How to depict things with words
The interface of meaning and perception
Hiromasa Mita
University of Tokyo / JSPS research fellow DC1

1. Introduction
This study aims to pave the way to properly capture the relation between semantic structures and perceptual information in a formal semantic representation. We humans can construct and communicate mental representations of objects with linguistic expressions by referring explicitly or implicitly to the objects’ color, shape, texture, etc. Such imagery aspects of natural language do not draw much attention from theoretical linguists since those aspects are typically regarded as irrelevant to the truth conditions of a sentence. Although some recent cognitive paradigms and NLP projects have gradually recognized the mapping of verbal expressions and visual information, for example, as a topic of language sciences, they have not yet provided any concrete answer as to how semantics is interfaced with perceptual representations. The current study will exemplify some interpretive phenomena that imply the interfacing mechanism and will explore how that should be formally treated in theoretical linguistics by modifying the Generative Lexicon framework (Pustejovsky 1995).

2. Image and Proposition
We will first look at some nominal predicate sentences which show that propositional computations can involve imagery information of objects in some cases. Let us consider the two declarative sentences in (1) below, where the adjective red predicates two different nouns:

(1) a. This is a red handkerchief. / b. This is a red flag.

Both sentences describe scenes observed by the speaker with a concrete object that is colored red, and there appears to be nothing more to say. However, a closer observation reveals that there are interpretive constraints on what kind of objects each sentence can depict. (1a), for example, can denote either a handkerchief whose whole surface is red, or one that is red for the most part but has a blue line as in the two pictograms on the left in Figure 1 below. On the contrary, it is remarkable that (1a) can only represent a wholly red-colored flag such as the flag on the upper right in Figure 1. For the lower left image, (1b) seems to be false as a real-world depiction, or at least undeterminable in truth value.

![Figure 1: Handkerchiefs and Flags](image)

Now the question arises as to how and why we extract the desired representation of objects from each sentence. Given that the difference in interpretation appears to occur at a propositional level and that these sentences are syntactically indistinguishable, what makes them differ must be the semantic structure of nouns. Moreover, any action-/event-based explanation fails to account for the difference in the manner of depiction of sentences. It is obviously implausible to insist that the difference in interpretation springs from that in action potentially associated with handkerchiefs and flags respectively. Considering these items, it can be hypothesized that our semantic processing at some stage retrieves the imagery information, or “a picture,” so to speak, of objects. In order to capture this mechanism within a linguistic framework, this study proposes an interfacing system that integrates perceptual information into semantic representation, by adjusting the Qualia Structure.

3. Perceptual Attributes and Interpretation of Similes
3.1 Unexpressed Grounds of Similes
Another example that shows how perceptive information interacts with logical computation is the simile. A simile is one form of metaphor, which has explicit markers of figurative comparison. (2) below shows Japanese simile expressions where three objects are compared figuratively to remon [lemon] with the simile marker -mitaina [like] in between:

(2) a. remon-mitaina kumo / b. remon-mitaina kousui / c. remon-mitaina shatsu
lemon-like cloud / lemon-like perfume / lemon-like shirt

Though the “vehicles” of those similes are the same (lemon), their “grounds” for “tenors” (cloud, perfume, and shirt) are different (for the terminology, see Richards 1936). To put it simply, (2a) refers to the lemon-like SHAPE of a cloud, whereas (2b) focuses primarily on the SCENT, similarly to saying a perfume that smells like lemon, and (2c) is typically read as describing the COLOR of a shirt. The interpretations may differ slightly, depending on the context or the individual’s preference, but the
important point is that from only the expressed vehicles and tenors, we readily figure out which attributes are at issue even with no explicit clues available about the grounds of simile. It can be said that semantic structures may contain imagery knowledge regarding how an object looks, smells, etc.

3.2 Similes and Computation of Anaphor

However, the observation in 3.1 alone may not be conclusive because understanding the grounds of each simile might be merely an elimination of vagueness by virtue of a pragmatic inference, irrelevantly to propositional semantics. Therefore, for more affirmative evidence, let us look at the examples in (3) below, each of which includes an adjective anaphor *souda* [be so] in the second sentence, its antecedent being the simile in the first sentence.


Consider what the anaphors can mean respectively. In general, anaphors copy the logical contents of their antecedents. In this case, it should be [be lemon-like] that is copied. Interestingly, however, *souda* in (3a) can be interpreted as [smells like lemon], while more crucially, *souda* in (3b) can only mean [smells like lemon] and never be read as [looks like lemon]. If interpretation of the grounds of simile was really based on a pragmatic inference, and not on semantic computation, one could readily obtain either SMELL or LOOK interpretations for (3b), because the anaphor *souda* would copy [be lemon-like] from the antecedent and underspecify anything more. However, (3b) does not, in fact, have a LOOK-LIKE reading, though, as shown in (2c), *remon-mitaina shatsu* does, in a neutral context. Therefore, from the observation about the anaphors in (3), it should be concluded that semantic computation of nouns sometimes involves the perceptual representation of their denotata at a logical level.

4. A Tentative Design of the Interfacing Structure

Considering the discussion so far, a detailed design of the interfacing system is presented here. In short, we need a device that connects perceptual information with conceptual (semantic, logical) components. Under the GL framework, the FORMAL quale of a lexical item represents information that “distinguishes the object within a larger domain” (Pustejovsky 1995:85). It is the most suitable quale for the description of perceptual characteristics pertaining to an object. (4) shows an exemplar schema, where perceptual information constitutes the lexical representation of a word.

![Diagram](image)

From the memory device, the slot Raw Imagery Content receives the unprocessed accumulation of perceptual data of the object that one has obtained empirically, which cannot be decomposed by any metalanguage, although the detail of our memory storage per se is still unknown. Next, the data of raw content is re-organized according to the sensory modalities and sub-modalities. This process occurs by virtue of functional mappings that relate the unprocessed image to each lexical item: HAVE COLOR relates the image of an object to a color term, for instance. The lineup of modalities is not universal but language-specific, since it does not reflect our physiological makeup per se, but is based on the inventory of sense vocabulary in each language.

By this schema, the interpretive behaviors of (1–3) can be explained. For (1), objects for which people discuss the design, such as flags, have multiple HAVE COLOR functions. The nominal a red flag specifies only one of those functions, the rest being null. This results in the interpretation that the flag is entirely red, without any other colors. In contrast, the lexical representation of [handkerchief] contains just one HAVE COLOR in the first place, hence a red handkerchief guarantees that it has a red color but says nothing about other colors. As to (2), [lemon] has both VISUAL, GUSTATORY, and Olfactory sub-qualia because of the empirical fact that we see, smell, and taste lemons, whereas [cloud] only has shape since we never touch, smell or hear clouds. Consequently, agreement of quale occurs between lemon and cloud, and (2a) only has a VISUAL reading. Finally, regarding (3), in each sentence, the sub-quaie for the SO-anaphor in the second sentence must be identical with that for its antecedent in the first sentence. This is how the LOOK-LIKE reading for (3b) is ruled out.
Reference