A Language codes

We use an internationally standard set of codes (ISO 639-3) to represent languages. Those for languages referenced in the main text are:

Language	ISO 639-3
Archi	aqc
Bondum Dogon	dbu
Cantonese	yue
Chippewa	ciw
German	deu
Italian	ita
Korean	kor
Mandarin Chinese	cmn
Martu Wangka	mpj
Nahuatl	nci
Russian	rus
Spanish	spa
Welsh	cym

B Feature importances

While it is encouraging that the diachronic acquisition sequence of black and white, red, and so on emerges from our operationalization of the basicness criteria, we may be interested in some diagnostic information about which features are not actually informative, as well as clues to the robustness of our findings.

To do this, we perform recursive feature elimination (Guyon et al., 2002). The resultant feature sets for correlating with basicness and acquisition sequence, respectively, are:

Basicness 9 features: Word concreteness, Count of compounding, Frequency of compounding, Cognate etymology, Derivation etymology, Google Ngram frequency, Google Ngram percentage adjectival, Penn Treebank percentage adjectival, and Affix presence. Gamma is 0.983, which closes the gap to perfect correlation by about 50%.

Acquisition sequence 6 features: Count of compounding, Frequency of compounding, Suffix derivation, Google Ngram frequency, Penn Treebank percentage adjectival, and Affix presence. Gamma is 0.988, which improves the gap to perfect correlation by about 70%.

In the latter case, the greatest benefit is derived from improving the rank of *orange*. Namely, it moves from position 24 to position 15 in the ranking, while the first six colors hold their positions. *Brown*'s position also improves, swapping with *gray*.

Interpreting this in light of our operationalization, we may ask which of the criteria are most pertinent. We keep three morphological features, one abstractness feature, and two hybrid features (the frequencies) which also convey salience and are tempered by the other features (so that, say, *gold* and *flesh* aren't inappropriately ranked—see §5.3). Thus, it seems that we cannot eliminate any of the three categories of features we have created, without harming our correlation.