Check for updates

RESEARCH ARTICLE

# A macroscopic view of Greek interwar poetry through Voyant tools

[version 1; peer review: 3 approved with reservations]

Anastasia Natsina

Research Lab for the History of Literature and Literary Genres, UCRC, University of Crete, Rethymno, Greece

**v1** **First published:** 28 Oct 2025, **5**:330
https://doi.org/10.12688/openreseurope.19956.1
**Latest published:** 28 Oct 2025, **5**:330
https://doi.org/10.12688/openreseurope.19956.1

## Abstract

### Background

This paper explores the application of digital humanities tools and methodologies to verify, qualify and extend the findings of traditional (close reading) literary criticism about Greek interwar poetry from a quantitative viewpoint.

### Method

The research uses Voyant Tools (http://voyant-tools.org), an open access, web-based text analysis environment to examine a corpus of over 600 poems by eight Greek interwar poets—namely, Karyotakis, Polydouri, Lapathiotis, Agras, Giofyllis, Filyras, Ouranis, and Papanikolaou—, both as a whole and through comparison among the poets. It uses word frequencies (WF), visualizations, as well as an analysis of collocates and contexts (CA) of frequent and/or significant words; it also discusses vocabulary density (VD) and average words per sentence (AWpS).

### Results

"Soul," "heart" and "eyes" rank highly in WF revealing the lyrical character of these post-symbolist works. Their melancholic mood is shown through the predominance of autumn and winter over spring and summer, as well as the outnumbering of night settings over daytime scenes. CA reveals that the melancholic mood is enhanced through the negation of positive terms, such as joy and light. The analysis of other thematic axes through WF and CA shows poetry presented not as high art but as lyrical expression, as well as the prominence of small-scale urban settings and enclosed spaces. WF

**Open Peer Review**

**Approval Status** ? ? ?

| | 1 | 2 | 3 |
|---|---|---|---|
| **version 1** 28 Oct 2025 | ? view | ? view | ? view |

1. **Hélder Fanha Martins**, Polytechnic University of Lisbon, Lisbon, Portugal
2. **Osamah Mohammed Alyasiri**, Al-Furat Al-Awsat Technical University, Karbala, Iraq
3. **Afendi Hamat**, National University of Malaysia, Bangi, Malaysia

Any reports and responses or comments on the article can be found at the end of the article.

and CA also uncovers individual differences between poets (e.g. Polydouri's emphasis on heart and love, Karyotakis's preoccupation with observation and gaze, as well as a characteristic discrepancy in his use of 'I' and 'you', Filyras's traditional use of simile, and Papanikolaou's modernist tendencies). VD and AWpS further highlight the poets' respective styles.

**Conclusions**

The findings provide quantitative support for existing literary criticism and offer some new insights into both the shared characteristics and distinctive features of Greek interwar poets.

**Keywords**
Voyant Tools, interwar Greek poetry, distant reading, quantitative methods, post-symbolism

This article is included in the Languages and Literature gateway.

This article is included in the Horizon Europe gateway.

**Corresponding author:** Anastasia Natsina (natsina@uoc.gr)

**Author roles: Natsina A**: Conceptualization, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101087269 (TALOS-Artificial Intelligence for Humanities and Social Sciences [TALOS-AI4SSH]).
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2025 Natsina A. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Natsina A. **A macroscopic view of Greek interwar poetry through Voyant tools [version 1; peer review: 3 approved with reservations]** Open Research Europe 2025, **5**:330 https://doi.org/10.12688/openreseurope.19956.1

**First published:** 28 Oct 2025, **5**:330 https://doi.org/10.12688/openreseurope.19956.1

## Introduction

This paper explores the possibilities opened by the open-access Voyant Tools (Sinclair & Rockwell, 2016, http://voyant-tools.org) for the distant/macroscopic reading of large literary corpora and applies them to the reading of a corpus of Modern Greek interwar poetry. Macroscopic (Jockers, 2013) or distant (Moretti, 2013) reading is an approach to literary studies that is ever evolving as more humanists develop programming skills and AI offers new possibilities. Although not based on a single widely shared literary theory, distant reading (as opposed to traditional close reading) provides exciting new insights on the history of literature and the study of literary genres as well as that of individual and group authors. Depending on the computational methods and the tools that are used, distance reading reveals various aspects of texts that go beyond close reading and, in several cases (such as Word2Vec or Principal Component Analysis), beyond immediate human perception and detailed empirical verification.

More specifically, this paper purports to verify and possibly qualify the findings of traditional (close reading) criticism about postwar Greek poetry from a quantitative perspective, as well as provide some new insights on the distinctiveness of different poets, using the multiple features of Voyant Tools.

## Method

Voyant Tools is a web-based reading and analysis environment for digital texts. It was created in 2003 by Stéfan Sinclair and Geoffrey Rockwell and has been regularly updated since then (after Sinclair's demise in 2020, Andrew MacDonald and Cecily Raynor have played an active role). These tools are neither the first nor the only ones of their kind – AntConc (https://www.laurenceanthony.net/software/antconc/) and WordSmith Tool (https://www.lexically.net/wordsmith/) are also well-known tools developed in the field of corpus linguistics that have been used for literary analysis (Al-Rawi, 2017; Wan, 2022). However, Voyant Tools are certainly among the most user-friendly for quantitative approaches without the need for programming skills. The possibilities they offer are numerous and can be applied to a unified or composite corpus (such as the different collections of a single author, or the complete works of different authors), thus facilitating a comparative or evolutionist reading.

Voyant Tools provide information that can be crucial on the level of semantics, such as word frequency (WF), immediate context and collocates for selected words. Moreover, they allow the tracing of how selected words trend in the entire corpus with insightful visualizations. Furthermore, and this is crucial for the study of literature, they offer the possibility to see a word within the context of the entire corpus, so one can practice a parallel close reading to avoid possible distant reading mistakes. The added possibility to extract all or parts of the information in various formats allows the recruitment of LLMs to offer insights on big data, such as long catalogues of different contexts for a specific word. In addition, Voyant Tools include useful features for the discussion of style, such as vocabulary density (VD), average words per sentence (AWpS) and readability index (RI). Other features of Voyant Tools, such as topic modelling for the automatic designation of major topics, tend to be rather less reliable when it comes to literary corpora. The usefulness of this suite of digital open-access and easy-to-use tools is not hard to understand, and it offers an excellent way to comprehend as well as qualify the original insights that quantitative methods can offer to the study of literature.

This paper is based on an open-access dataset created as part of a seminar in the Department of Philology at the University of Crete.[1] The slightly modified and richer corpus used here (accessible and ready-to-use on Voyant Tools: https://voyant-tools.org/?corpus=ffc9e171c756cdf2e60adb8c0d23711a) comprises over 600 poems by a group of interwar Greek poets, namely Tellos Agras, Fotos Giofyllis, Romos Filyras, Kostas Karyotakis, Napoleon Lapathiotis, Kostas Ouranis, Mitsos Papanikolaou, and Maria Polydouri. With the notable exception of Kostas Karyotakis, the most prominent figure of this group who is recognized as a major Greek poet, the interwar poets are often referred to collectively, with an emphasis on their shared features. Melancholy, pessimism, and existential anxiety, stemming among other sources from the frustration of national expansionist aspirations and the dire sociopolitical reality of Greek interwar, as well as an added emphasis on nostalgia and a *sotto voce* quality, all of which are ascribed to neoromanticism and/or neo/post-symbolism,[2] are the most frequently repeated features of these lyrical poets (Beaton, 1994: 122–125; Vitti, 2003: 364–365).

Voyant Tools offers the means to explore these semantic and stylistic similarities quantitatively, as well as to discern diverging tendencies among the different poets. The features of WF, the interpretation of the list of collocates and contexts of significant words, visualizations, VD, and AWpS will be used to this purpose. The same features may also prove useful as a means to compare the rest of the group to the central figure of Kostas Karyotakis, in terms of both style and semantics. Due to restrictions of space, this paper offers only a number of indicative observations, but Voyant Tools may offer material for a much more detailed analysis of themes and style.

## Results and discussion

Vocabulary is a fairly trustworthy index for the designation of theme as well as the 'atmosphere' or mood – in other words the sentiment surrounding the theme. Excluding stopwords – such as articles and pronouns or the verb 'to be' – the most

[1] The dataset, with detailed acknowledgments to the students who cooperated in its creation and the sources of the texts, can be found here: https://zenodo.org/records/12705940. The dataset does not include the work of Kostas Karyotakis, which was downloaded from CLARIN:EL (https://inventory.clarin.gr/corpus/1067).

[2] Aragis (2009) traces the first use of the twin terms neoromantics and neosymbolists for these poets to Stergiopoulos (1980). In her monograph, Filokyprou (2009) argues convincingly for the use of 'post-symbolists' for the same group. Papaleontiou (2009) offers an overview of the terms used to refer to this group in different contexts.

frequent word in the corpus is 'now' (τώρα, 357), positioning the poems repeatedly in the present, indicating a difference from past times that is concomitant with nostalgia (the word itself is not infrequent, νοσταλγ* 78). The very next ones are 'soul' (ψυχή*, 330 / but it jumps to an astounding 427 if we add the cognates by searching for ψυχ*), 'eyes' (μάτια, 319) and 'heart' (καρδιά*, 298), indicating the prevalence of sentiment and the importance of inner life. In these poems, heart and soul are practically interchangeable as the seat of sentiment and the innermost part of self, whereas eyes are considered explicitly or implicitly to be the expression of heart and soul. The most frequent collocate for all three words is the first-person personal pronoun (μου, 153, 83, 139 counts respectively) indicating a centering around personal feelings. The second most frequent collocate, the second-person personal pronoun (σου, 37, 58, 38 counts respectively), indicates a directionality of feeling towards another person. The next most frequent meaningful word is 'life' (ζωή*, 322), with 'my' (μου, 135) as its most frequent collocate, which supports the lyrical character of these poems. Very high in the catalogue of frequent words we find also 'hands' (χέρια, 196) and 'love' (αγάπη, 171). The contexts of 'hands' often indicate connection with a person, most frequently a loved one; its metaphorical uses center around power, protection, love, and care. This is my own assessment of the catalogue, but it is also one supported by the LLM Claude (2025) after exporting the data concerned in text format (since Voyant Tools allows for different forms of data export). Returning to the analysis, the overall idea given by these findings is a corpus of lyrical poems situated in the present, centering around the poetic persona and his/her sentimental world. The prevalence of these words further indicates that this preoccupation is not expressed in a subtle, evocative manner, but it is rather displayed in plain view, with direct reference to the seat of self and feeling.

If these findings give a fairly accurate picture of the thematic center of these poems – self and feeling, particularly love –, it is worth looking for the mood surrounding it, given the repeated mentions of melancholy in the bibliography concerning these poets. Surprisingly, the words 'light' (φως, 215) and 'joy' (χαρά, 198) are among the most frequent. Their most recurring collocate is again 'my'/'to me' (μου, 41 and 46 respectively), reinforcing the lyrical character of the texts. Interestingly, however, 'joy' is also paired with sorrow (λύπη 7, θλίψη 4). Turning to the immediate contexts, one finds that light is often that of the moon or the artificial light of a room; that it's dim, little or fading, scared or trembling, a pale agony. As for joy, it is sometimes sweet, even unmeasured, or someone is full of it. But it is also preceded by 'without'; joy is usually withered, passed, or lost, silent, unspoken, little and unexpected, short-lived, superficial, pale, a bloody rose, pain itself, flown away. So, light and joy, far from indicating positive feelings, are used within contexts that give them an opposite meaning. Indeed, one might suggest that light and joy are repeatedly evoked only to be frustrated. On the other hand, words at the other end of the spectrum seem to be less frequent when taken individually: darkness (σκοτάδ* 57), death (θάνατ* 79, θανάτ* 45) melancholy/ic (μελαγχολ* 17), sorrow/ful (λύπ* 47, λυπ* 47), sad/ness (θλιβ* 67, θλίψ* 49). But, added together, they make a sizeable quantity (405) that supports the perceived idea of these works as mostly melancholic. It seems, however, that these feelings are described more frequently negatively, through their opposites, as a lack or qualification of joy and light, rather than in immediate terms. So, even though the center of lyricism – heart and soul – is displayed in plain view, the moods and emotions surrounding it are presented more suggestively.

If we turn to seasons, that have an added symbolic value as well as an import on the mood of the poem, we will find an interesting pattern of use. Spring (άνοιξη*/άνοιξεσ 58) and summer (καλοκαίρι* 30) are used less frequently than autumn (φθινόπωρ* 25) and winter (χειμών* 77), i.e. 88 and 102 times respectively. But their contexts differ significantly. Spring is indeed very often sweet, full of light, bringing joy, sending all cares away, bringing its gifts, blossoming, light-weighted; but there is also a question as to its use, a certainty that it will end, that this is the last one, it's dying, it's being given to the hands of death. Spring is just another season in the year, a virgin no more, with no swallows, it stinks; springs are indifferent. Summer has more frequently negative associations: it's dying, over, fading, withered, passing, gone; what remains is the ashes of its fruits; it's when the lovers met, but now they are separating; the poet doesn't know why it even came; it's here yet it's still snowing. The positive associations are very rare indeed, limited to the sweet weather and its freshness. Here, one might discern a pattern of use that is fairly similar to that of 'light' and 'joy' – evoked mostly to be emptied of their positive meaning.

Autumn is the season linked par excellence to symbolism, due to its waning and fading qualities which are congruous with decadence; furthermore, the autumn's lack of vividness, its long shadows and half-lights are also linked to the suggestiveness much praised by the symbolists as the core of poetic expression (Polychronakis, 2016: 290–320). However, the word itself is not particularly frequent. But its contexts are, indeed, mostly congruous with waning and fading, melancholia and death: autumn passes through a life permeated with silence; it is associated with lies, nostalgia, sadness, loneliness; it fills the earth with withered leaves; it is announced by the deadly sound of the creaking leaves; it resides in the soul of the poet, it's in their voice; the poet will die in an autumn evening. The single positive connotation (golden autumns) is again used to mark lost happiness. As for winter, it is associated with exile, mourning, death, orphanhood, bitterness, *ennui*, nostalgia, fear, cold, black streets, night and paleness, decay; its positive associations are very rare: an association with the quiet pleasures of old age, the telling of fairytales marking the season, some sweet hours towards its end. Overall, with the partial exception of spring, for which there are almost as many positive associations as there are negatives, the same pattern is discernible: entities with a positive commonplace symbolic use (spring and summer) are imbued with negative meaning, whereas those with common negative associations (autumn and winter) are further corroborated as such within negative contexts.

Let's have a final look at the weather, so commonly associated with mood: Rain and its cognates (βροχ* 120, βρεχ* 52), are amongst the most frequent words of the corpus (172), with sun (ήλιο*147) coming a little behind. As expected, rain comes in almost entirely negative contexts with the rare exceptions of a refreshing, clean, or fertile spring rain. Other than that, it is repeatedly associated with cold, night, monotony, loneliness, frustration, melancholy, and death. The sun has a lot of positive associations, linking it mostly to hope, love, and joy – sometimes also playfulness. But it has almost as many negative ones: old memories drying their mold in the sun like cheap linen; Don Quixote showing his vain wound to the sun; a shadow won't let the poet see the sun; a life is lost following the setting sun; more generally, the sun is repeatedly setting, dead or dying. Again, overall, the negative associations of these elements overrun the positives.

Lastly, one final indication of mood is the vastly unequal uses of day/dawn (μέρα* 125, ξημερ* 6) and night/evening (νύχτ* 188/βραδ* 124). The difference between 131 and 312 occurrences respectively gives a clear image of the preferred time of the day for these neoromantic poems, that value the associations and suggestiveness of darkness more highly than those of the broad daylight.

However, other indications relating these poems to neoromanticism and post-symbolism are not as easy to spot. Filokyprou (2009: 83–157) discusses eight major themes in Greek post-symbolist poets. Some of them, like social reality, truth, self-determination, are not readily traceable through Voyant Tools. Others, like nature (and its counterpart, city), nostalgia and a wish to escape, poets and poetry, are more conducive to this sort of analysis, although not in a straightforward manner. 'Nature' itself (φύση* 47) has few counts, but if we added 'sea' (θάλασσα* 136), mountain/s (βουν* 53), tree/s (δέντρ* 92), garden/s (κήπ* 58), as well as all the specific types of trees/flowers, the image could be quite different. A high word frequency is always telling but a low frequency might not be telling the whole story. Of those, the prominence of sea, particularly in Ouranis and Papanikolaou, as well as its contexts, makes apparent its connection with a longing for travel and escape. As for the urban setting, often substituting nature as setting and theme (Filokyprou, 2009: 92), there is a very high frequency of street/s (δρομ* 273) but just two occurrences of avenue/s (λεωφόρ* 2), indicating a small scale; these poems operate on the level of neighborhood or small town, not the big city (of course, Athens was far from being a metropolis at the time). Houses (σπίτ* 141) and rooms (κάμαρ* 80) also have a marked presence, followed by windows (παράθυρ* 72). Contexts indicate that windows are typically the boundary between the inner space and the outside world – outsiders or neighbors peeking in, goblins threatening to break in; but mostly the poet looking at the world: the passers-by, the weather, the gray sky, the trees, the moon – contemplating or waiting; the windows are open or closed, letting the light in or not, indicating (often frustrated) hope or hopelessness.

Prying themes in this manner can provide some interesting insights. 'Poetry' (ποίη* 7) as such does not have much of a presence in this corpus. Poets (ποιη* 76) have a far more marked presence, confirming the self-centeredness of these works. Art (τέχνη* 24), sometimes with the meaning of craftmanship, has also a low frequency. On the other hand, writing and its cognates are far more frequent (γραφ* 47, γράφ* 28, γράμμα* 25), bringing the total to 100. And 'song' (τραγούδ* 134), most frequently a metaphor for poetry conceived through the musical emphasis placed on it by symbolism, is also found abundantly. So, this corpus doesn't speak so much of art and poetry as of poets and their songs and writing. Poetry is then more of an identity, a craft, a mode of lyrical expression, and less of a high art.

Turning to the specificities of each poet, it is interesting to take an indicative look at the dispersion of some of these findings across the corpus. Besides raw word count, Voyant Tools calculates relative frequencies to adjust for imbalances in document length, thus allowing for comparison among different parts of the corpus, in this case the work of different poets. The Figure below (Figure 1) shows the relative frequencies of the words 'soul', 'eyes', 'heart' and 'love', whose centrality became evident from the beginning. In comparison to everyone else, the poetry of Polydouri, recognized for the expression of strong emotions and for making of love a central theme, has indeed the highest frequency of 'heart' (καρδιά, in turquoise) and 'love' (αγάπη, in orange); Ouranis, whose lack of sentiment is distinctive (Filokyprou, 2009: 59), has the lowest use of both these words. Papanikolaou comes second in the use of 'heart' but he has the lowest use of 'love' compared to every other poet. Almost equally low is the use of 'love' in Lapathiotis, while 'heart' has the highest frequency. One could discern in these two gay poets a reluctance to speak of love *per se*, since theirs is 'a love that dare not speak its' name' as Wilde famously put it. Except for Giofyllis, whose work has a greater thematic variety that might account for a low use of 'love', the rest of the poets, i.e. Filyras, Ouranis, Karyotakis, and Agras, use the two words with about the same frequency - and there is also little difference amongst them.

When it comes to 'soul' (ψυχή, in fuchsia), Ouranis, Filyras and Giofyllis score a little higher than the rest, but not in a particularly distinctive manner compared to the group. However, Filyras, a poet with a serious mental health condition, uses 'soul' far more frequently than any of the other three words. Nonetheless, 'soul' is the word with the most even distribution across the corpus, which gives it a central role, supporting the hypothesis made initially by its particularly high frequency in the entire corpus. The same could almost be said about 'eyes' (μάτια, in purple), with the exception of Filyras and Lapathiotis that don't use it very much. It is interesting to note that Agras, who was also a fair and astute critic of the group, whose work (the longest in the corpus) expands over his long life and is particularly noted for a lack of center and a sense of incessant oscillation (Filokyprou, 2009: 31–32) scores average on the use of these words compared to the rest. He

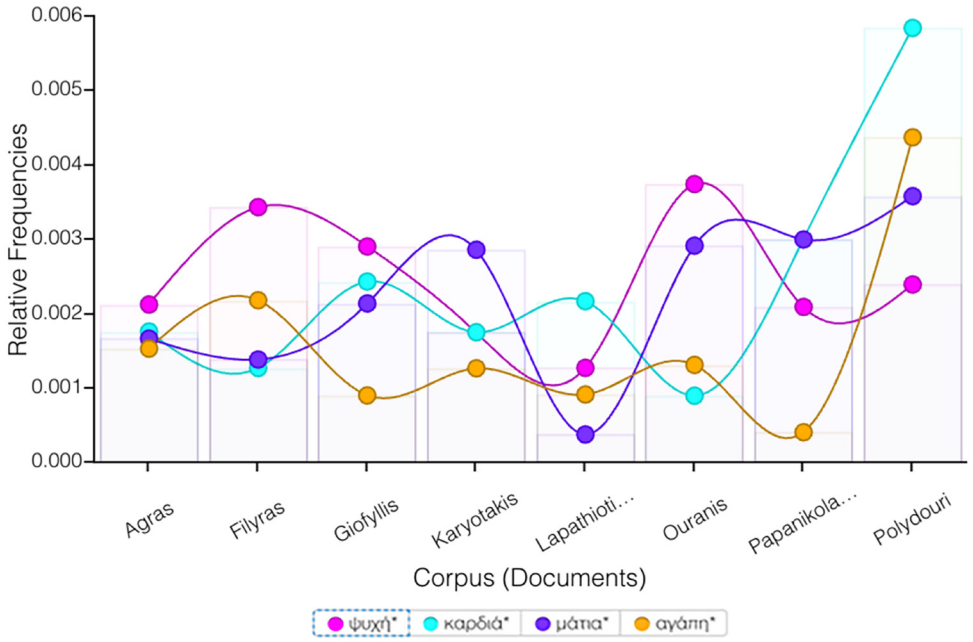


**Figure 1. Highly frequent words: soul, heart, eyes, love.**

doesn't even seem to have a very distinct preference for any one of them – but even so, his most frequently used word of the four (even by a small margin) is 'soul', the most typical for all of them. Karyotakis' use of these words is also rather average compared to the rest, and generally balanced. He makes an exception for 'eyes', which is his own most frequent word of the four (and generally his own most frequent meaningful word, with 41 occurrences); this is a word marked for its subtlety, as well as a certain distance, compared to the rest.

Interestingly, as seen in Figure 2, Karyotakis has also the highest combined frequency for 'seeing' and 'gaze' (βλέπ* in blue, βλέμμα* in green) compared to the rest – Papanikolaou follows suit, as he also does with the word 'eyes' (see Figure 1). This preoccupation with the gaze seems in tune with the symbolist distancing from immediate reality, where the artist, instead of talking about things themselves, talks about the act of observation (Polychronakis, 2016: 318). The contexts reveal that this is to do with the poet's own act of watching as with the others' perception of himself, whereas there is also no little self-reflection.

A graph that I find particularly revealing in terms of comparison is to do with the uses of the words 'I' (εγώ in fuchsia) and 'you' (εσύ, in blue) (Figure 3).

Although these words alone do not account for all instances of an I-thou relation and theme, they are fairly indicative. In almost all the poets, 'I' has a prevalence, but it is closely followed by 'you'. This is particularly the case of Polydouri with her focus on interpersonal relations. On the other end we find Papanikolaou, a poet moving away from lyricism, with practically null use of these words. Karyotakis stands out, as he has the most divergent use of the two, with a clear preference for 'I'. The second person seems to be rarely acknowledged in his corpus, accounting for a sense of loneliness and despair in his work. A similar pattern with a comparable feeling is also to be found to some extent in Lapathiotis, although far less acute.

Turning to stylistic observations, the use of simile (detectable through the use of 'like'/'as', 'σαν'/'σα', Figure 4) tends to be fairly similar across the corpus, with two notable exceptions: Filyras, the first one to publish a poem amongst these poets (in *Noumas* in 1903), displays the greatest use of it, whereas Papanikolaou, who became gradually an admirer and translator of modernist and avant-guard poetry (Bakogiannis, 2024: 51, 55–58), the least. Even allowing for some cases of alternative meanings of these words (such as 'when' and 'if'), which are secondary and rather infrequent, the difference is still significant. The simile is a trope used more broadly in older poetry, steadily giving way to metaphor, symbol and other expressive tropes. Filyras and Papanikolaou seem to hold the two ends of the spectrum that this poetry occupies between traditional tropes and modernist experimentations. Interestingly, Karyotakis scores average for both words. Unfortunately, Voyant Tools have a limitation when it comes to the detection of other major tropes in this poetry, such as irony and allusion.

Regarding vocabulary density, i.e. the variety of words used (Figure 5), Filyras, noted for his use of "sometimes irrelevant, and often ridiculous words" (Agras, 1942: 955), scores the

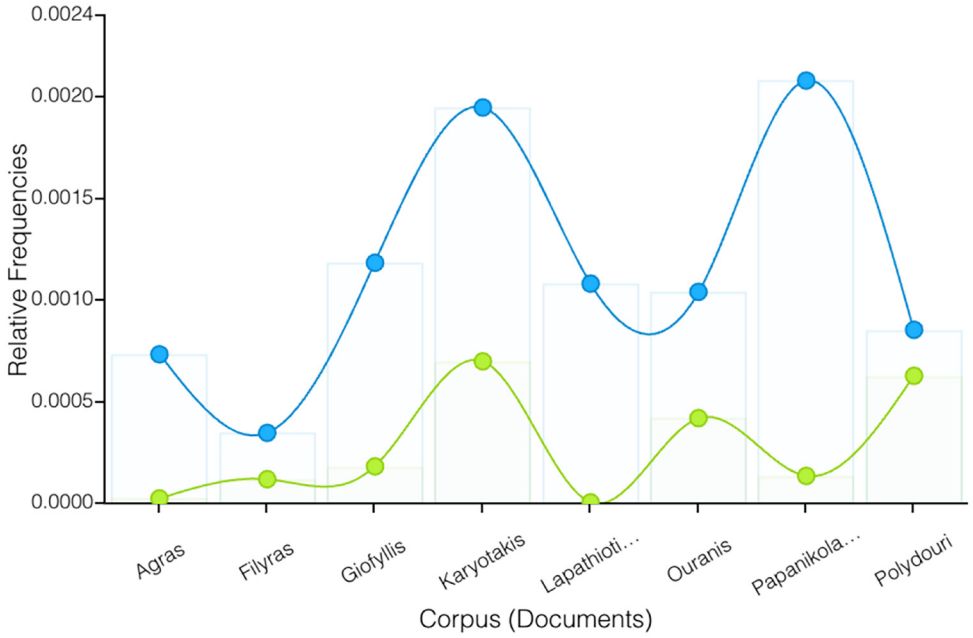


**Figure 2.** **Seeing and gaze.**

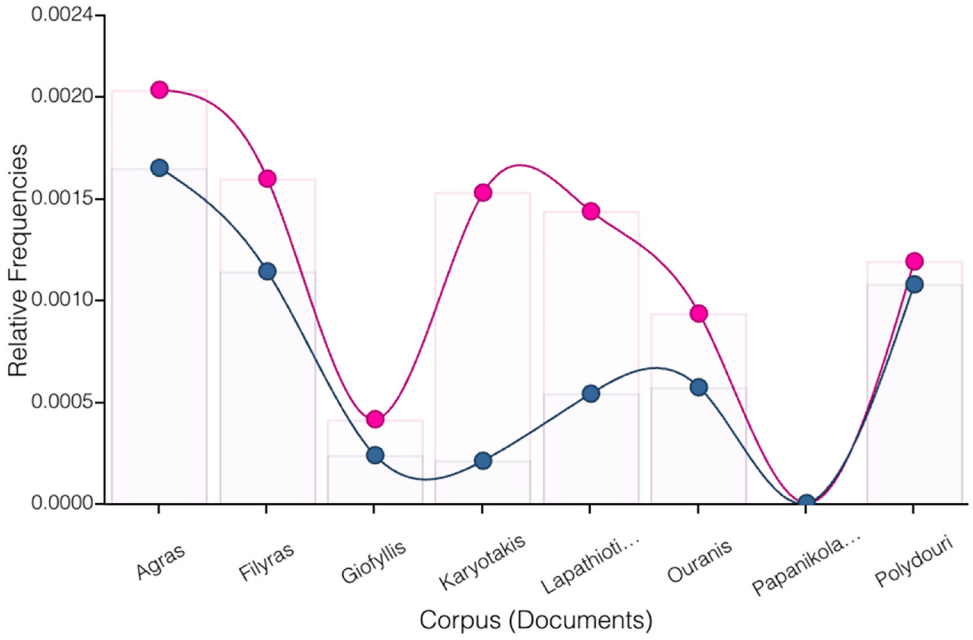


**Figure 3.** **Uses of 'I' and 'you'.**

highest; Agras, far more self-restrained, occupies the other end, demonstrating the least variety, whereas Giofylis and Karyotakis occupy the middle ground.

Finally, as regards sentence length, Ouranis scores the highest with 33,4 average words per sentence, whereas Karyotakis the lowest with 14,2.

Excepting the short sentence length, which accounts for greater immediacy and a rather modern style, and the divergent use of the words 'I' and 'you', which show the poet mostly alone and self-centered, on most other counts Karyotakis seems to occupy a middle ground among these poets. Voyant Tools by no means capture the depth of his poetry, much less explain the enduring appeal of his work, but they do give us a glimpse of

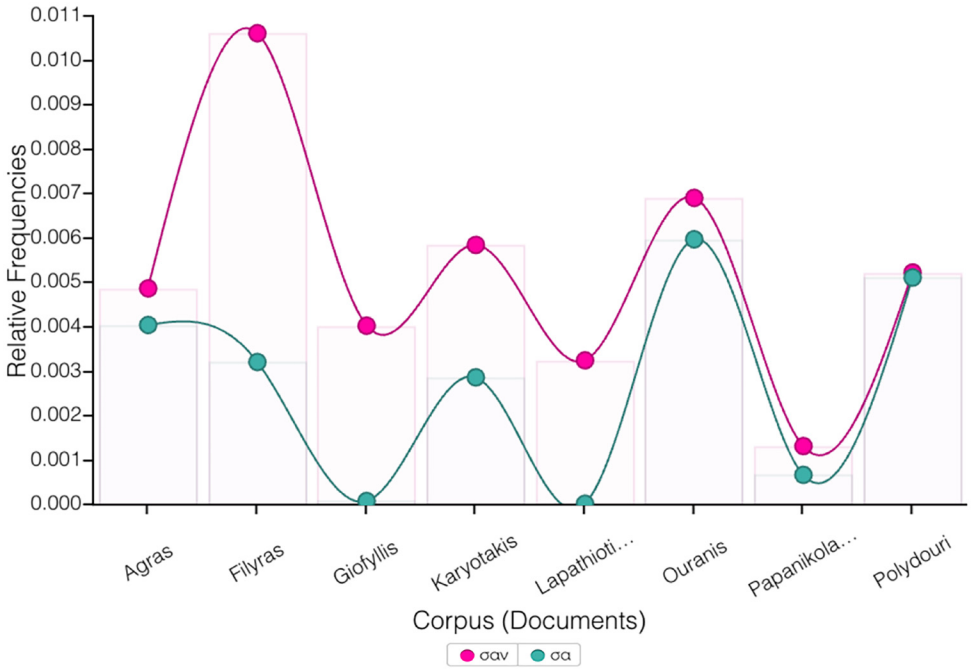


**Figure 4. Exploring the use of simile.**




**Figure 5. Vocabulary density and Average words per sentence.**

both his sides as representative and exceptional at the same time.

## Conclusion

Although this one remains an indicative exploration of this corpus, it does bring out some basic themes and the mood of these poems. It indicates their lyrical character, showing them to speak rather directly of their issues of concern (the heart and soul of the poets), but creating a melancholic mood through a lack, negation or qualification of words with typically positive connotations (such as joy, light, summer). Nature is present, particularly through the image of the sea, with rich connotations of escape. However, small-scale urban and the enclosed spaces of the room and the house are the most usual setting. Poetry is not a high art but a lyrical expression, a song that poets sing or scribble away. Finally, when peeking through the results further, one can discern indications for the balanced expression of Agras, the aloofness of Ouranis, the warmth of Polydouri, the thematic variety of Giofyllis, the loneliness of Lapathiotis, the modernism of Papanikolaou, the representative and exceptional sides of Karyotakis.

## Ethics and consent

Ethics and consent were not required.

## Data availability statement

**Natsina/Modern_Greek_Literature: v1**

https://doi.org/10.5281/zenodo.12705940 (StergiosCha & Natsina, 2024).

This project contains following datasets:

StergiosChatzikyriakidis/Modern_Greek_Literature-v1.zip

Data are available under license Creative Commons Attribution 4.0 International

## References


Agras T: **Ρώμος Φιλύρας.** *Νέα Εστία.* 1942; **32**: 955.

Al-Rawi MKS: **Using AntConc: a corpus-based tool to investigate and analyse the keywords in Dickens' novel *A Tale of Two Cities*.** *Int J Adv Res.* 2017; **5**(2): 366–372.
**Publisher Full Text**

Aragis G: **Νεορομαντικοί και νεοσυμβολιστές.** *Μικροφιλολογικά.* 2009; (26): 30–34.
**Reference Source**

Bakogiannis M: **Εισαγωγή στην ποίηση του Μήτσου Παπανικολάου.** In: M. Papanikolaou, *Τα ποιήματα 1916-1942.* (edited with an introduction by M. Bakogiannis), Athens: Smili, 2024; 15–60.

Beaton R: **An introduction to modern Greek literature.** Oxford: Clarendon Press, 1994.
**Reference Source**

Claude: **[Personal communication through Claude AI assistant].** 2025.

Filokyprou E: **Η γενιά του Καρυωτάκη. Φεύγοντας τη μάστιγα του λόγου.** Athens: Nefeli, 2009.
**Reference Source**

Jockers M: **Macroanalysis.** *Digital Methods and Literary History.* University of Illinois Press, 2013.
**Reference Source**

Moretti F: **Distant Reading.** London: Verso, 2013.
**Reference Source**

Papaleontiou L: **Μετασυμβολισμός και μετασυμβολιστές.** *Μικροφιλολογικά.* 2009; (26): 35–39.
**Reference Source**

Polychronakis D: **"Η ώρα των ποιητών": Το *Φθινόπωρο* του Κωνσταντίνου Χατζόπουλου και η φθινοπωρινή αισθητική του συμβολισμού.** In: K. Chatzopoulos, *Φθινόπωρο.* Herakleion: P.E.K., 2016; 275–463.

Sinclair S, Rockwell G: **Voyant tools.** 2016.
**Reference Source**

Stergiopoulos K: **Εισαγωγή. *Η Ελληνική Ποίηση. Η ανανεωμένη παράδοση, ανθολογία-γραμματολογία* (επιμ. Κ. Στεργιόπουλος). v.3.** Athens: Sokolis, 1980.

StergiosCha, Natsina A: **Natsina/Modern_Greek_Literature: v1.** [Data set]. Zenodo. 2024.
**http://www.doi.org/10.5281/zenodo.12705940**

Vitti M: **Ιστορία της νεοελληνικής λογοτεχνίας.** Athens: Odysseus, 2003.

Wan Z: **Corpus based analysis of literary works: a thousand splendid Suns.** *Advances in Social Science, Education and Humanities Research.* 2022; **673**: 126–131.
**Publisher Full Text**

# Open Peer Review

## Current Peer Review Status: ? ? ?

**Version 1**

Reviewer Report 22 January 2026

https://doi.org/10.21956/openreseurope.21587.r67237

© **2026 Hamat A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Afendi Hamat**
National University of Malaysia, Bangi, Malaysia

This article demonstrates how Voyant Tools can support a macroscopic/distant reading of a corpus of Greek interwar poetry, using frequency, context/collocates, and selected readability/stylistic indicators to comment on shared motifs and differences among poets. The piece is accessible and pedagogically useful, and the choice of corpus/topic is of genuine interest for readers who work on Greek modernism and DH.

**Strengths**

1. Clear motivation to combine distant reading with parallel close reading ("avoid possible distant reading mistakes"), which is a sensible stance for literary corpora.

2. Tool choice is appropriate for a low-barrier DH workflow, and the discussion is generally readable.

3. Data are at least partly traceable via the dataset/corpus links provided.

**Points that should be addressed to make the article scientifically sound (and easier to trust/replicate)**

1. Research aim/RQ needs sharpening (currently reads mostly as a demonstration).

Where: "This paper explores the possibilities…for the distant/macroscopic reading… and applies them…" (Introduction).

Suggested Fix: Add 1–2 explicit research questions (or clearly framed aims). Then structure Results/Discussion so each subsection answers them.

2. Corpus construction + preprocessing are under-described (replication gap).

Where: “The slightly modified and richer corpus used here…” (Method).

Suggested Fix: Specify what was modified (selection rules, cleaning/normalisation, document unit, metadata). A short corpus/preprocessing subsection or appendix is enough.

3. Stopword handling and Voyant settings should be reported explicitly.

Where: “Excluding stopwords – such as articles and pronouns…” (Results and discussion).

Suggested Fix: State which stopword list was used (default/custom; Modern Greek-specific), and report key Voyant parameters that affect outputs (tokenisation assumptions, collocate window, etc.).

4. LLM use (Claude) should be made transparent or removed from the evidential chain.

Where: “supported by the LLM Claude (2025) after exporting the data…” (discussion of “hands”).

Suggested Fix: If kept, describe what was exported, the prompt(s), and what Claude contributed (summary only vs interpretive inference). Otherwise, treat it as author interpretation supported by concordance evidence. If problematic, do not include.

5. Several interpretive moves go beyond what frequency data alone can support.

Where: “One could discern in these two gay poets a reluctance to speak of love…” (love/heart discussion). Also: “Filyras…with a serious mental health condition…” linked to high “soul” frequency.

Suggested Fix: Recast these as tentative hypotheses and/or add a small amount of textual evidence (a handful of concordance lines/collocate patterns) plus citations where biography is invoked.

Conclusion
The manuscript is promising as a DH-oriented contribution and could be valuable on ORE once the above methodological and evidential points are tightened. In its current form, I would recommend revisions focused on (i) explicit RQs/aims, (ii) corpus/preprocessing transparency, (iii) reporting of Voyant settings/stopwords, and (iv) moderating or evidencing stronger interpretive claims.

**Is the work original in terms of material and argument?**
Partly

**Does it sufficiently engage with relevant methodologies and secondary literature on the topic?**
Partly

**Is the work clearly and cogently presented?**
Partly

**Is the argument persuasive and supported by evidence?**

Partly

**If any, are all the source data and materials underlying the results available?**
Partly

**Does the research article contribute to the cultural, historical, social understanding of the field?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Digital Humanities, Corpus Linguistics and NLP.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 30 December 2025

https://doi.org/10.21956/openreseurope.21587.r65072

**© 2025 Mohammed Alyasiri O.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Osamah Mohammed Alyasiri**
Information Technology Engineering Department, Polytechnic College of Karbala, Al-Furat Al-Awsat Technical University, Karbala, Iraq

1. The paper should more explicitly document corpus construction, preprocessing steps, stopword handling, and Voyant parameter settings to ensure reproducibility.
2. A brief discussion of challenges specific to Modern Greek morphology and polysemy, and the implications for frequency-based analysis, would strengthen the methodological contribution.
3. Some interpretive claims (e.g., negation of positive concepts, pronoun asymmetry as evidence of loneliness) should be slightly moderated or supported with brief illustrative examples.
4. The role of the LLM (Claude) should be clarified, minimized, or moved to a footnote to avoid ambiguity regarding its analytical status.

**Is the work original in terms of material and argument?**
Yes

**Does it sufficiently engage with relevant methodologies and secondary literature on the topic?**
Yes

**Is the work clearly and cogently presented?**

Yes

**Is the argument persuasive and supported by evidence?**
Partly

**If any, are all the source data and materials underlying the results available?**
Yes

**Does the research article contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** The research interests include Artificial Intelligence, Knowledge Engineering, Data Mining, NLP, Pattern Recognition, Text Mining, Text Classification, Information Retrieval, Machine Learning, Feature Selection, Feature Extraction, Optimization, and AI Chatbots.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 11 December 2025

https://doi.org/10.21956/openreseurope.21587.r64467

**© 2025 Martins H.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Hélder Fanha Martins**
Polytechnic University of Lisbon, Lisbon, Portugal

Report

**Summary of the article**
The article **"A macroscopic view of Greek interwar poetry through Voyant tools"** investigates a corpus of over 600 poems by eight Greek interwar poets (Agras, Giofyllis, Filyras, Karyotakis, Lapathiotis, Ouranis, Papanikolaou, Polydouri) using the web-based text-analysis environment Voyant Tools. The stated aim is two-fold: (1) to verify and qualify existing close-reading criticism on this "post-symbolist" group from a quantitative perspective, and (2) to uncover new insights into both group traits and individual differences.
After a brief introduction to the concept of macroscopic/distant reading and to Voyant itself, the Method section outlines which features of Voyant are used: word frequencies (WF), collocates and contexts (CA), visualisations of trends, vocabulary density (VD), and average words per sentence (AWpS). The corpus derives from an open dataset created in a seminar at the University of Crete,

with Karyotakis' work sourced from CLARIN:EL; a link to a prepared Voyant corpus is also provided.
The Results and discussion section first focuses on the group as a whole. Excluding stopwords, the most frequent content words are "now", "soul", "eyes", "heart", "life", "hands", "love", suggesting a strongly lyrical focus on the self and affect. The author then explores mood via semantic fields and collocational patterns: "light" and "joy" are relatively frequent but typically appear in contexts of negation, loss or diminishment, whereas "darkness", "death", "sorrow", "melancholy" are less frequent individually but cumulatively substantial. Seasonal terms show autumn and winter overtaking spring and summer both in frequency and in their association with waning, death and sadness. Weather and time-of-day vocabulary (rain, sun, night vs day) are analysed similarly, reinforcing an overall melancholic atmosphere and a preference for night settings.
The study then turns to themes such as nature vs urban space and poetry itself. The author notes the relatively modest frequency of the explicit word "nature" but higher counts of sea, mountain, gardens, trees, etc., with the sea in particular linked to longing and escape. Urban environment is evoked through frequent references to streets, houses, rooms and windows, suggesting small-scale urban settings and enclosed domestic interiors rather than a modern metropolis. Poetry appears less as abstract "Art" than as the concrete activity of poets writing and singing "songs", highlighting lyrical expression over "high art".
In the final part, the author uses relative frequencies, VD, and AWpS to discuss differences between individual poets, visualised in Figures 1–5 (pages 7–8). Polydouri shows the highest relative frequency of "heart" and "love", which matches her reputation for intense emotive focus, whereas Ouranis uses these terms relatively little. Filyras, whose mental health struggles are well known, stands out for his especially frequent use of "soul" and of simile ("σαν/σα"), while Papanikolaou has minimal use of simile but higher VD and lower reliance on traditional tropes, consistent with his modernist tendencies. Karyotakis, the central figure of the group, appears "average" for many metrics, but shows a marked emphasis on eyes/seeing/gaze and a strong asymmetry between "I" and "you", which the author interprets as pointing to self-centredness and loneliness.
The Conclusion stresses that these are indicative findings but argues that they provide quantitative support for established critical views of the interwar/post-symbolist poets, while also suggesting more nuanced stylistic and thematic differentiations within the group.

**Major comments**

**Clarify corpus construction and preprocessing in more technical detail**

The article would benefit from a more explicit Method subsection that systematically details:
- Exact corpus composition: number of poems per poet; time span covered; whether complete collections are used or selected poems; criteria for inclusion/exclusion (e.g. occasional pieces, prose poems, fragments).
- Source editions: for each poet, specify which edition(s) or digital repositories were used, especially since Karyotakis' work comes from CLARIN:EL while the others derive from the seminar dataset.
- Pre-processing steps: how texts were cleaned; treatment of punctuation; handling of diacritics and spelling variation; whether any lemmatisation or stemming beyond the use of wildcards ("*") was applied.
- Stopwords: which stopword list was used for Modern Greek; whether it was adapted or custom-made; how pronouns (e.g. μου, σου) were handled when they function as both core vocabulary for lyric address and generic grammatical items.
Because the article's claims hinge on frequency and collocational patterns, these details are essential for reproducibility and for readers to fully assess the robustness of the findings.

**Specify Voyant Tools settings and collocation parameters**
The discussion mentions several Voyant features (WF, CA, VD, AWpS, visualisations) but does not fully specify the settings used. For example:
- What was the context window (in words or characters) for collocates?
- How were multi-word expressions treated?
- Were frequencies per poet normalised by token count or document count (poem as document?), and how exactly were "relative frequencies" computed for Figures 1–5?
- Were any stopword overrides or custom vocabularies employed in Voyant?
A short, explicit description (possibly with a brief table or appendix) would make the analyses more transparent and easier to replicate, in line with the journal's emphasis on methodological clarity and open data.
**Strengthen the methodological reflection on quantitative reading for Modern Greek**
The introduction gestures to the contrast between close reading and macroscopic/distant approaches, but the methodological reflection could go slightly deeper, especially given the linguistic specificity of Modern Greek:
- Briefly address the challenges that inflected morphology and rich derivation pose for word-frequency analysis and why the chosen approach (wildcards like ψυχ*, καρδιά* etc.) is adequate, while also acknowledging its limitations (e.g. inclusion of marginally related forms).
- Reflect on possible ambiguities in the data (e.g. homographs, multiple senses of words like φως, χαρά) and how they were handled or at least recognised.
- You might add a sentence or two on why Voyant (as opposed to other corpus tools) is particularly suitable for this corpus and this audience (e.g. ease of parallel close/distant reading, which you already hint at).
This would not require a major expansion but would signal a more sustained engagement with digital-humanities methodology.
**Moderate or further substantiate some interpretative claims**
Much of the interpretative work is thoughtful and well argued, but a few places would benefit from either slightly more evidence or more cautious phrasing:
- The assertion that entities with positive symbolic value (light, joy, spring, summer) are "repeatedly evoked only to be frustrated" is persuasive but strong. It would help to either provide a few representative Greek phrases (with translations) or to acknowledge explicitly that the reading is based on a substantial, though not exhaustive, sampling of contexts.
- Similarly, the conclusion that Karyotakis' divergence between "I" and "you" "shows the poet mostly alone and self-centered" could be nuanced by a short sentence noting that this is one possible interpretation of the pronoun distribution, which aligns with critical reception but does not, by itself, exhaust the complexity of his lyric "I".
- When linking autumn/winter and night to symbolist aesthetics and decadence, you might briefly bridge the gap between general symbolic associations and the specific quantitative patterns you observe in this corpus.
These are largely matters of tone, but they will make the argument feel even more rigorously grounded in the data.
**Clarify (or reconsider) the role of the LLM (Claude) in the analysis**
The article currently notes, in passing, that the author's reading of the collocational patterns for "hands" is "also one supported by the LLM Claude (2025) after exporting the data concerned in text format".
Given that this is a novel and potentially controversial methodological detail:
- Either provide a bit more information about how Claude was used (e.g. type of prompt; whether it was employed as a heuristic check rather than a decisive analytic tool), or

- Consider simplifying the reference so that the primary basis of interpretation is clearly your own reading of the contexts, with the LLM mentioned only in a footnote or omitted altogether.
As it stands, the single sentence mentioning Claude raises interesting questions but might distract readers who are not yet familiar with LLM-assisted text analysis.

**Minor factual/terminological clarification**

In the Introduction, there appears to be at least one slip where “postwar Greek poetry” is mentioned in a sentence that otherwise refers to the interwar poets; it would be helpful to check this for consistency and correct if it is a typo.

Minor comments and suggestions

These are not essential for scientific validity, but addressing them would improve clarity and readability.

1. Provide a brief contextual paragraph for non-specialists

Since the journal has a broad readership, a short paragraph early in the Introduction situating the eight poets (approximate dates, their place within Greek literary history, the political context of the interwar period) would help readers unfamiliar with Modern Greek literature.

2. Tables/appendices for frequencies

Consider adding a small table (either in the main text or as supplementary material) summarising the key frequencies discussed (e.g. for “soul”, “heart”, “eyes”, “love”, pronouns, seasons, night/day, simile markers) for each poet. This would complement Figures 1–5 and make the quantitative backbone of the argument more visible.

3. Terminology consistency

Ensure consistent use of key terms: “distant reading” vs “distance reading”, “macroscopic” vs “macroanalysis”, “post-symbolism” vs “post-symbolists”. A quick pass should resolve these small inconsistencies.

4. Stylistic and typographical polish

A light language edit would clear up occasional minor grammatical issues and tighten some long sentences, but overall the English is very good and the text is pleasant to read.

**Concluding assessment**

Overall, this is a coherent, clearly written and valuable study that successfully demonstrates how an open-access tool like Voyant can be used to explore a significant corpus of Greek interwar poetry, both confirming and nuancing existing close-reading scholarship. The work is solidly grounded in the literary tradition, uses openly available data and tools, and makes a meaningful contribution to the cultural and historical understanding of Modern Greek literature.

The revisions I recommend are, in my view, feasible and would significantly strengthen the article’s transparency and persuasive power, thereby enhancing its usefulness both to specialists in Greek literature and to scholars interested in digital approaches to poetry.

**Is the work original in terms of material and argument?**

Yes

**Does it sufficiently engage with relevant methodologies and secondary literature on the topic?**

Partly

**Is the work clearly and cogently presented?**

Yes

**Is the argument persuasive and supported by evidence?**
Partly

**If any, are all the source data and materials underlying the results available?**
Yes

**Does the research article contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Communication and Information sciences; Translations studies; Literature and Linguistics; Simulation Studies

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**