Measuring the Productivity of Noun-Deriving Suffixes across Languages: Greek *-tita* vs. English *-ness*¹

Theofanis Papoutsis

University of Athens
fanispapoutsis@hotmail.com

Abstract

This paper investigates the productivity of the Greek noun-forming suffix *-tita*, in comparison to its English counterpart *-ness*. To this aim, it exploits empirical data drawn from corpora and employs an array of corpus-based productivity measures. The analysis shows that, while both suffixes are considerably productive in relation to the derivational system of the relevant languages, *-ness* is interlingually more productive than *-tita*. Furthermore, both suffixes attach productively to adjectival bases; yet, bases belonging to other word-categories (e.g. nominal, pronominal, etc.) are also productively employed. Moreover, the combinations of adjectival bases with *-tita* and *-ness* exhibit varying degrees of productivity.

1. Introduction

The Greek suffix *-tita*² and the English *-ness* are similar, in that they both attach to adjectival bases and form abstract nouns which denote the same quality, property or state as that of the adjective itself (e.g. *kakós*³ 'bad' > *kakótita*

¹ I wish to thank Dr Katerina Stathi for her fruitful comments and suggestions as well as her encouragement. Thanks are also extended to Dr Dionysis Goutsos for providing me with the raw data from the Corpus of Greek Texts. Any errors are my own responsibility.

² The Greek data appear in broad phonemic transcription (with the lexical stress indicated over the relevant vowel). Since inflectional endings are irrelevant for the present discussion, they are not separated from the preceding derivational suffixes.

³ Adjectives in Greek, contra English, inflect for gender (masculine – feminine – neuter), case (nominative – genitive – accusative) and number (singular – plural). In this paper, the masculine-nominative-singular form is given only. Ditto for adjectival suffixes.

'badness' – *bad* > *badness*). But what exactly renders *-tita* and *-ness* interlingually equivalent?⁴ After all, in both languages, there are other noun-deriving suffixes that are tacked onto adjectives and thus are in rivalry with *-tita* and *-ness*. In Greek, these include *-osíni* and *-ia* (Ralli 2005: 149, Xydopoulos 2008: 184), while in English, they include *-ity* and *-ce/-cy* (Plag 2003: 88, 91–92). However, the use of both *-tita* and *-ness* is subject to the fewest possible structural constraints. More specifically, although *-ness* can be suffixed to virtually any adjective, *-ity* and *-ce/-cy* impose certain selection requirements. Thus, *-ity* is exclusively tacked onto adjectives of Latinate origin, prime examples being adjectives bearing the suffixes *-able/-ible*, *-al*, *-ic* or ending in the phonetic string [ɪd] (e.g. *readability*, *centrality*, *historicity*, *solidity*), while *-ce/-cy* attaches to adjectives in *-ant/-ent* (e.g. *efficient* > *efficiency*) and *-ate* (e.g. *adequate* > *adequacy*) (see Baayen & Lieber 1991, Plag 2003 for details). The Greek nominalising suffixes present a similar picture, in that *-osíni* is exclusively tacked onto non-suffixally derived adjectives which end in either the [-learned] inflection *-os* (e.g. *anixtómjalos* 'open-minded' > *anixtomjalosíni* 'open-mindedness') or the [+learned] *-on* (e.g. *isxiroyónomon* 'stubbornness'), *-ia* is added only to [+learned] adjectives in *-ís* (e.g. *evyenís* 'polite' > *evyénia* 'politeness'), while *-tita* is suffixed to practically any adjectival base, excluding adjectives in *-on* and *-ís* though (cf. Papoutsis 2012). Moreover, both *-tita* and *-ness* are intuitively felt to be significantly more productive than their rivals. Simply put, if a Greek or English speaker wishes to derive an abstract noun from an adjective, then s/he will resort to *-tita* and *-ness* as a first choice, respectively.

This intuitive feeling about the high productivity of *-ness* has been verified and quantified by a multitude of empirical evidence, drawn from corpora (see, among others, Baayen & Lieber 1991: 824–826, Baayen & Renouf 1996, Plag 2003, Plag et al. 1999), whereas regarding the productivity of *-tita* (as well as of any other Greek morphological process), there is practically no empirical evidence to this day.

Thus, the aim of this corpus-based study was threefold: (a) to verify and quantify our intuitions about the high productivity of both *-tita* and *-ness*; (b) to compare and contrast the productivity degrees of an interlingual pair of noun-forming suffixes; and (c) to apply the quantitative approach to morphological productivity to Greek data for the first time (and gauge any practical problems arising from this application).

⁴ The correspondence between *-tita* and *-ness* is discussed here from the perspective of morphology, not from that of translation, meaning that a given *-tita* word will not necessarily translate as a *-ness* word and vice versa. This is illustrated by *egirótita*, the English translation equivalent being an *-ity* formation, i.e. *validity*.

2. Data Collection, Processing and Categorisation

The Corpus of Greek Texts (henceforth CGT) served as a source of data on *-tita*. The CGT comprises approximately 30 million word tokens of present-day Greek language production (1990-onwards), including both written and spoken discourse (ratio 9/1). Being the most balanced and representative corpus, currently available for Greek (cf. Goutsos 2010), the CGT can fruitfully be exploited for morphological research purposes.⁵

Obviously, comparing affixes in different corpora raises issues of comparability regarding both the corpus size and the contained text-types (Baayen & Lieber 1991: 820). Unfortunately, no reliable parallel corpora for the Greek-English pair are freely available. Therefore, I opted for the freely available 425-million-word Corpus of Contemporary American English (henceforth COCA), which is a greatly balanced and representative collection of contemporary English (including both oral and written discourse) (Davies 2011) and thus is suitable for morphological research purposes.⁶ More specifically, in order to have an English corpus of comparable size to that of the CGT, the data on *-ness* were gathered from an approximately 30-million-word reduced version of the COCA (time period: January 2009-August 2010) (cf. Davies 2010).

Initially, raw data were extracted – by means of string searches – from the CGT word-frequency lists and the online version of the COCA. These raw data were then analysed manually, in order to exclude any irrelevant lexical items. Wherever necessary, I checked words in their context in the corpora used and consulted current dictionaries (namely, Babinotis (2002) and Institute for Modern Greek Studies (1998) for Greek; Stevenson (2010) and Merriam-Webster (2010) for English). Subsequently, the data were lemmatised, that is, all inflectional suffixes marking number and case were collapsed with their base forms (i.e. nominative-singular for Greek, singular for English).⁷ The end-product was a word-frequency list for *-tita* and one for *-ness*, respectively.

Finally, a lexical item ought to meet two basic requirements, so as to be considered as a token with *-tita* or *-ness*. Firstly, it needed to belong to the specific morphological process both morphologically and semantically. Hence, words such as *business* and *wilderness* were excluded from further scrutiny, since they are clearly not nominalisations. Secondly, the base should be either an independent word in Modern Greek or Modern English or a bound

⁵ The CGT is freely available at: <http://sek.edu.gr/index.php>.

⁶ The COCA is fully searchable at: <http://corpus.byu.edu/coca/>.

⁷ Greek nouns inflect for number (singular – plural) and case (nominative – genitive – accusative).

stem contained in at least one other formation. For example, the item *posótita* 'quantity' was included in the Greek data, as its stem occurs in other formations such as *posó(n)* 'amount', *pósos* 'how (much)', etc.

3. Productivity Measures

The corpus-based productivity measures used in this study originate in the work of Baayen and his colleagues (see, among others, Baayen 1992, 1993, 2008, Baayen & Lieber 1991, Baayen & Renouf 1996, Plag et al. 1999). In particular, three basic statistical measures were used. (i) The total number of tokens with a given affix (N^{aff}), i.e. how many times an affix occurs in a corpus. (ii) The number of types with an affix (V^{aff}), which shows the extent of use of a morphological category. (iii) The number of *hapax legomena* (or hapaxes for short) with that same affix (n_1^{aff}), i.e. the words that occur only once in the corpus. This number indicates how often a morphological process is employed to form new words, that is, words not previously evidenced in a corpus. Hapaxes are intimately tied to the corpus in which they appear and should not be thought of as necessarily true neologisms. However, there is compelling empirical evidence (Baayen & Lieber 1991, Baayen & Renouf 1996, Plag 1999, 2003, 2005, 2006) that a considerable number of neologisms do appear among the *hapax legomena*.

But what is the rationale behind regarding hapaxes as a reliable indication of the productivity degree of an affix? Crucially, the importance of hapaxes is that if language users wish to express a new concept, then they normally resort to productive word-formation rules (WFRs), for it is only a productive process that can give rise to new lexemes (Lüdeling & Evert 2005: 361).

Furthermore, it has been empirically found that a productive process is characterised by a large number of low-frequency lexemes (and many *hapax* and *dis legomena*) as well as a smaller amount of high-frequency words. Conversely, an unproductive process is characterised by very few low-frequency words (and even fewer *hapax* and *dis legomena*) and simultaneously a multitude of high-frequency items (Baayen & Lieber 1991, Lüdeling & Evert 2005).

Now, the principal derived measure of productivity is the so-called 'productivity in the narrow sense', namely the quotient of the number of hapaxes n_1 and the number of all tokens N :

$$(1) \quad P = n_1^{aff} / N^{aff}$$

Crucially, P is a probabilistic tool, which shows us the possibility of coming across new types with a certain affix, when we have sampled N tokens. A high value of P suggests a very productive affix, while a low value of P indicates the opposite (Baayen & Lieber 1991: 809-811, Plag 2003: 57). A great number of

hapaxes or/and the absence of many highly frequent words result in an increase of P . Conversely, the value of P is low, when there is a small number of hapaxes or/and a large number of high-frequency lexemes (Baayen & Lieber 1991: 815, Plag 2003: 57). These interrelations between P , n_1 and N perfectly match our intuitions about productivity, given that less productive WFRs are characterised by high-frequency words (Baayen & Renouf 1996, Plag 1999: 93ff.).

However, there are cases where P appears to give affix rankings which fail to match our intuitions. To cater for such problematic rankings, another derived measure has been proposed; the so-called 'global productivity' (P^*). The value of P^* is *prima facie* the result of a complex mathematical equation (see Baayen 1993: 192-193), but Plag (1999: 32) demonstrates that P^* actually equals the already familiar number of hapaxes:

$$(2) \quad P^* = n_1^{\text{aff}}$$

Note that P and P^* stand in a complementary relation; P is rather better applicable to discriminating productive from unproductive processes, whilst P^* is perhaps more appropriate for ranking productive processes (Baayen 1993: 194).

4. Results

4.1. Overall Productivity

First, let us explore the overall productivity of *-tita* and *-ness*, that is without taking into consideration the word-category of the base.

Table 1.

Productivity measures of *-tita* and *-ness*

Suffix	N	V	n_1	n_2	P
<i>-tita</i>	122,788	1,208	343	151	0.0027934
<i>-ness</i>	18,935	1,345	585	199	0.0308951

As shown in Table 1, *-ness* surpasses *-tita* in almost every measure. In particular, the English suffix has a greater extent of use V , more *hapax* and *dis legomena* (n_1 , n_2), and a higher value of P . The only index in which the Greek suffix exceeds the English one is the total number of tokens N . These figures lead us to conclude that *-tita* is less productive than *-ness*. However, this certainly does not mean that

-tita is not a productive suffix for the Greek language. In other words, just like *-ness* is very productive with regard to the English derivational system, so is *-tita* in relation to the Greek one.

This claim can be substantiated by examining the word-frequency distribution of both suffixes, which is presented in Figure 1. Recall that a great number of low-frequency types are typically indicative of productive WFRs. Indeed, the frequency distribution of both suffixes concerned is highly skewed to the left. Thus, we observe that the largest portion of the types in *-tita* and *-ness* have frequency 1-20 (i.e. 71.52% and 89.66%, respectively).

Figure 1. Summary of the frequency distribution of *-tita* and *-ness*

Moreover, focusing on the head of the frequency distribution, especially the first 15 ranks (see Figures 2&3), we see that the 585 hapaxes in *-ness* account for 43.4% of all different types *V*, the 199 types that occur twice represent 14.8% of *V*, etc. Therefore, roughly one in two types in *-ness* occurs only once in a 30-million-word corpus; a ratio which strongly indicates high productivity. Similarly, the 343 hapaxes in *-tita* account for 28% of all different types *V*, the 151 types that occur twice represent 12.5% of *V*, etc. Simply put, if one in three types in *-tita* occurs only once in a 30-million-word corpus, then this WFR has to be considered productive, too.

Figure 2. The head of the frequency distribution of *-tita*

Figure 3. The head of the frequency distribution of *-ness*

4.2. Productivity Ranking in Relation to the Word-Category of the Base

Now, let us examine the productivity of *-tita* and *-ness*, considering the word-category of the base.

Table 2.

Productivity measures of *-tita* in relation to the word-category of the base

Word-category of the base	<i>N</i>	<i>V</i>	n_1	n_2	<i>P</i>
Adjectival base	110,033	1,181	333	151	0.0030263
Nominal base	2,622	17	6	0	0.0022883
Pronominal base	10,134	10	4	0	0.0003947

As can be seen in Table 2, when combined to an adjectival base, *-tita* ($[[X_{\text{ADJ}}]-tita]_N$) exhibits by far the highest productivity indexes, that is $P=0.0030263$, 333 hapaxes, and 151 types occurring twice. The medial position in productivity ranking is occupied by the combination of nominal base and *-tita* ($[[X_N]-tita]_N$), since $P=0.0022883$ and $n_1=6$ (e.g. *laótita* 'people-ness' (< *laós* 'people'), *usiótita* 'essence-ness' (< *usía* 'essence')). Note that the eligibility of both adjectives and nouns as bases for *-tita* has already been mentioned in the relevant literature (see, among others, Ralli 2005, Xydopoulos 2008). Interestingly, according to the data, pronouns can also function as possible bases. Although $[[X_{\text{PRO}}]-tita]_N$ ranks third and last, with $P=0.0003947$ and only four hapaxes (e.g. *eaftótita* 'self-ness' (< *eaftós* 'self'), *eðikótita* 'mine-ness' (< *eðikós* (*mu*) 'mine')), yet it can be deemed as characteristic of the derivational power that *-tita* exhibits in present-day Greek.

As regards *-ness*, it has long been observed that it can be attached to virtually any adjective, while it is often tacked onto nouns, pronouns and phrases. Because of its great combinability, *-ness* is currently considered as the most productive suffix of the English language (Plag 2003: 92). Note that, on the basis of the COCA data, adverbials can also constitute eligible bases.

Table 3.
Productivity measures of *-ness* in relation to the word-category of the base

Word-category of the base	<i>N</i>	<i>V</i>	<i>n</i> ₁	<i>n</i> ₂	<i>P</i>
Phrasal base	21	12	10	1	0.4761904
Adjectival base	18,749	1,288	544	195	0.4226884
Nominal base	67	31	23	2	0.3432835
Adverbial base	29	7	5	0	0.1724137
Pronominal base	69	7	3	1	0.0434782

In Table 3, we notice that the combination of phrase and *-ness* scores extremely high in terms of *P*, which is against our intuition that it is more likely to form a new word in *-ness* using an adjective rather than a phrase. This incongruity between intuition and *P* perhaps lies in the exceptionally small number of types *N*, which is known for positively influencing the value of *P*. However, if solely the number of hapaxes is taken into account, then the emerging productivity ranking of the attested combinations is in better line with our intuitive feeling.

Thus, the combination $[[X_{\text{ADJ}}]\text{-ness}]_N$ is characterized by the highest productivity degrees, having a plethora of *hapax* (e.g. *deadness*, *pinkness*) and *dis legomena* (e.g. *keenness*, *waterproofness*). The combination $[[X_N]\text{-ness}]_N$ ranks second, having a significant number of words that occur only once (e.g. *boyness*, *brunneteness*). Interestingly, besides common nouns, proper names can also function as possible bases, be they person names (e.g. *Baldwin-ness*, *Gaga-ness*) or place names (e.g. *Hong-Kong-ness*, *Ticuaniness*). The third position is occupied by the combination $[[X_{\text{PHRASE}}]\text{-ness}]_{N'}$, having ten *hapax legomena* (e.g. *no-matter-what-ness*, *down-to-earth-ness*). The last two positions are taken by the combinations $[[X_{\text{ADV}}]\text{-ness}]_N$ and $[[X_{\text{PRO}}]\text{-ness}]_{N'}$, with five (e.g. *apart-ness*, *now-ness*) and three hapaxes (e.g. *me-ness*, *mine-ness*), respectively.

4.3. Productivity Ranking in Relation to Adjectival Bases

This section focuses on the productivity of the attested combinations of adjectival bases and *-tita* or *-ness*, which constitute by far the greatest portion of the data. Table 4 presents the attested combinations of adjectival suffixes and *-tita* in decreasing order of *P*, while Figure 4 summarizes their frequency distribution.

Table 4.Productivity measures of *-tita* in relation to adjectival bases

Adjectival base	<i>N</i>	<i>V</i>	n_1	n_2	<i>P</i>
<i>-alós/-ilós/-olós</i>	106	8	4	0	0.0377358
<i>-simos/-imos</i>	4,110	95	34	7	0.0082725
<i>-ikós/-akós</i>	27,964	657	200	97	0.007152
<i>-arós/-erós/-irós/-rós</i>	1,837	30	11	4	0.005988
<i>-inós</i>	967	9	2	0	0.0020682
<i>-tós</i>	11,374	75	17	13	0.0014946
<i>-os/-ís</i>	43,108	256	57	24	0.0013222
<i>-íos</i>	5,376	24	5	1	0.00093
<i>-éos</i>	4,315	14	2	3	0.0004634
<i>-tírios</i>	9,189	3	1	0	0.0001088
<i>-aléos</i>	21	2	0	0	0.0000
<i>-plós</i>	344	4	0	2	0.0000
<i>-teros</i>	1,316	3	0	0	0.0000
<i>-opós</i>	6	1	0	0	0.0000

Figure 4. Summary of the frequency distribution of *-tita* in relation to adjectival bases

First of all, notice that *-tita* in combination with adjectives in *-alós/-ilós/-olós* has the highest value for *P* in Table 4, which is in sharp conflict with our intuition, since the suffix *-alós/-ilós/-olós* is no longer used to form new words. Therefore, it would be better to take only n_1 and *V* into account; hence, having only four hapaxes (e.g. *xamilótita* 'lowness') and very few types (e.g. *dropalótita* 'shyness'), *-alótita/-ilótita/-olótita* ranks fairly low. Secondly, *-arós/-erós/-irós/-rós* and *-inós* in combination with *-tita* score surprisingly high in terms of *P*. This does not correspond to our intuitive feeling, considering that the particular adjective-forming suffixes are themselves rather moderately productive. So, if we use n_1 as a ranking criterion, these combinations score very low in productivity, a fact which is in better line with our intuition. Their extremely low productivity degree is also evidenced in the occurrence of a considerable number of high-frequency words (e.g. *nosirótita* 'morbidness' (63), *triferótita* 'tenderness' (204), *statherótita* 'stability' (1,419), *kaθimerinótita* 'dailyness' (837)).⁸

Therefore, the combinations of adjectives in *-simos/-imos* and *-ikós/-akós* with *-tita* occupy the top position in productivity ranking. *-simótita/-imótita* is more productive in the narrow sense (since $P=0.0082725$), but *-ikótita/-akótita* has more *hapax* (e.g. *aliθiakótita* 'truthfulness') and *dis legomena* (e.g. *diavrotikótita* 'corrosiveness') than the former (e.g. *xrisimopiisimótita* 'usability' (1), *ekθesimótita* 'showability' (2)). The high productivity degree of these co-occurrences is also evidenced in the presence of a large number of low-frequency items (e.g. *ereθisimótita* 'irritability' (3), *agrotikótita* 'agriculturalness' (4)). Note that this ranking seems to be related to our intuitive feeling that *-simos/-imos* and *-ikós/-akós* are themselves highly productive adjective-deriving suffixes.

The medial positions in our productivity ranking are occupied by the combination of *-tita* with deverbal adjectives in *-tós* as well as the combination of *-tita* with non-suffixally derived adjectives which bear the inflectional endings *-os* and *-ís*.⁹ Among the items occurring once and twice, we find: *anaklitótita* 'reversibility' (1), *prosforótita* 'suitability' (1), *amvlítita* 'bluntness' (1), *γρaptótita* 'writtenness' (2), *plusiótita* 'richness' (2), *platítita* 'wideness' (2). Note also the occurrence of a sizeable number of low-frequency formations (e.g. *ektelestótita* 'enforceability' (3), *proorótita* 'prematurity' (4)), which points to a non-negligible degree of productivity.

Turning to less productive combinations, *-iótita* scores extremely low in all productivity measures and shows many words at the high frequencies (e.g.

⁸ The number in parentheses following a word indicates the token frequency of the particular word in the corpora used.

⁹ The inflectional suffix *-ís* (spelt <ῖς> in Greek) mentioned here is the ending that appears in [+learned] adjectives such as *amvlís* 'blunt'. This ending should not be confused with the inflectional suffix *-ís* (spelt <ῖς> in Greek) which was mentioned in section 1 and which appears in examples such as *evγenis* 'polite'. In other words, adjectives with these endings belong to separate paradigms (see Ralli 2005 for details).

eoniótita 'eternity' (207)). Further, the co-occurrence of adjectives in *-éos* and *-tírios* with *-tita* is of negligible productivity. After all, only fourteen types in *-eótita* (e.g. *anangeótita* 'necessity' (927)) and three in *-tiriótita* (e.g. *iperðrastiriótita* 'hyperactivity' (14)) are attested in the entire CGT.

Finally, *-tita* is completely unproductive in combination with adjectives in *-aléos*, *-plós*, *-teros* and *-opós*. Moreover, the vast majority of the few types that are attested in the CGT are of high frequency (e.g. *polaplótita* 'multiplicity' (154), *iðieterótita* 'peculiarity' (1,062)). Note that, on an intuitive basis, these adjective-forming suffixes seem to be rarely used to coin any new words in Greek.

As for the English suffix, Table 5 lists all the attested combinations of *-ness* and adjectival base in order of decreasing *P*, whilst Figure 5 summarises their frequency distribution.

Table 5.

Productivity measures of *-ness* in relation to adjectival bases

Adjectival base	<i>N</i>	<i>V</i>	n_1	n_2	<i>P</i>
<i>-ern</i>	1	1	1	0	1.00
<i>-ose</i>	1	1	1	0	1.00
<i>-(i)an/-ean</i>	18	10	6	3	0.3333333
<i>-able/-ible</i>	33	14	11	0	0.3333333
<i>-ate</i>	13	6	3	1	0.2307692
<i>-al</i>	25	9	5	1	0.2
<i>-some</i>	18	8	3	3	0.1666666
<i>-ary</i>	25	8	4	1	0.16
<i>-ed</i>	507	133	76	19	0.1499013
<i>-ic/-ical</i>	22	6	2	1	0.1363636
<i>-ish</i>	183	32	16	5	0.0874316
<i>-y</i>	1,393	229	113	46	0.0811198
<i>-ese</i>	13	2	1	0	0.076923
<i>-ous</i>	1,057	108	60	20	0.0567644
<i>-ly</i>	397	41	22	5	0.0554156
<i>-ful</i>	495	61	21	8	0.0424242
<i>-less</i>	675	70	21	12	0.0311111
<i>-ive</i>	1422	99	40	12	0.0281293
<i>no suffix</i>	11,924	433	132	56	0.0110701
<i>-ing</i>	475	9	5	0	0.0105263
<i>-ant</i>	18	2	0	0	0.0000000
<i>-esque</i>	5	2	0	1	0.0000000
<i>-ward</i>	29	4	0	1	0.0000000

Figure 5. Summary of the frequency distribution of *-ness* in relation to adjectival bases

To begin with, three observations should be made. Firstly, if for an affix $P=1$, then that affix is practically dead. Thus, the combinations *-ernness* and *-oseness* are totally unproductive, since each has only one type with frequency one (*southernness*, *comatoseness*). Secondly, *-ateness* and *-eseness* score unexpectedly high in terms of P (surpassing even unambiguously more productive combinations such as *-liness*, *-fulness*, *-lessness*), considering that the Latinate suffixes *-ate* and *-ese* are not – on an intuitive basis – currently employed to form new adjectives. However, using n_1 as a ranking criterion, both *-ateness* and *-eseness* occupy the lowest positions, having three and one hapaxes (e.g. *passionateness*, *Japaneseness*), respectively. Thus, this ranking is in better line with our intuition. Thirdly, *-al*, *-some*, *-ary*, *-ic/-ical* in combination with *-ness* appear to be highly productive in terms of P , even exceeding combinations such as *-ishness*, *-iness*. However, if we assume n_1 as our sole ranking criterion, *-alness*, *-ariness*, *-someness* and *-ic/-icalness* are ranked fairly low, having five, four, three and two hapaxes (e.g. *spectralness*, *fragmentariness*, *bothersomeness*, *periodicalness*), respectively.

Thus, *-(i)anness/-eanness* and *-ableness/-ibleness* are by far the most productive combinations in terms of P , with the latter having more hapaxes (e.g. *disagreeableness*) than the former (e.g. *Italianness*, *Victorianness*). Crucially, both combinations do not show any high-frequency items. Next, the combination *-edness* is highly productive, since $P=0.1499013$, $n_1=76$ (e.g. *designedness*, *unpeopledness*) and $n_2=19$ (e.g. *open-handedness*, *pointedness*), and since it shows a multitude of types at the low frequencies (e.g. *small-mindedness* (4), *woundedness* (4)). *-ishness* and *-iness* are also significantly productive combinations. Although *-ishness* is more productive in the narrow sense, *-iness* is globally more productive, as it has more types appearing once (e.g. *hellishness*,

pepperiness) or twice (e.g. *childishness*, *legginess*) and shows more types at the low frequencies (e.g. *devilishness* (3), *achiness* (3)). This ranking possibly reflects the intuition that the native suffixes *-ed*, *-ish*, *-y* are often used to coin new words in English.

The combinations *-ousness*, *-liness*, *-fulness*, *-lessness* occupy the middle positions of the ranking. Interestingly, *-ness* is rather more productively suffixed to words in the Latinate *-ous* (e.g. *miraculousness* (1), *perilousness* (1)) rather than words in the native *-ly* (e.g. *dailiness* (1), *ungodliness* (1)). Moreover, *-ness* combines with the antonymic suffixes *-ful* and *-less* equally productively, as indicated by the identical number of hapaxes (e.g. *colorfulness*, *skillfulness*, *emotionlessness*, *worklessness*) and the similar frequency distribution (e.g. *cheerfulness* (9), *weightlessness* (9)).

The combination of *-ness* with adjectives in *-ive* (e.g. *associativeness* (1), *restiveness* (2)) and *-ing* (e.g. *amazingness* (1), *vivifyingness* (1)), as well as with non-suffixally derived adjectives (e.g. *goldness* (1), *spareness* (2)) are marginally productive. Interestingly, *-iveness* shows a sizeable number of types at the low-frequencies (e.g. *obsessiveness* (4)), whereas no-suffix-*ness* has many high-frequency types (e.g. *darkness* (872)).

Finally, the co-occurrence of *-ness* with the adjectival suffixes *-ant*, *-esque* and *-ward* is totally unproductive, in terms of both *P* and *n₁* (e.g. *unpleasantness* (12), *picturesqueness* (3), *backwardness* (17)). This appears to be related to our intuitive feeling that *-ant*, *-esque* and *-ward* do not give rise to any new formations in current English.

5. Conclusions

This final section discusses the main conclusions that can be drawn, concerning the suffixes examined and the application of the quantitative approach to productivity to Greek data.

We saw that *-tita* is a highly productive suffix for the Greek derivational system. Similarly, *-ness* is an extremely productive suffix, as regards the English derivational system. However, on an interlingual comparison, the latter surpasses the former in productivity degree.

How can this discrepancy be possibly interpreted? A mathematical explanation is that the appearance of a great number of high-frequency words in *-tita* leads to a low value for *P*. Conversely, the huge number of hapaxes in *-ness* results in a high value for *P*. A linguistic, albeit rather technical, explanation might be sought at the different textual make-up of the corpora used. The CGT and the COCA obviously differ not only in the contained texts, but also in the proportions of the contained texts them. Consider as an illustration the case of academic texts; the CGT contains both PhD theses and journal articles (14.67% of total corpus

size), whereas the COCA contains only PhD theses (20% of total corpus size) (cf. Goutsos 2010, Davies 2011). Differences such as these might possibly affect the results obtained by the application of the productivity measures.¹⁰ A related linguistic hypothesis is that the use of nominalizations might be more preferable in Greek than in English.¹¹ This preference may lead to the appearance of very frequent words, which in turn accounts for the low value for *P*. In any case, all of the above explanations are open to further discussion.

Another conclusion is that both *-tita* and *-ness* do not attach to adjectival bases only, but also to other categories of bases. Perhaps as an outcome of their high degree of productivity, the Greek suffix can combine with nouns and pronouns, while its English counterpart is affixed to nouns, pronouns, adverbs and phrases.

A further conclusion is that *-tita* and *-ness* are not suffixed equally productively to each and every adjective. In other words, some combinations of adjectival bases and the two noun-forming suffixes under study are extremely productive, some others are less productive and still others are completely unproductive. It might be argued then that the productivity degree of the combinations of *-tita* and *-ness* with any adjective is related to the productivity degree of the adjective-forming suffix itself. Recall, for instance, that *-simótita/-imótita* and *-ikótita/-akótita* are characterized by a high degree of productivity, whilst, according to our intuition, *-simos/-imos* and *-ikós/-akós* themselves are very productive as well. Similarly, the complete unproductiveness of *-antness*, *-esqueness*, *-wardness* appears to reflect the fact that the suffixes *-ant*, *-esque*, *-ward* are rarely used for adjective-formation in Present-Day English. In any case, this intuitive observation needs to be further scrutinized, in order to be substantiated by more empirical evidence.

As regards the reliability of the productivity measures used, it can be maintained that the index of 'productivity in the narrow sense' (*P*) can – to a lesser or greater extent – capture our intuitions on the productivity of *-tita* and *-ness*. Furthermore, in the few cases where *P* is in disagreement with our intuition, 'global productivity' ($P^* = n_i$) proves to be a more reliable measure. Additionally, it should be highlighted that the results obtained from applying the productivity measures to language data are indissolubly related to the corpora used (cf. Lüdeling & Evert 2005: 356, 361). In other words, if data are drawn from certain other corpora, there might be slightly variant productivity rankings. However, neither *-tita* nor *-ness* are anticipated to prove to be unproductive suffixes.

10 These problems might be eliminated if there were freely available parallel corpora for the Greek-English pair.

11 Thanks to Dr Dionysis Goutsos (p.c.) for bringing this possibility to my attention.

A practical issue needs to be raised. The CGT is currently the most balanced and representative corpus available for Greek; yet, it is neither lemmatised nor part-of-speech-tagged. This fact, along with the unambiguously rich inflectional system of Greek, poses certain difficulties, when the quantitative approach is to be applied to the category of verbs, namely the category which has by far the richest inflection in Greek. More specifically, Greek verb forms inflect for person (1st – 2nd – 3rd), number (singular – plural), tense ($\pm$ past), aspect ($\pm$ perfective), mood ($\pm$ imperative) and voice (active – passive). Therefore, if one wished to explore the productivity of roughly synonymous verb-forming suffixes such as *-áro* (e.g. *sutáro* ‘to shoot’) and *-ízo* (e.g. *kapnízo* ‘to smoke’),¹² s/he would have to collect all the relevant tokens by hand, taking into consideration the morphological manifestation of the aforementioned grammatical categories. Clearly, such an endeavour would be extremely time-consuming (if not daunting!). Thus, the availability of lemmatized and morphologically annotated corpora is deemed vital to the study of the productivity of WFRs in languages with so-called rich inflectional morphology.

Despite these practical difficulties, though, one of the most important advantages of the quantitative approach to morphological productivity is the fact that if another researcher uses the very same set of data and follows the very same methodology, then s/he is bound to come up with the same results.

12 As it is often the case with the productivity of synonymous suffixes, selection requirements come into play; thus, *-áro* is typically tacked onto loan bases (cf. Mela-Athanasopoulou 2002: 443), *-ízo* attaches to native ones.

References

- Baayen, H.R. 1992. Quantitative aspects of morphological productivity. In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 109-149.
- Baayen, H.R. 1993. On frequency, transparency and productivity. In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1992*. Dordrecht: Kluwer, 181-208.
- Baayen, H.R. 2008. Corpus linguistics in morphology: morphological productivity. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics: An International Handbook*, vol. 2. Berlin: Mouton De Gruyter, 899-919.
- Baayen, H.R. and R. Lieber. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29: 801-844.
- Baayen, H.R. and A. Renouf. 1996. Chronicling the *Times*: productive lexical innovations in an English newspaper. *Language* 72(1): 69-96.
- Babiniotis, G. 2002. *Λεξικό της νέας ελληνικής γλώσσας*. Athens: Centre for Lexicology.
- Davies, M. 2010. corpus.byu.edu: five announcements/new resources. Retrieved 8 September 2010 from corpora@listserv.byu.edu
- Davies, M. 2011. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25: 447-65.
- Goutsos, D. 2010. The Corpus of Greek Texts: a reference corpus for Modern Greek. *Corpora* 5(1): 29-44.
- Institute for Modern Greek Studies. 1998. *Λεξικό της κοινής νεοελληνικής*. Thessaloniki: Institute for Modern Greek Studies.

Lüdeling, A. and S. Evert. 2005. The emergence of non-medical *-itis*: corpus evidence and qualitative analysis. In S. Kepser and M. Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. Berlin: Mouton de Gruyter, 315-333.

Mela-Athanasopoulou, E. 2002. English loan words in Modern Greek. *Studies in Greek Linguistics* 22: 440-450.

Merriam-Webster. 2010. *Merriam-Webster Online Dictionary*. <http://www.merriam-webster.com>

Papoutsis, T. 2012. On the combinability and productivity of the Greek noun-forming suffixes *-tita* and *-sini*: evidence from corpora. In *Proceedings of the 6th Athens Postgraduate Conference of the Faculty of Philology, University of Athens*, vol. A. Athens: University of Athens, 267-277.

Plag, I. 1999. *Morphological Productivity: Structural Constraints on English Derivation*. Berlin & New York: Mouton de Gruyter.

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

Plag, I. 2005. Productivity. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, vol. 10. Boston: Elsevier, 121-128.

Plag, I. 2006. Productivity. In B. Arts and A. McMahon (eds.), *The Handbook of English Linguistics*. Oxford: Blackwell, 537-556.

Plag, I., C. Dalton-Puffer and H.R. Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2): 209-228.

Ralli, A. 2005. *Μορφολογία*. Athens: Patakis. .

Stevenson, A. (ed.). 2010. *Oxford Dictionary of English*. Oxford: Oxford University Press.

Xydopoulos, G.I. 2008. *Λεξικολογία: Εισαγωγή στην ανάλυση της λέξης και του λεξικού*. Athens: Patakis.