Affix Order, Vienna, 15-16 January 2011 [submitted].

Efthymiou, Angeliki, Fragaki, Georgia, and Angelos Markos. 2012. "Productivity of verb forming suffixes in Modern Greek: A corpus-based study." *Morphology*. DOI: 10.1007/s11525-012-9202-4

Fradin, Bernard, Dal, Georgette, Grabar, Natalia, Namer, Fiammetta, Lignon, Stéphanie, Tribout, Delphine, and Pierre Zweigenbaum. 2008. "Remarques sur l' usage des corpus en morphologie." *Langage* 171:34-59.

Gaeta, Livio, and Davide Ricca. 2003. "Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data." *Rivista di Linguistica* 15(1):63-98.

Giannoulopoulou, Giannoula. 2000. Morphosemantic comparison of affixes and confixes in Modern Greek and Italian. PhD Dissertation, University of Thessaloniki. [In Greek]

Goutsos, Dionysis. 2010. "The Corpus of Greek Texts: A reference corpus for Modern Greek." *Corpora* 5(1):29-44.

Hay, Jennifer. 2000. *Causes and Consequences of Word Structure.* PhD Dissertation, Northwestern University, Evanston, Illinois.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2010. *Introducing Morphology.* Cambridge: Cambridge University Press.

Mela-Athanasopoulou, Elizabeth. 2007. "The polysemy of -*ize* derivatives and the Modern Greek counterpart -*pio*." In Eleni Agathopoulou, Maria Dimitrakopoulou, and Despina Papadopoulou (eds) *Selected papers on Theoretical and Applied Linguistics*, 17th International Symposium, 157-167. Thessaloniki: Monochromia.

Moser, Amalia. 1994. "The interaction of lexical and grammatical aspect in Modern Greek." In Irene Philippaki-Warburton, Katerina Nicolaidis, and Maria Sifianou (eds), *Themes in Greek Linguistics*, 137-144. Amsterdam/Philadelphia: Benjamins.

Plag, Ingo. 1999. *Morphological Productivity. Structural Constraints in English Derivation.* Berlin/New York: Mouton de Gruyter.

Plag, Ingo. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Plag, Ingo. 2006. "Productivity." In Bas Aarts, and April McMahon (eds), *Handbook of English Linguistics*, 537-556. Oxford: Blackwell.

Plag, Ingo, Dalton-Puffer, Christiane, and Harald Baayen. 1999. "Morphological productivity across speech and writing." *English Language and Linguistics* 3(2): 209-228.

Rainer, Franz. 2005. "Constraints on Productivity." In Pavol Stekauer, and Rochelle Lieber (eds), *Handbook of Word-Formation*, 335-352. Dordrecht: Springer.

Ralli, Angeliki. 2005. *Morphology*. Athens: Patakis. [In Greek].

Ralli, Angela. 2011. "Morphology in language contact: Verbal loanblend formation in Asia Minor Greek (Aivaliot)." In Thomas Stolz, Martine Vanhove, Hitomi Otsuka and Aina Urdze (eds), *Morphologies in Contact* (Studia Typologica), 185-201. Berlin: Akademie Verlag.

Scott, Mike. 2008. *WordSmith Tools*, version 5. Liverpool: Lexical Analysis Software.

Triandafyllidis Dictionary. 1998. *Standard Modern Greek Dictionary*. Institute of Hellenic Studies, Aristotle University of Thessaloniki [In Greek].

Tsakou, Eleni. 2010. Derivative verbs of Modern Greek without correspondence between form and meaning. The case of prefixed verbs ending in -*ízo*, -*iázo* and -*óno*. MA thesis, Democritus University of Thrace, Greece [In Greek].