In the corpus analysis, we are building an inventory of specific VACs identified in the COBUILD Grammar Patterns volume (Francis, Hunston, & Manning, 1996) using the BNC. We define searches against part-of-speech categories and dependency relations produced by three different parsers (RASP [Briscoe et al. 2006], Stanford [de Marneffe et al 2006] and C&C [Curran et al. 2007]) and take a consensus of whether a sentence matches the VAC pattern. Then for each VAC, such as V across n, we record the distribution of the verb types, their token frequencies and sentence contexts. We determine the degree to which the distributions are Zipfian (e.g. come 474 ... spread 146 ... throw 17 ... stagger 5). Statistical analyses (MI, Delta-P, Chi-Square) examine the associations between verbs and constructions (e.g. scud, skitter, sprawl, flit have the strongest association with V across n). WordNet and other semantic resources are used to measure the cohesion of the types in each distribution (e.g., semantic fields TRAVEL and MOVE most frequent for V across n). These data allow us to make predictions regarding language users’ knowledge of verbs in constructions.

In psycholinguistic experiments we use free association tasks to test these predictions. We have native and non-native speakers of English think of the first word that comes to mind to fill the V slot in a particular VAC frame. The range of the verbs that they generate, and their speed of access, inform us about the representation of these VACs in the human mind. For each VAC, we compare the results from the experiments and the corpus analysis in terms of verb selection preferences. This research demonstrates the productive synergy of corpus linguistic and psycholinguistic methods and findings.

References


