

Deconstructing subcategorization: Conditions on insertion vs. position

Laura Kalin & Nicholas Rolle

Princeton University & Leibniz-Zentrum Allgemeine Sprachwissenschaft

NELS 51, 6-8 November 2020, UQAM

1. Introduction: The notion of *subcategorization* has been utilized for decades to model certain idiosyncratic behaviors of lexical items. We hone in on one particular use of the term subcategorization, as relates to properties of *individual exponents* (morphs); see e.g. Lieber 1980; Kiparsky 1982; Selkirk 1982; Inkelas 1990; Paster 2006; Yu 2007; Bye and Svenonius 2012; McPherson 2019. Our goal is to show that, even at this fine level of granularity, subcategorization must be bifurcated into two separate and universally ordered mechanisms.

2. Background: Subcategorization has been used to subsume two types of restrictions on exponents. The first type—which we refer to as **CONDITIONS ON INSERTION** (hereafter **COINs**)—constrain *when/whether* an exponent is allowed to be inserted, by placing an environmental pre-condition on it. Consider the 3rd person possessive prefix in Tzeltal, which has two suppletive allomorphs, (1). (Data from Paster 2006, 59, who cites Slocum 1948, 80.)

- (1) a. 3.POSS *y*- : __V (≈ “*y* must be before a vowel”) e.g. *y-ahwal*, ‘his ruler’
 b. 3.POSS *s*- : __C (≈ “*s* must be before a consonant”) e.g. *s-mul*, ‘his sin’

COINs are employed for allomorphy (e.g., Halle and Marantz 1993, Paster 2006, Embick 2015), as well as for morphological compatibility and gaps (e.g., Jensen 1990, Booij 1993).

The second type of subcategorization for individual exponents—which we refer to as **CONDITIONS ON POSITION** (hereafter **COPs**)—regulate *where* an exponent is positioned, in particular, when it does not appear in its expected place. In Chamorro, the verbalizing affix *-um-* must appear before the first vowel in the stem, resulting in infixation. This too is expressed via subcategorization, (2). (Data from Yu 2007, 89, who cites Topping 1973, 185.)

- (2) VBZR *-um-* : __V (≈ “*um* must be before a vowel”) e.g. *tr<um>isti* ‘become sad’

COPs are most notably employed to model unexpected constituency disruption, including infixation (Yu 2007) and types of ‘special clitics’ (Zwicky 1997).

3. The puzzle: Given a particular subcategorization frame for a particular exponent, how does the grammar know whether this frame can be satisfied by displacing the exponent (as in Chamorro) or not (as in Tzeltal)? Where in the grammar is the COIN vs. COP distinction encoded? Most of the literature does not contend with these questions.

One answer is offered by Yu (2007, 229), who proposes that COINs and COPs are two sides of the same theoretical coin. If a language allows for constituent disruption, then a subcategorization frame may be satisfied by displacement (as a COP). However, if a language doesn’t allow constituent disruption then a subcategorization frame must be satisfied in situ (as a COIN). Non-compatibility results in a gap, which may be filled by a complementary form (an allomorph). Despite its elegance, we show that such an proposal cannot be right.

4. Proposal: Subcategorization encompasses two formally independent components, (3).

- (3) **Subcategorization bifurcation**: An exponent may include a Condition on Insertion and/or a Condition on Position, which are (i) formally independent of one another and (ii) universally ordered (COIN \gg COP)

5. Arguments: We present four arguments in support of our proposal, laid out below.

Argument 1 We saw in (1)-(2) that COINs and COPs can refer to at least some of the same types of elements in their frames. However, we hone in on two major differences. First, COINs can refer to specific segments and fine-grained phonological features in their frames (see, e.g., Paster 2005, 2006, 2009), whereas COPs cannot (Yu 2007:218ff); thus there is no infix that positions itself with respect to (e.g.) the closest stop or labial. Second, COINs can refer to an idiosyncratic class of roots (see, e.g., Carstairs 1987, Bobaljik 2000), whereas COPs cannot. Thus while a set of roots can condition (e.g.) special participial allomorphy (-*en* in English), no idiosyncratic set of roots can condition exponent placement.

Argument 2 A single exponent may concurrently have a COIN and a COP. Consider the intransitive marker in Puyuma, which has two suppletive allomorphs: the form *em* when the stem is obstruent-initial (a COIN), *me* elsewhere. Independent of this, *me-* is a prefix while *-em-* is an infix, linearizing after the first consonant (a COP). (Data from Teng, 2008, 26-7.)

- (4) a. s<**em**>anan ‘to stray (intransitive)’ b. **me**-languy ‘to swim (intransitive)’

Argument 3 In cases where an exponent has both a COIN and a COP, the COIN is satisfied at a derivationally earlier point than the COP. Consider the instrumental marker in Nancowry, expounded as *an* when it combines with a monomorphemic stem (a COIN), (5), *in* otherwise, (6). Both are infixes. (Data from Paster 2006, 167, citing Radhakrishnan 1981.)

- (5) a. káp ‘to bite’ → k<**an**>áp ‘tooth’ b. rúk ‘to arrive’ → r<**an**>úk ‘a vehicle’
c. lé? ‘to catch’ → l<**an**>é? ‘an object to catch with’
- (6) a. kaʔáp ‘to close’ → k<**in**>ʔáp ‘a trap’ b. tikó? ‘to prick’ → t<**in**>kó? ‘pin’
c. sahuáj ‘cool’ → s<**in**>huáj ‘something that cools’

The COP for *-an-* places it after the initial consonant, whereas the COP for *-in-* places it after the first vowel, thereby creating vowel hiatus which is resolved by deleting this vowel.

Nancowry provides a classic ordering argument due to opacity. The loss of the first vowel of the stem in (6) means that, on the surface, *-in-* and *-an-* have an identical distribution: they are both in bisyllabic words, appearing after the first consonant and before main stress. The choice of *-an-* vs. *-in-* (regulated by COINs) must therefore be made at a derivational point when the *underlying* form of the stem is visible, and only once the exponent is chosen is it possible to know what the COP is. (See Kalin 2019 for related cases and arguments.)

Argument 4 Our final argument comes from a typological gap: no language has an exponent with a subcategorization frame that may be satisfied as either a COIN or a COP. Thus, there is no language *Tzeltal Prime* where the COINs of *s* and *y* can optionally be satisfied as COPs instead of COINs, producing variant forms *a<s>hwal* and *m<y>ul* alongside those in (1).

5. Implications of (3): There must be an intermediate level of representation after syntax but before phonetic implementation, where COINs are evaluated. Therefore, a strictly parallel “P with M” model (McCarthy and Prince, 1993a,b) is insufficient. Further, our proposal supports models which deny a role of surface optimization in dictating allomorph choice (in line with, e.g., Paster 2006; Bye 2008) and runs counter to those which allow it (Mester 1994; Kager 1996; Mascaró 1996, 2007; Wolf 2008; Bennett 2017; de Belder 2020, *i.a.*).

6. Future work: To conclude, we discuss further directions suggested from this research, including (i) the fine timing of post-syntactic operations, (ii) the *representation* and *evaluation* of subcategorization frames, and (iii) *why* COINs and COPs have the properties they have.