Open-class vocabulary and the granularity of lexical meaning

Omer Preminger (omer@lingsite.org)

Introduction: A cornerstone of much of contemporary formal semantics (especially, but not exclusively, in those traditions rooted in the work of Richard Montague) is a methodological skew towards focusing on "logical vocabulary" (e.g. *every*, *some*, *the*). The reason is that accounting for the interpretation of open-class items (e.g. *beauty*) will likely involve the same mechanisms and principles involved in the interpretation of logical vocabulary, along with the added challenge of formulating a theory of the conceptual encyclopedia.

I will argue that this (exceedingly reasonable) methodological choice has created something of an illusion when it comes to the syntax-semantics mapping—namely, that the syntactic terminal plays a privileged role in the "lookup" of lexically-stored meanings (cf. Heim & Kratzer 1998: 43, 48; *i.m.a.*).

Claim: There are at least some natural languages (e.g. languages of the Semitic family) in which open-class vocabulary seldom, if ever, involves a mapping between individual syntactic terminals and lexical meanings.

Evidence: In his discussion of modern Hebrew, Aronoff (2007) notes the existence of consonantal roots like */k-{b,v}-ʃ/*, which in combination with different templates, runs the gamut of meanings from 'pickles' to 'roads' to 'conquest':

(1) a. */k-{b,v}-ʃ/+CaCuC:* *kvuʃim* ('pickles')
b. */k-{b,v}-ʃ/+CCiC:* *kviʃ* ('road')
c. */k-{b,v}-ʃ/+Ci(C)CuC:* *kibuʃ* ('conquest')

What is less often remarked upon, however, is that even with so-called "well-behaved" roots, the grammatical state of affairs is actually no different. Consider the root */x-ʃ-{b,v}/* (etymologically: */ħ-ʃ-b/*), for example. This root is "well-behaved", in that all of the verbs and nouns derived from it have something to do with cognition or computation. But even here, there is no way to actually predict which specific meaning goes with which combination of */x-ʃ-{b,v}/* and a particular template:

(2) a. */x-ʃ-{b,v}/+CaCaC:* *xaʃav* ('think')
b. */x-ʃ-{b,v}/+CiC(C)eC:* *xiʃev* ('calculate')
c. */x-ʃ-{b,v}/+hiCCiC:* *hixʃiv* ('consider')

That is, there is no predictive way to tell what each of the items in (2a–c) will mean. (This is not to say that the choice of verbal template is not related to other properties of the verb, besides its encyclopedic meaning, such as its Voice. In the examples in (2), however, each of the meanings could in fact have occurred in any of the other templates. See Kastner 2020, and references therein, for discussion.)

On the assumption that **(i)** consonantal roots are grammatically real entities of the mental grammar of Hebrew speakers; and **(ii)** all combinatorics that are not exclusive to the phonology or exclusive to the semantics take place in syntax (i.e., no multiplicity of generative engines; Marantz 1997, *i.m.a.*), it follows that:

(3) Expressions like (1a–c, 2a–c) involve mappings from *sets* of syntactic terminals to lexical meanings.

The sets in question would involve, minimally, the root plus the n^0 corresponding to each nominal template in (1); and the root plus the v^0 corresponding to each verbal template in (2). It is crucially these sets, and not the individual terminal, that must be associated to a lexically-listed meanings in Hebrew, if we seek to capture the meaning mappings in (1–2).

Importantly, as noted, there is absolutely nothing exceptional about the state of affairs shown in (2), as far as Hebrew is concerned. This is essentially the state of affairs for *every root in the language* that can combine with at least two distinct derivational templates. We can therefore assume, without loss of generality, that this is so even when it comes to those few roots that only ever combine with one derivational template (say, */k-l-{b,v}/+CeCeC = kelev* 'dog'). That is, even for such roots, we could still assume that the meaning arises from a joint mapping from root+template to a lexical meaning, as we must anyway assume for the cases in (1–2). Essentially, then, every open-class item in Hebrew is an "idiom", insofar as that term is taken to denote many-to-one mappings from syntactic terminals to lexical meanings (à la the English *kick the bucket*).

That said, consigning this to the terminological bin of "idiomaticity" risks losing sight of something important. The term "idiom" is typically understood to indicate a marked departure from the linguistic norm—an expression whose syntax-semantics mapping is unusual. But these many-to-one mappings in a

language like Hebrew are not a departure from the norm; they *are* the norm, at least as far as open-class items are concerned. Plainly put: under this usage of "idiom", more or less the entire open-class lexicon of Hebrew would consist of idioms only.

Beyond Semitic: The lesson from Hebrew is that there seems to be nothing wrong with a natural language in which more or less the entire open-class lexicon is populated by "idioms" (i.e., many-to-one mappings from syntactic terminals to lexical meanings). What might this teach us about semantic interpretation more generally? One logical possibility is that Semitic languages are "special"—and that in other languages, like English, an expression like *dog* represents a one-to-one mapping of syntactic terminal to lexical meaning. However, given that English n^0 is often phonologically null, the phonological spellout of the root $\sqrt{\text{DOG}}$ alone would be string-identical to the spellout of this root and n^0 together. English would therefore look the same even if it worked exactly like Hebrew does (i.e., if open-class items always involved, at minimum, a joint mapping from root+categorizer to lexical meaning, and never a mapping from the root alone).

There is in fact some evidence that this is on the right track. Consider the following expressions:

(4) a. in cahoots
b. dribs and drabs
c. newfangled
d. underwhelm
e. short shrift

Clearly these are, in some sense, "frozen" expressions. However, we would not want to say that it is an accident that *in cahoots*, for example, has exactly the form of an English PP whose head noun is plural; that it could just as easily have been *s-in-cahoot* or *cahoot-in-s*. Or that it is an accident that in *short shrift*, the adjective *short* occurs in the same position as it does in a compositional noun phrase like *short film*. (For related discussion, see Bruening 2018.) Indeed, it seems that all of the expressions in (4) are entirely ordinary from the perspective of their syntax. If semantic interpretation systematically associates lexical meanings with syntactic terminals, we would expect that, e.g., *fangle*, when it occurs not in the context of the adjective *new* and participial *-d*, would be interpretable, according to whatever its listed lexical meaning is. But this is not so for this example or any other underlined element in (4). (At least some of these, e.g. *whelm* and *shrift*, had such interpretations in earlier stages of the language. But for current speakers of English, they do not.)

Note that this is not about "idiomaticity" per se, at least not the way that term is usually understood. In a terminal-centric approach to semantic interpretation (e.g. Heim & Kratzer 1998), idiomaticity amounts to the ability to assign to particular non-terminal nodes (e.g. the verb phrase *kick the bucket*) a meaning which overrides the one that would arise from applying the compositional rules of interpretation. This is crucially not what is going on in (4a–e): these expressions cannot be given a compositional meaning at all.

If, on the other hand, semantic interpretation associates lexical meanings with *sets* of syntactic terminals (as is independently necessary, at least for languages like Hebrew), then (4a–e) become far less of an oddity. It just happens to be the case that the terminal $\sqrt{\text{FANGLE}}$ participates in no lexical mappings involving a set smaller than $\{a^0, \sqrt{\text{NEW}}, v^0, \sqrt{\text{FANGLE}}, \text{Ptpl}^0\}$; whereas $\sqrt{\text{DOG}}$, for example, happens to participate in a lexical mapping as small as $\{n^0, \sqrt{\text{DOG}}\}$.

Implications: As discussed in the introduction, formal semantics often concentrates on "logical vocabulary", on the assumption that the mechanisms uncovered there will in any event be necessary when confronting the (presumably harder) issue of open-class meaning. Importantly, though, this is a choice of *opportunity*—there is no principle (that I am aware of) that says that logical vocabulary provides a more faithful window into the workings of semantic interpretation than open-class vocabulary does. If we have nevertheless managed to learn something about semantic interpretation by examining open-class vocabulary, then those conclusions should, by parity of reasoning, inform our understanding of the semantic interpretation of logical vocabulary.

The assumption that lexical meaning is associated with individual syntactic terminals faces significant obstacles in a language like Hebrew, and is possibly incorrect even for a language like English. We should now carry over these conclusions to the semantic treatment of logical vocabulary. Specifically, we should revisit the interpretation of logical vocabulary with the starting point that *individual* syntactic terminals need not be—and perhaps never are—the locus or grain-size of lexical meaning.