Quantifying Greek rhyme

Nina Topintzi, Konstantinos Avdelidis & Theodora Valkanou

Aristotle University of Thessaloniki
topintzi@enl.auth.gr, *kon.avdel@gmail.com*, *teoval@hotmail.com*

Abstract
In this paper we report on *Greek Rhyme (GrR)*, the first freely accessible pilot database on rhyme in Greek poetry. The database contains several poems and is constantly expanding. For the purposes of this project, specially designed algorithms have been constructed for automatic rhyme detection and classification according to numerous parameters, including rich and imperfect rhymes, among others. *GrR* facilitates the study of patterns of rhyme within the work of a single poet and enables comparison with the rhyme schemata found in other poets. It is thus an invaluable tool for both linguists and philologists with interest in poetic meter.

Keywords: rhyme, database, detection algorithms, rich rhyme, imperfect rhyme

1 Introduction

The concept of *rhyme* refers to the sound correspondence between one or more syllables from – usually –the last stressed vowel up to the end of the line, e.g. kr<u>ína</u> – elaf<u>ína</u> 'lilies - doe' (see also Κοκόλης/Kokolis 1993: 26). Although related, poetic rhyme and phonological rime are different. The latter refers to the syllable nucleus and any coda consonants following it, that is, the syllable portion that excludes the onset. Our focus, however, in the current paper is on poetic rhyme.

Rhyme is rather under-studied in comparison to other metrical components (Köhnlein & van Oostendorp 2014). Greek rhyme in particular, is virtually unexplored, at least from a linguistic point of view, given that the single study exclusively devoted to the phenomenon is a philological treatise by Κοκόλης/Kokolis

(1993), which outlines and exemplifies certain patterns of rhyme, excludes many others, and supplies no information regarding their frequency.

In this paper we attempt to provide a preliminary answer to this issue. To accomplish that, however, it is necessary that we access relevant poetic data in a more systematic and quantifiable way. Towards that end, a major step has been the construction of a database of Greek rhyme, which we report on here. In what follows, we discuss several methodological and technical aspects of this project, alongside some initial results. These lay out aspects of a previously uncharted territory, offering a quantitative treatment of various Greek rhyming schemes that will allow us to evaluate, complement, and challenge philological studies with newly gained insights. We also contend that a more comprehensive understanding of Greek rhyme may prove significant for a fuller understanding of the general phonology of Greek.

The article is structured as follows. Section 2 presents general information on the *Greek Rhyme* project (internal structure, content) and delineates the basic architecture of the project-specific algorithms for automatic rhyme detection. The parameters taken into consideration are defined and discussed. Section 3 provides some preliminary results pertinent to the whole corpus available at this time, and to specific poets. Section 4 showcases how the quantitative-linguistic approach outlined here may complement literary studies by highlighting the preponderance of imperfect rhymes in Palamas and confirming impressionistic assessments that Mavilis' poetry is full of rich rhymes (Σπαταλάς/Spatalas 1935: 20). Section 5 offers some concluding remarks.

2 The *Greek Rhyme* project

Greek Rhyme (GrR) (http://greek-rhyme.web.auth.gr/) is the first freely accessible pilot database concerned with the phenomenon of rhyme in Greek. It has been funded by the AUTh Research Committee (grant made to the first author; code: 93330). The GrR website consists of: (a) a small scale database with poems from diverse poets, which gets to be regularly updated through an extension of the library of poems, (b) a series of original rhyme detection and classification algorithms, especially devised for the current project, (c) administrator tools for manual error correction (*currently in progress*), (d) graphical user interface, (e) information and instructions on how to use

the site, and (f) statistic tools. Administrators have access to all components, whereas regular users have access to the latter three components.

At the time of writing (February 2018), the database contained 5,779 lines from the following poems:

- *Ύμνος εις την Ελευθερία* του Δ. Σολωμού (*Hymn to Freedom* by D. Solomos)
- The collection *Πυθμένες* του Κ. Βάρναλη (*Depths* by K. Varnalis)
- The collection *Ο Πόνος του Ανθρώπου και των Πραμάτων* του Κ. Καρυωτάκη (*The Pain of Man and Things* by K. Karyotakis)
- 23 sonnets by L. Mavilis
- *Ο Δωδεκάλογος του Γύφτου* του Κ. Παλαμά (*The twelve Words of the Gypsy* by K. Palamas)
- A collection of short poems by A. Valaoritis, G. Seferis, M. Polydouri, A. Sikelianos

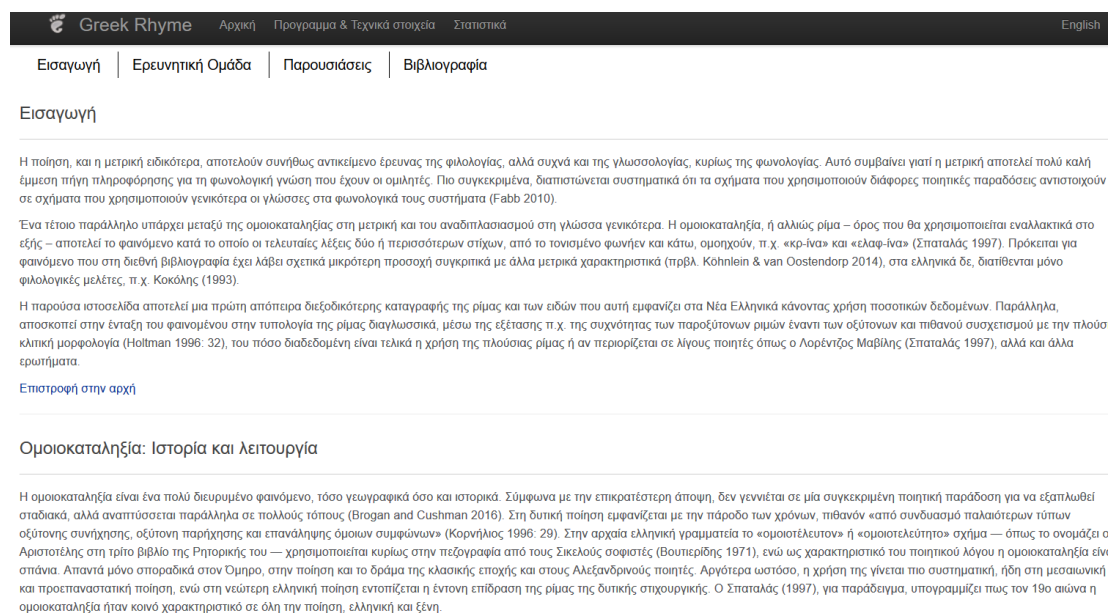A screenshot from the website's main page is presented below.

Figure 1. Screenshot from GrR main page

2.1 Stages in rhyme detection and classification

The goal of this project has been the automatic detection of rhyme pairs (**RP**s) and their classification according to certain criteria. A RP by definition includes two lines. However, while line [a] forms a RP with line [b] when they rhyme, in reality, the

rhyme of line [b] with [a] also forms a RP. In that sense, 2 lines form 2 RPs. This bidirectional effect is encoded in the text later on, along with the statistics we have implemented. The criteria utilized assume the internal structure depicted in Fig.2.
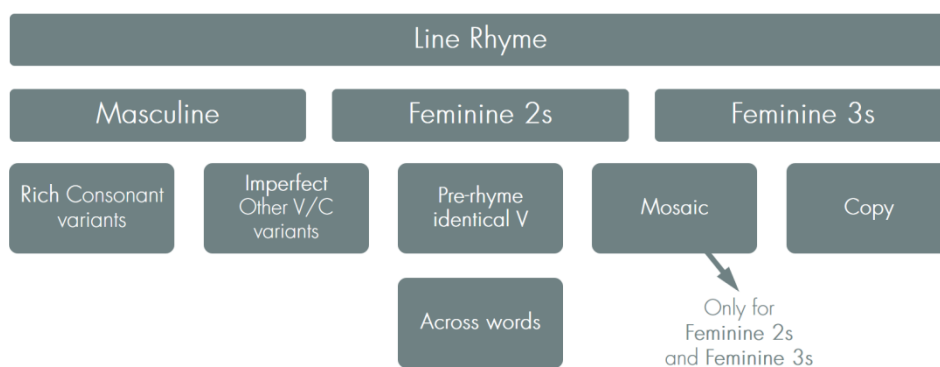

Figure 2. Rhyme structure in Greek Rhyme

The first main categorization classifies rhymes in 3 groups[1]: (i) masculine rhymes (M), where the rhyme domain contains the finally stressed vowel and everything to its right, (ii) feminine-2 rhymes (F2), where the rhyme domain contains the stressed vowel on the penult and everything to its right, and (iii) feminine-3 rhymes (F3), where the rhyme domain contains the stressed vowel on the antepenult and everything to its right. When no other factors come at play, then such rhymes are dubbed **prototypical** or **pure**, since they constitute the simplest rhyme types possible in a 1-, 2- or 3-syllable domain.

For Κοκόλης/Kokolis (1993), the only book-length study to date on Greek rhyme, pure rhyme is basically the only type of rhyme worthy of study, as his remarks reveal: "Είναι γεγονός..., θα συναντήσουμε, ανάμεσα στις άλλες ομοιοκαταληξίες, τις κανονικές, και **κάποιες ρίμες μη ικανοποιητικές**, που θα μπορούσαμε να τις ονοματίσουμε «μερικές ομοιοκαταληξίες»" (ibid.: 34; emphasis added is ours), and then later, "η ομοιοκαταληξία είναι *πλήρης ομοηχία*" (-ibid.: 37) ["it is a fact that, beyond other rimes, we will encounter some non-satisfactory rimes, which we could call 'partial rhymes" and "rime is perfect sound match"].

We disagree with such a conclusion on the basis of work on rhyme from different traditions, e.g. English, where a number of other (partial) rhyme variants are

---

[1] These rhymes are traditionally known in the Greek literature as οξύτονη, παροξύτονη and προπαροξύτονη ομοιοκαταληξία.

admitted (e.g. Lennard 1996; Miller & Greenberg 1981, among many others). In fact, Hobsbaum (1996: 46), views partial rhyming (pararhyme) as an achievement in poetry, rather than an incomplete feature or a mistake. Interestingly, a similar view to ours is independently proposed for Greek by Ψάλτη/Psalti (2017), who admits and defines several other partial rhyme types based on the poetry of Nasos Vagenas, one of the many poets that admittedly employ such rhyme patterns (ibid.: 314).

With this in mind, we claim that several rhyme variants are possible according to the parameters below, which are presented in detail with their specific values in §2.2[2]. These include **Rich rhyme** (onsets of stressed syllable match in the RP), **Pre-rhyme identical V**(the vowel in the prestressed syllable is identical across RPs; this parameter is detected within and across words), **Mosaic rhyme** (the rhyme domain stretches across words), **Imperfect rhyme** (a vowel or consonant within the rhyme alternates in RPs), **Copy** (the rhyme appears as a fully reduplicated string across RPs). Since a mosaic rhyme spans word boundaries, it can only be relevant for F2 and F3 rhymes.

The process of rhyme detection and identification proceeds as follows. Poems are first manually inserted to the database by the administrators. Each poem undergoes pre-processing; lines are first syllabified, and then orthographic-to-phonetic-transcription (in SAMPA) follows. Synchronous multi-layered representation (Wd, σ, cluster, phoneme) is produced per line.

Each line is then analysed according to hierarchical rhyme detection rules. The rules are structured in such a way so that they identify RPs, when these fit any of the criteria outlined above. Rules are operator-based and involve successive comparison steps of respective line representation nodes in reverse order. That means that they scan a line from right-to-left, pick relevant elements, e.g. vowels in syllables, and compare them with the respective elements of other lines in a pre-set domain (by default the domain is 4 lines). If all comparison steps in the queue are true, then a match is attained and an RP is recorded. An example of a rule sequence, alongside with its interpretation and (mis-)matches is presented in (1).

---

[2] We do not claim that these factors are the only relevant ones nor that there is absolute consensus in their definition (esp. for imperfect rhymes). They are however among the predominant ones across different rhyme traditions and the definitions implemented here are fairly standard.

(1)   Rule sequence example

Comparison step definition		Interpretation
`syllable.r_index.linecluster.type`	`→0`	Each line last syllable
`cluster.phono`	`→ V`	cluster types are vowels and
`cluster.f_index.syllable`	`→op:eq`	are phonemically same and
	`→op:and:lte:1`	are both first or second in σ

Step match: εντροπαλ**ή** [edropa'li] - π**εί** ['pi]

Step mismatch: εντροπαλή [edropa'li] - ακαρτερούσες [akarte'ruses]

The end user can then access the database-wide basic statistics that report the frequency of the various rhyme patterns or filter the results across a number of categories (Poet, Poem, Collection, Section in Collection, Rhyme Type) and combinations thereof (see §3). Thus, for instance, one may choose to see how frequent rhyming of any type is in Karyotakis' 14 poems included here (almost 95% of the lines have some rhyme), and then, for example, zoom in specifically on the F2 rhymes and their frequencies across all poems (55.20%) or within independent sections. For instance, the two poems of *Ελεγεία και Σάτιρες* present some type of rhyme in 90% of the time, with F2 being observed in the majority (58.83%).

2.2 Rhyme patterns and exemplification

The algorithms were designed so that they identify aspects of rhyme that are commonly cited in the literature as important to the phenomenon (Abrams & Harpham 2015; Adams 1997; Κοκόλης/Kokolis 1993; Lennard 1996; Σταύρου/Stavrou [1930] 2004), with the exception of the factor 'Pre-rhyme Identical V'. In English for example, this is not a feature usually considered part of the rhyme, although scholars such as Lennard (1996: 191) and Miller & Greenberg (1981: 110) note that rhymes may include more sounds before the last stressed. As far as Greek rhyme is concerned, scholars such as Κοκόλης/Kokolis (1993) assert that rich rhyme also covers the preceding vowel before the stressed V, and everything in-between (ibid.: 42-43). Given that, during our readings of the material, at least some poets, e.g. Solomos, favored assonance of the pre-rhymal vowel, we decided to include this parameter as well.

The components of rhyme that we considered alongside the settings they may take are listed below with their respective examples. The abbreviations for the patterns in question appear in different typeface to facilitate readability. Recall that words appear transcribed in SAMPA (for IPA correspondence, see https://www.phon.ucl.ac.uk/home/sampa/grk-uni.htm) and that the RP is indicated through underlining. In the case of rich rhymes, boldface indicates matching consonants and italicization mismatching consonants in partial rhymes. Capitalization of vowels is reserved to mark identical pre-rhymal vowels.

(2)   *Rhyme factors under consideration*

a. **Rich Rhyme:** onsets of the stressed σ match in RP

   i. total rich rhyme with singleton or complex onsets, i.e. `TR-S` (aliso**m**éno – Gra**m**éno) and `TR-CC` (a**vj**í – na**vj**í)

   ii. partial rhyme rich rhyme with singleton & complex onset, i.e. `PR-C1` (**s**tómata – **s**ómata) or `PR-C2` (*t***r**ízun – fove**r**ízun)

   iii. partial rich rhyme with complex onsets, i.e. `PR-CC1` (pi*x*t*á* – vra*x*n*á*) or `PR-CC2`

b. **Pre-rhyme Identical V**: vowel in prestressed σ is identical across RPs (shown in capitals)

   i. `IDV`: Anáfti – Astráfti

   ii. `IDV-2W` (across words): tOvóli – ecítOdóli

c. **Mosaic rhyme**: rhyme stretches across words

   `MOS`: Dóz mu --fóz mu

d. **Imperfect rhyme**: V or C within rhyme alternates (shown through double underlining)[3]

   i. `IMP-V`: stressed V differs across RPs (xánete – j\ínete)

   ii. `IMP-C`: one or more Cs after stressed V differ across RPs (ksaf**n**ízi – tex**n**íti)

e. **COPY**: rhyme as full reduplication (Den íne – pan íne)

---

[3] There is one more pattern of imperfect rhymes that the system captures, abbreviated as **IMP-0**. This identifies cases where within the rhyme domain, there is a consonant-zero alternation, so that a consonant appears in one RP, but is missing in the other. When this occurs finally in codas, as in πιστοί – χαρείς, we have the *IMP-0F* variant. When it occurs medially, as in ωραίοι – κατατρεμένοι, we find the *IMP-0M* variant.

Let us work through an example using the 2nd stanza of Seferis' *Arnisi* (3). Line numbering appears in the first column, followed by numeric arrangement in stanzas (and internal line numbering). The third column presents the line in Greek. At the end of the line, a number appears in brackets. The digit (1) signals that there is a single RP for this line *later* in the poem, while (0) that there is none. Finally, we get the SAMPA transcription of each line. Note that RPs are identified only the first time they are encountered. Thus, a (0) marking does not necessarily signal the lack of a rhyme. This choice of rhyme representation is justified on multiple grounds: first, it allows easy detection of redundant or false-detected rhymes; second, it permits single appearance in statistics, and third, when manual corrections are applicable, there is only one single editing position.

(3)   *Seferis' Arnisi, lines 5-8*

5	2/1	Πάνω στην άμμο την ξανθή	(1)	'pa-no 'stin 'a-mo tin ksan-'Ti
6	2/2	γράψαμε τ' όνομά της·	(1)	'Gra-psa-me 'to-no-'ma tis
7	2/3	ωραία που φύσηξεν ο μπάτης	(0)	o-'re-a pu 'fi-si-ksen o 'ba-tis
8	2/4	και σβήστηκε η γραφή.	(0)	ce 'zvi-sti-ceiGra-'fi

Upon clicking on the digits in brackets that indicate a RP, a greyed-out text unfolds that presents the matching line(s), as well as the type of rhyme(s) identified.

(4)   *RPs and rhyme classification of (3)*

5	2/1	Πάνω στην άμμο την ξανθή	(1) 'pa-no 'stin 'a-mo tin ksan-'Ti
8	2/4	και σβήστηκε η γραφή.	M
		M(IDV)	
6	2/2	γράψαμε τ' όνομά της·	(1) 'Gra-psa-me 'to-no-'ma tis
7	2/3	ωραία που φύσηξεν ο μπάτης	F2
		F2(IDV(IDV-2W), MOS)	
7	2/3	ωραία που φύσηξεν ο μπάτης	(0) o-'re-a pu 'fi-si-ksen o 'ba-tis
8	2/4	και σβήστηκε η γραφή.	(0) ce 'zvi-sti-ce i Gra-'fi

Line 5 thus rhymes with 8 and line 6 with 7. The former RP involves a masculine rhyme with identical pre-rhymal V, i.e. M-IDV, as in ksAn'T<u>i</u> -GrA'f<u>i</u>. The latter is a F2 rhyme (starting from the penultimate vowel in the line), and is MOS, as it straddles

word boundaries 'tonO'm<u>a tis</u> - O 'b<u>atis,</u> with pre-rhymal vowel identity across the words (IDV-2W).

In some cases, multiple RPs are identified. For instance, line 213 below forms RPs with both lines 214 and 216, as exemplified in (5).

(5)   *Lines 213, 214 and 215 (Palamas, The twelve Words of the Gypsy – Speech 8)*

213	θεοδώρητες και νίκες δίνει	(2) Te-o-'Do-ri-tes ce 'ni-ces 'Di-ni
214	του Βασιλιά σου, Ρωμιοσύνη! F2	F2
216	κ' η Ανατολή, μεστή και η Δύση, F2(IMP-C, TR-S)	F2

3 Preliminary results

3.1 General Remarks

In the current corpus of 5779 lines, the majority, i.e. 4936 (85.41%) presented some type of rhyme. This is anticipated given that: (a) our sample applied a loose criterion for inclusion of poems in our database, namely that they demonstrate *some* rhyme within the poem, but not necessarily throughout the poem, (b) we have permitted a wide range of subtypes of rhymes (e.g. rich and imperfect), beyond the prototypical `M,F2` and `F3` ones, for the reasons explained in §2.1.

With this in mind, let us proceed with some preliminary results regarding our corpus of lines with some rhyme (i.e. the 4936 ones). The majority of rhymes exhibits the F2 pattern (54.06%), followed by the M ones (44.94%), with only 1.01% of the rhymal lines displaying the F3 pattern. Given the rarity of F3 rhymes, very little can be said about them at this stage, so unless explicitly mentioned, we will focus on M and F2 rhymes.

Recall from (2) that the rhyme factors taken into consideration involve *rich, imperfect, copy* and *mosaic* rhymes, as well as *identity of the pre-rhymal vowel*. The following preferences arise, alongside their frequencies.

(6)   *Frequency results over the whole rhyme corpus*

   a. Breakdown of the M rhymes (= 44.94% of the rhymes in the corpus)

Imperfect (25.42%) > Pre-rhyme IDV (11.13%) > Rich (6.95%) > Copy (0.28%)[4]

b. Breakdown of the F2 rhymes (= 54.08% of the rhymes in the corpus)

Imperfect (16.78%) > Pre-rhyme IDV (15.97%) > Rich (11.60%) > Mosaic (4.82%) > Copy (1.67%)

The findings are comparable resulting in the following hierarchy of rhyme type preference: Imperfect > Pre-Rhyme IDV > Rich > Mosaic > Copy. We should be aware, though, that the percentage of imperfect rhymes may be somewhat exaggerated, especially in M rhymes, since the algorithms classify RPs of the type ['les - 'pas] with alternation of the stressed vowel as IMP-V. Any similar V-alternation will be deemed IMP-V. In addition, the algorithms identify co-existing types of imperfect rhymes. For example, the pair *γιορτής - ζωντανά* [j\or'tis - zoda'na] (Palamas, Λόγος H' lines 183, 185) is classified as M(IMP-V, IMP-0F). Strictly speaking, these are imperfect rhymes, but it is less clear whether they are perceived as such. Given the lack of studies on the perception of rhyme in Greek, we have decided to incorporate multiple variants of imperfect rhymes with the caveat that these might later need to be modified.

At present, however, our system provides the statistical tools to tease apart the types of imperfect rhymes and allows us to focus on specific types only, if desired. Thus, the first panel in (7) includes the, more marginal, IMP-0F and IMP-0M patterns, whereas the second one ignores them. There are two percentages in both panels; the first one demonstrates the frequency of the pattern in question within the total 5779 lines of the current corpus, while the second shows the frequency within the lines that present imperfect rhyme. This is why e.g. IMP-C occurs in 19.87% of the IMP cases when all IMP patterns are considered (1st panel), but in 30.29% when IMP-0F and IMP-0M are omitted (2nd panel). What is common, though, is that IMP-C is found in 7.23% of the total corpus.

---

[4] Note that the percentages here and below do not necessarily add up, e.g. the sum of IMP+IDV+Rich+Copy = is 43.78% and not 44.94%. This is a result of the fact that an overlap between categories may emerge that is not reflected by looking at the percentages of the individual categories (see also example (5) for illustration).-

(7) *Imperfect rhymes throughout the corpus*   *Query without IMP-0F, IMP-0M*

Lines	5779.00		
Imperfect +	2102.10	36.37%	
IMP-0F	576.93	9.98%	27.45%
IMP-0M	146.33	2.53%	6.96%
IMP-C	417.60	7.23%	19.87%
IMP-V	961.23	16.63%	45.73%

Lines	5779.00		
Imperfect +	1378.83	23.86%	
IMP-0F	0.00	0.00%	0.00%
IMP-0M	0.00	0.00%	0.00%
IMP-C	417.60	7.23%	30.29%
IMP-V	961.23	16.63%	69.71%

3.2 On specific poets

In this section we present some of the findings regarding the preferred rhyme patterns of specific poets (and specific poems)[5]. We start with *Ύμνος είς την Ελευθερία* by D. Solomos (hereafter referred to as *Imnos*). The poem contains 632 lines and consists of quatrains with RPs being formed between lines 1-3 and 2-4. The former RP is always M, whereas the latter is – almost – always F2, with the exception of a single F3 RP. Our system performed remarkably well, having identified a rhyme pattern in 616 lines and missing a few RPs (2.5% of them actually) as a result of syllabification issues. For instance, lines 213 and 215 end in the sequences *άδεια* and *μαυράδια* which rhyme, but the former has been syllabified as ['a-Di-a] and the latter as [ma-'vra-Dj\a] leading to a mismatch. Misidentifications like these will be manually corrected as soon as the rhyme-editing tools – currently underway – have been constructed. The percentages below refer to the whole poem of 632 lines, but are based on RPs identified in 616 of these lines. We anticipate negligible differences in comparison to the full RP corpus of 632 lines.

(8)   *General outlook of Solomos*

        a. Pure rhymes

Lines	632.00	
Type +	400.00	63.29%
M	212.00	33.54%
F2	188.00	29.75%
F3	0.00	0.00%

---

[5] In this section no reference is made to the following poets in our database: Valaoritis, Sikelianos, Polydouri, and Seferis, because at present only a few lines per poet are available.

b. Major patterns of complex rhymes

Lines	632.00	
Type +	616.00	97.47%
M	306.00	48.42%
Pre-rhyme Identical V +	74.00	11.71%
Rich +	30.00	4.75%
F2	308.00	48.73%
Pre-rhyme Identical V +	84.00	13.29%
Rich +	34.00	5.38%
Mosaic +	8.00	1.27%
Copy +	2.00	0.32%
F3	2.00	0.32%
Pre-rhyme Identical V +	2.00	0.32%
Rich +	2.00	0.32%

As is evident (8a), most of the rhymes in *Imnos* (63.29%) are pure, in the sense that they are ***prototypical***, matching just the rime vowel and everything that follows it. There is a slight preference for M pure rhymes over F2 ones, too. No pure F3 rhymes emerge. (8b) demonstrates the distribution of the M, F2 and F3 rhymes in the poem, also detailing the more complex patterns. Thus, M and F2 rhymes are roughly split within the poem, with just two lines (1RP) displaying a F3 rhyme. This one involves lines 151 and 153 and the RP ['sto-ma-ta] – ['so-ma-ta]. In terms of the more complex patterns attested in *Imnos*, our intuition has been correct; 25.32% of the lines exhibit pre-rhymal V-identity in contrast to 10.45% that presents rich rhyme. Mosaic rhymes and copies are exceedingly rare. Notably, *Imnos* admits no imperfect rhymes at all.

31 poems constitute the present corpus of Varnalis' poems. The vast majority of the lines present some rhyme. However, unlike Solomos, almost 92% of all rhyme instances are of the F2 type, with less than 8% exhibiting a M pattern. Only one RP displays the F3 type. Almost all prototypical rhymes are of the F2 type. The reverse holds for the M rhymes, but these are admittedly very few, as mentioned before.

(9) *Varnalis: General outlook*

a. main rhyme types

Lines	480.00		
Type +	432.00	90.00%	
M	34.00	7.08%	7.87%
F2	396.00	82.50%	91.67%
F3	2.00	0.42%	0.46%

b. pure rhymes

Lines	480.00		
Type +	163.00	33.96%	
M	2.00	0.42%	1.23%
F2	161.00	33.54%	98.77%
F3	0.00	0.00%	0.00%

Turning to complex rhymes now, we just consider the patterns of F2 rhymes, given that this is the predominant type. The hierarchy Pre-Rhyme IDV > Imperfect > Rich > Mosaic emerges. Almost 30% of all F2 rhymes display Pre-Rhyme IDV, rendering this preference even stronger than in Solomos' *Imnos*. A feature that stands out in comparison to other poets is Varnalis' affinity for MOS rhymes, which are generally highly rare otherwise.

(10) *Varnalis: Complex rhymes*

Lines	480.00		
Type +	432.00	90.00%	
F2	396.00	82.50%	91.67%
Rich +	67.33	14.03%	15.59%
Pre-rhyme Identical V +	127.00	26.46%	29.40%
Imperfect +	79.33	16.53%	18.36%
Mosaic +	58.00	12.08%	13.43%

The present corpus of Karyotakis contains 14 short poems. These make 265 lines in total, with the vast majority (≈95%) illustrating some type of rhyme. Unlike Varnalis, Karyotakis brings back some balance in his choice of M vs. F2 rhymes. Over half of the RPs are of the F2 type, and about 40% of them are of the M type. While still highly rare, Karyotakis admits fractionally more F3 rhymes than other poets, however none of them is of the prototypical 'pure' type, as demonstrated in (11b). About 1/3 of the total lines present prototypical rhyming, and this seems to be more common in the case of F2 rhymes than that of M rhymes.

(11)   *Karyotakis: General outlook*

a. main rhyme types

Lines	265.00		
Type +	250.00	94.34%	
M	102.00	38.49%	40.80%
F2	138.00	52.08%	55.20%
F3	10.00	3.77%	4.00%

b. pure rhymes

Lines	265.00	
Type +	97.00	36.60%
M	27.00	10.19%
F2	70.00	26.42%
F3	0.00	0.00%

A closer look at the complex rhymes reveals a few interesting points. First, almost all F3 RPs involve COPY, i.e. rhyme repetition, with only one pair displaying MOS. Second, Karyotakis' rhymal strategy is not uniform across the remaining M and F2 rhymes. When it comes to M rhymes, the rhyme preference is Imperfect > Rich > Pre-Rhyme IDV. While the former two are roughly equally frequent, identity of the pre-rhymal V is clearly less important. In F2 rhymes though, this latter factor becomes the most prominent. Conversely, Imperfect rhymes are highly dispreferred. Rich rhymes are less common, but still come second in preference. The resulting hierarchy then is: Pre-Rhyme IDV > Rich > Imperfect. There are also a few instances of MOS rhymes, but these by definition are impossible in M rhymes.

(12)   *Karyotakis: Complex rhymes*

Lines		265.00		
Type +		250.00	94.34%	
M		102.00	38.49%	40.80%
	Rich +	36.00	13.58%	14.40%
	Imperfect +	39.67	14.97%	15.87%
	Pre-rhyme Identical V +	21.67	8.18%	8.67%
F2		138.00	52.08%	55.20%
	Rich +	22.00	8.30%	8.80%
	Mosaic +	14.00	5.28%	5.60%
	Pre-rhyme Identical V +	40.00	15.09%	16.00%
	Imperfect +	6.00	2.26%	2.40%

4 Gaining (literary) insight from GrR

In this section, we consider two cases that are particularly interesting from a literary point of view and demonstrate *GrR's* important contribution. One involves Palamas' *The twelve Words of the Gypsy* (hereafter *Gypsy*), the longest poem in our corpus; the other is on the sonnets of Mavilis. We show that the findings of *GrR* can prove illuminating in uncovering hardly noticed patterns of rhyming, as in the *Gypsy* case, and in quantitatively confirming former observations that so far remained merely impressionistic, as in the case of Mavilis.

4.1 Palamas' imperfect rhyming

The *Gypsy* is a lengthy poem of 4260 lines[6], primarily written in free trochee, while free iamb is also common. Nonetheless, various rhyme patterns emerge within the poem, even when stanzas are not visually separated from each other with stanza breaks; rhyme occurs predominantly in lines 2/4 in quatrains, lines 2/5 in quintains, lines 3/6 in sestets, and either 4/7 or 3/7 in septets (Πολίτης/Politis [1972] 2014: 61-65).

Ψάλτη (2016) offers a much more elaborate examination of the metrical aspects of the *Gypsy*. With reference to rhyme, she identifies *pure* rhymes in 23.49% of the poem (Ψάλτη/Psalti 2016: 529), but notes that once less canonical rhymes are considered, then these double to almost 50% (ibid.: 532). Our database results *partly* agree with these findings.

(13)   *Palamas: General outlook*

a. pure rhymes

Lines	4260.00	
Type +	1129.00	26.50%
M	627.33	14.73%
F2	497.67	11.68%
F3	4.00	0.09%

b. main rhyme types

Lines	4260.00	
Type +	3376.00	79.25%
M	1846.00	43.33%
F2	1486.33	34.89%
F3	43.67	1.03%

---

[6] This is 3 lines more than the 4257 that Ψάλτη/Psalti (2016) mentions, presumably due to use of different editions.

While the figure in (13) of pure rhymes is comparable (26.50%), the one found once all main types of rhyme are taken together, is far greater, being very close to 80%. Most rhymes in addition are of the M type. When it comes to the complex individual patterns, Palamas displays consistency in most factors. For example, the percentage of Rich and Pre-Rhyme IDV rhymes is roughly equivalent in both M (Rich: 5.41%; Pre-Rhyme IDV: 10.62%) and F2 rhymes (Rich: 5.01%; Pre-Rhyme IDV: 10.32%). The rather remarkable feature of this poem, we argue, is thus the preponderance of IMP rhymes. Almost half the lines exhibit some form of imperfect rhyme. The table below demonstrates how these are distributed across different sub-patterns and main rhyme types. The two most frequent sub-patterns, i.e. IMP-0F and IMP-V are most common in M rhymes, whereas the less common IMP-0M and IMP-C are more frequent in F2 rhymes.

(14) *Palamas: Imperfect rhymes*

Lines		4260.00	
Imperfect +		1967.33	46.18%
IMP-0F		572.00	13.43%
	Type +	572.00	13.43%
	M	500.00	11.74%
	F2	72.00	1.69%
	F3	0.00	0.00%
IMP-0M		112.33	2.64%
	Type +	112.33	2.64%
	M	0.00	0.00%
	F2	110.33	2.59%
	F3	2.00	0.05%
IMP-C		365.00	8.57%
	Type +	365.00	8.57%
	M	23.33	0.55%
	F2	327.67	7.69%
	F3	14.00	0.33%
IMP-V		918.00	21.55%
	Type +	918.00	21.55%
	M	708.00	16.62%
	F2	204.00	4.79%
	F3	6.00	0.14%

4.2 Quantifying qualitative observations: *The case of Mavilis*

Our rhyme detection algorithms also allow us to quantitatively verify or discard impressionistic generalizations that have been reported in the literature. For instance, Σπαταλάς/Spatalas (1935: 20) states that Mavilis is fond of rich rhymes (15)[7]. We are now able to test this observation. As it turns out, after the examination of 23 sonnets by Mavilis, indeed over half of the total lines in the corpus (56.94%) manifest rich rhymes. Their vast majority are of the total TR-S type. Almost 10% exhibit a partial PR-C2 (2a.ii) pattern, where onset identity is ensured between a singleton onset and the second member of an onset cluster. The other total rich rhyme possibility, i.e. TR-CL, appears only occasionally, with the other partial types being highly rare (PR-CC2) or virtually unattested (PR-C1, PR-CC1). An interesting result made visible through this computation of rich rhymes is that when it comes to partial rhymes, Mavilis will – with one exception – always opt for a rich rhyme favouring preservation of the second member of the onset cluster, presumably the more sonorous one, and not of the first one.-

(15)   *Rich rhymes in Mavilis' corpus of 322 lines*

Rich +	183.33	56.94%
TR-S	130.33	40.48%
TR-CL	13.00	4.04%
PR-C1	1.00	0.31%
PR-C2	31.00	9.63%
PR-CC1	0.00	0.00%
PR-CC2	8.00	2.48%

All Mavilis' rhymes in this corpus are of the F2 type. There are no M or F3 rhymes, but this may be a limitation imposed by the sonnets' form. In this context, 31.37% of the (F2) rhymes are of the pure type, a figure that amounts to half of the pure rhymes found in Solomos, but is on a par with other poets such as Varnalis and Karyotakis. The complex rhyme patterns, with the inclusion of the rich rhyme, just discussed, are summarised as follows.

---

[7] These 23 sonnets add up to 322 lines, but our system has identified 330 RPs. This is because in some cases, multiple RPs for a single line have been identified, cf. (5).

(16)   *Mavilis' complex rhyme patterns*

F2	330.00	102.48%
Imperfect +	41.33	12.84%
Rich +	183.33	56.94%
Pre-rhyme Identical V +	105.00	32.61%
Mosaic +	20.00	6.21%
Copy +	13.00	4.04%

Overall, Mavilis' poetry, as presented here, is much less rigid rhyme-wise compared to the rhymes found in Solomos' *Imnos*. Imperfect rhymes are not uncommon, whereas mosaic rhymes and copies are occasionally present too. As noted before, by the standards of the fairly strict form of *Imnos*, pre-rhymal V identity is quite eminent in Solomos; in Mavilis, it is common too, but possibly not as common as one would have expected given the more expanded rhyme patterns he otherwise admits. What stands out instead, as explained, is the prevalence of rich rhymes.

5 Conclusion

The *Greek Rhyme* (GrR) project is the first attempt to access poetic rhyme data in a more systematic and quantifiable way. In its current form, it includes a database of Greek rhyme data, a set of algorithms for the automatic detection of various rhyme types, different GUI for administrators and casual users, as well as numerous statistic tools. The result has been the quantifiable visualization of the rhyme patterns found among (some) Greek poets, as well as the possibility to test for accuracy qualitative assessments that have been reported in the literature (cf. Section 4).

GrR is of course far from complete or error-free, but this has been anticipated given the limited funds available for its initial construction. Future funding should allow us to apply important improvements, additions and modifications, all enabled by provisions that have already been made for GrR's future expandability. For example, the findings that imperfect and rich rhymes are very common in Palamas and Mavilis, respectively, are interesting per se, but raise other questions which in the

present state and capacity of the database, cannot be answered as yet; for instance, are there preferred alternating and matching segments in rich and imperfect rhyme, respectively? And if so, are these random or do they fall into particular groups that can be viewed in the light of natural classes (cf. Kern 2015), thus offering further empirical evidence for the existence of certain theoretical concepts? Consequently then, the scope of this preliminary study is far more sweeping, extending well beyond the metrics-philology to the metrics-phonology interface. Further funding to support continuation of the project will enable us to address such and similar issues.

References

Abrams, M.H. & G.G. Harpham (2015). *A glossary of literary terms* (11th ed.). Boston: Cengage.

Adams, S. (1997). *Poetic designs: An introduction to meters, verse forms, and figures of speech*. Ontario: Broadview.

Hobsbaum, P. (1996). *Metre, rhythm and verse form*. London: Routledge.

Kern, G. (2015). *Rhyming grammars and Celtic phonology*. PhD dissertation, MIT.

Köhnlein, B. & M. van Oostendorp (2014). The relevance of feminine rhyme for phonological theory. In R. Kager, J. Grijzenhout & K. Sebregts (eds.), *Where the principles fail. A Festschrift for Wim Zonneveld*. Utrecht: UIL-OTS, 113-121.

Κοκόλης, Ξ.Α. / Kokolis X.A. (1993). *Η ομοιοκαταληξία. Τύποι και λειτουργικές διαστάσεις*. Αθήνα: Στιγμή.

Lennard, J. (1996). *The poetry handbook: A guide to reading poetry for pleasure and practical criticism*. Oxford: Oxford University Press.

Miller, R. & R.A. Greenberg (1981). *Poetry: An introduction*. London: MacMillan.

Πολίτης, Λ. / Politis L. (2014). *Μετρικά* (Α′ έκδοση 1972). Αθήνα: Μορφωτικό Ίδρυμα Εθνικής Τραπέζης.

Σπαταλάς, Γ. / Spatalas, G. (1935). *Τα σονέττα του Λορέντζου Μαβίλη (Εισαγωγή – Σχόλια στην έκδοση)*. Αθήνα: Δικαίος.

Σταύρου, Θ. / Stavrou Th. (2004). *Νεοελληνική μετρική* (Α′ έκδοση 1930). Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών/Ίδρυμα Μανόλη Τριανταφυλλίδη.

Ψάλτη, Μ. / Psalti, M. (2016). Τροπές και τρόποι του παλαμικού *πολυτροπικού* στίχου: Το παράδειγμα του *Δωδεκάλογου του Γύφτου*. Στα *Πρακτικά του Γ′ Διεθνούς Συνεδρίου Η ποίηση και η ποιητική του Κωστή Παλαμά εβδομήντα χρόνια από το θάνατό του*. Αθήνα: Ίδρυμα Κωστή Παλαμά, 519-554.

Ψάλτη, Μ. / Psalti, M. (2017). Σχόλια στην τυπολογία της νεοελληνικής ρίμας. Η περίπτωση του Νάσου Βαγενά. Στα *Πρακτικά του 8ου Συνεδρίου Μεταπτυχιακών Φοιτητών και Υποψηφίων Διδακτόρων του τμήματος Φιλολογίας του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών*, τόμος Α′. Αθήνα: ΕΚΠΑ, 314-331.