Does constituent disambiguation facilitate compound interpretation?

Melanie J. Bell
Anglia Ruskin University
melanie.bell@anglia.ac.uk

In English, nominal compounding is a very productive means of word formation and hence of lexical expansion. Recent additions to the Oxford English Dictionary online include e.g. *bucket list* – 'a list of things that a person hopes to experience or achieve before they die (or kick the bucket)' – and *trout pout* – 'unnaturally swollen lips resulting from the injection of excessive collagen into the lips in a cosmetic procedure intended to enhance their appearance'. As these examples demonstrate, although the compound meaning bears some relation to the meanings of the constituent nouns, the relation between them is unexpressed and must therefore be inferred by the hearer or reader. In some cases, the constituent nouns themselves seem to suggest a default reading; for example a compound like *polo umpire* might be expected to lead to context-free consensus about the meaning, because an umpire officiates in the playing of games, and polo is a game. However, in many cases, compound interpretation depends on other sources of information, including not only deixis and explicit explanation, but also context and world knowledge.

This study explores the role of immediate linguistic context in disambiguating newly-coined compounds. All noun-noun strings occurring within a sentence were extracted from the prose fiction section of the British National Corpus. This section was chosen to reduce the risk of selecting compounds whose interpretation would rely on deixis, specialised technical knowledge or highly time-specific information as might be expected in e.g. newspapers. To create a set of hapaxes, the sample was then reduced to items that occurred only once in the whole corpus, and not at all in ukWaC, a much larger corpus of more than 2 billion words. A random selection of 80 hapaxes was then examined in their sentential context. This led to the hypothesis that, in the majority of cases, the compound could be disambiguated on the basis of the

immediate sentence, usually through disambiguation of the head noun. For example, a police trailer could potentially be either a type of vehicle used by the police or an advertisement for a police procedural TV series. But in the sentence *Outside the incongruous police trailer she paused and knocked smartly on the door*, the words *outside* and *door* make it clear that the vehicle reading is required.

In a second stage, adult native speakers of English produced free paraphrases for a subset of the items, presented first in isolation and then in their sentential contexts. Preliminary results suggest that the range of interpretations can largely be categorised on the basis of the senses of the constituents; furthermore, inclusion of the immediate sentence does often shift interpretations towards a particular sense of the head. However, there is much less evidence that the immediate sentence leads readers to converge on a single reading, and overall the results suggest that language users may routinely tolerate considerable vagueness in compound interpretation.