

Abstract

Past research has demonstrated cross-linguistic, cross-modal, and task-dependent differences in neighborhood density effects, indicating a need to control for neighborhood variables when developing and interpreting research on language processing. The absence of a centralized database of neighborhood information has led to the inconsistent identification of neighbors, particularly across languages. The goals of the present paper are two-fold: (1) to introduce CLEARPOND (Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities), a centralized database of phonological and orthographic neighborhood information, both within and between languages, for five commonly-studied languages – Dutch, English, French, German, and Spanish; and (2) to show how CLEARPOND can be used to compare general properties of phonological and orthographic neighborhoods across languages in order to determine where and how languages differ in respect to their neighborhoods.

CLEARPOND allows researchers to input a word or list of words and obtain their phonological and orthographic neighbors, neighborhood densities, mean neighborhood frequencies, word lengths by number of phonemes and graphemes, and spoken-word frequencies. Neighbors can be defined by substitution, deletion, and/or addition, and the database can be queried separately along each metric or summed across all three. Neighborhood values can be obtained both within and across languages, and outputs can optionally be restricted to neighbors of higher frequency. To enable researchers to more quickly and easily develop stimuli, CLEARPOND can also search by features, generating lists of words that meet precise criteria, such as a specific range of neighborhood sizes, lexical frequencies, and/or word lengths.

CLEARPOND is freely-available to researchers and the public as a searchable, online database and for download at <http://clearpond.northwestern.edu>.