A Mathematical Foundation of Gradient Harmonic Grammar Modeling of Greek Lexical Stress

M. Kanetidis[1], I. Mittas[1], K. Kosmidis[1], A. Revithiadou[1], G. Markopoulos[2], P. Argyrakis[1], M. Tsouchnika[1], Ei. Apostolopoulou[1,4], V. Apostolouda[1]

[1]AUTh, [2]University of the Aegean, [3]UiT

This study explores the complexities of stress assignment in the Greek lexical stress system, focusing on the phonology-morphology interface. Greek has lexically determined stress (Revithiadou 1999) (cf. languages with phonologically determined stress, Hayes 1995). Variability in stress assignment is influenced by factors like word size and inflection/noun class. In Greek, stress may fall on any of the last three syllables, as in [ˈðaskalos] 'teacher', [kuˈɲaðos] 'brother-in-law', [ceraˈvnos] 'thunder'. Prior studies suggest that stress assignment in adult native speakers is affected by word size and noun class, with a preference for antepenultimate stress in trisyllabic nouns ending in *-os* or *-o* (e.g., Apostolouda 2018). These findings align with data from lexical resources like the *2002 Reverse Dictionary* and an annotated version of *Clean* (Propopapas et al. 2012) (Apostolouda 2018).

To model this variation, we use *Gradient Harmonic Grammar* (GHG) (Smolensky & Goldrick 2016), which is built on a system of weighted constraints. The employed constraint set includes DEP ("do not insert structure"), MAX ("do not delete structure"), TROCHEE ("display stressed-unstressed pattern"), and ALIGN-R ("align trochee with the right edge"). Each constraint bears a weight, i.e., a numerical value that reflects its importance in the evaluation of linguistic forms. Higher weights lead to greater penalties in case of a violation. The degree of violation is also influenced by the *activity level*, i.e., the strength, of a stress property. The total penalty score incured by a stress pattern corresponds to its *harmony*, which translates into a probability of occurrence. GHG ranks the possible stress patterns according to their harmony.

This study provides a mathematical framework for computing stress probabilities, offering a detailed modeling of stress distribution across noun classes based on their endings. We construct three sets of GHG based on stress patterns in the Lexicon, the annotated Clean corpus, and a lexical database, i.e., *HelexKids* (Terzopoulos et al. 2016), respectively. These combine to give a fine-grained view of stress assignment across Greek noun classes. Significantly, our mathematical approach reveals a flexibility in constraint weights, suggesting novel insights into constraint behavior not previously thought in GHG models.

This interdisciplinary approach emphasizes the importance of integrating linguistic theory into computational models and applying complex systems theory to real-world challenges. Our findings highlight GHG's strength as a stochastic model for capturing variable stress patterns, paving the way for AI systems to produce more natural speech and better interpret prosody.