

Journal of Integrated Information Management

Vol 8, No 2 (2023)

Jul-Dec 2023



Lemmatization and sentiment analysis of Greek political tweets during the pre-election period of 2023

Elisavet Chantavaridou, Sarantos Kapidakis

Copyright © 2023, Elisavet Chantavaridou, Sarantos Kapidakis



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

To cite this article:

Chantavaridou, E., & Kapidakis, S. (2023). Lemmatization and sentiment analysis of Greek political tweets during the pre-election period of 2023. *Journal of Integrated Information Management*, 8(2), 37–43. Retrieved from <https://ejournals.epublishing.ekt.gr/index.php/jiim/article/view/37820>

Lemmatization and sentiment analysis of Greek political tweets during the pre-election period of 2023

Elisavet Chantavaridou^{1,2}, Sarantos Kapidakis¹

¹Department of Archival, Library & Information Studies, University of West Attica, Athens, Greece

²Library, University of Macedonia, Thessaloniki, Greece

echantavaridou@uniwa.gr [ORCID: 0000-0002-0614-1213], sarantos@uniwa.gr [ORCID: 0000-0002-8723-0276]

□

Article Info

Article history:

Received 28 November 2023

Received in revised form 12 December 2023

Accepted 21 December 2023

<http://dx.doi.org/10.26265/jiim.v8i2.37820>

Abstract:

Purpose – The present work discusses findings of research on Twitter (now X) data of the two politicians running for Prime Minister, as well as their press secretaries, during the pre-election period of May 2023 in Greece.

Design/methodology/approach – We collected the tweets (posts on Twitter/X) posted by the two main candidates running for Prime Minister in Greece, as well as their press secretaries, during the pre-election period of May 2023. The four sets of tweets were lemmatized, and sentiment analysis was performed using SentiStrength for Greek, a sentiment dictionary developed for Greek political short text/ tweets.

Findings – Results revealed the importance of positive sentiment when posting on Twitter; they also revealed the different approaches of mentioning people vs. locations of the two politicians who were running for Prime Minister.

Originality/value – Writing positive tweets during a pre-election period can lead to victory in Greece. Focusing excessively on the opponent and other political figures is not a way to win voters in Greece in 2023. Instead, making references to geographic locations in Greece is preferred. Furthermore, SentiStrength for Greek is used for the first time.

Index Terms — political tweets, lemmatization, Greek language, data visualization, sentiment analysis.

I. INTRODUCTION

On 21st May 2023, national elections were held in Greece. Although the pre-election period had started much sooner unofficially, the official date was 22nd April 2023. On this date, the Prime Minister of Greece, Kyriakos Mitsotakis, visited the President of the Republic and requested the proclamation of elections [1]. The two political figures that we focused on are Kyriakos Mitsotakis and Alexis Tsipras. Both of them were running for Prime Minister and were the main leading candidates. During the pre-election period, Kyriakos Mitsotakis was running as the then Prime Minister

of Greece (2019-2023) [2] and as the President of the Nea Demokratia Party. Alexis Tsipras was running as the President of the SYRIZA Progressive Alliance Party, as the former leader of the main opposition during 2019-2023 and as a former Prime Minister (2015-2019) [3]. Nea Demokratia won the elections in May, collecting 40.79% of the total votes, while SYRIZA came second with 20.07% of the total votes [4]. Besides the two main primary opponents, our research examined two more political figures and their tweets; those of the press secretaries of each party. Popi Tsapanidou was the press secretary of SYRIZA, and Akis Skertsos was the press secretary of Nea Demokratia at the time [5]. The addition was thought in case the press secretaries were using different (i.e. harsher) language than their party leaders. The four Twitter usernames/ accounts that we focused on were @askertsos, @atsipras, @kmitsotakis and @tsapanidou. The period of examination was from 22nd April 2023 until 21st May 2023 19:00 when voting stopped. With the help of lemmatization, a process that converts each word into its dictionary-based form, otherwise known as lemma [6], we aimed to compare and research the language used by Kyriakos Mitsotakis and Alexis Tsipras in order to understand the preferences of the voters given the outcome of the 21st May 2023 elections. Our research aimed to extract information from these political texts from the scope of view of information science, not to conduct political analysis.

II. METHODOLOGY

A. Twitter posts and their pre-processing

To compose our corpus of tweets we used our Academic Researcher status [7] and collected tweets via Twitter API v2. During the specific timeframe that we set (22/4-21/5) Kyriakos Mitsotakis posted 139 tweets, Alexis Tsipras posted 185 tweets, Popi Tsapanidou posted 6 tweets and Akis Skertsos posted 6 tweets.

Before proceeding with lemmatization, we excluded two tweets from the corpus. One from Alexis Tsipras because the tweet consisted only of a URL, and the other was from Kyriakos Mitsotakis because the tweet had been written in English and therefore did not serve our purpose of Greek

□□□

language lemmatization. As for the tweets of the two press secretaries, they were so few that no conclusions could be drawn. However, we will include them in our data findings below to satisfy one's curiosity.

The table below (Table 1) shows the final number of tweets included in the corpus.

Table 1 Number of tweets examined

Politician	No of tweets
Akis Skertsos	6
Alexis Tsipras	184
Kyriakos Mitsotakis	138
Popi Tsapanidou	6

B. Lemmatization

After pre-processing the four sets of tweets, we used a lemmatization service provided by *clarin:el*, the central inventory of language resources and services in Greece. The service is called *ILSP Lemmatizer* and is a language-processing tool for Modern Greek texts [8]. One has to create an account first and then run the service by uploading zip files that contain text/plain files. The output is two files, one in .csv and one in .xmi format. *ILSP Lemmatizer* works by segmenting text into paragraphs, sentences and tokens; it then annotates tokens using a Part of Speech (POS) tagger (e.g. *Ad* for adverbs, *PtNg* for negative particles) [9]. Past research on lemmatized Twitter data has focused on user tweets during the World Cup Soccer tournament [10].

In order to process the lemmatized data we used Excel spreadsheets and counted the weight/frequency of each lemma in each set of tweets. What the lemmatizer does is return the lemma of each word, for example the word-forms *running*, *ran* and *run*, all return the lemma *run* (verb form in first person). In the case of pronouns, nouns, adjectives, etc., the token returns lemmas in singular number only (not plural), masculine gender and in nominative case (ονομαστική). For example, the phrase «οι εκλογές» (the elections) consists of two tokens: *οι*, *εκλογές*. Once lemmatized, *οι* becomes *ο* and *εκλογές* becomes *εκλογή* (*election*). Lemmatization is a good way of discovering words that often come up in a text or classifying verbs, adverbs, etc., of a text. Before moving on to lemma weights we performed one extra step, excluding stop words. Stop words appear too frequently in texts but carry little importance in information retrieval [11] or, in this case, political meaning. Based on previous work [11], we expanded the list of stop words. We eliminated numbers, quotation marks, punctuation marks, URLs and the following lemmas in Greek followed by their English translation: *ο* (the), *σου* (in the), *και* (and), *με* (with), *από* (from), *σε* (to), *να* (to), *είμαι* (am), *εγώ* (I), *για* (for), *πως* (how), *που* (that), *ότι* (that), *μου* (my), *τι* (what), *πώς* (how), *ή* (or), *αυτός* (he), *ένας* (someone), *έχω* (have), *ως* (as), *πιο* (more), *μέσα* (in), *κ.* (mr., mrs.). As described earlier, lemmatization returns lemmas in the first person, so this cannot be identified as self-referencing. Auxiliary verbs like *έχω* were used as stop words since they

assist in forming past tenses or passive voice.

The remaining lemmas that each politician used in their tweets are intriguing in many ways. First, they show the focus of each politician in terms of politics, and secondly, they show the different methodologies of each politician. Our findings show that Alexis Tsipras focused greatly on Kyriakos Mitsotakis and what the latter said/ did. Alexis Tsipras referred to many people in his tweets whether they belonged to his party or other parties, whether alive or deceased. On the contrary, Kyriakos Mitsotakis did not refer to other people except for very few. What he did instead was mentioned many places and geographic locations. Alexis Tsipras also referred to cities and places, but although his tweets were bigger, he referred to fewer places. The tables below (Tables 2-5) list the people and places mentioned by the two primary opponents. One note that could be made here is that some surnames were lemmatized as nouns during the lemmatization process and therefore were not included in the list of surnames initially. This happened when the surname was identical to a noun (e.g. *Λινού*, *Κεραμέως*); in English, the same thing would happen with surnames like *Forest* or *Baker*. The top 10 most popular lemmas in Tsipras' tweets were: *δεν* (not-116), *θα* (will-109), *Μητσοτάκης* (Mitsotakis-80), *λέγω* (say-72), *γιατί* (why-39), *ΣΥΡΙΖΑ* (syriza-38), *κάνω* (do-36), *μπορώ* (can-35), *θέλω* (want-33) and *αλλαγή* (change-32). The top 10 lemmas in Mitsotakis' tweets were: *θα* (will-72), *Ελλάδα* (Greece-54), *όλος* (whole-45), *μισθός* (wage-38), *υγεία* (health-32), *μεγάλος* (big-32), *πολύς* (a lot-30), *χώρα* (country-29), *καλός* (good-28) and *σήμερα* (today-27).

Table 2 Names of people mentioned by Alexis Tsipras, their frequency and their capacity

SURNAME	FREQUENCY	POLITICAL PARTY / CAPACITY
Mitsotakis	80	Nea Demokratia
Maximou	9	PM Office
Varoufakis	8	MERA25
Karamanlis	7	Nea Demokratia
Androulakis	5	PASOK
Skertsos	5	Nea Demokratia
Georgiadis	4	Nea Demokratia
Kerameos	2	Nea Demokratia
Linou	2	SYRIZA
Papandreou	2	PASOK
Schäuble	2	German politician
Scholz	2	Chancellor
Vlachou	2	Prosecutor
Chatzidakis	1	Nea Demokratia
Gennimata	1	PASOK
Gerotziapas	1	SYRIZA
Glezos	1	SYRIZA

Iliopoulos	1	SYRIZA
Juan Ramón Rocha	1	Football player
Katrougkalos	1	SYRIZA
Katseli	1	PASOK
Koutsoumpas	1	KKE
Nikiforos Fokas	1	Ship
Nikoltsiou	1	Journalist
Oruç Reis	1	Ship
Patsis	1	Nea Demokratia
Plevris	1	Nea Demokratia
PM	1	Nea Demokratia
Pretenteris	1	Journalist
Rammos	1	Judge/President
Santorinios	1	SYRIZA
Savvopoulos	1	Composer
Staikouras	1	Nea Demokratia
Theodorakis	1	Random
Trump	1	US President
Tsakres	1	SYRIZA
Tzima	1	Journalist
Voridis	1	Nea Demokratia

Table 3 Names of people mentioned by Kyriakos Mitsotakis, their frequency and their capacity

SURNAME	FREQUENCY	POLITICAL PARTY / CAPACITY
Papastergiou	1	Mayor
Arnaoutoglou	1	TV Presenter
Beleris	1	Albanian Politician
Karamanlis	1	Nea Demokratia
Gennimata	1	PASOK

Table 4 Geographic locations mentioned by Kyriakos Mitsotakis and their frequencies

GEOGRAPHIC LOCATION	FREQUENCY
Greece	54
Europe	8
Evros	3
Elefsina	3
Thessaloniki	3
Crete	3
Moria	3
Attica	2
Mytilene	2
Perama	2
Skaramagas	2

Chalkida	2
Athens	1
Aegean (sea)	1
Albania	1
Arta	1
Volos	1
France	1
Zografou	1
Heraklion	1
Thebes	1
Thissio	1
Thrace	1
Kastoria	1
Kypseli	1
Lavrio	1
Lesbos	1
Livadeia	1
Makedonia (region)	1
Nikaia	1
Xanthi	1
Patras	1
Piraeus	1
Rethymno	1
Rhodes	1
Schimatari	1
Chania	1
Himara	1
Hersonissos-Neapoli	1

Table 5 Geographic locations mentioned by Alexis Tsipras and their frequencies

GEOGRAPHIC LOCATION	FREQUENCY
Tempi	14
Greece	11
Europe	6
Athens	3
Thessaloniki	3
Crete	3
Serres	3
Turkey	3
Giannena	2
Igoumenitsa	2
Agrinio	1
America	1
Bulgaria	1
Thissio	1
London	1
Patras	1
Pontos	1

Trikala	1
Ymittos	1
Chalkidiki	1

III. DATA VISUALIZATION

Long tables of data that include words and numbers can be wearing on the eyes. Several tools can help visualize data and make them more presentable and more accessible to comprehend. Such tools are maps. Datawrapper¹ is a website that allows one to create maps with added markers based on their data. We used Datawrapper to illustrate the locations mentioned by Mitsotakis and Tsipras in their tweets (fig.1-2), as well as the frequency of name references by Alexis Tsipras (fig.7). The spread of the locations shows that Mitsotakis focused not only on Athens but on more remote locations as well (e.g. Mytilene, Xanthi). This seems reasonable as national elections were at stake. Mitsotakis used the word Greece 54 times and the word Europe only 8. Tsipras, on the other hand, referred to Greece only 11 times and to Europe 6 times. The lemma Tempis also appeared a lot (14 times) in Tsipras' tweets. It is the location of a terrible train accident in 2023 and put much pressure on the government. Mitsotakis did not refer to Tempis. The frequency of references to Europe and Greece from a left-wing and a right-wing politician should be studied in terms of political meaning, but this is not the purpose of the current research.

Word clouds are tools that visualize words with bigger fonts if they appear more times in a text. We used the services of WordArt.com² in order to create the word clouds from the most popular lemmas that appeared in the tweets after the elimination of stop words. Figures 3-6 illustrate the most frequently used lemmas in the tweets of the four politicians. Figure 7 illustrates the significant number of attention that Alexis Tsipras placed on the face of his opponent, Kyriakos Mitsotakis, whom he mentioned either as Prime Minister or by using his surname or by using the word Maximou (the location of the PM's office). Almost 60% of name references that Tsipras made were about Mitsotakis. Very few references were made to members of SYRIZA, his party, as illustrated in yellow in Fig.7. Blue indicates the names of people who are/have been candidates of Nea Demokratia.

[Spread of geographic locations in Greece mentioned by Kyriakos Mitsotakis]

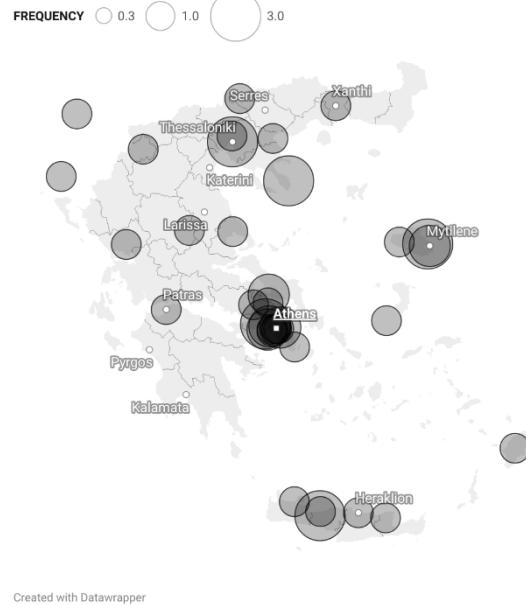


Fig. 1 Spread of geographic locations in Greece mentioned by Kyriakos Mitsotakis

[Spread of geographic locations in Greece mentioned by Alexis Tsipras]

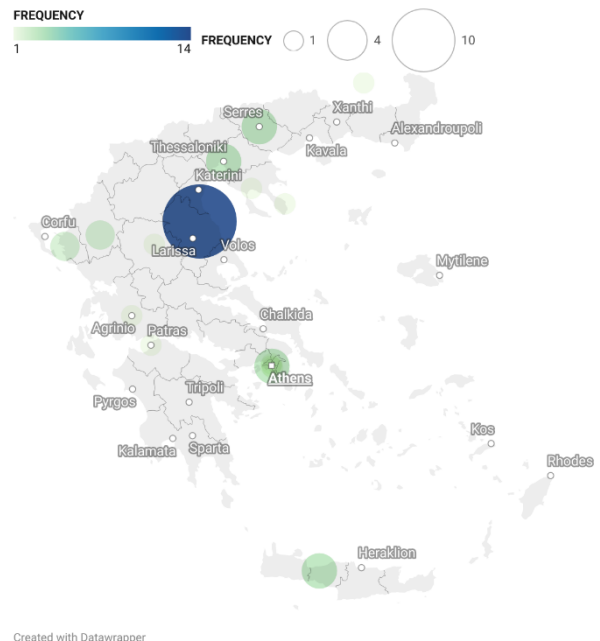


Fig. 2 Spread of geographic locations in Greece mentioned by Alexis Tsipras

1 <https://www.datawrapper.de/>

2 <https://wordart.com/>

(not positive) to 5 (extremely positive). Negative strength is measured from -1 (not negative) to -5 (extremely negative). Table 6 shows that Mitsotakis used more positive sentiment than Tsipras and that Tsipras used more negative sentiment than Mitsotakis. The average numbers for Skertsos and Tsapanidou cannot be considered since their tweets were very few.

Table 6 Average positive and negative sentiment of tweets posted between 22/4/2023 and 21/5/2023

Username	Positive Sentiment	Negative Sentiment
@askertsos	1,88	-1,66
@atsipras	1,64	-1,84
@kmitsotakis	2,02	-1,23
@tsapanidou	2,16	-1,83

V. CONCLUSION

Our research focused on political tweets in the Greek language during a pre-election period in Greece. Twitter, or X now, is used by politicians as a means for campaigning [17] or even developing a peace discourse [18]. We aimed to extract information from text using natural language processing tools, like lemmatizers, and to perform sentiment analysis on tweets in order to evaluate our Greek version of SentiStrength and to discover if positive or negative sentiment prevailed in the elections. Indeed, lemmatization can help information to surface, not only by uncovering the most frequently used lemmas, which can further be analyzed politically, but also by grouping names of people and places. Lemmatization of Greek political tweets seems to be working sufficiently, perhaps due to the fact that political leaders use correct grammar and syntax. Results could be different if lemmatization was run on user-generated content that involves text normalization for abbreviations, spelling mistakes, etc. [19]. *ILSP Lemmatizer* helped us discover that Alexis Tsipras referred to many people from different political parties. His primary focus was his opponent, Kyriakos Mitsotakis, and members of the Nea Demokratia party. It is worth noticing that during the pre-election period Alexis Tsipras did not stress his party's values enough or those of his candidates. Nearly 60% of his references to other people revolved around Kyriakos Mitsotakis. This communication policy did not go well for him, judging from the outcome of the elections. Kyriakos Mitsotakis, on the other hand, hardly mentioned other people. His communication policy included referencing many cities and regions in Greece, which seems reasonable during national elections. Alexis Tsipras referenced Greek cities in his tweets, but on a smaller scale. The conclusion is that personalizing tweets is not a way to win elections—at least in 2023. Approaching voters by referencing their cities is a better communication policy.

One lemmas lemmatization brought to the surface was the negating particle $\delta\epsilon\nu$ (not). This word was used a lot by Alexis Tsipras and hardly at all by Kyriakos Mitsotakis. Sentiment analysis confirms this negative attitude of Alexis Tsipras since his average score of positive sentiment was

smaller than Mitsotakis'. In contrast, his average score of negative sentiment was bigger than Mitsotakis'. SentiStrength's algorithm addresses negating words in a file called *NegatingWordList.txt*. This file exists both in the English and Greek language version. Words in the file reverse the polarity (+, -) of subsequent words, therefore the phrase "not happy" would be classified as negative. Lemmatization also revealed that both politicians used the future particle $\theta\alpha$ (will), which makes sense in a time of seeking votes. Sentiment analysis was performed using our Greek version of SentiStrength and this corpus of tweets helped us to evaluate our work positively. Another conclusion is that national elections in Greece are leader-dependent. One person carries the weight of the whole campaign, which can explain why the press secretaries of the two parties did not tweet enough.

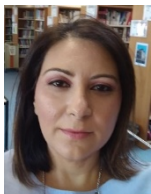
We believe that our research and conclusions can benefit future political campaigns and political scientists who can repeat the process for similar corpora.

VI. REFERENCES

- [1] Newsroom, "PM requests dissolution of Parliament, proclamation of elections | eKathimerini.com". Available: <https://www.ekathimerini.com/news/1209353/pm-requests-dissolution-of-parliament-and-proclamation-of-elections/>
- [2] "The Prime Minister | Kyriakos Mitsotakis". Available: <https://www.primeminister.gr/en/the-prime-minister/kyriakos-mitsotakis>
- [3] "Alexis Tsipras - Greek PM, Reforms, Syriza | Britannica". Accessed: Jan. 11, 2024. [Online]. Available: <https://www.britannica.com/biography/Alexis-Tsipras>
- [4] "National elections – May 2023". Accessed: Jan. 11, 2024. [Online]. Available: <http://ekloges-prev.singularlogic.eu/2023/may/v/home/en/index.html>
- [5] "Νίκος Ρωμανός / Νέος εκπρόσωπος Τύπου της Νέας Δημοκρατίας", Αυγή. Accessed: Jan. 13, 2024. [Online]. Available: https://www.avgi.gr/politiki/462169_neos-ekprosopos-typou-tis-neas-dimokratias
- [6] S. Bengesio, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets", *IEEE Access*, vol. 11, pp. 11811–11826, February 2023, doi: [10.1109/ACCESS.2023.3242290](https://doi.org/10.1109/ACCESS.2023.3242290).
- [7] "Twitter data for academic research". Accessed: Jan. 11, 2024. [Online]. Available: <https://developer.twitter.com/en/use-cases/do-research/academic-research>
- [8] "CLARIN:EL - ILSP Lemmatizer". Available: <https://inventory.clarin.gr/tool-service/694>
- [9] P. Prokopidis, B. Georgantopoulos, and H. Papageorgiou, "A suite of NLP tools for Greek", 10th International Conference of Greek Linguistics, 2011. Accessed: Jan. 14, 2024. [Online]. Available: <http://archive.ilsp.gr/en/infoprojects/meta?view=publication&task=show&id=897>
- [10] R. Patel and K. Passi, "Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning", *IoT*, vol. 1, no. 2, Art. no. 2, Dec. 2020, doi: [10.3390/iot1020014](https://doi.org/10.3390/iot1020014).
- [11] F. Lazarinis, "Lemmatization and stopword elimination in Greek web searching", in *Proceedings of the 2007 Euro American conference on Telematics and information systems*, Faro Portugal: ACM, May 2007, pp. 1–4. doi: [10.1145/1352694.1352757](https://doi.org/10.1145/1352694.1352757).

- [12] D. Vilares, M. Thelwall, and M. A. Alonso, "The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets", *Journal of Information Science*, vol. 41, no. 6, pp. 799–813, Dec. 2015, doi: 10.1177/0165551515598926.
- [13] G. Kalamatianos, D. Mallis, S. Symeonidis, and A. Arampatzis, "Sentiment analysis of greek tweets and hashtags using a sentiment lexicon", in *Proceedings of the 19th Panhellenic Conference on Informatics*, Athens Greece: ACM, Oct. 2015, pp. 63–68. doi: 10.1145/2801948.2802010.
- [14] A. Tsakalidis, S. Papadopoulos, R. Voskaki, K. Ioannidou, C. Boididou,, A.I. Cristea, M. Liakata and Y. Kompatsiaris, "Building and evaluating resources for sentiment analysis in the Greek language", *Lang Resources & Evaluation*, vol. 52, no. 4, pp. 1021–1044, Dec. 2018, doi: 10.1007/s10579-018-9420-4.
- [15] Saeedeh Momtazi, "Fine-grained German Sentiment Analysis on Social Media", *LREC*, pp. 1215–1220, Jan. 2012.
- [16] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment Strength Detection for the Social Web", *Journal of the American Society for Information Science and Technology (Print)*, vol. 63, no. 1, pp. 163–173, Jan. 2012.
- [17] "Tweeting to Power: The Social Media Revolution in American Politics". Accessed: Jan. 14, 2024. [Online]. Available: <https://eds.p.ebscohost.com/eds/ebookviewer/ebook/ZTAwMHh3d19fNjU3OTAwX19BTg2?sid=da7201f2-3e5e-4635-b1c1-1213704185f7@redis&vid=16&format=EB&rid=1>
- [18] C. Karayianni and C. Psaltis, "Tweet for peace: Twitter as a medium for developing a peace discourse in the hands of the Greek-Cypriot and Turkish-Cypriot leaders", *OIR*, vol. 47, no. 5, pp. 925–943, Aug. 2023, doi: 10.1108/OIR-03-2022-0161.
- [19] S. Schulz, G. de Pauw, O. de Clercq, B. Desmet, V. Hoste, W. Daelemans and L. Macken, "Multimodular Text Normalization of Dutch User-Generated Content", *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, pp. 1–22, Jul. 2016, doi: 10.1145/2850422.

VII. AUTHORS



Elisavet Chantavaridou is a PhD Candidate at the Department of Archival, Library and Information Studies of the University of West Attica. She holds a BA degree from the Library Science Department of the former Technological Educational Institute of Thessaloniki. She also holds a MSc degree in Information and Library

Studies from the University of Strathclyde in Glasgow, Scotland. Her second BA degree is from the Open University of Cyprus in Studies in Hellenic Culture (direction: Literature). She is currently a librarian at the University of Macedonia in Thessaloniki, Greece.



Sarantos Kapidakis is a professor at the Department of Archival, Library & Information Studies of the University of West Attica, in the field of "Informatics with emphasis on Digital Libraries and Electronic Publication". He was appointed associate professor in 2002 and professor in 2009 at the

Department of Archive, Library and Museum Sciences at the Ionian University, and director of the Laboratory on Digital Libraries and Electronic Publishing. His current research interests include digital libraries, electronic publishing, architectures, usability, interoperability, metadata, Dublin Core and information retrieval.