

Corpus evidence for systematicity in English compounds

Beth Levin, Lelia Glass, Dan Jurafsky
Stanford University

The puzzle Studies of English noun-headed compounds frequently highlight the many, often apparently idiosyncratic relationships between the head and modifier. For example, a *pumpkin bus* can refer to a bus that transports pumpkins or one that becomes a pumpkin at night (Downing 1977). These meanings are more than a function of a compound’s parts, invoking rich cultural knowledge. These properties make compounds hard to analyze compositionally. Thus, Dowty’s Montagovian analysis links the head and modifier by an “appropriately classificatory relation R ” (1979:319); however, it is unclear what relations are “appropriately classificatory,” or how speakers identify them in context. We show how R is enriched in context, identifying cognitively salient notions that influence how speakers categorize entities and thus choose compound names to refer to them based on how they interact with them.

In particular, we argue that distinct head-modifier relations are used depending on whether the referent of a compound is construed as an *artifact*, created by humans for a purpose, or as a *natural kind*, existing independently of humans. This well-known philosophical distinction may be metaphysically contentious, but we argue it has cognitive and linguistic consequences. We claim that we conceptualize and interact with artifacts and natural kinds in distinct ways. Artifacts are conceptualized in terms of an “associated event” (Nichols 2008) of their creation or use (*sewing* for a quilt, *writing* for a pen), whereas natural kinds are understood in terms of an abstract “essence” (Bird & Tobin 2009, Kripke 1972, Keil 1993), a set of distinguishing properties such as origin, habitat, or appearance. Thus, when a modifier is used in both an artifact and a natural kind name, its relation to the head is necessarily different. For example, *butter* evokes a specific color and texture in the natural kind compound *butter bean* – properties highlighting the referent’s essence – but is understood as the entity being spread – a participant in the associated event – in the artifact compound *butter knife*.

Our study We present evidence from a corpus study of over 1500 endocentric English noun-noun and adjective-noun compounds. The compounds, scraped from online databases and retailer inventories, were drawn from two conceptual domains: food and cooking (utensils; cakes/cookies; greens/legumes); and jewelry and precious minerals (bracelets, necklaces, rings, earrings; gemstones, corals, ebonies). Both domains include multiple types of artifacts and natural kinds to illuminate patterns within as well as across these categories.

We coded each compound for its head-modifier relation, using categories extended from those used by Wisniewski & Love 1998; then we aggregated these relations into three meta-relations according to whether they (i) evoked an associated event directly or indirectly, (ii) described the environment the entity is found in, or (iii) described perceptually salient properties of the entity. Drawing on Brown (1999), we also recognized a “borrowed” meta-relation for modifiers borrowed from another language, as in *adzuki bean* (from Japanese). Three-word [X_1 [X_2 N]] compounds were coded both for the relation of X_1 to [X_2 N] and the relation of X_2 to N; three-word [[X_1 X_2] N] compounds were simply coded for the relation of [X_1 X_2] to N. The results are summarized here, rounded to the nearest percentage point:

| Metarelations | NATURAL KINDS | | ARTIFACTS | | |
|------------------|----------------|-----------|-----------|---------------|---------|
| | Greens/legumes | Gemstones | Utensils | Cakes/cookies | Jewelry |
| Perceptual | 43% | 61% | 10% | 14% | 61% |
| Environmental | 30% | 25% | 1% | 5% | 1% |
| Associated event | 8% | 5% | 80% | 73% | 34% |
| Borrowed | 13% | 1% | 0% | 2% | 1% |
| Other | 6% | 8% | 9% | 6% | 4% |
| | N=342 | N=323 | N=283 | N=350 | N=377 |

Artifacts vs. natural kinds Across the board, the head-modifier relations depend systematically on whether the compound names an artifact or a natural kind. 62% of artifacts are named for an associated event – *canning funnel* is designed for an event of canning, *refrigerator cookie* is created via a cooling event. In contrast, natural kinds tend to be named for their environmental (27%) or perceptual (51%) properties: origin (*Madagascar opal*), habitat (*field green*), and appearance (*kidney bean*). These preferences are predicted since artifacts are created by humans (in an event) for a human use (another event), whereas natural kinds are found in nature and categorized by properties that distinguish them from one another. We also found that borrowed modifiers are more prevalent in natural kind – especially plant (13%; *adzuki bean*) – names – than artifact names. We hypothesize that the modifier is borrowed because it is taken to capture the essence of the compound’s referent. We now examine distinct preferences for head-modifier relationships among the natural kind subtypes and the artifact subtypes.

Differences within natural kinds The head-modifier relations for both plants and gemstones often make reference to their place of origin (*West Africa sorrel*, *Madagascar opal*) or perceptual qualities (*kidney bean*, *blue diamond*). However, distinct perceptual qualities matter to plants and gemstones. Plants are named for distinctive features (8%; *shell bean*), general appearance (11%; *kidney bean*) or color (16%; *green bean*), whereas gemstones are overwhelmingly named for color (38%; *black amber*, *blue diamond*) – perhaps because gemstones are recruited for adornment, for which color is especially relevant.

Differences within artifacts Utensils are often named for an event of *use* (64%; *canning funnel*). Cakes and cookies, instead, tend to be named for an event of *creation* (67%; *refrigerator cookie*), perhaps because they share a common function – serving as food –so they cannot be distinguished on the basis of use. In contrast, although jewelry is an artifact, it patterns like a natural kind in that it is often named for perceptual features such as distinctive part (50%; *cross necklace*); in fact, gemstones and jewelry are named for perceptual qualities in equal proportions (61% each). However, jewelry names like artifact names may evoke an event (34%; *leather bracelet*, referring to the material used in its creation, or *opera necklace*, worn for a particular occasion). We suggest that jewelry names straddle the artifact/natural kind boundary because the perceptual features of jewelry matter for its use.

Significance Our study demonstrates that the head-modifier relation used in a compound depends on how we conceptualize its referent, and, specifically, on whether we construe this referent as an artifact or natural kind. Within those categories, the head-modifier relation further depends on how we interact with the referent and the specific events and qualities made salient by that interaction. That this categorization is a matter of linguistic construal and not objective reality is shown by the names given to entities that might strictly speaking be natural kinds, but are cultivated or raised for a purpose, and thus named like artifacts, as in *pie pumpkin*, where the modifier refers to the entity that the head is used to create.

More generally, we have argued that the head-modifier relation pertinent to a given compound makes sense if the question asked is not “What are the possible interpretations of this compound?”, but “What is the entity that this compound names?”, that is, by considering the nature of the compound’s referent. In so doing, we have provided insight into what is meant by saying, as Dowty does, that the head-modifier relation is “appropriately classificatory.”

Selected References Bird, A. & E. Tobin. 2009. Natural Kinds. *Stanford Enc. of Philosophy*. • Brown, C.H. 1999. *Lexical Acculturation in Native American Languages*. • Downing, P. 1977. On the Creation and Use of English Compound Nouns. • Dowty, D. 1979. *Word Meaning and Montague Grammar*. • Keil, F.C. 1993. *Concepts, Kinds, and Cognitive Development*. • Kripke, S. 1972. Naming and Necessity. • Nichols, L. 2008. Lexical Semantic Constraints on Noun Roots and Noun Borrowability. • Wisniewski, E.J. & B.C. Love. 1998. Relations versus Properties in Conceptual Combination.