A new lexicalization constraint?

0. On the basis of data from Latin, Modern Greek, and Romance, I argue that, while natural language lexicons allow verbs with a "copular", "*BE* + PROPERTY" paraphrasis (cf. Lat. *rubere* 'be read'), they do so provided that PROPERTY is transient, and not permanent—in familiar terms, involving a S(tage-)L(evel), rather than an I(ndividual-)L(evel) predicate **[I]**. I show, more generally, that the languages of my sample—or English, for that matter—do not allow the synthetic expression of any IL predicate **[II]**. Adopting a Distributed Morphology approach to the interfaces of syntax, I propose that the seemingly copular verbs of Latin, Modern Greek, etc., are, in fact, unergative verbs, therefore including a (Davidsonian) event, and that true copular predications, eventless, involve too much underlying structure to be packed into a (synthetic) verb **[III]**.

[I] At least since the 1970s, a strand of research in formal linguistics has focused on patterns of lexicalization, including recurrent and systematic gaps in natural language lexicons: the fact that lexical items—notably verbs—do not have all the range of meanings and associated syntactic properties that they could logically have (Carter 1976, Hale & Keyser 1993ff., Rappaport Hovav & Levin 2010, a.o.). I report here on what seems to be a new such constraint, at least in a sample of languages.

My observation starts out from Latin. Many verbs in this language standarldy receive a "copular", "*BE* + PROPERTY" paraphrasis in the dictionaries (cf. Lewis & Short 1879)

(1)   a. *arere*: 'be dry'   d. *ualere*: 'be strong, healthy'   g. *calere*: 'be warm'
      b. *umere*: 'be wet'   e. *esurire*: 'be hungry'   h. *aegrotare*: 'be ill'
      c. *tumere*: 'be swollen'   f. *frigere*: 'be cold'   i. *candere*: 'be brilliant'

However, there seems to be a restriction so that PROPERTY is always transient, rather than permanent; in more standard terminology, it corresponds to an SL predicate, rather than an IL one (Milsark 1974, Carlson 1982, Kratzer 1995, a.o.). Thus, all the paraphrases in (1) involve SL predicates. Moreover, common IL adjectives like the *altus* 'high, tall' or *romanus* 'Roman', either allow a verbalization with a change of state interpretation (and thus not the "copular" interpretation focused on here) or do not allow any verbalization at all:

(2)   a. *altus* 'high, tall': *inaltare* 'raise, exalt', *'be high, tall'
      b. *breuis* 'short': *breuiare* 'abridge, shorten, abbreviate', *'be short'
      c. *romanus* 'Roman': no derived verb
      d. *pulcher* 'beautiful': no derived verb

Concentrating on *rubere* 'be red', IL/SL diagnostic tests like embeddability under a perception verb (3), or a phase verb (4), or compatibility with spatiotemporal modifiers (5), show that this and similar verbs like those in (1) are SL rather than IL:

(3)   Cernis   et   aestivo   mora   *rubere*   die. [Prop. 4, 2, 13]
      see.2SG also summery.ABL mulberry.ACC.PL *rubere*(INF) day.ABL
      'You can also see the mulberry blushing in the summer day.'

(4)   Per   herbas   matutina   *rubent*   [lumina]. [Lucr. 5, 457]
      through grass.ACC   morning.NOM.PL rubere.PL   light.NOM.PL
       'On the grass shines with reddish shine the morning light.'

(5)   Ubi   *rubere*   coeperit   corpus. [Cels. 3, 27]
      as_soon_as   rubere(INF)   begin.FUT.3SG   body.NOM
      'As soon as the body begins to get red.'

In a nutshell, *rubere* cannot be used as an equivalent of *rubeus esse* 'be red', predicating an *essential* property of red things, as *stative* definitions such as Lewis & Short's (1879) or Haverling's (2003) could lead us to conclude:

(6)   Humanus   sanguis   *rubet*.   ≠ 'Human blood is red.'
      human.NOM.M.SG   blood.NOM.SG   *rubere*.3SG

I extend the observation to Catalan, Italian, French, and Modern Greek. In Catalan, for instance, there is a productive process of derivation of this kind of verbs via the suffix *-ej* (Oltra-Massuet & Castroviejo 2013), but even in the cases where the input could be an IL predicate, like *pla* 'flat', the resulting verb is not interpreted as involving an IL predicate:

(7) Aquí el   camí plan-ej-a.
     here the path flat-*ej*-1STCONJ.3SG
     'Here the path is kind of flat / gets flat.'

(8) #Encara n'hi   ha   que pensen que la Terra   plan-ej-a.
      still   PTVE=LOC have.3SG that think.PL that the Earth   flat-*ej*-1STCONJ.3SG
      (Intended) 'There are still those who think that the Earth is flat.'

On the basis of this evidence, I propose the Property Verbalization Constraint (PVC):

(9) Languages may have intransitive verbs expressing SL predicates, but not IL predicates.

[II] Importantly, there are reasons to think that the PVC holds more generally. Thus, while languages have some transitive verbs involving IL predicates, like *know*, *own*, or *fear*, these verbs are obligatorily transitive, and hence the resulting predicates themselves cannot be synthetic (univerbal), suggesting a stronger version of the PVC (a version which I will not explore here):

(10) *Peter knows. [Only ok with a deictic or anaphoric interpretation]
       *Peter owns/fears.

[III] Keeping to the phenomena in **[I]**, and working within a Distributed Morphology approach to syntax and its interfaces, where syntax feeds the generation of morphological products, I suggest an explanation in terms of the Spell-Out theory developed in Embick (2010). Certain functional heads in the syntax trigger the Spell-Out of chunks of structure, and one same exponent cannot straddle two domains of Spell-Out. Following Roy (2013), a.o., I assume that copular predications involve a functional head, Pr, projecting the external argument and taking a whole property-denoting projection as its complement. Pr cannot select roots, since these do not denote properties (Acquaviva 2009, Borer 2013). Assuming that Pr and the embedded categorizing head (say, *a*), head their own Spell-Out domains (signalled with curly brackets below), Pr and its complement are predicted not to form a Spell-Out domain. Instead, Pr will end up spelled out together with the Tense-Aspect-Tense morpheme(s), yielding an overt tensed copula in some languages:

(11) {TAM [$_{PrP}$ DP$_{Subj}$ [$_{Pr'}$ Pr {$_{aP}$ a HIGH}]]}: cf., e.g., *Mount Everest is high*.

By contrast, the Latin and Catalan pseudostative verbs examined in **[I]** are unergative verbs, involving an event—Davidsonian states in Maienborn's (2003) terminology. As such, they involve a *v* head encoding an event, and a Voice head that projects the external argument (Kratzer 1996). Unlike Pr, Voice is, following Wood & Marantz (2017), part of the extended projection of its complement, together with TAM, i.e., it does not trigger Spell-Out. As a result, TAM, Voice, *v* and the root can be spelled out together as a single synthetic verb:

(12) {TAM [$_{VoiceP}$ DP$_{Subj}$ [$_{Voice'}$ Voice [$_{vP}$ v RUB]]]}: Lat. *rubere* "be read", cf. (3)–(4)

Selected references

Carlson, G. N. 1982. Generic Terms and Generic Sentences. *Journal of Philosophical Logic* 11. 145–181. **Carter**, R. J. 1976. Some Constraints on Possible Words. *Semantikos* 1. 27–66. **Embick**, D. 2010. *Localism versus globalism in morphology and phonology*. MIT Press. **Hale**, K. & S. J. **Keyser**. 1993. On argument structure and the lexical expression of syntactic relations. In Kenneth H. & S. J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 53-109. MIT Press. **Kratzer**, A. 1995. Stage-Level and Individual-Level Predicates. In G. N. Carlson & F. J. Pelletier (eds), *The Generic book*, 125–175. The University of Chicago Press. **Roy**, I. A. 2013. *Non-verbal predications. Copular sentences at the syntax-semantics interface*. OUP.