Ontological modeling of morphological entities, allomorphy and representation in Modern Greek derivation

Nikos Vasilogamvrakis	Michalis Sfakakis	Giannoula Giannoulopoulou	Maria Koliopoulou
Ionian University	Ionian University	NKUA	NKUA
Corfu, Greece	Corfu, Greece	Athens, Greece	Athens, Greece
l20vasi@ionio.gr	sfakakis@ionio.gr	giannoulop@ill.uoa.gr	mkoliopoulou@gs.uoa.gr

Abstract

In the present article, we ontologically explore the entities of Modern Greek (MG) morphology as well as the variety of their allomorphic and representational relationships. The aim of this modeling is to fully enable the representation of lexical data in the MMoOn ontology and to propose an interactive allomorphy framework for MG derivation. According to this, interconnected allomorphy paradigms and derivational rules are placed inside the ontology, engulfing both the Permanent and Dynamic lexicon so that lexical data can be generated automatically and be morphologically justified. In respect of the morphological entities representation, different examples are presented to elaborate how allomorphy or morphological semantics affect them, as they show different or identical phonetic, morphemic and orthographic forms.

1. Introduction

Modern Greek (MG) is a synthetic inflectional language that presents a variety of morph types participating in complex morphological structures. Moreover, a significant characteristic is that it engulfs several non-transparent or phonologically unjustified allomorphic forms partly originated from Ancient Greek (AG) or based on AG roots. In order to explore language derivational processes, it is necessary to identify the different types of morphs, especially the stem and affix concepts and their subcategories. But it is equally important to look into these entities under the phenomenon of allomorphy involved in MG derivation and place it within suitable derivational environments (Melissaropoulou & Ralli, 2009) for creating a framework towards the generation of new forms.

In what follows, in section 2, we explore the different morphological entities of MG participating in derivation and then we focus on the types of allomorphy and propose a framework in which it can operate and be modeled. Then, we present the different representational aspects of these entities that justify the MMoOn ontology conceptual analysis. Finally, in section 3, we conclude on the topic.

2. Morpho-Ontological analysis

2.1. MG morphological typology

Morphemes or more precisely their realizations, *morphs*, are divided into two broad categories: *free* and *bound* (Booij, 2012; Ralli, 2005; Spencer, 2017). Free morphs are mono-morphemic words, either of *grammatical* or *lexical* nature, while bound cannot stand alone as free words and can be either roots, stems, affixes, confixes (Giannoulopoulou, 1999) or bound stems (Ralli, 2005).

Roots are the keystones of a lexeme but as Ralli postulates (Ralli, 2005), a root concept in MG cannot easily be located because roots are traced back in AG lexical forms. It would be more sensible, then, to use a *Stem* concept that may be either a *Base* (an initial stem) (e.g. *χορ-* (*xor-*) > *χορός* (*xorós*) 'dance') or an *Affixed Base* (e.g. *χορεύ-* (*xorév-*) > *xorévo* 'to dance').

Affixes are bound morphs that append to bases, operating as "satellites", to form new affixed bases according to their *categorial signature* (Ralli, 2005). Affixes are divided into *Prefixes* when they precede

Proceedings of the 4th Int. Workshop on Resources and Tools for Derivational Morphology (DeriMo 2023), pages 71–80,
Dubrovnik, Croatia, 5-6 October 2023.

(e.g. *δια-* in *δια-δρασ-* (*δia-δras-*)), or *Suffixes*, when they follow stems. Prefixes may also precede words, thus forming new words (e.g. *δρω* (*δro*) 'to act' > *δια-δρώ* (*δia-δró*) 'to interact'). Suffixes may in turn be of *Derivational* (e.g. *-ευ-* in *χορ-εύ-ω* (*xor-év-o*) 'to dance') or *Inflectional* (*-ω* in *χορεύ-ω*) nature.

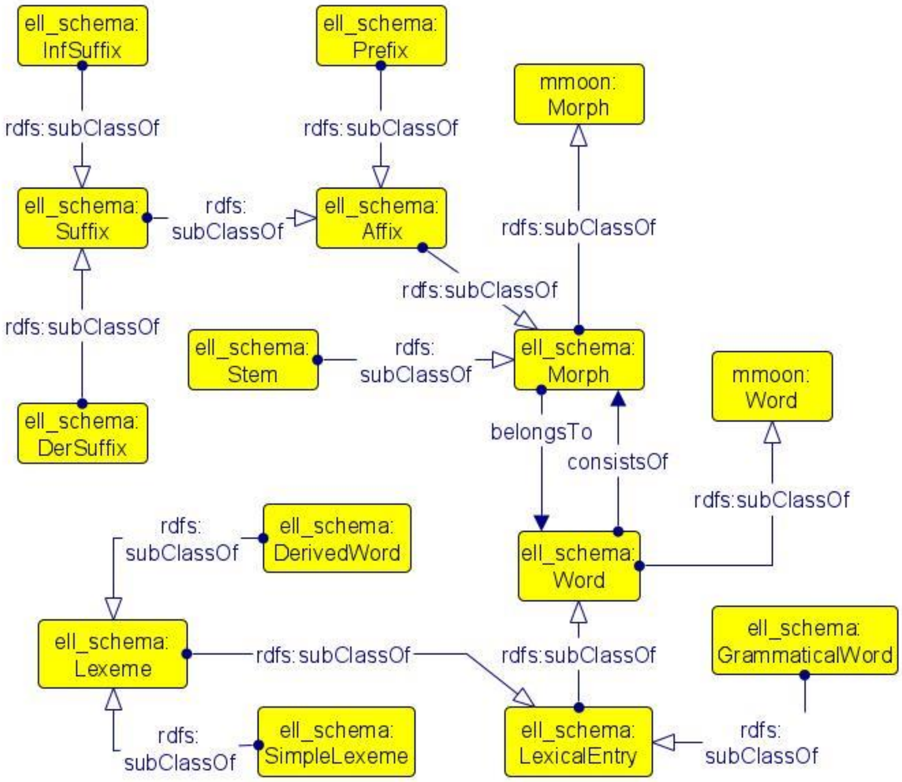

Figure 1. ell_schema morphological entities embedded into the MMoOn model

Confixes (Anastasiadi-Symeonidi, 1986; Giannoulopoulou, 1999), *Bound stems* (Ralli, 2005, 2007, 2012) or bound morphs of *neo-classical compounds* (Booij, 2012), as they are named, are a special group of morphs found as constituents in dual-structured forms of scientific or other vocabularies, usually coming from AG or Latin (e.g. *δol-o-plókos* 'schemer', *γloss-o-loγía* 'linguistics', *meta-mondernismós* 'post-modernism' etc.). However, because these are rather placed between derivation and composition areas and because of their functional and semantic peculiarities, they are not analyzed or represented here as they will be considered at a later stage of analysis when decisions on data processing are to be made.

Words[1] can be either composed by a series of morphs (multi-morphemic) or consist of just a single morph (mono-morphemic) with no further morphological analysis. Mono-morphemic words can be *Grammatical* (e.g. conjunctions *όταν* (*ótan*) 'when', *και* (*ke*) 'and') or *Lexical* (usually loan words from foreign languages (e.g. *taxi* > *ταξί* (*taksí*)). Multi-morphemic words are always finalized by an

[1] Compounds are also regarded as word types but they are not part of this research.

inflectional suffix, even an unrealized one (Ø) (e.g. *μητέρα-* (*mitéra-*) > *μητέρα* (*mitéra*) 'mother') and can be *Simple Lexemes* (e.g. *χορ-ός* (*xor-ós*) 'dance') or *Derived Words* (e.g. *χορ-ός* (*xor-ós*) > *χορ-εύ-ω* (*xor-év-o*) 'to dance'). The former uses a base and the latter an affixed base, which in both cases are finalized by an inflectional suffix.

Based on the previous conceptual analysis, in Figure 1, we identify the related classes in the MMoOn ontology (Klimek et al., 2020) and develop the specific *ell_schema*[2] embedded into it. We further add two classes: *ell_schema:DerSuffix* and *ell_schema:InfSuffix* as subclasses of *ell_schema:Suffix*. For the moment, we leave the *Stem* concept as it is, considering its subdivision in due time. We have chosen MMoOn, as already done before (Vasilogamvrakis et al., 2022; Vasilogamvrakis & Sfakakis, 2022), because it has been a comprehensive domain ontology for the representation of morphological language data (Klimek et al., 2019) and because it has been used as a template for the development of the Ontolex Morphology Module[3].

2.2. Allomorphy

Allomorphy is the morphological phenomenon according to which a morpheme that is realized by a morph has more than one form with the same meaning. This morph variant[4] is found in different morphological environments, that is why allomorphs stand in complementary distribution within words. Allomorphy can be basically of two types: a) *morpho-phonological*, when the change depends on some still-existent morpho-phonological rule[5] (e.g. *κλεβ-* (*klev-*) ~ *κλεφ-* (*klef-*) ~ *κλεψ-* (*kleps-*) of the simple lexeme *κλέβ-ω* (*klev-o*) 'to steal') and b) *morphological* or *grammatical*, when the occuring allomorph is grammatically dependent and unpredictable (e.g. *σώμα-* (*sóma-*) ~ *σωματ-* (*somat-*) of the noun *σώμα* (*sóma*) 'body' or the AG form *κλοπ-* (*klop-*), an additional allomorph to *κλεβ-* ~ *κλεφ-* ~ *κλεψ-*) and it engulfs either bases or affixes alone or their combinations as affixed bases. An excessive type of allomorphy can also occur in forms, which substitute absent *lexical* realizations in inflection (e.g. *είδ-α* (*íð-a*) 'I saw, which is the aorist word form of *βλέπ-ω* (*vlép-o*) 'I see'). These forms are usually considered as instances of *suppletion* and, therefore, not true allomorphs as they do not show any phonological or semantic similarity (Ralli, 2005).

The representation of allomorphy in MG derivation is central because it triggers the creation of new derivatives (Karasimos, 2011) and offers connectivity between them. This is evident, in Figure 2, in the morpheme-based analysis[6] of *αγαπ-ώ* (*αγαp-ó*) 'to love' and its derivative *αγαπη-τ-ός* (*αγαpi-t-ós*) 'beloved', where their bases *αγαπ-* (*αγαp-*) and *αγαπη-* (*αγαpi-*) are allomorphs to each other.

---

[2] The ell_schema current version can be reached at: https://github.com/nvasilogamvrakis/mmoon_project/blob/main/ell_schema/ell_schema_03.owl.

[3] https://github.com/ontolex/morph/.

[4] Allomorphs are related to each other with appropriate morpholexical rules, which normally depict the morphological environment in which an allomorph occurs (Karasimos, 2011; Ralli, 2005).

[5] For Ralli (2005), true allomoprhs are synchronically unjustified and unpredictable forms and not those derived by phonological rules.

[6] The MG morpheme-based analysis is elaborated in Vasilogamvrakis & Sfakakis (2022).

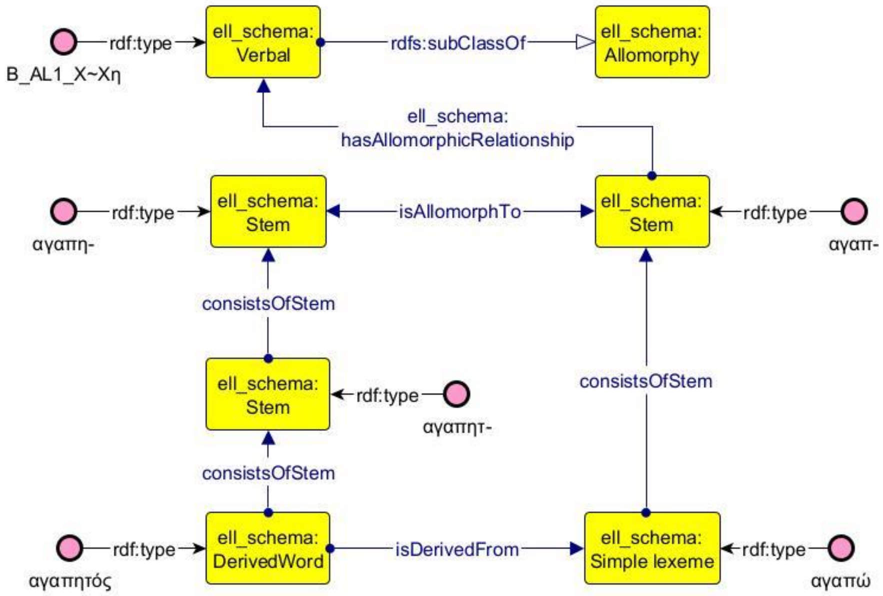

Figure 2. Interconnection between words through the allomorphs *αγαπ-~αγαπη-*, belonging to paradigm *B_AL1_X~Xη*

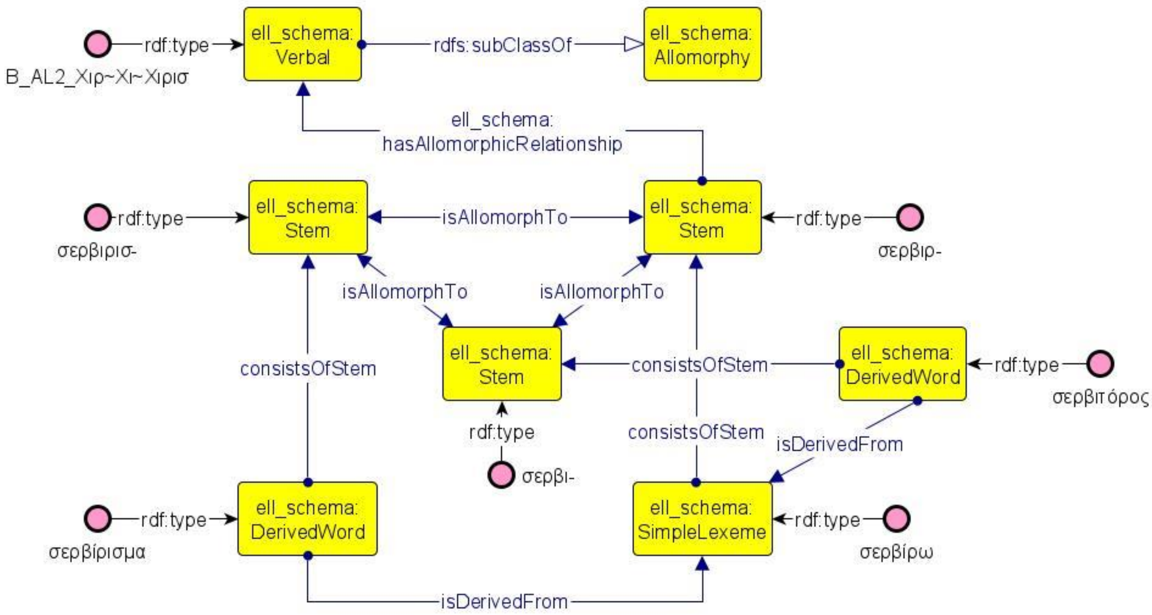

Figure 3. Allomorph instances *σερβιρ-~σερβι-~σερβιρισ-* adapted to MG, belonging to paradigm *B_AL2_Xιρ~Xι~Xιρισ*

Allomorphy can also occur in cases of loans from foreign languages. For example, in Figure 3, we show that the base *σερβιρ-* (*servir-*) of *σερβίρ-ω* (*servír-o*) 'to serve' (*servir* from French) is allomorph to *σερβι-* (*servi-*) of *σερβι-τόρος* (*servi-tóros*) ‘waiter’ (*servi-tore* from Italian) and to *σερβιρισ-* (*serviris-*) of *σερβίρισ-μα* (*servíris-ma*) ‘serving’ (Karasimos, 2011; Ralli, 2005).

Furthermore, since allomorphs stand in complementary distribution, forms like *αγαπώ* (*agapó*) / *αγαπάω* (*agapáo*) (Present, 1st Person, Singular) of Figure 4, emerged by *Reanalysis* of the AG contracted forms, are rather considered free variants (Ralli, 2005) and not true allomorphs. In the same figure, we also observe that the stem variant *αγαπα-* (*agapa-*) is specifically combined with the variant inflectional suffix *–γα* in *αγάπα-γα* (*agápa-ga*) whereas *αγαπ-* (*agap-*) with the variant *-ούσα* (*-usa*) in *αγαπ-ούσα* (*agap-úsa*) in Imperfect. We, therefore, create a new *ell_schema:hasFreeVariant* object property (OP) to connect the two morph entities, which we extend to also connect the two word lemma forms (*ell_schema:Morph* or *ell_schema:Word* as domain and range of the OP *ell_schema:hasFreeVariant*).

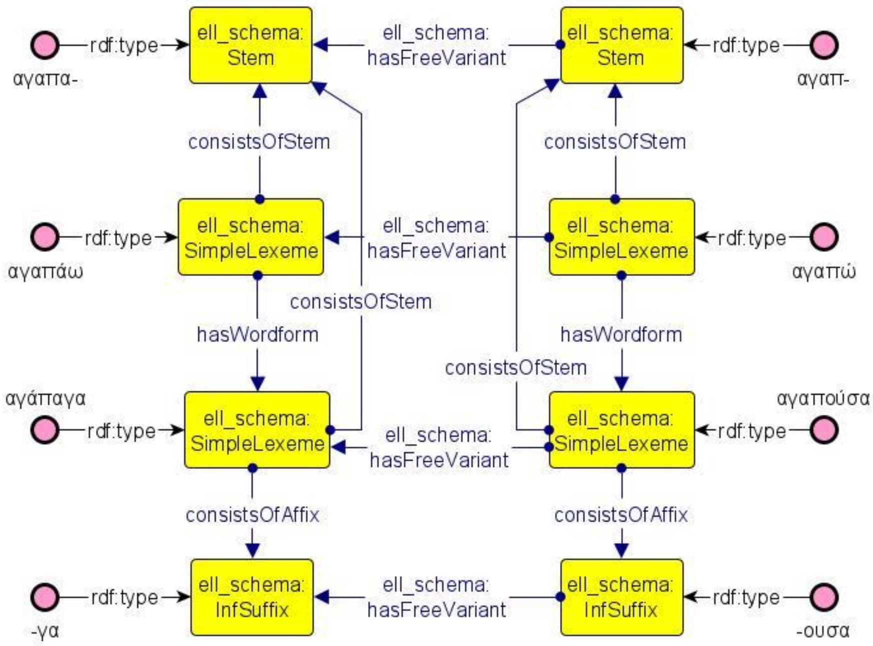

Figure 4. Interconnection between free variants via the *ell_schema:hasFreeVariant* OP

Allomorphy framework

The insertion of rules in the ontology does not contradict the assumption of some linguists that the Mental or Permanent Lexicon may include, next to morphological lemmas and non-transparable words, the dynamic area of word construction, i.e. the grammar or morphology (Kiparsky, 1982; Lieber, 1980; Selkirk, 1982). As presented in Vasilogamvrakis et al., 2022, the kind of morphological rules inserted in the ontology are rather descriptive, i.e. a top-down element that clusters similar lexical data. However, these rules, as it will be shown next, can be leveraged for modulating an appropriate pipeline workflow for generating new forms.

In a computational-based approach, allomorphy is categorized into nominal, verbal and prefixal according to the affected lexico-grammatical category (Karasimos, 2011). Each of these categories encapsulates a series of allomorphy paradigms[7], which are destined to operate as Regex patterns to bootstrap a morphological analyzer. These patterns are combined with appropriate computational rules placed within a specific morphological environment so as to predict the allomorphic change of a word.

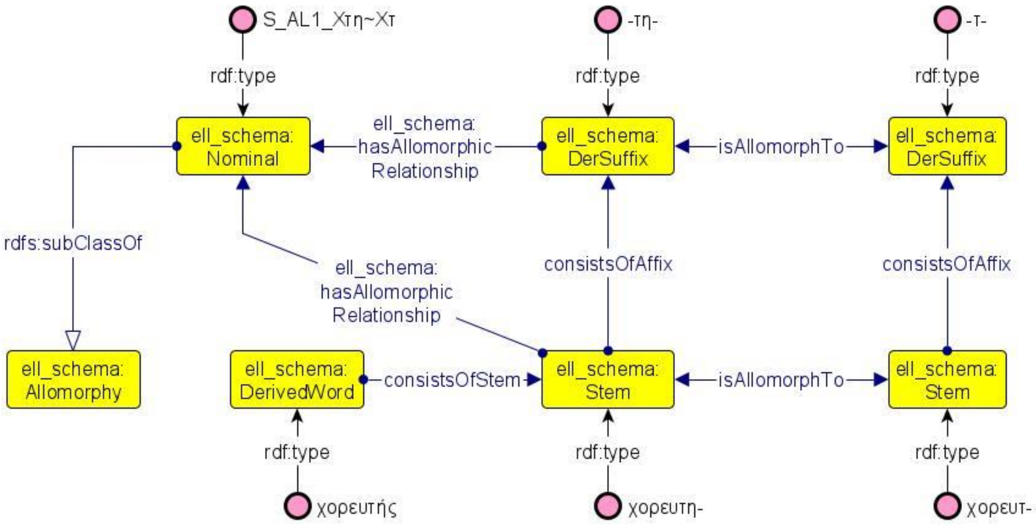

Figure 5. Interconnection between allomorph derivational suffixes *-τη-* ~ *-τ-* and between their attached stems, belonging to paradigm *S_AL1_Xτη~Xτ*

In a similar manner, we want to create allomorphy paradigms as morpholexical rules (Karasimos, 2011; Ralli, 2005) and relate them to specific derivational environments according to suffix-driven selectional restrictions (Melissaropoulou & Ralli, 2009). To host allomorphy paradigms, we introduce a new *ell_schema:Allomorphy* class in the core MMoOn schema, which, for the moment, we subdivide into *ell_schema:Verbal* and *ell_schema:Nominal* subclasses (Figures 2, 3 and 5). Although all variant forms are allomorphs to each other, which is represented in the ontology, the allomorphy paradigm is linked only to the basic morph lemma[8] (*σερβιρ-*) and not to its alternative forms (*σερβι-* ~ *σερβιρισ*) (Booij, 2012; Karasimos, 2011). For doing so, we add an *ell_schema:allomorphic_relationship* OP, with *ell_schema:Morph* as domain and *ell_schema:Allomorphy* as range (Figures 2, 3 and 5). We represent this specific allomorphy paradigm starting with the base (B) paradigm number and an X character for the common lexical part, followed by each variant with the symbol ~ in between (*B_AL1_X~Xη or B_AL2_Xιρ~Xι~Xιρισ*)[9]. We choose this inclusive pattern, adhering to the common morphological representation of allomorphs (Ralli 2005) but alternative ways may be also considered in the course of

[7] We chose the term 'paradigm' instead of 'class' so that it is distinguished from the ontological term 'class'.
[8] This forms the initial lexical entry of the derivational family (*σερβιρ-* > *σερβίρ-ω*).
[9] The given paradigm numbers are arbitrary.

the research. Similar is the modelling for allomorph suffixed bases (S), in Figure 5 (e.g. *S_AL1_Xτη~Xτ* for the derivational suffix -*τη*- ~ -*τ*-, preceded by the common lexical part X).

In order for an allomorphy paradigm to operate as a data classification module, an additional built-on programming pipeline should be implemented, based on pattern matching queries, which are sent to a core Lexicon component. According to the modelling of Figure 6, a verbal allomorphy paradigm (*B_AL1*) finds matches by its instance (*X~Xη*) inside the Lexicon of lemmas and clusters them according to the common lexical part X, making a unique set of related bases (e.g. *αγαπ~αγαπη:* set1). Then, every term of the set replaces the placeholder AL of the derivational word-pairs based on suffixation rules (Melissaropoulou & Ralli, 2009; Vasilogamvrakis et al., 2022), which are simultaneously validated against the existent lemmas of the Lexicon.

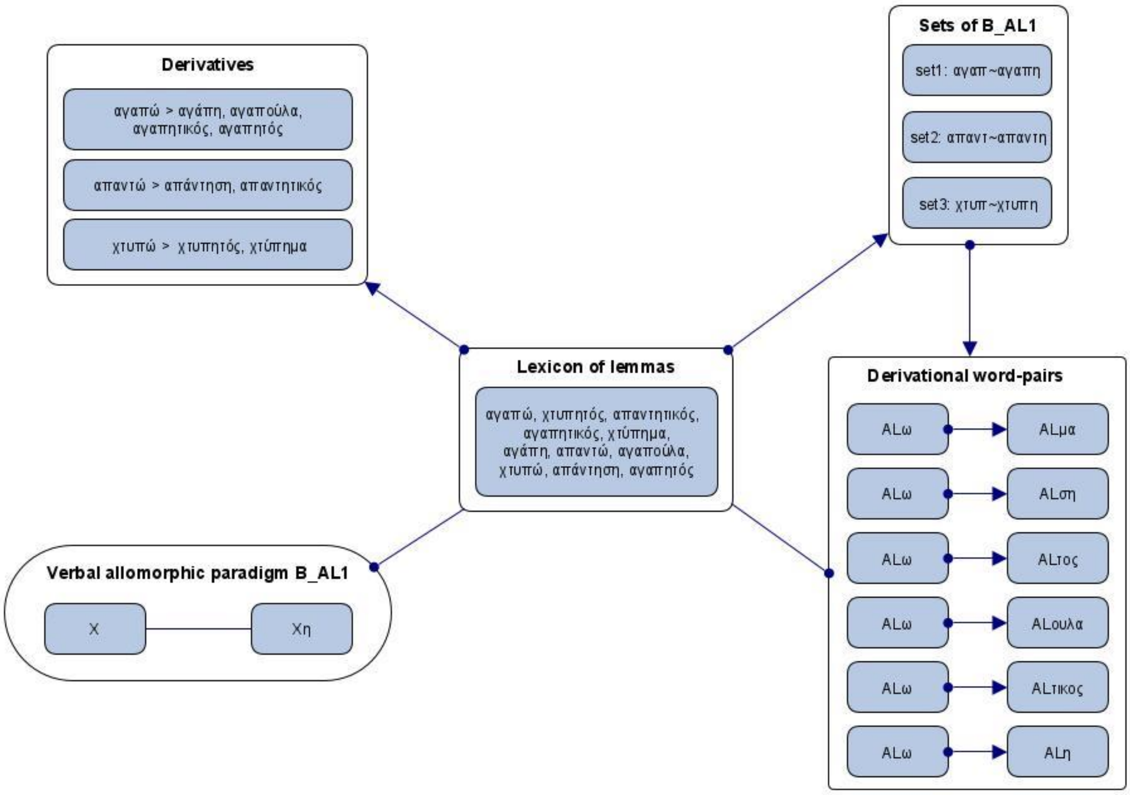

Figure 6. Provisional pipeline model for creating derivatives based on allomorphy paradigms

As a filtering rule, the placeholder of the source word always uses the common allomorph (*αγαπ*), whereas the placeholder of the target word may use all available allomorphs of the paradigm (*αγαπ~αγαπη*). For example, for the *αγαπ~αγαπη* set1 of *B_AL1* paradigm, the derived words, *αγάπ-η* (*aγáp-i*) 'love', *αγαπ-ούλα* (*aγap-úla*) 'sweetheart', *αγαπη-τικός* (*aγapi-tikós*) 'lover' and *αγαπη-τ-ός* (*aγapi-t-ós*) 'beloved' will be generated from the simple lexeme *αγαπ-ώ* (*aγap-ó*) 'to love', within a specific derivational environment of word-pairs, and after validated against the Lexicon of lemmas. Apparently, a derivational word-pair can be combined with more than one allomorphy paradigm, which makes the model particularly economical.

This modeling reproduces the theoretical assumption that the Dynamic lexicon (morphology) applies rules to the Permanent lexicon to generate or re-analyze derivational structures, placing the ontology at the centre of this operation. However, it would be wise, here, to stress that until we test the model's effectiveness upon real lexical data, it is likely that it will be modified to optimize performance and consistency and is, therefore, considered provisional.

2.3. Representation

With regard to the representation of form, the MMoOn provides the class *Representation* as domain of the data properties (DP): *morphological, phonetic* and *orthographic representation*. The usability of this class is evident mostly in cases of allomorphy or homonymy.

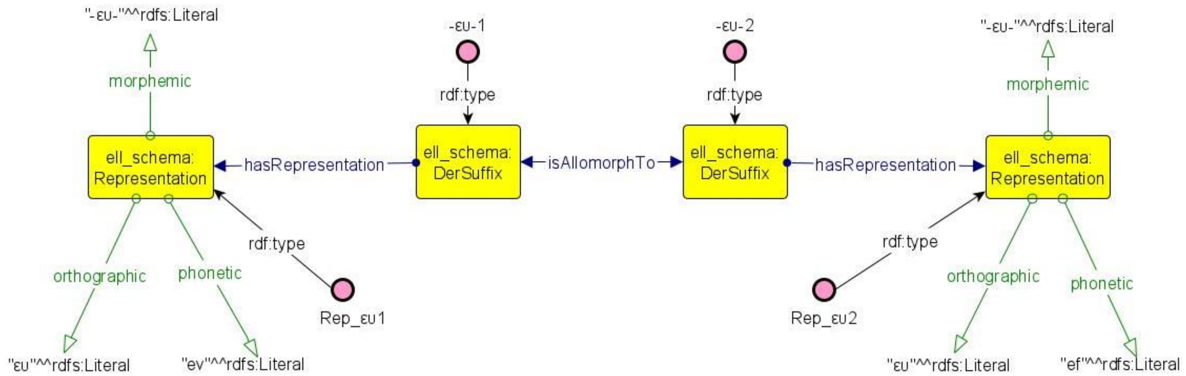

Figure 7. Representation of allomorphs (with allophones) with different *Representation* instances

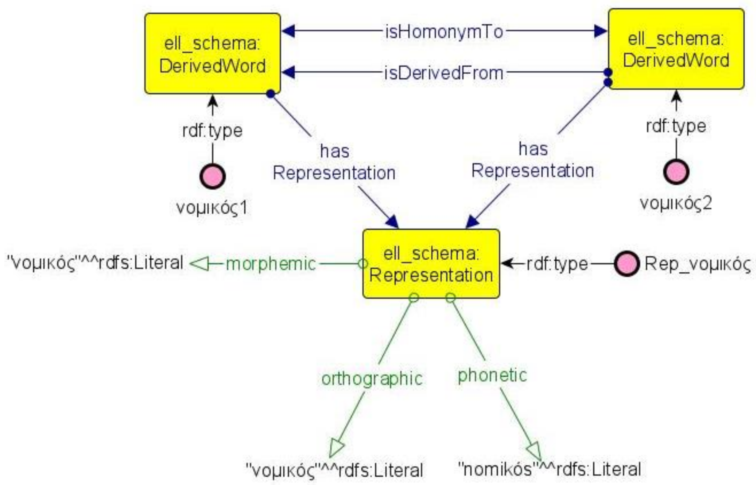

Figure 8. Representation of homonyms with the same *Representation* instance

Except for those cases explored previously, *allomorphy* can also occur when there are variant phonetic realizations of a phoneme (*allophones*) within a morph. Accordingly, in Figure 7, each of the derivational suffixes *-ευ-1* and *-ευ-2* retains different *Representation* instances, *Rep_ευ1* and *Rep_ευ2*, because of their different phonetic transcription (*ev* and *ef* respectively). This is better understood when both are seen as constituents of their belonging words within a very common MG derivational chain e.g. *χορ-εύ-ω* (*xor-év-o*) 'to dance > *χορ-ευ-τή-ς* (*xor-ef-tí-s*) 'dancer'. Their phonologically-based allomorphic interconnection is captured by the symmetrical OPs *is allomorph to*, having, at the same time, a common morphemic and orthographic representation *-ευ-* value.

On the other hand, *homonymy* occurs when there are similarly spelled (homographs) and pronounced (homophones) morphs or words while having different lexical or grammatical meanings. For example, as shown in Figure 8, the two different words *νομικός1* (*nomikós1*) 'juristic' and *νομικός2* (*nomikós2*) 'lawyer' are also marked as Adjective and Noun respectively. Each word is connected to the other with a symmetrical *is homonym to* OP, while both of them have a common *Representation* instance *Rep_νομικός* and identical morphemic, orthographic and phonetic representation values. Furthermore, they have a derivational relation, as the second word *νομικός2* is derived from the first *νομικός1* by *Conversion*.

3. Conclusion

In the present article, we ontologically analyzed the types of MG morphological entities participating in derivational structures, justifying their presence in the MMoOn *ell_schema* ontology. In particular, we focused on the stem and affix concepts and their subclasses because we showed that these entities are affected by the phenomenon of allomorphy. We additionally provided evidence that the latter impacts significantly on derivational processes and, for that reason, we modeled and placed it within certain derivational environments so that it is functional and can generate new lexical forms. This framework is actually consistent with the postulation that the Lexicon can incorporate both morphological rules and lexical data and we assigned the ontology that role. Finally, we showed how morphological semantics or certain allomorphy types can affect the representational aspects of morphs or words.

4. 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

Anastasiadi-Symeonidi, A. (1986). *Neology in Standard Greek (in Greek)*. Aristotle University of Thessaloniki.

Booij, G. (2012). *The grammar of words: An introduction to linguistic morphology*. Oxford University Press.

Giannoulopoulou, G. (1999). *Morphosemantic comparison between affixes and confixes in Modern Greek and*

Italian (in Greek) [Ph.D. Diss., Aristotle University of Thessaloniki (AUTH). School of Italian Language and Literature]. http://hdl.handle.net/10442/hedi/32432

Karasimos, A. (2011). *Computational processing of allomorphy in Modern Greek derivation (in Greek)* [Ph.D. Diss., University of Patras]. http://hdl.handle.net/10442/hedi/32458

Kiparsky, P. (1982). *Lexical Morphology and Phonology*. https://web.stanford.edu/~kiparsky/Papers/Lexical%20Morphology%20and%20Phonology.pdf

Klimek, B., Ackermann, M., Brümmer, M., & Hellmann, S. (2020). MMoOn Core—The Multilingual Morpheme Ontology. *Semantic Web*, *4*, 1–30. http://www.semantic-web-journal.net/system/files/swj2549.pdf

Klimek, B., McCrae, J. P., Bosque-Gil, J., Ionov, M., Tauber, J. K., & Chiarcos, C. (2019). Challenges for the Representation of Morphology in Ontology Lexicons. *Proceedings of the eLex 2019 Conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.*, 570–591. https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_33.pdf

Lieber, R. (1980). *On the Organization of the Lexicon* [Ph.D. Diss.]. MIT.

Melissaropoulou, D., & Ralli, A. (2009). Combinatoriality of Derivational Suffixes in Modern Greek: A First Approach (in Greek). *Studies in Greek Linguistics*, *29*, 97–107. http://ins.web.auth.gr/images/MEG_PLIRI/MEG_29_97_107.pdf

Ralli, A. (2005). *Morfologia (in Greek)*. Patakis.

Ralli, A. (2007). *The composition of words (in Greek)*. Patakis.

Ralli, A. (2012). Deverbal Compounds with Bound Stems. In *Compounding in Modern Greek* (pp. 201–220). https://www.angelaralli.gr/sites/default/files/Neoclassical%20compounds%20with%20bound%20stems.pdf

Selkirk, E. (1982). *The syntax of Words*. MIT Press.

Spencer, A. (2017). Morphology. In *The Handbook of Linguistics* (pp. 211–233). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119072256.ch11

Vasilogamvrakis, N., Koliopoulou, M., Sfakakis, M., & Giannoulopoulou, G. (2022). Testing the Word-Based Model in the Ontological Analysis of Modern Greek Derivational Morphology. In S. Chiusano, T. Cerquitelli, R. Wrembel, K. Nørvåg, B. Catania, G. Vargas-Solar, & E. Zumpano (Eds.), *New Trends in Database and Information Systems* (pp. 572–584). Springer International Publishing. https://doi.org/10.1007/978-3-031-15743-1_52

Vasilogamvrakis, N., & Sfakakis, M. (2022). A Morpheme-Based Paradigm for the Ontological Analysis of Modern Greek Derivational Morphology. In E. Garoufallou, M.-A. Ovalle-Perandones, & A. Vlachidis (Eds.), *Metadata and Semantic Research* (pp. 389–400). Springer International Publishing. https://doi.org/10.1007/978-3-030-98876-0_34