

Anna Corazza, Simonetta Montemagni and Giovanni Semeraro (dir.)

Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016
5-6 December 2016, Napoli

Accademia University Press

Reassessing inflectional regularity in Modern Greek conjugation

Stavros Bompolas, Franco Alberto Cardillo, Marcello Ferro, Claudia Marzi and Vito Pirrelli

DOI: 10.4000/books.aaccademia.1721
Publisher: Accademia University Press
Place of publication: Torino
Year of publication: 2016
Published on OpenEdition Books: July 26, 2017
Series: Collana dell'Associazione Italiana di Linguistica Computazionale
Electronic EAN: 9788899982546

OpenEdition Books

http://books.openedition.org

Electronic reference
BOMPOLAS, Stavros ; et al. *Reassessing inflectional regularity in Modern Greek conjugation* In: *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016: 5-6 December 2016, Napoli* [online]. Torino: Accademia University Press, 2016 (generated 06 octobre 2023). Available on the Internet: <http://books.openedition.org/aaccademia/1721>. ISBN: 9788899982546. DOI: https://doi.org/10.4000/books.aaccademia.1721.

The text only may be used under licence CC BY-NC-ND 4.0. All other elements (illustrations, imported files) are “All rights reserved”, unless otherwise stated.

Reassessing inflectional regularity in Modern Greek conjugation

Stavros Bompolas
University of Patras, Greece
stavros.bompolas@gmail.com

Marcello Ferro
ILC-CNR Pisa, Italy
marcello.ferro@ilc.cnr.it

Claudia Marzi
ILC-CNR Pisa, Italy
claudia.marzi@ilc.cnr.it

Franco Alberto Cardillo
ILC-CNR Pisa, Italy
francoalberto.cardilllo@ilc.cnr.it

Vito Pirrelli
ILC-CNR Pisa, Italy
vito.pirrelli@ilc.cnr.it

Abstract

Paradigm-based approaches to word processing/learning assume that word forms are not acquired in isolation, but through associative relations linking members of the same word family (e.g. a paradigm, or a set of forms filling the same paradigm cell). Principles of correlative learning offer a set of dynamic equations that are key to modelling this complex dynamic at a considerable level of detail. We use these dynamic equations to simulate acquisition of Modern Greek conjugation, and we compare the results with evidence from German and Italian. Simulations show that different Greek verb classes are processed and acquired differentially, depending on their degrees of formal transparency and predictability. We relate these results to psycholinguistic evidence on Modern Greek word processing, and interpret our findings as supporting a view of the mental lexicon as an emergent integrative system.

Secondo l'approccio paradigmatico allo studio dell'elaborazione e dell'apprendimento lessicali, le parole di una lingua non sono acquisite in isolamento, ma attraverso legami associativi tra membri della stessa famiglia morfologica, la cui dinamica è modellata dalle equazioni dell'apprendimento correlativo. Il presente contributo offre una serie di esperimenti nei quali l'apprendimento del sistema verbale del greco moderno è simulato come un processo di auto-organizzazione dinamica di parole memorizzate in modo concorrente. I risultati mostrano chiari effetti di interazione dinamica tra trasparenza e regolarità morfologica nell'acquisizione di classi di forme del verbo greco.

1 Introduction

Issues of morphological (ir)regularity have traditionally been investigated through the prism of morphological competence, with particular emphasis on aspects of the internal structure of complex words (Bloomfield, 1933; Bloch, 1947; Chomsky and Halle, 1968; Lieber, 1980; Selkirk, 1984; among others). Within this framework, one of the most influential theoretical positions is that morphologically, phonologically, or/and semantically transparent words are always processed on-line through their constituent elements, whereas irregular, idiosyncratic (non-transparent) forms are stored (and retrieved) as wholes in the lexicon (Pinker and Prince, 1994). Likewise, Ullman and colleagues (1997) assume that the past tense formation of regular verbs in English requires on-line application of an affixation rule (e.g. *walk* > *walk+ed*), while irregular past tense forms, involving stem allomorphy (e.g. *drink* > *drank*), are retrieved from the lexicon.

Modern Greek introduces an interesting variation in this picture. First, stem allomorphy and suffixation are not necessarily mutually exclusive processes, but coexist in the same inflected forms (e.g. *lin-o* 'I untie' > *e-li-s-a* 'I untied', *aɣap(a)-o* 'I love' > *aɣapi-s-a* 'I loved'). Secondly, affixation rules may select unpredictable stem allomorphs: *aɣap(a)-o* 'I love' > *aɣapi-s-a* 'I loved', *for(a)-o* 'I wear' > *fore-s-a* 'I wore', *xal(a)-o* 'I demolish' > *xala-s-a* 'I demolished'. These cases suggest that inflectional (ir)regularity is not an all-or-nothing notion in Greek. Different inflectional processes may compound in the same words to provide a challenging word processing scenario (Tsapkini et al., 2004). From this perspective, Modern Greek offers the opportunity to test traditional hypotheses of grammar and lexicon interaction in word processing and learning, to explore the potential of single, distributed mechanisms in addressing word processing challenges (Alegre and Gordon, 1999; Baayen, 2007).

1.1 The evidence

Modern Greek conjugation is stem-based, each fully inflected verb form requiring obligatory suffixation of person, number and tense markers that attach to either a bare or a complex stem in both regular (*aɣap-o* 'I love' ~ *aɣapis-a* 'I loved') and irregular verbs (*pern-o* 'I take' ~ *pir-a* 'I took'). Unlike English speakers, Greek speakers must always resort to an inflectional process to understand or produce a fully inflected form, no matter how regular the form is (Terzi et al., 2005: 301).

Classifying a Greek verb as either regular or irregular thus requires observation of the stem formation processes on whose basis agreement and tense suffixes are selected. Accordingly, it is assumed that the presence or absence of the aspectual marker is a criterion for assessing the degree of regularity of a Greek verb. In particular, so-called "sigmatic" past-tense forms (e.g. *aɣap-o* ~ *aɣapis-a*) are traditionally considered to be regular, in that they involve a segmentable marker (*-s-*) combined with phonologically predictable or morphologically systematic stem-allomorphs. Asigmatic past-tense forms (e.g. *pern-o* ~ *pir-a*), in contrast, exhibit typical properties of irregular inflection, since they involve unsystematic stem allomorphs (in some cases suppletive stems), and no segmentable affixes marking perfective aspect. This distinction has also been supported by psycholinguistic evidence (Stamouli, 2000; Tsapkini et al., 2001, 2002a,b,c, 2004; Mastropavlou, 2006; Varlokosta et al., 2008; Stavrakaki and Clahsen, 2009a,b; Stathopoulou and Clahsen, 2010; Stavrakaki et al., 2012; Konstantinopoulou et al., 2013; among others), suggesting that sigmatic past-tense forms are typically produced on-line by rules, and asigmatic forms are stored and accessed from the mental lexicon.

However, careful analysis of the Greek verb system appears to question such a sharp processing-storage divide. In particular, Greek data provide the case of a mixed inflectional system where both stored allomorphy and rule-based affixation are simultaneously present in the formation of past tense forms. Ralli (2005) provided a classification of verb paradigms which is based on two criteria; firstly, the presence vs. absence of the sigmatic affix and, secondly, the presence vs. absence of (systematic) stem allomorphy. As a result, we can define the following three classes (see also Tsapkini et al., 2001, 2002a,b,c, 2004):

(i) an affix-based class, requiring the presence of the aspectual marker *-s-*, and including verbs with a predictable phonological stem-allomorph (e.g., *lin-o* 'I untie' ~ *e-li-s-a* 'I untied', *ɣraf-o* 'I write' ~ *e-ɣrap-s-a* 'I wrote');

(ii) a mixed class where active perfective past tense forms are produced by affixation of the aspectual marker *-s-* to a systematic morphological stem-allomorph (e.g., *mil-o* 'I speak' ~ *mili-s-a* 'I spoke');

(iii) an idiosyncratic verb class whose forms are based on non-systematic stem-allomorphy (requiring stem-internal alternation or suppletion) or no stem-allomorphy at all, and no (sigmatic) aspectual marker (e.g., *pern-o* 'I take' ~ *pir-a* 'I took', *tro-o* 'I eat' – *e-fag-a* 'I ate', *krin-o* 'I judge' – *e-krin-a* 'I judged').

It should be noted that, in regular Greek verbs, transparency/systematicity and predictability are not mutually implied. The morphologically-conditioned allomorphy of class-(ii) verbs requires a systematic pattern of perfective stem formation, namely $X_{(a)}$ ~ X + V (e.g. *aɣap(a)-* > *aɣapi-*), where 'X' is a variable standing for the bare stem, 'V' stands for a vowel, and the subscripted '(a)' indicates an optional 'a', forming a Modern Greek free variant of the imperfective stem (e.g. *aɣapo* ~ *aɣapao*, see Ralli, 2005, 2007). The variable *V* in the perfective stem can be instantiated as an *i*, *e* or *a*, and cannot be predicted from the bare stem. On the other hand, the phonologically-conditioned allomorphs of class-(i) verbs (e.g. *lin-* > *e-li-s-*) are the outcome of exception-less phonological rules, which nonetheless obfuscate a full formal correspondence (transparency) between the imperfective stem and the perfective stem.

Evidence from language acquisition and experimental psycholinguistics shows that perception of formal transparency between imperfective and perfective Greek stems plays a prominent role in human word processing strategies (Tsapkini et al., 2002c: 116, 2004: 616; Stavrakaki and Clahsen, 2009a: 117; Stathopoulou and Clahsen, 2010: 872). More specifically, lack of full formal nesting between imperfective and perfective stems (compare *aɣap-o* 'I love' ~ *aɣapi-s-a* 'I loved' *vs.* *ðulev-o* 'I work' ~ *ðulep-s-a* 'I worked') appears to have an extra processing cost (Tsapkini et al., 2002c: 116).

To sum up, analysis of Greek data offers evidence of graded levels of morphological regulari-

ty, based on the interaction between formal transparency (degrees of stem similarity) and (un)predictability of stem allomorphs. The evidence appears to question a dichotomous view of storage vs. rule-based processing mechanisms. In fact, no sharp distinction between affix processing and allomorph retrieval can account for the interaction of formal transparency and predictability in Greek word processing. On the one hand, rule-based mechanisms are called for to account for transparency effects of stem allomorphy on word processing. On the other hand, storage is required if the *same* allomorphs cannot be predicted. In the remainder of this paper, we test the hypothesis that this evidence is compatible with a parallel processing architecture (a Temporal Self-organising Map) where processing and storage are in fact mutually implied.

2 TSOMs

Temporal Self-organising Maps (TSOMs, Ferro et al., 2011; Marzi and Pirrelli, 2015; Pirrelli et al., 2015) are unsupervised artificial neural networks that learn to dynamically memorise input strings as chains of maximally-responding processing nodes (*Best Matching Units* or *BMU*s), whose level of sensitivity to input symbols in specific contexts is a continuous function of the distributional regularities of the input symbols during training. In a TSOM, each processing node has two layers of synaptic connectivity: an input layer, connecting the node to the current input stimulus (e.g. the letter of a written word), and a (re-entrant) temporal layer, connecting the node to all other nodes.

Given the *BMU* at time *t*, the temporal layer encodes the expectation of the current *BMU* for the node to be activated at time *t+1*. The strength of the connection between consecutively activated *BMU*s is trained through the following principles of correlative learning (compatible with Rescorla-Wagner (1972) equations):

Given the input bigram *ab*, the connection strength between *BMU* of *a* at time *t* and *BMU* of *b* at time *t+1* will

- increase if *a* often precedes *b* in training (entrenchment)
- decrease if *b* is often preceded by a symbol other than *a* (competition).

The interaction between entrenchment and competition in a TSOM accounts for important dynamic effects of self-organisation of stored words (Marzi et al., 2014, 2016). In particular, high-frequency words tend to recruit specialised (and stronger) chains of *BMU*s, while low-frequency words are responded to by more "blended" (and weaker) *BMU* chains. In what follows, we report how well a TSOM can accommodate the complexity of the Greek verb system, by controlling factors such as word frequency distribution, degrees of inflectional regularity and word length.

2.1 The experiment

To allow pairwise comparison with existing experimental evidence on German and Italian (Marzi et al., 2016), the Greek training dataset was designed to contain 50 top-ranked paradigms by cumulative token frequency, for a total of 750 verb forms, whose frequency distributions were sampled from the FREQcount section of the Greek SUBTLEX-GR corpus (BCBL, 2016). From each paradigm, 15 inflected forms were extracted: the full set of present indicative (6) and simple past tense (6) forms, and the singular forms of simple future (3). As we were mainly interested in effects of global paradigm-based organisation of active voice indicative forms, we excluded paradigms with systematic gaps, impersonal verbs, and deponent verbs. We included high-frequency paradigms with suppletive forms or/and non-systematic allomorphy (Ralli, 2007, 2014) as attested in the training set.

The dataset was administered to a 42x42 node map for 100 learning epochs. Word frequencies in the training data were a function of the real word frequency distribution in the reference corpus, fitted in the 1-1000 range. To control for random variability, we repeated the experiment 5 times.

For each repetition, we then assessed how well the map could acquire the 750 input forms, using the task of Word Recall as a probe. Word recall is defined as the process of retrieving a word form from its chain of *BMU*s. Successful recall is possible if inter-node connections on the temporal layer are finely tuned to the distribution of symbols in the training data. The more accurate the re-entrant temporal coding is, the easier for the map to retrieve the symbols of a word in their appropriate order. We make the further reasonable assumption that a word is acquired by a TSOM when the map is in a position to recall the word accurately and consistently from its *BMU* chain. Average recall accuracy at epoch 100 turned out to be considerably high: 99.6 % (std = 0.1%).

3 Data analysis

Results were analysed using Linear Mixed Effects (LME) models with experiment repetitions and training items as random variables.

Figure 1 shows the marginal plot of the interaction between word length and regular vs. irregular verb classes for German, Italian and Greek, using an LME model fitting word learning epochs, with (log) word frequency, inflectional class and word length as fixed effects. In German and Italian, the distinction between regular and irregular paradigms is based on the criterion of absence vs. presence of stem allomorphy across all forms of a paradigm (Marzi et al., 2016). In Greek, we consider regular all paradigms showing a sigmatic perfective stem, and irregular those with an asigmatic perfective stem.

Unlike German and Italian (Figure 1, top and middle panels), where irregulars tend to be acquired systematically later than length-matched regulars are, and no significant interaction is found, Greek data (Figure 1, bottom panel) show an interesting crossing pattern: shorter irregulars are acquired earlier than length-matched regulars of comparable frequency, but long irregulars are acquired later than long regulars.

Marzi and colleagues (2016) account for earlier learning epochs of both German and Italian regulars as an effect of stem transparency on cumulative input frequencies. With German and Italian regular verbs, stems are shown to the map consistently more often, since they are transparently nested in all forms of their own paradigm. This makes their acquisition quicker, due to specialised chains of stem-sensitive *BMUs* getting more quickly entrenched. Once a stem is acquired, it can easily be combined with a common pool of inflectional endings for tense and agreement, simulating an effect of (nearly) instantaneous (or paradigm-based, as opposed to item-based) acquisition. In contrast, Greek verb classes always present stem allomorphy throughout their paradigms, no matter whether allomorphy is systematic, phonologically motivated or unsystematic. In regular verbs, where perfective stem formation requires *-s-* affixation, perfective stems are systematically longer than their imperfective counterparts, and are acquired after them. Nonetheless, since imperfective stems are redundantly embedded in perfective stems, learning a long regular perfective form is easier (i.e. it takes a comparatively shorter time) than learning an irregular perfective form of comparable length. This is, again, a regularity-by-transparency effect, and explains why long regular forms tend to be acquired (on average) more easily than long irregular forms.

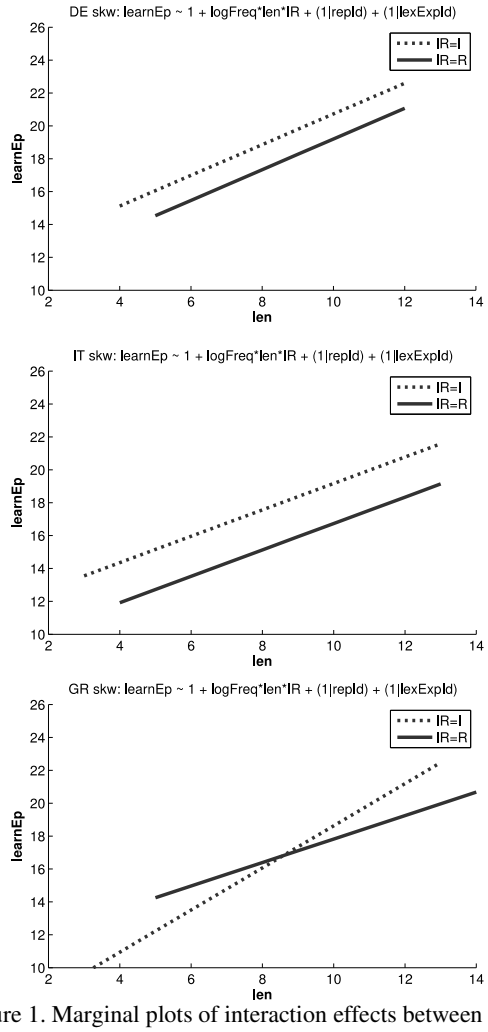


Figure 1. Marginal plots of interaction effects between word length and inflectional regularity in an LME model fitting word learning epochs in German (top), Italian (middle) and Greek (bottom). Solid lines = regulars, dotted lines = irregulars.

To further investigate the impact of degrees of formal transparency on the processing of Greek verb forms, we conducted an LME analysis of the interaction between word length and classes of (ir)regularity in word recall (Figure 2). When we control for length, regular verbs with systematic morphological allomorphs (e.g. *ayap(a)-o ~ ayapi-s-a*, solid line in the plot) are recalled more easily than regular verbs with phonological allomorphs (e.g. *ðulev-o ~ ðulep-s-a*, dashed line in the plot). Notably, both classes are easier to recall than asigmatic (irregular) verbs (dotted line in the plot), which show, in most cases, formally more opaque allomorphs (e.g. *pern-o ~ pir-a*). As shown by the difference in slope between the solid line and the other two lines of Figure 2,

facilitation increases with word length, supporting our interpretation of the crossing pattern in the bottom panel of Figure 1.

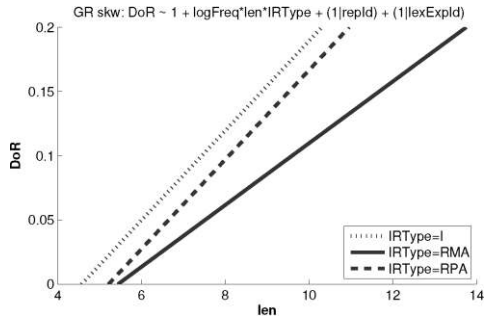


Figure 2. Marginal plot of interaction effects between length (x axis), and degrees of stem regularity in an LME model fitting Difficulty of Recall (y axis) by TSOMs trained on Greek verb forms

4 Conclusions

Data analysis highlighted a hierarchy of regularity-by-transparency effects that appear to have consequences on morphological processing. In particular, the evidence offered here emphasises the role of formal preservation of the stem (or stem transparency) in the paradigm as a key facilitation factor for morphological processing.

Our case study is focused on a distinguishing characteristic of Greek conjugation: all verb paradigms, both regulars and irregulars, involve (unpredictable) stem allomorphy in past-tense formation. Hence, the difference between regular and irregular verbs could not be attributed to the presence or absence of stem allomorphy as is the case with other languages, such as English and Italian (and, to a lesser extent, German), but rather to the type of stem allomorphy itself. The finding (Figure 2) that stem-final systematic change, as in the case of regulars, led to significantly easier recall than stem-internal vowel changes and non-systematic/non-predictable stem-final change, as is the case of irregulars, lends support to the conclusion that the type of stem allomorphy is what determines the different levels of morphological regularity in Greek. Crucially, this seems to involve a regularity-by-transparency interaction, with predictability playing second fiddle.

Our data meet growing psycholinguistic evidence pointing in the same direction, to emphasise the importance of formal redundancy for speakers' perception of morphological structure. Furthermore, it paves the way to the definition of a performance-oriented notion of inflectional regularity that may ultimately cut across traditional competence-based classifications, which presuppose a hardly tenable subdivision of work between storage and processing.

References

Alegre, M. and P. Gordon. 1999. Frequency effects, and the representational status of regular inflections. *Journal of Memory, and Language*, 40: 41-61.

Baayen, H. R. 2007. Storage, and computation in the mental lexicon. In: G. Jarema and G. Libben (eds.), *The Mental Lexicon: Core Perspectives*. Amsterdam: Elsevier, 81-104.

BCBL. 2016. *SUBTLEX-GR: The corpus*. Donostia: Basque Center on Cognition, Brain, and Language. [Retrieved from: http://www.bcbl.eu/subtlex-gr/, 12-04-2016].

Bloch, B. 1947. English verb inflection. *Language*, 23: 399-418.

Bloomfield, L. 1933. *Language*. New York: Henry Holt, and Co.

Chomsky, N. and M. Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Ferro, M., Marzi, C. and V. Pirrelli. 2011. A Self-Organizing model of word storage and processing: implications for morphology learning. *Lingue e Linguaggio*, X(2): 209-226.

Konstantinopoulou, P., Stavrakaki, S., Manouilidou, C. and D. Zafeiriou. 2013. Past tense in children with focal brain lesions. *Stem-, Spraak- en Taalpathologie*, 18(1): 90-94.

Lieber, R. 1980. *On the organization of the lexicon*. PhD thesis. Cambridge: MIT.

Marzi, C., Ferro, M. and V. Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio*, XIII(2): 263-290.

Marzi, C. and V. Pirrelli. 2015. A neuro-computational approach to understanding the Mental Lexicon. *Journal of Cognitive Science*, 16(4): 493-534.

Marzi, C., Ferro, M., Cardillo, F. A. and V. Pirrelli. 2016. Effects of frequency, and regularity in an integrative model of word storage, and processing. *Italian Journal of Linguistics*, 28(1): 79-114.

Mastropavlou, M. 2006. *The effect of phonological saliency, and LF-interpretability in the grammar of Greek normally developing, and language impaired children*. Ph.D. thesis. Thessaloniki: Aristotle University of Thessaloniki.

Pinker, S. and A. Prince. 1994. Regular, and irregular morphology, and the psychological status of rules of grammar. In: S. D. Lima, R. L. Corrigan and G. K. Iverson (eds.), *The reality of linguistic rules*. Amsterdam: Benjamins, 321-351.

Pirrelli, V., Ferro, M. and C. Marzi. 2015. Computational complexity of abstractive morphology, In: M. Baerman, D. Brown and G. Corbet (eds.), *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press, 141-166.

Ralli, A. 2005. *Morfologia* [Morphology]. Athens: Patakis.

Ralli, A. 2007. On the role of allomorphy in inflectional morphology: evidence from dialectal variation. In: G. Sica (ed.), *Open problems in linguistics, and lexicography*. Milano: Polimetrica, 123-152.

Ralli, A. 2014. Suppletion. In: G. K. Giannakis (ed.), *Encyclopedia of Ancient Greek language, and linguistics*, Vol. 3. Leiden, Boston: Brill, 341-344.

Selkirk, E. 1984. *Phonology, and Syntax*. The MIT Press.

Stamouli, S. 2000. Simfonia, xronos ke opsi stin eliniki idiki glosiki diataraxi [Agreement, tense, and aspect in specific language impairment in Greek]. In: *Proceedings of the 8th symposium of the Panhellenic Association of Logopedists*. Athens: Ellinika Grammata.

Stathopoulou, N. and H. Clahsen. 2010. The perfective past tense in Greek adolescents with Down Syndrome. *Clinical Linguistics & Phonetics*, 24(11): 870-882.

Stavrakaki, S. and H. Clahsen. 2009a. The perfective past tense in Greek child language. *Journal of Child Language*, 36: 113-142.

Stavrakaki, S. and H. Clahsen. 2009b. Inflection in Williams Syndrome. The perfective past tense in Greek. *The Mental Lexicon*, 4: 215-238.

Stavrakaki, S., Koutsandreas, K. and H. Clahsen. 2012. The perfective past tense in Greek children with specific language impairment. *Morphology*, 22: 143-171.

Terzi, A., Papapetropoulos, S. and E. D. Kouvelas. 2005. Past tense formation, and comprehension of passive sentences in Parkinson's disease: Evidence from Greek. *Brain, and Language*, 94: 297-303.

Tsapkini, K., Jarema, G. and E. Kehayia. 2001. Manifestations of morphological impairments in Greek aphasia: A case study. *Journal of Neurolinguistics*, 14: 281-296.

Tsapkini, K., Jarema, G. and E. Kehayia. 2002a. A morphological processing deficit in verbs but not in nouns: A case study in a highly inflected language. *Journal of Neurolinguistics*, 15: 265-288.

Tsapkini, K., Jarema, G. and E. Kehayia. 2002b. The role of verbal morphology in aphasia during lexical access: Evidence from Greek. In: E. Fava (ed.), *Clinical linguistics, and phonetics: Theory, and applications in speech pathology, and therapy*. Amsterdam/Philadephia: John Benjamins, 315-335.

Tsapkini, K., Jarema, G. and E. Kehayia. 2002c. Regularity revisited: Evidence from lexical access of verbs, and nouns in Greek. *Brain, and Language*, 81: 103-119.

Tsapkini, K., Jarema, G. and E. Kehayia. 2004. Regularity re-revisited: Modality matters. *Brain, and Language*, 89: 611-616.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. and S. Pinker. 1997. A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, 9(2): 266-276.

Varlokosta, S., Arhonti, A., Thomaidis, L. and V. Joffe. 2008. Past tense formation in Williams syndrome: evidence from Greek. In: A. Gavarro and M. Freitas (eds.), *Proceedings of GALA 2007*. Cambridge: Cambridge Scholars Publishing, 483-491.

Rescorla R. and A. Wagner. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: A. H. Black and W. F. Prokasy (eds.), *Classical conditioning ii*. New York: Appleton-Century-Crofts, 64-99.