

Appendix 8

Technical and Methodological Discussion of HCA Analysis Results

The specific HCA methodology used was Unweighted Pair Group Method with Arithmetic Mean (UPGMA). For a discussion of this and related techniques, see Andrews, Nicholas O., and Fox, Edward A. 2007, “Recent developments in document clustering” (Technical Report TR-07-35, Computer Science, Virginia Tech). Although not strictly necessary, the preprocessed corpus was normalized by slicing it into 1000 characters slices. To verify association robustness, additional trials were run with non-normalized texts and slice lengths of 100 and 500 characters. All trials produced almost identical clusterings of the query terms, suggesting that our results are extremely robust.

As noted in the main text, with only a few exceptions, the tree relations in our dendrogram results of control terms are precisely what we expect given the well-understood semantic relations between these terms. For instance, we see the seasons clustering together (pictured in lighter grey near the top of the diagram), with the two classic opposites (summer and winter) sharing the most basic node, but then joining with spring the next node up, and finally *yue* 月, which refers to both the “moon” and “month.” We also see most of the scalar opposites and complementary pairs (*yinyang* 陰陽, Heaven/earth *tiandi* 天地, king/minister *wangchen* 王臣, above/below *shangxia* 上下) clustering tightly together.

The exceptions to this pattern are easily explainable as a result of either common compounds or semantic ambiguity. For instance, “son” (*zi* 子) clusters with “lord” (*jun* 君), rather than with “father” (*fu* 父), as one might expect. This, however, reflects the high frequency of the compound term *junzi* 君子 (“gentleman, lit. son of a lord”) in the corpus. “Big” (*da* 大) has been pulled out of its expected tight fit with “small” (*xiao* 小) because of its common appearance in the compound *dafu* 大夫 (the name of a government rank), with “human/person” (*ren* 人) being similarly pulled into the orbit of “person/servant/spouse” (*fu* 夫) by the common compound form “lady/married woman,” *furen* 夫人.

In some cases, as in Studies 1 and 2, we could have done a better job of picking specifically-defined and semantically unambiguous pairs. For instance, “run/walk” (*zou* 走), which we had initially envisioned as an endemic function pair for “human/person” (*ren* 人), appears quite far away in the tree. This is almost certainly because 1) *many* species besides humans *zou* 走 ‘walk, run,’ and 2) *ren* 人 actually only rarely refers to ‘humans’ as a distinct species, commonly referring to “people” in general. Our much more narrowly defined and unambiguous endemic function pair, “bird-fly” (鳥飛), appears, as expected, tightly linked at the first node. *Ting* 聽 ‘to hear,’ which we had expected to pair with the appropriate organ (*er* 耳 ‘ear’), instead clusters with minister and king. This seems odd until one considers that a common alternative meaning of *ting* is “to obey,” the central imperative of minister-king relations.