# Dual-route Connectionist Model of Greek Spelling

**Ioanna Katidioti (ikatidio@phs.uoa.gr)**
Graduate Program in Cognitive Science, Athens University Campus
GR-15771 Athens, Greece

**Ian C. Simpson (isimpson@ugr.es)**
Departamento de Psicología Evolutiva y de la Educación, Campus de Cartuja s/n
18071 Granada, Spain

**Athanassios Protopapas (aprotopapas@phs.uoa.gr)**
Department of Philosophy & History of Science, University of Athens
GR-15771 Athens, Greece

**Keywords:** connectionist model; spelling; phoneme-grapheme mapping

## Introduction

We created a dual-route connectionist model of Greek spelling. The model maps sequences of phonemes to corresponding sequences of graphemes, using a sublexical and a lexical route, i.e., phonographemic information and word knowledge, respectively. It is based on the model of Houghton and Zorzi (2003), but handles words up to 5 syllables long, with full connectivity between the syllables. Greek has 37 phonemes and 84 graphemes related via 118 mappings with 80,3% consistency (spelling) (Protopapas & Vlahou, 2009). Model architecture is as follows:

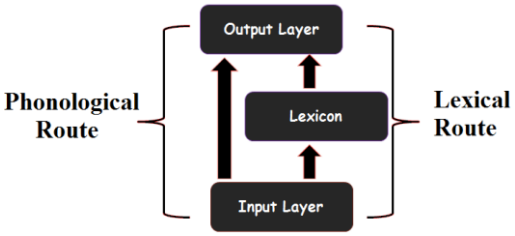


Figure 1: Dual-route model of spelling

## Input-Output Representation

The representation is syllabic and nucleus-centered. There are 4 consonant slots on each side of the vowel. The orthographic slots are occupied by graphemes, not letters.

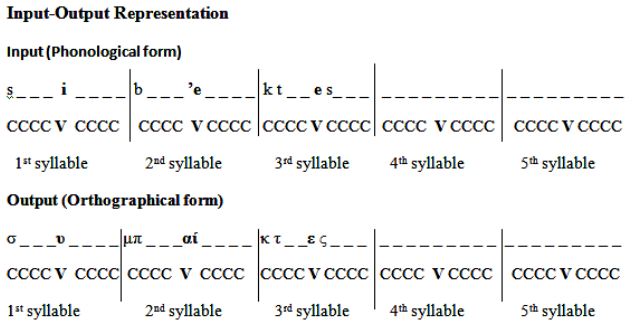


Figure 2: Input and output representation

## Training and parameters

To simulate spelling development using children's data, we trained the model to a corpus of 30,391 words from elementary school books. The model was trained for 30 epochs, with learning rate 0.02 and no weight pruning. During spelling, feedback was set to a value of 0.2.

### Results

Using both routes, the entire training set is spelled correctly. Using only the phonological route, 65.2% of the training set is spelled correctly and almost all errors are phonologically plausible. By adding a small contribution from the lexical route we were able to simulate Grade 3-4 children's data of 48 words. In the simulation, 13 out of 14 mistakes were the same as those made by the children, and 11 of these were the most typical.

## Problems

The model made two kinds of phonologically implausible mistakes: it spelled /s/ inside 19 words with "ς" (which is only used word-finally) and it also omitted a grapheme in a few words. In addition, the model has two problems: (a) the number of cycles needed to compute the output don't always correspond to the difficulty of the word and (b) certain palatal consonants were consistently misspelled (e.g. /ca/ as "κα" instead of "κια").

## Empirical validation

Greek has a number of ambiguous phonemes, the alternative spellings of which appear with different frequency (Protopapas & Vlahou, 2009). For example, in our training corpus, the phoneme /o/ is spelled with the letter "ο" 74% of the time and with "ω" 26%. Due to frequency-sensitive training the model usually spells the ambiguous phonemes with the highest-frequency grapheme. However, due to asymmetries in the distribution of consonant-vowel co-occurrences, this is not always the case. That is, the model will use the less frequent graphemic variant of a phoneme when more likely in the particular phonographemic context. If the model corresponds to human spelling performance,

children should also be more likely to choose the less frequent graphemes in the same contexts.

To test this prediction, we created two groups of 39 nonwords each, with ambiguous phonemes (o, e, i and g). **Group A** included nonwords spelled by the model with a low-frequency grapheme ("ω", "αι", "οι", "υ" and "γγ"). This was accomplished by inspection of the model's weights, choosing consonants with strong weight connections to target graphemes. **Group B** included similar nonwords (same number of phonemes and consonant-vowel structure) that were spelled by the model with the high-frequency graphemic alternative ("ο", "ε", "ι" and "γκ"). For example, nonwords /xo΄θafo/ and /mo΄ðamo/ were spelled by the model as "χωθάφο" and "μοδάμο", respectively (note the ω/o difference in the second position).

### Participants

177 students of the elementary Grades 5-6 participated in the experiment. Each child spelled 39 nonwords dictated by the experimenter.

### Results

The relative proportion of frequent vs. infrequent grapheme used by the children in each nonword group was examined for each phoneme using generalized linear mixed-effects models in R (function lmer of package lme4). The interaction of item group (A vs. B) by grapheme frequency (high vs. low) was significant in every case (i.e., for every phoneme tested), indicating that participants wrote more Group A items with a low-frequency grapheme than Group B items.

- For /o/ (ο-ω): $\beta = -2.87$ z = −4.7, $p < .0005$
- For /e/ (ε-αι): $\beta = 2.22$ z = 4.77, $p < .0005$
- For /i/ (ι-οι): $\beta = -2.04$ z = −2.82, $p = .005$
- For /i/ (ι-υ): $\beta = -2.12$ z = −4.62, $p < .0005$
- For /g/ (γκ-γγ): $\beta = .76$ z = 2.92, $p = .004$

## Discussion

The model spells known words perfectly, based on the lexical route. When only the phonological route is used, almost all errors are phonologically plausible. The model also simulates children's data successfully. We created nonwords using the model's weights in order to promote the use of low-frequency graphemes for ambiguous phonemes. Children were influenced by the context of ambiguous phonemes, which indicates that the frequency of phoneme-grapheme co-occurrence affects spelling. In conclusion, our model is a useful tool for exploring the development and difficulties of Greek spelling

## References


Houghton, G. & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20,* 115–162.

Mouzaki, A., & Protopapas, A. (Eds.) (2010) *Ορθογραφία: Μάθηση και διαταραχές [Spelling: Learning and disorders]*. Athens: Gutenberg.

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods. 41*, 991–1008.