

The Linguistic Distribution of Relational Categories

Micah Goldwater
Northwestern University

Jon Willits
University of Wisconsin

Abstract: Behavioral research has distinguished relational categories (e.g., barrier), which are defined by relations among entities, from feature-based categories (e.g., vegetable), which are defined by sets of descriptive features intrinsic to entities (e.g., Gentner & Kurtz, 2005; Rein, Goldwater, & Markman 2010). Corpora research has demonstrated that category structure is reflected in their distribution in natural language texts (e.g., Willits, 2009). The current project connects these two lines of research by examining the distributions of both kinds of categories. Findings include: Feature-based categories' distributions are more similar to each other than relational categories' are to each other. Relational categories appear in more diverse contexts than feature-based; however relational categories are "anchored" by a single frequent collocate to a greater degree than are feature-based categories. We discuss relations between corpus measures and behavioral ratings and consider theoretical implications for category representation.