Appendix Six

Methodological Discussion Regarding Endemic Function Pairs

As noted in the main text, when it comes to endemic function, the pattern of sharp fall-off as we lose semantic specificity, although clear in the sample case (bird::fly), is a bit muddied in our full dataset, breaking down with our other two pairs, people::walk and water::flows. This may be because endemic functions are just not as clearly signalled as contrastive or part-whole pairs. It is, in our opinion, more likely the result of methodological errors. Water::flow is a well-chosen endemic function pair, but it was probably a mistake to pick "fire" (*huo* 火) as the semantically-related control, since it also commonly appears in conjunction with water in the list of the Five Phases (*wuxing* 五行). In other words, as in the case of summer::spring, we made a mistake in choosing a semantically-related term that also commonly appears with the target in compounds. It probably would have been better to choose a term related to 'wetness' or 'moisture' instead.

In the case of 'people/humans,' *ren* 人 is an extremely high-frequency term that also appears in many compounds. As noted in the topic modeling study below, it is common enough to have been included in our "stop list" of highly pervasive terms with limited semantic use. Moreover, as we note in the hierarchical clustering study, many species besides people "walk/run" (*zou* 走)*,* so we should have chosen a species-verb pair that was more specific, as in bird-fly. Finally, one of the compounds *ren* appears in is *renxing* 人性; we, therefore, erred in picking our semantically-related term in this case as well, since the chosen term is not just semantically-related, but also linked with the target term in a common compound or set phrase, as with water-fire and the seasons.

These are all issues that can, and will, be untangled in follow-up studies. We choose to report our original results not only in the interest of transparency, to also to show how challenging this sort of work can be, and why it requires experts with a deep knowledge of the corpus under study.