Computational prediction of Greek Nominal Allomorphy

Athanasios Karasimos
Academy of Athens
akarasimos@academyofathens.gr

Abstract
In Theoretical Morphology, an interest in the topic of stem allomorphy has been renewed the last two decades. On the other hand, the computational treatment of allomorphy is still a huge challenge since the first systematic attempts on predicting allomorphy with machine learning techniques. The goals of this paper are to predict the allomorphic changes and to show the essential contribution of various morphological, phonological and semantic characteristics. Therefore, we use a MaxEnt model to identify the weights of these characteristics that are directly dependent on allomorphy and help design a predictive model. Our model is based on AMIS and its overall accuracy was 86.49%. To improve the system, we tried a more rational approach to achieve a better performance (the correct prediction was raised to 91.43%).

Keywords: maximum entropy, supervised morphology learning, allomorphy, derivation, Modern Greek, AMIS, AlloMantIS

1. Introduction

In Theoretical Morphology, the interest in the topic of stem allomorphy has been renewed by Aronoff's (1994) work, which led to novel descriptions of inflectional and derivational phenomena in work by Booij (1997), Maiden (2004), Pirelli and Battista (2000a, 2000b), Stump (2001) and Thornton (1997), among others. The main aim of Aronoff's work and later research is the notion that the significance of a lexeme is not a single phonological representation, but an array of indexed stems, which may stand in relations ranging from identity through semiregular/ irregular phonological alternation to full suppletion. It is pointed out that, beyond the theoretical challenges of the phenomenon, allomorphy remains a serious problem for morphological parsing that must be solved.

On the other hand, the goal of Computational Morphology is to create programs which can produce an output that matches as closely as possible the analysis that would be given by a morphologist. More specifically, an Unsupervised Morphology Learning Model (UMLM) accepts only huge corpora and tools for analysis as input, without the use of a lexicon and morphological (or phonological) rules for a particular language (Goldsmith 2001; Hafer & Weiss 1974; Harris 1955, 1967). As part of the criticism of Unsupervised Morphology Learning Models for their failure to deal with

Greek allomorphy, Karasimos (2009) has argued that probably only a supervised morphology learning model is more likely to successfully face allomorphy. Nevertheless, neural network approaches have been shown to overcome several problems that the earlier approaches faced (Malouf 2016; Mikolov & Zweig 2012; Sundermeyer, Schlüter & Ney 2012, 2015). The computational treatment of allomorphy has been a huge challenge since the first systematic attempts on predicting allomorphy with machine learning techniques (Ling & Marinov 1994; Pinker & Prince 1988; Rumelhart & McLelland 1986, among others).

2. Re-visiting allomorphy

This study is couched in a theoretical framework centered on the morpheme, treating allomorphy as a morphological phenomenon which places derivation at a separate level of word formation process. Therefore, we adopt here Ralli's (1994, 1999, 2000, 2005, 2007, 2008) framework and thereby modify it. Allomorphs are defined with morphological criteria. More specifically, allomorphy, as defined by Lieber (1982), is the study of morpheme variants which share such lexical information as semantic representation and argument structure, but which differ unpredictably and arbitrarily in the phonological forms and in the morphological environments in which they occur (e.g., *vima* ~ *vimat* 'step N ', *vrisk*~ *vrik*~ *vre* 'find'). Additionally, we introduce the term *phonomorph*, for phonologically-driven 'allomorphy' which is not an actual allomorph but a product of a (morpho-)phonological rule. In addition, we classify allomorphy into various categories based on its variant forms.

To enhance our theory, we use the *rules of allomorphy treatment,* which is a model of registry information (inflection, derivation, and compounding) for adjectival, nominal, and verbal allomorphs, given that the morpholexical rules in Lexicon (Lieber 1982) include only inflectional information for allomorphs. These rules are placed outside of the Lexicon, in the area of Grammar and assign to each allomorph the proper information of their morphological environments of appearance. Based on Greek data, we also define the *Principle of Allomorphic Behavior,* according to which all morphemes exhibit the same allomorphic behavior, i.e., the presence or absence of allomorphy in all word formation processes without dependence on any specific process (Karasimos 2009). Finally, we propose the *single allomorphy selection constraint*, according to which a correlation between derived words with a common base is ensured by the participation of only one form of the lexeme-base (this constraint prevents the appearance of all allomorphs of each morpheme as basis of a nominal derived word (Καρασίμος 2011)). This constraint applies to all nominal and adjectival bases and to ex-contracted verbs without exceptions, usually satisfying the optimal syllabic structure (CV). Based on all the aforementioned principles, rules and constraints, we try to create a theoretical and effective model of computational processing of allomorphy with a concrete prediction system.

3. Computational linguistics: Supervised morphology learning and maximum entropy

3.1 Unsupervised morphology learning vs. supervised morphology learning

As opposed to the computational analyses on syntax, research on computational morphology has been relatively poor. According to Roark and Sproat (2007), the absence of a corpus of morphologically annotated words put a burden to the development of a machine learning morphological system that could confront successfully a morphologically-complex analyzer such as the one proposed by Koskenniemi (1983). However, close to the dawn of the new millennium, the interest in statistical models of morphology —more specifically of unsupervised (or lightly supervised) morphology learning from annotated corpora— has been rapidly increased. Unfortunately, most of the unsupervised morphological models have quite problematic performance when they are tested with morphologically-rich languages (for more discussion about this, see Καρασίμος 2011).

The strong opponents of the above methods are the supervised morphology learning models. These approaches are considered to provide more benefits and advantages than other theories; some popular implementations are the rule-based models, the probabilistic-stochastic ones, and the connectionist ones. The main behind-the-scenes idea of each implementation is to extract some generalised standards, rules, principles, and behaviors of a group of training data. More precisely, they describe the relationship between input data and obtained results presented as a set of examples; the algorithm usually learns the training data to predict what will be asked of each new imported data. Nowadays, the most dominant approaches are the Maximum Entropy Learning, the Memory-based Learning and the Transformation-based Learning.

3.2 Maximum entropy learning approach

For the purposes of this paper, we follow the Maximum Entropy Learning Approach (MELA) that was extracted from the corresponding mathematic theory. The first objective of this approach is to determine the statistical datasets that can capture the behavior of a random process, i.e. the feature selection of our training data. Then given all these statistics, the second objective is to include these features in a precise process model —a model that can predict the future exported processing— i.e. the final choice of this model. According to the supporters of MELA all the known and unknown, regular and irregular words are using the same strategy, since they constitute another feature in the general model of probability. This strategy offers great potentials to treat allomorphy, which is considered as something irregular, a marginal synchronic junk pile and a relic.

The MaxEnt framework offers a mathematically sound way to build a probabilistic model for SOI (Subject-Object Identification) which combines different linguistic cues. The approach of Dell’ Orletta et al. (2007) uses constraints on the prediction of Subject and Object in Italian and Czech by resorting to the technique of Maximum Entropy. Based on their concept, we attempt to test a model for the Greek nominal allomorphy in derivation. Our goals are to predict the allomorphic changes and to show the essential contribution of various morphological, phonological, and semantic characteristics. The aim of this model is to identify the weight of these characteristics that are directly dependent on allomorphy, in order to help design a predictive model. This model is not only destined for nominal stem allomorphy, but also for nominal derivational allomorphy.

4. The AMIS experiment for nominal derivation

4.1 Introduction

Our model is based on AMIS, which is a parameter estimator for maximum entropy models (Berger, Della Pietra & Della Pietra 1996). It is freeware and benefits from linguistic feature sets (Yoshida 2006); given a set of events as training data, the program outputs parameters that optimise the likelihood of the training data. Given a set of events as training data, the program outputs parameters that optimise the likelihood of the training data. Usually it was used in syntactic annotated corpora to extract processing data for morphosyntactic analysis and it is a significant risk to test AMIS with a morphological model.

4.2 Linguistic interpretations as feature sets

The diachronic research points out that allomorphy is usually relics of non-active phonological and morphological rules and changes in a language, more specifically in Greek. Therefore, we make the assumption that the Greek words perhaps “include” the necessary information to a system with minimal supervision to predict whether a stem or a word has allomorphs, and if so, what kind of allomorphy. We maintain that the stochastic models seem to be more suitable to satisfy the requirements of a model with linguistic feature sets. These characteristics are functions type-f_{xn} (λ,Σ), where a particular item χ_i is tested for the word-attribute λ, which is included in a feature set Σ. For this MaxEnt model, we chose different types of features that contain morphological, phonological and semantic dimensions of the distributions of nominal allomorphy (in allomorphic classes *ACx*).

Our characteristics are 8 (for more information, see Καρασίμος 2011):

(i) *Allomorphic Class* (8 classes of different nominal allomorphic behavior[1]), as the main characteristic that is under survey to discover the connection with the other characteristics,

(ii.) *Inflectional Class* (8 classes based on Ralli’s (1994) model; two for masculine nouns, two for feminine nouns and four for neutral nouns),

(iii) *Syllables* (up-to-6 syllables),

(iv) *Stress* (3 levels – ultimate, penultimate, antepenultimate),

(v) *Last characters* (up-to-4 characters),

(vi.) *Last Syllable types* (3 last syllables in reverse order),

(vii) *Animacy* (yes/ no) and

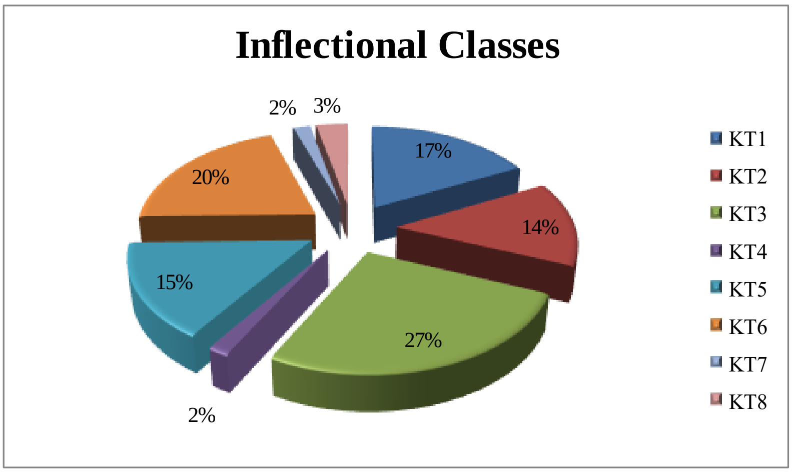
(vii.) *Origin – Calque* (Greek, English, French, German, Turkish, Italian, Slavic and Arabic).

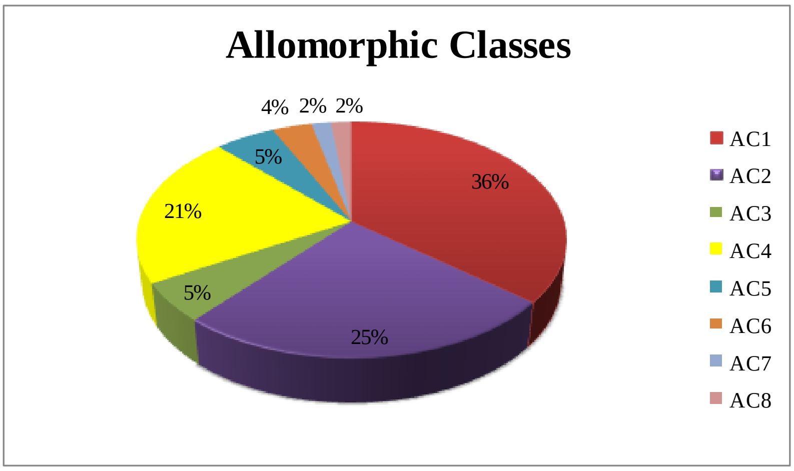
[1] More specifically, the classes are the following:

AC1 – zero allomorphy (e.g. άνθρωπ-ος ‘man’),
AC2 – α-deletion (e.g. καρδιά~ καρδι-, ‘heart’),
AC3 – η-deletion (e.g. ψυχή~ ψυχ- ‘soul’),
AC4 – ι-delection (e.g. παιδί~ παιδ- ‘child’),
AC5 – δ-addition (e.g. παπά-ς~ παπαδ- ‘priest’),
AC6 – τ-addition, (e.g. κύμα~ κυματ- ‘wave$_N$’)
AC7 – η-deletion/ε-replacement (e.g. πόλη~ πολ-~ πολε- ‘city’) and
AC8 – δ-addition/η-deletion (τεμπέλη-ς~ τεμπεληδ-~ τεμπελ- ‘lazy’).

4.3 Training data

This Greek model of maximum entropy was trained on a corpus of 4,677 inflected nouns (neither derived nor compound nouns), a sufficient sample of all eight inflectional classes; training data contain inflected nouns (stem and inflectional suffixes), which are not derived by other words or have derivational suffixes that are synchronically morphological opaque. Based on electronic dictionary of Triantafyllidis, all the nouns were manually imported and every feature of the model was checked with the help of the dictionary. From our data only 34.5% of nouns do not display allomorphy; therefore, the amount of allomorphs is quite high in the Greek language. AMIS produced weights for more than 20,000 features. It is expected that a model with more than 20,000 features for weights is quite heavy statistically due to the uncountable combinations of syllables and characters increased exponentially the size of our sets.





Graph 1: Statistics from our nominal training data based on inflectional and allomorphic classes

To evaluate the effectiveness of our model, a testing corpus with derived nouns that have any kind of stem, at least one nominal derivation suffix (in the most right part of the word) and an inflectional suffix, was created. This second corpus contains 2,755 carefully selected nouns to cover the full range of features and all the nominal derivational suffixes. We created ALLOMANTIS[2], a morphological prediction analyzer for nominal allomorphy, which takes an input imported data from our training corpora on AMIS. ALLOMANTIS replaces each word characteristic with the proper weight given by the training corpus from AMIS. The analyzer multiplies the weights of all attributes for each candidate allomorphic class and proceeds with the one with the largest result of multiplication; according to the model of maximum entropy, this is the winner and is identified by the ALLOMANTIS as the proper allomorphic class.

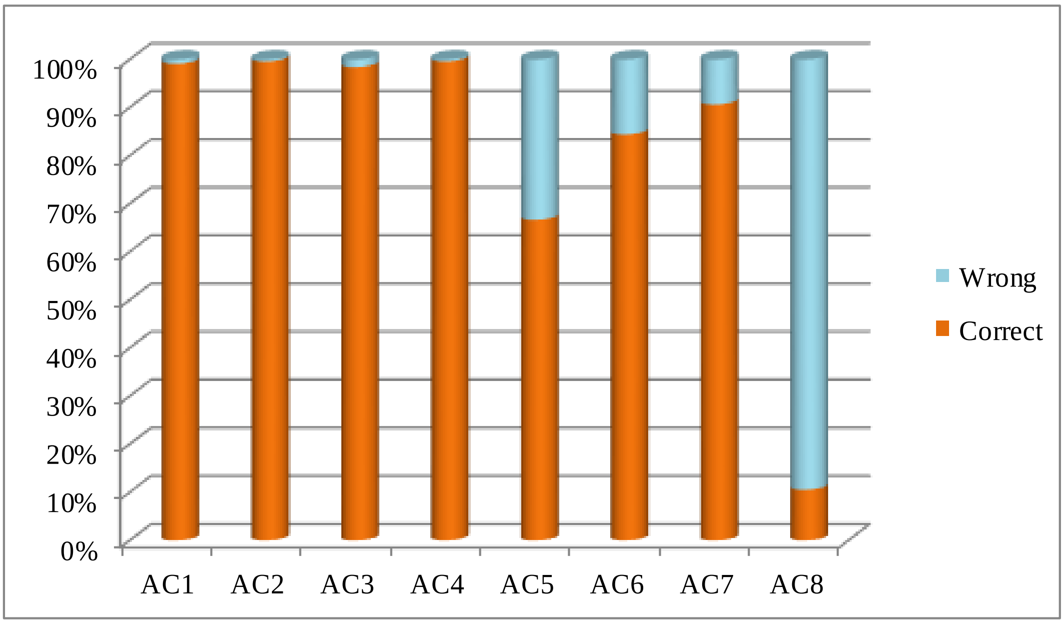
AC2	Positive affection weights		Negative affection weights	
	Syllable3_τζας	5,33E+01	Syllable1_δε	1,38E-01
	Syllable3_τη	1,24E+01	Syllable3_δι	1,29E-01
	Syllable2_χια	1,19E+01	Syllable3_μα	1,20E-01
	Syllable2_μπας	1,01E+01	Character2_μ	1,16E-01
	Syllable1_φιρ	9,93E+00	Syllable3_λο	7,79E-02
	Syllable2_γας	8,13E+00	Syllable3_ρο	7,70E-02
	Syllable2_γκας	7,46E+00	Character2_υ	7,10E-02
	Syllable2_ρια	6,88E+00	Syllable3_ο	6,16E-02
	Syllable1_μνα	5,90E+00	Syllable3_νι	3,67E-02
	Syllable3_βια	5,46E+00	Character1_η	4,37E-03
AC3	**Positive affection weights**		**Negative affection weights**	
	Syllable2_τρης	5,41E+02	Character3_θ	1,96E-01
	Syllable3_ρη	1,37E+02	Syllable2_τα	1,78E-01
	Syllable1_μεγ	7,62E+01	Syllable2_να	1,64E-01
	Syllable2_δης	5,31E+01	Syllable3_δα	1,24E-01
	Syllable3_πης	4,28E+01	Character2_σ	1,09E-01
	Syllable2_ντης	3,54E+01	Stress_antipenultimate	9,18E-02
	Syllable3_δης	2,87E+01	Syllable2_μα	7,96E-02
	Syllable3_φης	2,58E+01	Syllable3_τα	7,05E-02
	Syllable4_χη	2,19E+01	Origin_italian	2,43E-02
	Syllable1_ζη	2,05E+01	Origin_turkish	1,98E-02

Table 1: Sample of the assigned positive and negative affections weights by AMIS algorithm

[2] The blend ALLOMANTIS is a combination of 'αλλομορφία' (allomorphy) and 'μάντης' (seer, prophet) and the capital letters refer to the initials of the maximum entropy program AMIS.

4.4 Results

The overall accuracy (recall) of the model was 86.49% with the failure rate up to 13.51%. A detailed analysis of the model for each allomorphic class is shown in the following graph. More than 90% was achieved in several classes, as in the AC1 96.92%, AC2 95.97% and AC4 91.2%, whereas the two classes with the lowest percentage was AC5 (64.37%) and AC8 (9.28%), with the latter rates considered to be a strong flaw of (from) the average success.



Graph 2: Recall performance for each allomorphic class

	Precision	Recall	Failed
Allomorphic Class 1	99,78	99,09	0,91
Allomorphic Class 2	100	99,59	0,41
Allomorphic Class 3	99,05	98,43	1,57
Allomorphic Class 4	100	99,6	0,4
Allomorphic Class 5	69.78	66,67	33,33
Allomorphic Class 6	88,74	84,42	15,58
Allomorphic Class 7	91,11	90,64	9,36
Allomorphic Class 8	25,25	10,53	89,47
Average	93,46	91,43	8,57

Table 2: Precision and Recall performance for each allomorphic class

To improve the system, we tried a more rational approach to achieve a better performance. In the previous version of the AlloMantIS, we numbered syllables from left to right (i.e. ‘uranos’ ‘sky’ ου-syl1, ρα-syl2, νος-syl3), while in the updated version we followed the stress strategy for spelling, i.e, the ultimate syllable was numbered as first, the penultimate as second and so forth. The result of the upgraded version of ALLOMANTIS was the rise of the correct prediction to 91.43% with the failed cases to 8.47%. Indeed, the first four allomorphic classes reached 100%, but AC8 remained in a tragically low threshold (10.53%), as well as AC5 with 66.67% of erroneous estimations, since both of them are similar cases of Turkish nouns and derivational suffixes that have a slightly different allomorphy in derivation (two allomorphs, i.e. kanape~ kanapeδ ‘sofa’ vs. three allomorphs, i.e. bakali~ bakaliδ~ bakal ‘grocer). If we try to manually change the importance of stress values for these two groups, then the accuracy of the system will reach almost 100%. Additionally, some trials were performed to reduce the accompanied characteristics with the aim to create a less complex model with a similar high performance; nevertheless, it was only possible to eliminate a couple of features without affecting the overall performance.

5. Conclusions: Re-visiting computational prediction of derivational allomorphy

It is noteworthy that our model was trained by a corpus of inflected nouns (not created by the process of derivation and compounding) with the small number of the dataset and tested / evaluated by a corpus with derived nouns, since we tried to make our task more difficult. This choice was not arbitrary as it was based on Καρασίμος (2011) argument that the nominal derivational suffixes display similarities with nominal stems / roots, they participate in the same inflectional classes, and thus they exhibit the same allomorphic behavior. ALLOMANTIS correctly predicted allomorphy for more than 91% of the derived nouns of the testing corpus. It is expected that if ALLOMANTIS is trained with a corpus of inflected and derived nouns, then the prediction accuracy rate will be much higher. It was considered necessary in this primary testing stage of our model to provide a minimal help from the training corpus.

The probabilistic language models, the supervised machine learning algorithms and modern linguistic theory models appear to support the viewpoint of language processing, which is the result of a dynamic and on-line grammatical analysis of conflicting constraints. Extending this reasoning means that certain morphological phenomena or processes can be a result of a combinatorial analysis of morphological features that are (sometimes) assisted from data of other languages (phonology, semantics, etc.). We totally agree with Dell' Orletta et al. (2007) that we anticipate that this kind of research is bound to shed light on the integration of performance and competence factors in language study; additionally, it will make mathematical models of language increasingly able to accommodate richer and richer language structures, thus putting explanatory theoretical accounts to the test of a usage-based empirical verification. Therefore, we are planning to test the limits of other machine learning approaches of non-probabilistic nature, where the generalizability might be higher, and probably the problematic class of loans and calques will be overcome.

In this MaxEnt experiment, it is inferred how the existence of morphologically annotated corpora is essential for the effective conduct of morphological experiments in Greek. We have shown that a (supervised) probabilistic model applied to a corpus with quite rich annotated words can extract some basic principles that can be the keystone to construct a computational model to process the "unpredictable" and hard-to-deal phenomenon of allomorphy.

References

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. (Linguistic Inquiry Monograph 22). Cambridge, MA: MIT Press.

Berger, A., S. Della Pietra & V. Della Pietra (1996). A maximum entropy approach to natural language processing. *Computational Linguistics* 22 (1): 37-71.

Booij, G. (1997). Allomorphy and the autonomy of morphology. *Folia Linguistica* XXXI/1-2: 25-56.

Dell' Orletta, F., A. Lenci, S. Montemagni & V. Pirrelli (2007). Corpus-based modeling of grammar variation. In A. Sansò (ed), *Language Resources and Linguistic Theories*. (Materiali Linguistici 59). Milano: Franco Angeli, 38-55.

Goldsmith, J. (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27(2): 153-198.

Hafer, M. & S. Weiss (1974). Word segmentation by letter successor varieties. *Information Storage and Retrieval* 10: 371-385.

Harris, Z. (1955) [1970]. From phoneme to morpheme. *Language* 31: 190-222.

Harris, Z. (1967) [1970]. Morpheme boundaries within words: Report on a computer test. *Transformations and Discourse Analysis Papers* 73. University of Pennsylvania.

Καρασίμος, Α. (2011). *Υπολογιστική επεξεργασία της αλλομορφίας στην παραγωγή λέξεων της Νέας Ελληνικής*. Διδακτορική διατριβή, Πανεπιστήμιο Πατρών.

Karasimos, A. (2009). Comments on unsupervised morphology learning model: The case of Greek allomorphy. Paper presented at the *19th International Symposium in Theoretical and Applied Linguistics* (ISTAL19). 3-5 April 2009, Thessaloniki.

Koskenniemi, K. (1983). *Two-level morphology: a general computational model for word-form recognition and production*. Ph.D. Thesis, University of Helsinki.

Lieber, R. (1982). Allomorphy. *Linguistic Analysis*. Volume 10, Number 1, 1982, pp. 27-52.

Ling, C.X. & M. Marinov (1994) A symbolic model of the nonconscious acquisition of information. *Cognitive Science* 18: 595-621.

Maiden, M. (2004). When lexemes become allomorphs – On the genesis of suppletion. *Follia Linguistica* XXXXCIII/3-4: 227-256.

Malouf, R. (2016). Generating morphological paradigms with a recurrent neural network. *San Diego Linguistic Papers* 6: 122-129.
Mikolov, T. & G. Zweig (2012). Context dependent recurrent neural network language model. In *Proceedings of Speech Language Technology*: 234-239.
Pinker, S. & A. Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Mehler (eds), *Connections and Symbols (Special Issue: Cognition)* 28(1-2): 73-193.
Pirrelli, V. and M. Battista (2000a). The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Italian Journal of Linguistics (Rivista di Linguistica)* 12(II): 307-380.
Pirrelli, V. and M. Battista (2000b). On the interaction of paradigmatic and syntactic stem alternation in Italian conjugation. *Acta Linguistica Hungarica* 47(1-4): 289-314.
Ράλλη, Α. (2005). *Μορφολογία*. Αθήνα: εκδόσεις Πατάκη.
Ράλλη, Α. (2008). *Η Σύνθεση λέξεων: διαγλωσσική μορφολογική προσέγγιση*. Αθήνα: εκδόσεις Πατάκη.
Ralli, A. (1994). Feature representation s and feature-passing operations in Greek nominal inflection. *Proceedings of the 8*th *Symposium on English and Greek Linguistics*. Thessaloniki: English Department of Aristotle University of Thessaloniki, 19-46.
Ralli, A. (1999). Inflectional features and the morphological module hypothesis. In A. Kakouriotes (ed), *Working Papers in Theoretical and Applied Linguistics* 6. Thessaloniki: English Department of Aristotle University of Thessaloniki, 111-142.
Ralli, A. (2000). A Feature-based analysis of Greek nominal inflection. *Glossologia* 11-12: 201-228.
Ralli, A. (2007). On the role of allomorphy in inflectional morphology: Evidence from the dialectal varieties of Lesvos, Kydonies and Moschonisia. In G. Sica (ed), *Open problems in Linguistics and Lexicography*. Milano: Polimetrica, 123-154.
Roark, B. & R. Sproat (2007). *Computational Approaches to Morphology and Syntax*. Oxford: Oxford University Press.
Rumelhart, D. & J. McClelland (1986). On learning the past tense of English verbs. In J. McClelland & D. Rumelhart (eds), *Parallel Distributed Processing* 2. Cambridge, MA: MIT Press, 216-271.
Stump, G. (2001). *Inflectional Morphology*. Cambridge: Cambridge University Press.
Sundermeyer, M., R. Schlüter & H. Ney (2012). LSTM neural networks for language modeling. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech 2012)* vol. 1. Portland, Oregon: International Speech Communications Association, 195-197.
Sundermeyer, M., R. Schlüter & H. Ney (2015). From feedforward to recurrent LSTM neural networks for language modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23: 517-529.
Thornton, A.-M. (1997). Stem allomorphs, suffix allomorphs, interfixes or different suffixes? On Italian derivatives with antesuffixal glides. In G. Booij, S. Scalise & A. Ralli (eds), *Proceedings of the 1st Mediterranean Meeting of Morphology*, 86-98.
Yoshida, K. (2006). *AMIS – A maximum entropy estimator for feature forests. AMIS Manual*. Department of Computer Science, University of Tokyo. Available at: http://www.nactem.ac.uk/amis/manual.html [accessed: 11/07/2017].