Minimal Pairs and Functional Loads of Sound Contrasts Obtained from a List of Modern Greek Words

Constandinos Kalimeris, Stelios Bakamidis

Institute for Language and Speech Processing (ILSP),
Epidavrou & Artemidos 6, Marousi, 151 25, Athens, Greece
{c_kal, bakam}@ilsp.gr

Abstract

This paper reports on the initial results of our investigation into the distribution of speech sounds across the lexicon of Modern Greek (MG). The data we discuss ultimately derive from the list of orthographic word-types of a large general corpus of written MG. The orthographic word-types were automatically transcribed into their respective citation forms. Minimal pairs were automatically extracted from the resultant list of citation forms. The Functional Load (FL) of each sound opposition was computed as a function of (a) the length of citation forms, (b) the position of each sound contrast within citation forms and (c) the number of minimal pairs pertinent to each opposition in question. The body of data yielded by this study will be used for further research in MG phonology as well as for the improvement of the performance of Automatic Speech Recognition applications.

Index Terms: minimal pair, functional load, contrastive distribution, phoneme inventory, Modern Greek

10.21437/Interspeech.2007-352

1. Introduction

The research work reported herein was originally undertaken in the hope of facilitating phonological investigations of Modern Greek (MG) [1]. Theoretical phonological analyses of MG abound but not uncommonly reflect views based on sparse data ([2] – [7]). To the best of our knowledge, a comprehensive body of data capable of supporting robust (i.e. data-driven) phonological analyses of MG has heretofore been non-existent. This paper reports on the building of such a body of data and attempts some tentative, yet data-driven, initial statements relevant to the phonology of MG.

The paper is organized as follows: Section 2 outlines a number of basic concepts. Section 3.1 discusses the study's source (1st order) data. Section 3.2 discusses the study's canonical allophonic (2nd order) data. Section 3.3 discusses and presents the combinatory (3rd order) data obtained from the processing of the 2nd order data. Section 3.4 discusses and presents the Functional Loads (FLs) of pairs of sound segments in contrastive distribution across our wordlist and the FLs of candidate phonemes (4th order data). Section 4 lists and briefly discusses our initial phonological observations. Section 5 discusses our limitations and prospects for future research. Finally, Section 6 concludes the paper.

2. Basic concepts

Every written word can be said to have a typical phonetic realisation, a sort of mean which underlies all its instantiations in speech irrespective of inter- and intra-speaker variation within a given linguistic code. This representative phonetic mean may be called the word's *citation form* or *canonical form* (cf. [8], [9]). It is commonly represented in writing as a string of phonetic symbols (representing discrete speech sounds, or *sound segments*, or *phones*) belonging to a special alphabet like the IPA [10]. The citation forms we use consist of the following set of 36 symbols: A = {p, t, c, k, b, d, ɟ, g, f, θ, ç, x, v, ð, ʝ, ɣ, s, z, ɱ, m, n, ɲ, ŋ, ɾ, l, ʎ, e, i, a, o, u, é, í, á, ó, ú}. Set A is assumed to be sufficient for the representation of all citation forms relevant to the standard variety of MG.

The less predictable the occurrence of a symbol within the citation forms of a language, the more likely it is that the symbol is a *phoneme* of that language. The more predictable the occurrence of a sound symbol relevant to its neighbouring symbols is, the more likely it is that the symbol in question represents a conditioned phonetic variant (or *allophone*) of a phoneme of that language. Ideally, phonemes are in *contrastive distribution* within canonical forms and allophones in *complementary distribution* (the terms are used in their linguistic sense [8]). It is obvious that the notions "phoneme" and "allophone" are not absolute but relevant to phone distribution within a particular body of data.

We take all symbols in set A (except the last 5) to be legitimate *candidate* phonemes of MG. The last 5 accented symbols are to be taken as shorthand notations. For example, [á] in a citation form denotes a phone with the segmental quality of an [a] which, additionally, happens to be the most prominent nuclear element of the syllable which bears the lexical stress of that citation form: [patáta]=[pa.ˈta.ta] ('potato'), [gáida]=[ˈgai.da] ('bagpipe'), [trizðiástato]=[triz.ˈðia.sta.to] ('3D').

In this work, a pair of citation forms (i.e., a pair of strings of phonetic symbols, say $w_1 = [a_1, a_2, a_3, \ldots, a_l]$ and $w_2 = [b_1, b_2, b_3, \ldots, b_l]$), is defined as a *minimal pair* iff:

$$l_{w_1} = l_{w_2} = l \quad (1)$$

where l_{w_1} and l_{w_2} are the respective lengths of w_1 and w_2, and

$$\begin{array}{ll} a_i = b_i & \forall\, i \in [1, \ldots, k-1] \cup [k+1, \ldots, l], \\ a_i \neq b_i & i = k \end{array} \quad (2)$$

where a_i is the i-th character of string w_1, and b_i is the i-th character of string w_2. Note that, in this work, we consider all pairs of citation forms satisfying conditions (1) and (2) as legitimate minimal pairs, irrespective of the membership of each a_i and b_i in traditional linguistic categories such as "vowel" and "consonant". This decision reflects our intention to produce FLs which will be as free from theoretical bias (i.e. as much data-driven) as possible. Hence, not only are [á**ɾ**ma]~[á**l**ma] ('tank'~'leap'), [p**i**náo]~[p**o**náo] ('starve'~'hurt') and [st**í**los]~[st**ó**los] ('pole'~'fleet') considered legitimate minimal pairs, but also [ó**a**si]~[ó**p**si] ('oasis'~

‘aspect’) and [p**ú**]~[p**u**] (‘where?’~‘that’). (The tilde symbol ‘~’ denotes sounds in contrastive distribution.)

The nature of our data (a wordlist, as opposed to a corpus) further forces us to adopt an even broader (linguistically-functionally speaking) view of the minimal pair: the members of a legitimate minimal pair cannot only correspond to different lexemes (as in the above examples), but also to different inflected forms of the same lexeme, or to forms of different lexemes which are closely related through morphological derivation. Hence, [tró**m**e]~[tró**t**e] (‘$\text{eat}_{\text{pres-1st-plural}}$’~‘$\text{eat}_{\text{pres-2nd-plural}}$’) and [ʝenik**ó**s]~[ʝenik**ú**s] (‘generally’~ ‘$\text{general}_{\text{acc-plural}}$’) are also considered legitimate minimal pairs.

3. Data

3.1. Orthographic (1st order) data

These derive from the list of word-types (unique orthographic words forms) of HNC [11]. Only strings consisting exclusively of Greek orthographic characters and bearing no more than one stress marks were retained. To achieve maximal source-data validity [1], the remaining orthographic forms were further filtered with the help of a morphological lexicon and a spelling-checker software [12], [13]. A stress diacritic was manually added to all non-clitic monosyllable forms corresponding to content words. The above procedures yielded a final list of about 210,000 orthographic word-types.

3.2. Allophonic (2nd order) data

The orthographic data were automatically transcribed into their respective canonical allophonic forms using an upgraded version of *PHONEMIA* [14]. The transcriber produced about 201,000 different canonical forms. The numerical difference between the transcriber’s input and output sets is due to homophony. The orthographic-to-allophonic form ratio is R≅1.046.

The system’s output underwent a single normalization process relevant to the so-called “prenasalization” of voiced stops of MG: all instances of [mb], [nd], [ɲɟ], [ŋg] clusters were respectively simplified to [b], [d], [ɟ], [g] to avoid loss of combinatory information [1]. The potential magnitude of such loss can be illustrated by the case of the word <κόμπος> ‘knot’. The canonical form [kó**mb**os] cannot combine with any other legitimate MG form to yield minimal pairs. The normalized form [kó**b**os], however, belongs to the 7-member *paradigm* [kó(b|ɾ|k|l|m|n|p)os] which, according to (3), for $\mu = 7$ and $\nu = 2$, yields $\sum = 21$ minimal pairs:

$$\sum\nolimits_{\nu}^{\mu} = \frac{\mu!}{\nu!(\mu-\nu)!} \tag{3}$$

where μ is a number of discrete objects (here, legitimate MG canonical forms) and $\sum$ is the number of possible non-ordered ν-tuples (here, pairs) that these objects can form. [kóbos] is jointly responsible for 6 of the 21 (29%) relevant minimal pairs.

3.3. Minimal Pairs (3rd order data)

Around 149,500 minimal pairs were automatically extracted from the allophonic data. These are distributed unevenly across the 630 (theoretical) combinations of phones in contrastive distribution. The distribution of numbers of minimal pairs per sound contrast is given in Table 1a.

3.4. Functional Loads (4th order data)

A comprehensive review of the literature relevant to the quantification of the notion of Functional Load (FL) can be found in [15]. For our purposes, we adopt the following view of FL: the FL of a certain sound contrast operating within a lexicon must reflect the portion of that lexicon’s (quantifiable) *ability to convey information* through utilization of that contrast. Thus, the contrast [p~t] operating in the MG lexicon serves, among other things, to keep utterances like [**p**íno] (‘drink’) and [**t**íno] (‘tend’) apart. If MG stopped making that distinction, the words corresponding to such pairs of spoken signals would begin to sound the same, i.e., they would become homophones.

In this work, we define the functional load $FL[x\sim y]_{w_i,w_j}$ of two candidate phonemes x and y in contrastive distribution within two specific canonical forms $w_i = [a_1, a_2, a_3, \ldots, a_l]$ and $w_j = [b_1, b_2, b_3, \ldots, b_l]$ (i.e. within a specific minimal pair) as follows:

$$FL[x\sim y]_{w_i,w_j} = k / l \tag{4}$$

where l is the length of either w_i or w_j and k is the distance of the site of the sound contrast (measured in number of sound symbols) from the beginning of either w_i or w_j . Hence, the $FL[x\sim y]_{w_i,w_j}$, with x = [k] and w_i = [s**k**azmós] (‘silence’), and y = [p] and w_j = [s**p**azmós] (‘spasm’), is 2 / 7 = 0.286. By contrast, $FL[x\sim y]_{w_i,w_j}$, with x = [k] and w_i = [sína**k**si] (‘gathering’), and with y = [p] and w_j = [sína**p**si] (‘synapse’), is 5 / 7 = 0.714. The fact that the FL of the [k~p] contrast in the second pair is *greater* than the one in the first makes good sense: in the second case, the sound string [sína...], which needs to be retained in memory before it is disambiguated by the occurrence of either [k] or [p], is *longer* than the respective string [s...] in the first case. It is conceivable that an FL notion comparable to the one defined herein could be used to improve the performance of Automatic Speech Recognition (ASR) applications.

The functional load $FL[x\sim y]$ of a contrast between candidate phonemes x and y is defined as the sum of the FLs of every minimal pair pertinent to that contrast divided (i.e., normalised) by the sum of the FLs of all minimal pairs pertinent to all combinations of all candidate phonemes:

$$FL[x\sim y] = \frac{\sum FL[x\sim y]_{w_i,w_j}}{\sum FL[r\sim s]_{w_i,w_j}}, \quad (\exists\, x, y \in A, x \neq y) \wedge (\forall\, r, s \in A, r \neq s) \wedge (\forall\, w_i, w_j \in L, (1) \wedge (2)\text{: true}) \tag{5}$$

where L is the lexicon of MG (in our case, our wordlist).

Finally, we may define the functional load $FL[x]$ of a candidate phoneme x as the sum of all functional loads pertaining to each contrast of that phoneme and each of all the other members of A:

$$FL[x] = FL[x\sim y] + FL[x\sim r] + \ldots + FL[x\sim s], \quad \forall\, x, y, r, \ldots, s \in A, x \neq y \neq r \neq \ldots \neq s \tag{6}$$

A notion such as “the FL of a *single* phoneme” (as opposed to the FL of a *contrast* between *two* phonemes) is not self-contradictory, since phonemes are theoretical constructs which, by virtue of their very definition, are inherently contrastive entities [8], [9], [16].

The FLs of the various oppositions between candidate phonemes can be found in Table 1b. The FLs of the candidate phonemes of set A (symbols 1 – 31) are outlined in Figure 1.

Table 1a (top-right half). *Numbers of Minimal Pairs pertinent to each pair of candidate phonemes in contrastive distribution across the wordlist. Counts ≥ 100 in grey background. Relevant sound classes (left-to-right): (voiceless and voiced) plosives and fricatives; sibilants; nasals; liquids; (front and back) vowels (unstressed and stressed).*

○	p	t	c	k	b	d	ɟ	g	f	θ	ç	x	v	ð	ʝ	ɣ	s	z	ɱ	m	n	ɲ	ŋ	ɾ	l	ʎ	e	i	a	o	u	é	í	á	ó	ú	○
p	\	391	214	602	186	156	27	35	677	257	175	217	300	343	157	103	492	92	0	634	510	30	1	272	351	40	69	79	98	61	16	1	1	0	1	1	p
t	.001	\	300	334	72	288	35	30	303	2360	180	146	301	239	168	112	1316	88	0	1374	885	29	0	430	622	32	32	104	62	24	28	0	1	1	6	0	t
c	.001	.001	\	44	23	26	8	0	155	201	135	4	59	136	169	7	137	69	0	269	147	7	4	110	232	6	3	21	32	3	1	0	0	0	0	0	c
k	.002	.001	.000	\	93	65	3	67	280	155	8	791	280	128	40	154	300	51	6	229	291	7	0	186	270	29	105	83	83	46	17	0	0	0	0	0	k
b	.001	.000	.000	.000	\	42	32	22	114	49	16	61	76	92	37	31	88	46	0	137	62	9	0	134	125	10	13	12	7	6	0	0	0	0	0	0	b
d	.000	.001	.000	.000	.000	\	9	28	64	48	81	30	34	93	81	17	607	64	0	845	189	10	0	116	152	35	6	8	1	9	3	0	0	0	0	0	d
ɟ	.000	.000	.000	.000	.000	.000	\	1	12	12	12	2	6	15	20	1	19	10	0	20	43	1	0	24	17	4	0	0	0	0	0	0	0	0	0	0	ɟ
g	.000	.000	.000	.000	.000	.000	.000	\	8	10	0	34	40	5	1	52	21	20	0	24	24	4	0	49	30	2	9	13	7	10	5	0	0	0	0	0	g
f	.002	.001	.000	.001	.000	.000	.000	.000	\	170	120	160	180	166	92	86	230	126	0	279	214	10	0	186	237	16	38	52	37	17	5	0	0	0	0	0	f
θ	.001	.015	.001	.000	.000	.000	.000	.000	.000	\	134	91	162	187	100	54	158	56	0	223	177	7	0	334	312	6	41	24	32	8	4	0	1	0	1	0	θ
ç	.001	.001	.000	.000	.000	.000	.000	.000	.000	.001	\	6	60	117	100	5	100	42	0	138	80	6	0	96	173	7	12	27	9	0	1	0	0	0	0	0	ç
x	.001	.000	.000	.003	.000	.000	.000	.000	.000	.000	.000	\	103	114	30	69	156	67	0	127	81	20	0	97	83	32	42	28	68	46	6	0	0	0	0	0	x
v	.001	.001	.000	.001	.000	.000	.000	.000	.001	.000	.000	.000	\	138	81	67	152	62	0	258	208	12	0	168	185	11	26	12	63	50	7	0	0	0	0	0	v
ð	.001	.001	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	\	106	40	302	148	0	281	340	46	0	290	396	34	18	20	18	19	8	0	0	0	0	0	ð
ʝ	.000	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	12	134	86	0	197	112	9	0	115	133	5	58	60	18	15	3	0	0	0	0	0	ʝ
ɣ	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	99	56	0	138	64	17	0	144	107	12	39	39	24	29	6	0	0	0	0	0	ɣ
s	.001	.009	.001	.001	.000	.004	.000	.000	.001	.001	.000	.000	.001	.001	.001	.000	\	3260	0	1010	11446	24	1	479	531	28	73	273	400	224	100	3	5	2	4	3	s
z	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.022	\	0	96	127	20	0	179	235	22	7	17	3	25	0	0	0	0	0	0	z
ɱ	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	0	5	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	ɱ
m	.002	.009	.001	.001	.000	.006	.000	.000	.001	.001	.000	.000	.001	.001	.001	.000	.006	.000	.000	\	1743	11	0	352	349	41	25	44	28	9	2	0	2	1	0	0	m
n	.002	.006	.001	.001	.000	.001	.000	.000	.001	.001	.000	.000	.001	.001	.000	.000	.094	.001	.000	.012	\	65	0	596	537	37	20	57	41	45	33	0	0	0	0	0	n
ɲ	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	0	34	23	33	2	7	1	2	0	0	0	0	0	0	ɲ
ŋ	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	6	0	0	0	1	0	0	0	0	0	0	0	0	ŋ
ɾ	.001	.002	.001	.001	.000	.000	.000	.000	.001	.001	.000	.000	.001	.001	.000	.001	.002	.001	.000	.001	.003	.000	.000	\	946	34	22	71	21	14	10	0	0	0	0	0	ɾ
l	.001	.003	.001	.001	.000	.001	.000	.000	.001	.001	.001	.000	.001	.001	.000	.000	.002	.001	.000	.001	.003	.000	.000	.004	\	55	15	36	11	7	2	0	0	0	0	0	l
ʎ	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	1	4	0	6	0	0	0	0	0	0	ʎ
e	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	4931	6511	4118	2570	2	1	2	2	5	e
i	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.036	\	6501	9886	5134	0	1	1	2	3	i
a	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.003	.000	.000	.000	.000	.000	.000	.000	.000	.000	.047	.049	\	7509	5826	4	5	6	3	6	a
o	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.031	.081	.060	\	8595	0	0	1	0	1	o
u	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.020	.042	.046	.069	\	2	2	0	0	4	u
é	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	\	4125	805	926	804	é
í	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.030	\	1393	2863	1975	í
á	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.009	\	3608	2798	á
ó	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.006	.022	.029	\	6086	ó
ú	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.006	.015	.023	.049	\	ú
○	p	t	c	k	b	d	ɟ	g	f	θ	ç	x	v	ð	ʝ	ɣ	s	z	ɱ	m	n	ɲ	ŋ	ɾ	l	ʎ	e	i	a	o	u	é	í	á	ó	ú	○

Table 1b (bottom-left half). *Functional Loads of pairs of candidate phonemes in contrastive distribution across the wordlist. FLs rounded off to the 3rd decimal. FLs ≠ 0 in grey background. Relevant sound classes as in Table 1a.*

4. Initial Observations and Discussion

Our 3rd and 4th order data lend themselves to sophisticated statistical analysis. This will be pursued in a future publication. For lack of space here, it will suffice to make a number of simple observations based on a first inspection of our data:

The data bears out the traditional distinction between vowels and consonants. Generally, vowels appear to do more distinction work than consonants. Note, however, the unique distribution of [s] (a characteristically turbulent voiceless sound), which forms a large number of pairs with nasals, liquids and even unstressed *vowels* (all of which are characteristically periodic and sonorous sounds). A comparative study of the distribution of [s] within words in a number of languages may help to explain the long acknowledged problematic behaviour of [s] in the universal *sonority hierarchy* or *sonority scale* of sounds [9], [16]. The potential of [s] functioning (even if only marginally) as syllabic nucleus in the lexical and phrasal phonology of MG certainly warrants further exploration (c.f. [ˈko.ði.ks̩] > [ˈko.ði.kas] 'code', [ˈpi.ks̩ ˈla.ks̩] 'with punches and kicks').

The 30 most frequent sound contrasts of the 446 attested in our data (Tables 1a and 1b) involve minimal pairs consisting either of two inflexional types of the same lexeme, or of types closely related through the process of morphological derivation, or of two type-variants related to MG *diglossia* (e.g. [x**θ**és] ~ [x**t**és], 'yesterday') [5], [6].

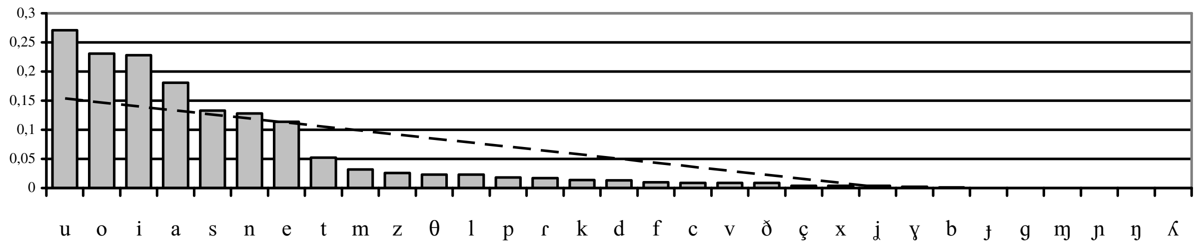

Figure 1: *Functional Loads of candidate phonemes and linear trendline of FL values.*

This is reflected in the arrangement of Figure 1: the first 11 candidate phonemes are involved in such pairs. The distributions of minimal pair frequencies and of the FLs of candidate phonemes are likely to change dramatically if such pairs are filtered out. A first attempt towards this direction will be reported in a forthcoming paper.

Voiced plosives appear to be the least exploited sound class in MG, with the exception of [d], which participates in a large number of inflexional pairs. Their phonemic status is questionable and requires further investigation.

For [ç] and [ʝ], which are traditionally considered to be allophones of the phonemes /x/ and /ɣ/ respectively, it holds that $FL_{[ç]} = FL_{[x]}$ and $FL_{[ʝ]} > FL_{[ɣ]}$. This is almost certainly due to the *CiV problem*, treated in [4], [14] and, in greater detail, in [1]: instances of [ç] and [ʝ] in MG can also occur as allophones of the phoneme /i/.

5. Present limitations and future prospects

As alluded above, minimal pairs may not be of equal importance if viewed across the entire linguistic system of MG and not within the relatively narrow bounds of its lexicon or, indeed, of its phonological sub-component. Our future plans include the indexing and subsequent categorization of minimal pairs according to the lexicographical and morphosyntactic characteristics of their lexical members (e.g. part-of-speech, number, gender, case, person, tense, aspect, mood, etc.) and, following that, the re-computation of the FLs of the various sound contrasts and candidate phonemes.

Furthermore, there are plans for the exploitation of the data yielded by this study in the field of ASR applications.

6. Concluding Remarks

No final conclusions can be drawn on the basis of the data presented herein prior to their thorough analysis and subsequent re-categorisation: we hope that this study will only be the first in a series of future investigations in the distribution of speech sounds in MG. If anything, we believe that this study represents a promising first step towards the conduct of more robust (i.e. data-driven) phonological analyses of MG.

7. Acknowledgements

The first author would like to thank Mr John Higgins [17] for his encouragement back in summer 2002.

8. References

[1] Kalimeris, C., "A study for the extraction of phonological information from the Hellenic National Corpus (HNC) with a view to producing Minimal Pairs and computing Functional Loads for Modern Greek", unpublished MSc thesis, "Technoglossia", University of Athens and National Technical University of Athens, 2004. In Greek.

[2] Kotropoulos, K., Mavromatidou, P. and Pitas, I., "Phones and phonemes of Modern Greek", Proc. of the 21th Annual Meeting of the Linguistics Dept. of the University of Thessaloniki, 2000. In Greek.

[3] Householder, F.W., Kazazis, K., Koutsoudas, A., "Reference Grammar of Literary Dhemotiki", in the International Journal of American Linguistics, Supplement to Vol. 30, part II, 1964.

[4] Setatos, M., Phonology of Modern Greek, Thessaloniki, Papazisis, 1974. In Greek.

[5] Mackridge, P., The Modern Greek Language, Oxford, Oxford University Press, 1985.

[6] Holton, D., Mackridge, P., Philippaki-Warburton, I., Greek: A Comprehensive Grammar of the Modern Language, London, Routledge, 1997.

[7] Klairis, C., Bambiniotis, G., Grammar of Modern Greek, Athens, Ellinika Grammata, 2004. In Greek.

[8] Crystal, D., A Dictionary of Linguistics and Phonetics, 4th edition, Oxford, Blackwell, 1997.

[9] Laver, J., Principles of Phonetics, Cambridge, Cambridge University Press, 1994.

[10] International Phonetic Association (ed.), Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet, Cambridge: Cambridge University Press, 1999.

[11] Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., et al., "Design and implementation of the online ILSP Greek Corpus", Proc. of the 2nd International Conference on Language Resources and Evaluation, 1737-1742, 2000. HNC Online, available at: http://hnc.ilsp.gr/en/.

[12] Gavrilidou, M., Lambropoulou, P, Mantzari, E., Roussou, S., "The Morphological Level of a Computational Lexicon", Proc. of Panhellenic Conf. on New Information Technology, Athens, 1998. In Greek.

[13] SIMFONIA, Spelling Checker Software, © 2000 ILSP, http://www.ilsp.gr/correct_eng.html .

[14] Bakamidis, S. and Carayannis, G., "PHONEMIA: a phoneme transcription system for speech synthesis in Modern Greek", Speech Communication 6,2, 159-170, 1987.

[15] Surendran, D., "The Functional Load of Phonological Contrasts", unpublished MSc thesis, Department of Computer Science, University of Chicago, 2003.

[16] Spencer, A., Phonology: Theory and Description, Oxford, Blackwell Publishers, 1996.

[17] Higgins, J., "Minimal Pairs for English RP", March 16, 2007. [Online]. Available: http://myweb.tiscali.co.uk/wordscape/wordlist/index.html. [Accessed: May 29, 2007].