The Ontological Approach of Modern Greek Morphology (short paper)

Nikos Vasilogamvrakis[1]

[1]*Department of Archives, Library Science and Museology, Ionian University, Corfu, Greece*

Abstract

This article comprises a brief overview of my PhD research proposal investigating the ontological approach of Modern Greek (MG) morphology. Its main objective is to study contemporary onto-linguistic models in order to form an onto-morphological tool for MG morphological analysis. The research was motivated by the lack of an ontologically holistic approach based on the Semantic Web (SW) paradigm to represent MG morphology. After a brief review on the current ontological setting within the Semantic Web, the respective morphological framework is determined and placed into the *Strong Lexicalist* theory justified by MG morpheme-based nature. Following this, main research questions are defined and the methodology of the research is presented as an itinerary process between ontological development, theory and lexical data testing. Finally, the article concludes with some preliminary research results based on a morpheme-based analysis of indicative MG lexical data in the MMoOn ontological model.

Keywords

Modern Greek morphology, ontologies, Semantic Web, Linguistic Linked Open Data

1. Introduction

The present PhD research was motivated by the lack of an ontological representation and analysis of MG morphology that integrates the Semantic Web (SW) paradigm. Therefore, it aims to:

- study the current ontological models and form an optimal representational paradigm for MG morphology in full or in part (derivation and/or inflection and/or composition)
- check, condense, resynthesize or enrich MG morphological theory, where appropriate, under the framework of the ontological representation
- establish a consistent ontological model, which would ideally represent theory and its instantiations (data) sufficiently and in separate but interconnected levels
- create an information retrieval (IR) tool for supporting query expansion (QE) e.g. the productivity of a derivational template, the frequency of a specific morpheme etc.

2. Relation of the work to the state of the art in the field

Morphology focuses on the least meaningful entities within words, called *morphemes*[2], as well as how words are composed (word formation) or inflected. Except for the extensive linguistic framework that also includes empirical analysis, the study of language has been triggered by informational models, among which is the ontological paradigm [1]–[3], that has given totally new perspectives to language studies as a more effective and functional tool. Language ontological representation has, indeed, proven to underpin multiple areas of language analysis such as lexicography [4], [5], language annotation and theory representation [2], ontology-based information extraction (OBIE), text linkage and Information Retrieval [6], [7] or NLP applications [2].

TPDL2022: 26th International Conference on Theory and Practice of Digital Libraries, 20-23 September 2022, Padua, Italy
EMAIL: l20vasi@ionio.gr
ORCID: 0000-0001-9121-2436

[2] The derived word *xor-ef-ti-s* 'dancer' for example consists of four morphemes, one stem (*xor-*) and three suffixes (*-ef-, -ti-, -s*).

For MG morphology, in particular, the major attempts of representation to date have been mostly machine readable dictionaries (MRDs) [8]–[12] with little reference to morphological theory [9], [13] and with their focus on inflection to a large extent. Other resources are more user-oriented, incorporating lexical representations (lexica) but most of them are not freely accessible nor do they follow an ontological model or a linguistic theory and additionally are available by different formats and mediums.

On the other hand, more purposeful steps towards language ontological analysis within the SW have been made in the last decade by the LLOD community, resulting in the creation of representational models such as the Ontolex-lemon [14], [15], OLiA [16], GOLD [17] and LexInfo [18], [19]. However, even though all of these models deal with morphological information, they are far from granular nor do they focus on sub-lexical analysis or on derivational morphology. To fill this gap, the MMoOn[3] model has been formed lately, focusing exclusively on language morphology and taking a morpheme-based approach that puts the morpheme concept at the center of analysis [20].

3. Theoretical approach

In terms of theory, the research is situated within the Lexicon and is sufficiently interpreted by *Strong Lexicalism* [21]–[27], that regards morphology as a separate and autonomous field in relation to syntax. Additionally, it adopts a binary relational pattern [28] of combinatorial morphology [24] as well as the *Lexical Morphology* theory, which explains word formation as a layered hierarchical process [29], [30]. The adoption of Lexicalism has been due to the morpheme prioritization as a steady meaningful entity within words and in the Lexicon, a fact that also aligns with the complexity MG creates lexical structures and its rich morphemic typology. The emphasis on the morpheme concept lies, additionally, in the need to justify non-or-hard-transparent words due to their origin from Ancient Greek (AG) (e.g. *ríγnimi* 'to break' > *ríγ-ma*[4] 'breach'). This intrinsic relation reveals the MG language allomorphic nature, which also requires the management of words as strings of distinct morphemes.

4. Research questions

From the previous discussion the following research questions are due to be explored:

- the coverage of the available models and especially of the selected base-model on indicative areas such as: morphemic typology and limits, diachronic analysis, interoperability between morphological levels (inflection, derivation, composition), allomorphy, semantic and grammatical meaning, morphemic relations, morpho-phonological processes, models of word formation, theory representation etc.
- the consistency of the base-model as well as of the MG ontological instance to the former and how this can be expressed in an ontology language e.g. OWL
- Usability issues via realistic use cases. How for example SPARQL can be used to form simple or complex queries for postulating morphological axioms (e.g. the extent of productivity of specific morphemes, which the most frequent derivational pattern is etc.)
- How would the ontology be populated so that it is overall tested and evaluated towards MG lexical data? Can an automated or semi-automated way be leveraged according to related implementations?

[3] https://github.com/MMoOn-Project/MMoOn.

[4] The phonological transcriptions are based on the International Phonetic Alphabet (IRA) (cf. https://www.internationalphoneticassociation.org/content/ipa-chart).

5. Research methodology and techniques applied

The methodology of this research moves iteratively between morphological theory, primary lexical data and the MG ontological instance and will prompt continuous tests for the extension, reforming and functionality of the latter. The research will go through the following stages:

- Review of current ontological models for language morphology
- Review of MG contemporary morphological theory with arguable viewpoints in the field. It is possible that this check-up will redefine morphological areas in view of the ontological representation
- Development of the ontological model through: a) schematic or tabular depiction of MG morphological peculiarities b) extension and development of the model according to coverage, consistency and usability criteria with the assistance of an ontology editor (e.g. Protégé) c) assessment of the ontological model with sufficient lexical data via query expansion (QE) or inferencing

6. Stage of progress and preliminary results

In the research so far, the morpheme-based approach has been explored for MG morphology and tested on the MMoOn model, which was selected as most appropriate to host the MG ontological instance [31]. In Figure 1, we ontologically analyze the structural units participating in a MG common concatenative pattern, i.e. *-τη-ς* (*-ti-s)* > *-τ-ικ-ος (t-ik-os)*, applied to two different lexical bases: *καλλιεργη-* (*kallierji-)* ∼ *καλλιεργ- (kallierγ-)*[5] and *χορευ- (xoref-)* ∼ *χορευ- (xorev-)* of the derived words *καλλιεργητής* 'cultivator' (*kallierjitís*) > *καλλιεργητικός* 'cultivating' (*kallierjitikós*) and *χορευτής* 'dancer' (*xoreftís*) > *χορευτικός* 'dancing' (*xoreftikós*). We do this by just using the classes *mmoon:Word* and *mmoon:Morph* (and their subclasses) in binary formation structures and leveraging the inverse *consistsOf* ↔ *belongsTo* object properties (OP) [31].

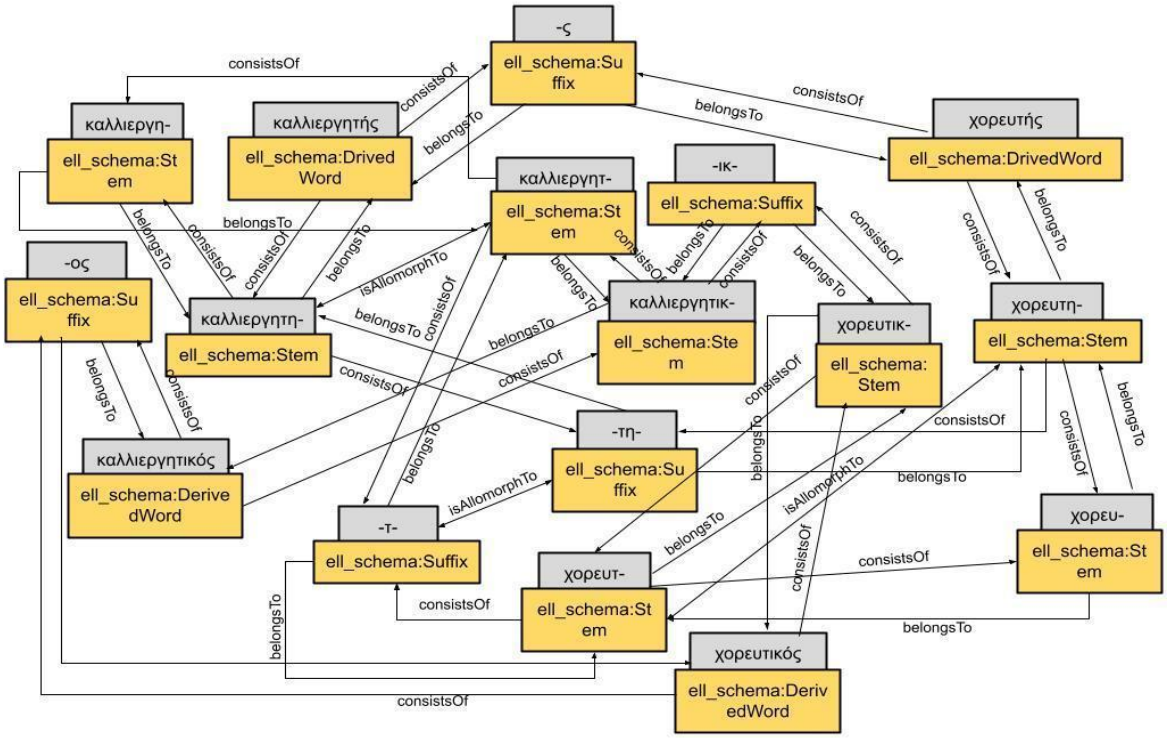

Figure 1: MMoOn ell_schema onto-lexical representation and analysis

[5] The ∼ symbol denotes the allomorphic relation (given by the *isAllomorphTo* object property) between two morphemes.

As to the following steps in the research, other approaches to word formation are to be explored (e.g. word-based) and contrasted to the morpheme-based.

7. Acknowledgements

This research was supported by the project: "Activities of the Laboratory on Digital Libraries and Electronic Publishing of the Department of Archives, Library Science and Museology".

References

[1] J. Bosque-Gil, J. Gracia, E. Montiel-Ponsoda, and A. Gómez-Pérez, "Models to represent linguistic linked data," *Nat. Lang. Eng.*, vol. 24, no. 6, pp. 811–859, Nov. 2018, doi: 10.1017/S1351324918000347.

[2] P. Cimiano, C. Chiarcos, J. P. McCrae, and J. Gracia, *Linguistic Linked Data: Representation, Generation and Applications*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-30225-2.

[3] A. C. Schalley, "Ontologies and ontological methods in linguistics," *Lang. Linguist. Compass*, vol. 13, no. 11, p. e12356, 2019, doi: 10.1111/lnc3.12356.

[4] S. Bosch, T. Eckart, B. Klimek, D. Goldhahn, and U. Quasthoff, "Preparation and Usage of Xhosa Lexicographical Data for a Multilingual, Federated Environment," presented at the The 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, Japan, Miyazaki, 2018.

[5] T. Declerck, M. Siegel, and S. Racioppa, "Using OntoLex-Lemon for Representing and Interlinking German Multiword Expressions in OdeNet and MMORPH," in *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, Florence, Italy, Aug. 2019, pp. 22–29. doi: 10.18653/v1/W19-5104.

[6] G. Ganino, D. Lembo, M. Mecella, and F. Scafoglieri, "Ontology population for open-source intelligence: A GATE-based solution," *Softw. Pract. Exp.*, vol. 48, no. 12, pp. 2302–2330, 2018, doi: 10.1002/spe.2640.

[7] Università Cattolica del Sacro Cuore, "LiLa: Linking Latin," *LiLa: Linking Latin*, 2021. https://lila-erc.eu/ (accessed May 25, 2021).

[8] P. Gakis, T. C. Panagiotakopoulos, K. Sgarbas, and C. Tsalidis, "Design and implementation of an electronic lexicon for Modern Greek," *Lit. Linguist. Comput.*, vol. 27, no. 2, pp. 155–169, 2012, doi: https://doi.org/10.1093/llc/fqs002.

[9] E. Papakitsos, M. Grigoriadou, and G. Philokyprou, "Modelling a Morpheme-based Lexicon for Modern Greek," *Lit. Linguist. Comput.*, vol. 17, no. 4, pp. 475–490, 2002.

[10] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, and E. Zavitsanos, "Ontology Population and Enrichment: State of the Art," Jan. 2011, pp. 134–166. doi: 10.1007/978-3-642-20795-2_6.

[11] K. Sgarbas, N. Fakotakis, and G. Kokkinakis, "A Straightforward Approach to Morphological Analysis and Synthesis," in *Proceedings COMLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries, Kato Achaia, Greece, September 22-23, 2000*, 2000, pp. 31–34. [Online]. Available: https://arxiv.org/ftp/cs/papers/0112/0112010.pdf

[12] Δ. Σ. Μπαλτζής, Δ. Δούκα, Σ. Κολαλάς, Ε. Ευμοιρίδου, and Ά. Αλεξάκης, "Ένα καινοτόμο ηλεκτρονικό λεξικό της Νέας Ελληνικής Γλώσσας: πρώτο μέρος: μορφολογικό λεξικό," in *Ο ελληνικός κόσμος ανάμεσα στην εποχή του Διαφωτισμού και στον εικοστό αιώνα: πρακτικά του Γ΄ Ευρωπαϊκού Συνεδρίου Νεοελληνικών Σπουδών (ΕΕΝΣ), Βουκουρέστι 2-4 Ιουνίου 2006*, 2006, vol. 2, pp. 341–354. Accessed: May 25, 2020. [Online]. Available: https://www.eens.org/eens-congress-access/?main__page=1&main__lang=de&eensCongress_cmd=showPaper&eensCongress_id=149#_ftn1

[13] Μ. Γαβριηλίδου, Π. Λαμπροπούλου, Έ. Μάντζαρη, and Σ. Ρούσσου, "Προδιαγραφές για ένα υπολογιστικό μορφολογικό λεξικό της Νέας Ελληνικής," presented at the 3ο Διεθνές Γλωσσολογικό Συνέδριο για την Ελληνική Γλώσσα, Αθήνα, 1997.

[14] B. Klimek, J. P. McCrae, J. Bosque-Gil, M. Ionov, J. K. Tauber, and C. Chiarcos, "Challenges for the Representation of Morphology in Ontology Lexicons," in *Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.*, 2019, pp. 570–591. [Online]. Available: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_33.pdf
[15] J. P. McCrae, J. Bosque-Gil, J. Gracia, and P. Buitelaar, "The OntoLex-Lemon Model: Development and Applications," in *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference : Lexicography from Scratch*, 2017, pp. 587–597.
[16] C. Chiarcos and M. Sukhareva, "OLiA – Ontologies of Linguistic Annotation," *Semantic Web*, vol. 6, no. 4, pp. 379–386, Aug. 2015, doi: 10.3233/SW-140167.
[17] S. Farrar and D. T. Langendoen, "An OWL-DL Implementation of Gold," in *Linguistic Modeling of Information and Markup Languages: Contributions to Language Technology*, A. Witt and D. Metzing, Eds. Dordrecht: Springer Netherlands, 2010, pp. 45–66. doi: 10.1007/978-90-481-3331-4_3.
[18] P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek, "Towards Linguistically Grounded Ontologies," in *The Semantic Web: Research and Applications*, Berlin, Heidelberg, 2009, pp. 111–125. doi: 10.1007/978-3-642-02121-3_12.
[19] P. Cimiano, P. Buitelaar, J. McCrae, and M. Stintek, "LexInfo: A declarative model for the lexicon-ontology interface | Elsevier Enhanced Reader," *J. Web Semant.*, vol. 9, no. 1, pp. 29–51, 2011, doi: 10.1016/j.websem.2010.11.001.
[20] B. Klimek, M. Ackermann, M. Brümmer, and S. Hellmann, "MMoOn Core - The Multilingual Morpheme Ontology," *Semantic Web*, vol. 4, pp. 1–30, 2020.
[21] A. M. Di Sciullo and E. Williams, *On the Definition of the Word*. Cambridge, Mass: MIT Press, 1987.
[22] S. Lapointe, "A Theory of Grammatical Agreement," Ph.D. Diss., University of Massachusetts, Amherst, 1980.
[23] R. Lieber, "On the Organization of the Lexicon," Ph.D. Diss., MIT, 1980.
[24] A. Ralli, *Morfologia (in Greek)*. Athens: Patakis, 2005.
[25] A. Ralli, "Morphology in Greek Linguistics: A State-of-the Art," *J. Greek Linguist.*, vol. 4, pp. 77–130, 2003.
[26] E. Selkirk, *The syntax of Words*. Cambridge, Mass: MIT Press, 1982.
[27] Α. Ράλλη, "Η μορφολογία ως αυτόνομο τμήμα της γραμματικής," in *Γλώσσης Χάριν. Τόμος αφιερωμένος από τον Τομέα Γλωσσολογίας στον καθηγητή Γεώργιο Μπαμπινιώτη*, Α. Μόζερ, Ed. Ελληνικά Γράμματα, 2008, pp. 141–156. Accessed: Jun. 09, 2020. [Online]. Available: https://www.angelaralli.gr/sites/default/files/Morphology%20as%20an%20%20Independent%20%20Grammatical%20%20Module%20%28in%20Greek%29.%20%20.pdf
[28] R. Jackendoff, "Morphological and Semantic Regularities in the Lexicon," *Linguist. Inq.*, vol. 7, pp. 89–150, 1975.
[29] P. Kiparsky, *Lexical Morphology and Phonology*. 1982. Accessed: Oct. 25, 2020. [Online]. Available: https://web.stanford.edu/~kiparsky/Papers/Lexical%20Morphology%20and%20Phonology.pdf
[30] K. P. Mohanan, *The Theory of Lexical Phonology. Studies in Natural Language and Linguistic Theory*. Dordrecht: D. Reidel, 1986.
[31] N. Vasilogamvrakis and M. Sfakakis, "A Morpheme-Based Paradigm for the Ontological Analysis of Modern Greek Derivational Morphology," in *Metadata and Semantic Research*, Cham, 2022, pp. 389–400. doi: 10.1007/978-3-030-98876-0_34.