INTERPRETING NOVEL COMPOUNDS
Martin Schäfer and Melanie J. Bell
Anglia Ruskin University

Although it is well known that compounds encountered for the first time are open to a variety of plausible interpretations (e.g. Downing 1977), it is equally ‘acknowledged that the nature of the compounded elements ... does have some effect in circumscribing the range of meanings conceivable for any given attributive compound’ (Bauer, Lieber & Plag 2013: 474). What is unclear, is how the lexical semantics of the constituents interacts with other sources of information, such as encyclopaedic knowledge and context. Theoretical and philosophical accounts of compound interpretation range from those that rely mainly on pragmatic inference (e.g. Levinson 2000), to those that attribute a bigger role to word meanings (e.g. Asher 2011). Experimentally, there is evidence that not all compounds are equally ambiguous or difficult to interpret when they are presented in isolation, suggesting that constituent identity may be more constraining for some compounds than for others (e.g. Coolen et al. 1991). Studies giving subjects a choice of possible paraphrases have also revealed differences between compounds, in terms of strength of preferred reading, plausibility and difficulty of interpretation (Gagné et al. 2005; Zlatev et al. 2010; Smith et al. 2014). Furthermore, there is evidence for interdependence between the strength of an interpretation in isolation and its availability in the context of a single sentence or paragraph (Gagné et al. 2005, Zlatev et al. 2010, Middleton et al. 2011). Recent eye-tracking studies have also shown a relationship between comprehension of compounds in context and their ease of interpretation out of context (Cohen & Staub 2014, Smith et al. 2014). However, all previous studies have used artificially constructed novel compounds, and it is unclear whether the results are generalisable to compounds actually encountered in natural language. It is also unclear exactly why some compounds are easier to interpret than others, and what determines the range of possible meanings for any given item. We fill these gaps by using attested novel compounds to investigate what properties of a compound’s constituents affect its ease and range of interpretation out of context.

We used the following procedure to select novel noun noun compounds. First, we identified noun noun strings that occur only once in ukWaC, a 2 billion word corpus representing British English. In order to be able to model ease and diversity of interpretation using semantic properties of the constituent families, we restricted our search to nouns for which Bell & Schäfer (2016) provide family-level semantic annotations. We checked each uniquely occurring string in its sentential context, and excluded those that were not actually compounds. We also excluded proper names and quotative or unclear usages. Native speakers of British English provided paraphrases of the meanings of each of these compounds presented in isolation, and rated the difficulty of providing a meaning on a 10-point Likert scale. The paraphrases provided were subsequently coded and grouped according to whether or not they seemed to represent the same reading. This coding was undertaken independently by two trained raters, who exhibited a high level of agreement (r = 0.76, p < 0.001).

Initial results, based on 16 paraphrases for each of 22 compounds, suggest that context free readings may vary even more widely than previously supposed. Out of the 22 new combinations, only three achieved majority interpretations, with ten subjects agreeing on the meaning of flea reaction (reaction to fleas), and nine each on the meanings of kangaroo egg (egg laid by a kangaroo) and ivory wall (a wall made of ivory). Half of the combinations had more than seven interpretations, with only five or fewer subjects agreeing on the most common reading. The least diversity was shown by ivory wall, with just three different interpretations, and the greatest diversity by engine egg, with 12 distinct interpretations. These findings are unexpected if one supposes that specific properties of the meanings of the lexical
items involved drive the interpretation out of context (e.g. Asher 2011). Nevertheless, the number of different readings was strongly correlated with average difficulty of interpretation ($r = 0.77$, $p < 0.001$), i.e. the more difficult a compound was found to interpret, the greater the number of suggested interpretations. There was also a small negative correlation between perceived difficulty of giving a meaning and degree of convergence on a particular reading. The greater the number of participants who suggested a particular meaning for a given compound, the lower the average difficulty rating given by those participants for that compound ($r = -0.23$, $p < 0.01$). This suggests that, at least for some participants and for some compounds, certain readings arise more easily than others, providing support for the notion of default readings in the sense of Jaszczolt (2016), who regards default meanings as those that arise without conscious inferential effort. Furthermore, such readings are more likely than others to be shared between language users, which suggests that they may also be default (though weakly so) in the ‘presumptive meaning’ sense of Levinson (2000). We will present models of variation in meaning and ease of interpretation using our full dataset and with properties of the constituent families as predictors, to shed light on what determines the extent to which any given compound approaches a default reading.

References


