NO ESCAPE FROM CATEGORIZATION: AN INSIDER'S VIEW OF COMPOUNDS

Vitor A. Nóbrega[1]

[1]Universidade de São Paulo, São Paulo, São Paulo, Brasil

Abstract

There has been a surge of syntactic research on compounding, joining a large literature on the nature of roots and phase theory. In an attempt to probe into the syntactic domain for idiosyncratic interpretation and to account for categorial exocentricity, disappearance of subcategorization, and lexical integrity effects, some recent studies on compounding have argued that root compounds are made up of two free acategorial roots directly merged in syntax, without undergoing categorization. The main goal of such an approach is to extend the phase domain in order to maintain two uncategorized roots awaiting further Merge operations. When a category head is merged on the top of this structure, it will trigger its Spell-Out, and as a result, both roots will (i) receive a single category status, (ii) be identified as a single syntactic object for the purposes of extraction and binding, and (iii) be assigned a non-compositional interpretation. In this article, we argue that root categorization should not be analyzed as an optional derivational step. By exploring compounding in Brazilian Portuguese, we identify a handful of phenomena that challenge the assumption that root compounds are made up of two bare roots. We propose that categorial exocentricity, subcategorization, and lexical integrity effects can be straightforwardly accounted if we assume that the unifying characteristic of compounds is the presence of a category head merged on the top of two categorized roots. We claim that non-compositional domains are not determined by categorization. Following Harley (2014), we admit that non-compositionality is assigned at LF through a set of LF instructions associated with roots in a particular syntactic environment.

Keywords: Root Compounding; Non-compositionality; Lexical Integrity; Phase Theory; Distributed Morphology.

1. Introduction

There has been a surge of syntactic research on compounding joining a large literature on the nature of roots and phase theory. In an attempt to probe into the syntactic domain for idiosyncratic interpretation, as exemplified in (1), and to account for categorial exocentricity (2), disappearance of subcategorization (3), and lexical integrity effects (4), some recent studies have argued that root compounds (RootCs) are made up of two free acategorial roots directly merged in syntax without undergoing categorization (Zhang, 2007; Zwitserlood, 2008; Bauke, 2013, 2014, 2016; Borer, 2013; De Belder, 2017, a.o.).

A.   Semantic non-compositionality

(1)   Brazilian Portuguese (BP)
       samba+canção
       samba+song
       'boxers'

B.   Categorial exocentricity
       (i.e., the compound's overall category differs from those of its constituents).

(2)   BP
       bate+bate   (V+V → N)
       hit+hit
       'bumper cars'

C.   Disappearance of subcategorization
       (i.e., the selection properties of a predicate within a compound are not satisfied).

(3)   BP
       a. *bate+bate*   b. *O   João   bateu o   carro*
          hit+hit   DET John   hit   the car
          'bumper cars'   'John hit the car'

D.   Lexical integrity effects
       (i.e., the impossibility of moving (4) or pronominalizing only one of the compound's members).
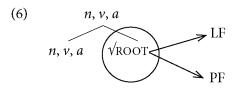
(4)   BP
       a. *bolsa+família*
          financial.aid+family
          'social welfare program to assist families in poverty; family allowance'

b. **Família,, várias bolsas-família, foram cortadas no último ano.*
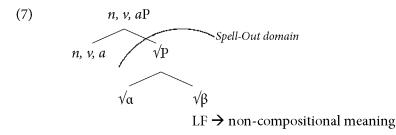'Family, many financial.aids were cut last year'.

In syntactic theories of word formation like Distributed Morphology (DM), lexical roots —i.e., primitives bearing conceptual content— are essentially category neutral (Marantz, 1995, 1997). They are categorized by combining with a category-assigning head (viz., *n*[oun], *v*[erb], *a*[djective]), as illustrated in (5):

(5)   *n, v, a*
   n, v, a   √ROOT

Following a phase-based approach to interpretation, most works in DM admit that roots are not interpreted independently, since they never constitute a syntactic phase (Marantz, 2001, 2008; Arad, 2003, 2005, a.o.). Thus, once a root is categorized —i.e., once a root is merged with a category head—, it is dispatched to the phonological (PF) and semantic (LF) interfaces —as described in (6)—, and receives an interpretation, which can be idiosyncratic. This interpretation is then carried along throughout the derivation.

(6)   *n, v, a*
   n, v, a   √ROOT → LF
                        √ROOT → PF

In root compounding, the merger of two uncategorized roots would extend the phase domain, and consequently it would maintain two uncategorized roots awaiting further Merge operations in the derivation. When a category head is merged on the top of this "compound structure" (√P), both roots are shipped together to LF, as depicted in (7). In this context, independent interpretation is not expected; hence both roots can be assigned a non-compositional meaning, as is the case with the nominal RootC in (1). Additionally, the structure in (7) would bear a single categorial status and serve as a single syntactic object (SO), thus neither of the two roots, √α and √β, could be independently manipulated. This would account for the phenomena exemplified in (2), (3), and (4).

(7)   *n, v, aP*
   n, v, a   √P   Spell-Out domain

         √α   √β
      LF → non-compositional meaning

In this article, we question whether the merger of two or more bare roots can give rise to well-formed SOs. We depart from the assumption that acategorial roots are defective syntactic primitives, since they are feature-less items and lack a pre-specified content (Arad, 2003, 2005; Acquaviva & Panagiotidis, 2012; Harley, 2014; Panagiotidis, 2011, 2014, 2015, 2020, a.o.). Each root must therefore be independently merged with a category head before being sent to the interpretive interfaces. By reviewing Bauke's (2013, 2014, 2016) contrast between Germanic and Romance nominal RootCs, we show that the generic structure in (7) does not account for a handful of morphosyntactic and morpho-phonological phenomena in BP nominal RootCs. We also revisit Zhang's (2007) discussions on categorial exocentricity, subcategorization, and lexical integrity, arguing that these phenomena can be straightforwardly accounted if we assume that the unifying characteristic of compounds is the presence of a category head merged on the top of two categorized roots (Nóbrega, 2014, 2015; Nóbrega & Miyagawa, 2015; Nóbrega & Panagiotidis, 2020). We also indicate that (7) cannot explain a set of interpretive effects observed in compounding more generally.

The article is laid out as follows. In Section 2, we discuss the identity of categorization in root compounding. By evaluating the set of phenomena listed above, we provide a body of evidence that root categorization should not be considered an optional derivational step. Romance RootCs, exemplified with data from BP, indicate that both roots must be categorized individually. We primarily review Bauke's (2013, 2014, 2016) interpretive distinctions between Romance and Germanic nominal RootCs, pointing out where the author's structural account fails. Subsequently, we explore an alternative solution to explain the three phenomena examined by Zhang (2007). In Section 3, we expand our analysis to account for the assignment of a non-compositional interpretation to RootCs. We argue that non-compositionality in compounding is dissociated from categorization. Following Harley (2014), we assume that any type of idiosyncratic interpretation is assigned at LF through a set of LF instructions associated with roots and the overall syntactic environment they are in. Finally, in Section 4, we present our final remarks.

2. On the identity of categorization

2.1 *The assignment of non-compositional meanings to complex structures*

A parametric distinction is commonly made between compounding in Germanic and Romance languages (see Di Sciullo & Williams, 1987; Snyder, 1995, 2001; Roeper, Snyder, & Hiramatsu, 2002; Roeper & Snyder, 2005; Di Sciullo, 2005; Delfitto, Fábregas, & Melloni, 2011; a.o.). In recent works, Bauke (2013, 2014, 2016) re-assesses this distinction, reconsidering the two tendencies frequently pointed out about nominal root compounding in these language groups, namely: (i) Germanic languages tend to produce nominal RootCs that are

compositional, productive, and recursive, while (ii) Romance languages hardly ever create nominal RootCs productively, and the existing forms tend to display a fixed interpretation, which can quite often be expressed by simple nouns, as illustrated with the examples in (8):

(8)   a. French   *homme+grenouille*   *plongeur*
      b. Spanish   *hombre+rana*   *buceador(a)*
      c. Italian   *uomo+rana*   *sommozzatore*
                     'frogman'

                                            (Bauke, 2014 p. 22)

Focusing on German nominal RootCs, Bauke identifies an intra-language variation, showing that these two tendencies in fact can co-exist in a single system. German displays two patterns of nominal root compounding, which can be distinguished by the following morphological and interpretive properties:

(9)   Patterns of nominal root compounding in German
      a. **Pattern #1**:
         • Nominal RootCs made up of two bare lexical items combined without any intervening inflectional material;
         • They are non-recursive, non-compositional, and non-productive.
      b. **Pattern #2**:
         • Nominal RootCs displaying inflectional material in compound internal positions, which is attached to the compound's first constituent member;
         • Whenever these inflectional markers occur, the resulting compound is compositional and productive, and allows for a range of alternative interpretations — the so-called 'weak compositionality' (see Pirrelli, 2002).

To illustrate the two patterns in (9), let us take the examples in (10):

(10) German (Bauke, 2014, p. 26-27)
      a. *Landstraße*
         country+road
         'country road'
      b. *Landsmann*
         country.GEN+man
         'compatriot'
         or 'man who loves the countryside'
         or 'man who advocates for the conservation of the countryside', etc.

c. *Landeskirche*

country.GEN+church

'national church'

or 'church that is associated with the country',

or 'church that shows the country's typical architecture', etc.

d. *Landerspiel*

country.PL+match

'match between two national teams'

or 'game that involves knowledge about certain countries',

or 'game that is typically played in certain countries',

or 'game that is characterized by customs of a certain country', etc.

The compound in (10a) consists of the combination of two roots: *Land* 'country' and *straße* 'street'. Since this compound does not involve any internal inflectional material, it can only display —according to Bauke—, one fixed interpretation, thus exemplifying the pattern in (9a). This pattern parallels the Romance RootCs in (8). As Bauke (2014, p. 28) points out, *Landstraße* is "a very specific type of road" and it cannot be interpreted as "any kind of road that runs through the countryside." This lexicalized type of interpretation is quite different from that of the compounds in (10b) – (10d). Although the examples in (10b) – (10d) display a preferred reading, a number of alternative interpretations exist alongside. For instance, although the preferred reading of *Landsman*, in (10b), is 'compatriot', it does not exclude the emergence of a set of additional interpretations, such as 'man who loves the countryside', and 'man who advocates for the conservation of the countryside'. This interpretational flexibility is assumed to be due to the presence of inflectional material attached to the first constituent, generally a plural or genitive marker. This latter type illustrates the pattern in (9b).

Bauke (2014, 2016) indicates that the assignment of alternative interpretations in German RootCs can be attested even in cases where the preferred interpretation has a strong tendency for an idiosyncratic meaning. (11a), for example, has no alternative interpretation available. (11b), on the other hand, despite being more drifted, can also refer to (i) 'a castle that is built in the shape of a bed', and (ii) 'an arrangement of several beds that resemble a castle'.

(11)   German (Bauke, 2014, p. 28)

a. *Bettlaken*

bed+sheet

'bedsheet'

b. *Bettenburg*

bed.PL+castle

'big ugly hotel with lots of rooms'

Based on these empirical observations, Bauke puts forth a morpho-semantic generalization about the interplay between the morphological and interpretative properties of German nominal RootCs, which can be synthetized as in (12):

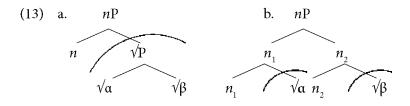(12) **Bauke's generalization on German nominal root compounding**
"As long as an inflectional marker is available [in compound internal positions], a compositional reading can still be retrieved; once the inflectional marker is lost, the interpretation of the compound is fixed and does not allow for productive alternatives" (2014, p. 30).

In an attempt to account for the inter- and intra-language variation observed in root compounding, and for the morpho-semantic generalization in (12), Bauke proposes that compositional, recursive, and productive RootCs result from the merger of two independently categorized roots. Following Marantz (2001, 2008), Bauke argues that the inflectional marker attached to the first constituent member is a categorizing head that has the properties of a phase head. Consequently, inflectional markers coincide with a phase that triggers cyclic Spell-Out, as depicted in (6). Thus, the element to which the inflectional marker is attached undergoes independent interpretation at LF and allows for a compositional reading. On the other hand, compounds with a non-compositional interpretation, such as Romance and the Germanic pattern in (9a), would have two roots merged without undergoing categorization. In this case, independent interpretation is not expected —as pointed out earlier—, which would explain their idiosyncratic and non-productive character.

In Table 1, below, we summarize Bauke's observations and structural distinctions to each compound type:

Table 1. Summary of Bauke's proposal.

	Romance and German #1	German #2
Interpretation	Fixed, lexicalized	Recoverable compositional interpretation
Productivity	Non-productive	Productive
Recursion	Non-recursive	Recursive
Syntactic structure	Two bare roots (13a)	Two categorized roots (13b)

(13) a. nP   b. nP

With respect to Romance languages, Bauke recovers the recurrent assumption that "novel compounds are hardly ever formed productively" and "when speakers of a Romance language form a novel endocentric nominal RootC, the result

requires an explanation of the meaning of this compound" (2014, p. 22). This claim, however, is not entirely correct. Although Romance nominal RootCs are not recursive (as opposed to English and German RootCs; e.g., *restaurant coffee cup*), they are not peripheral as the literature often suggests, and most newly coined forms do not necessarily display a fixed, lexicalized interpretation. BP speakers, for example, easily coin novel RootCs with compositional and straightforward readings, such as the ones listed in (14):

(14)   BP

      a. *video+depoimento*
         video+testimony
         'video testimony'

      b. *pastor+deputado*
         pastor+congressman
         'pastor congressman'

      c. *empreiteiro+político*
         contractor+politician
         'contractor-politician'

      d. *pauta+bomba*
         schedule+bomb
         'tariff bomb'

      e. *aula+debate*
         class+debate
         'debate class'

      f. *juiz+estrela*
         judge+star
         'a judge who seeks fame; who got famous'

All nominal RootCs in (14) have at least two interpretations. Alongside their attributive reading —i.e., a reading involving a modification relation—, they can also display a coordination reading, as illustrated in (15):

(15)   BP

    pastor-deputado lit. pastor-congressman
    i. <u>Attributive reading</u>: 'a pastor who, in addition to being a pastor, has another parallel occupation, congressman';
    ii. <u>Coordinate reading</u>: 'pastor and congressman'.[1]

This interpretational flexibility indicates that their internal structure is not entirely opaque. Furthermore, it is relevant to highlight that Romance non-compositional RootCs may also display inflectional material occurring in compound internal positions, similarly to what has been observed in German pattern #2. Examples are listed in (16). Oppositely to the German pattern #2, the inflectional material attached to the compound's first member does not trigger

a whole range of alternative interpretations, as noticed particularly with the German RootCs in (11).

(16) BP

a. *samba+canção*   → *samba-s+cançõ-(es)*
   samba+song   samba-PL+song-PL
   'boxers'   'boxers.PL'

b. *bolsa+família*   → *bolsa-s+família-(s)*
   finantial.aid+family finantial.aid-PL+family-PL
   'family allowance' 'family allowances'

c. *cara+metade*   → *cara-s+metade(s)*
   face+half   face-PL+half-PL
   'soul mate'   'soul mates'

d. *banana+maçã*   → *banana-s+maçã-(s)*
   banana+apple   banana-PL+apple-PL
   'Latundan banana'   'Latundan bananas'

e. *meia+calça*   → *meia-s+calça-(s)*
   sock+pants   sock-PL+pants-PL
   'pantyhose'   'pantyhoses'

The examples in (16) contradict Bauke's structural account for the German pattern #2 in (13b). The RootCs in (16) evidence that (i) BP nominal RootCs cannot be the result of the merger of two uncategorized roots, and that (ii) non-compositional RootCs must also have their roots independently categorized in order to allow for the insertion of inflectional markers in their first constituent members.

To further verify the plausibility of Bauke's account, let us explore four different contexts. As the first context, let us consider Romance non-compositional RootCs, such as the ones in (16), admitting —as Bauke suggests— that Romance RootCs are made up of two uncategorized roots. Since LF assigns a fixed meaning to this type of SO, we can set aside any discussion on the establishment of different grammatical relations between its constituent members, such as attribution or coordination. Bauke's account, however, is not able to explain the following PF facts: first, it would preclude the insertion of inflectional markers in the first constituent member of BP non-compositional RootCs (e.g., *cara-s metade-(s)* lit. face-PL+half-PL 'soul mates'). Second, if RootCs bear a single category head, they should display a single primary stress (Marvin, 2002, 2013), a theoretical expectation that conflicts with the phrasal stress pattern of Romance RootCs (Nespor, 1999), as illustrated with the examples in (17):

(17) BP

a. *video+depoiménto*
   video+testimony
   'video-testimony'

Italian

b. *capo+stazióne*

  chief+station

  'stationmaster'

Spanish

c. *emprèsa+fantásma*

  firm+ghost

  'cover-up company'

Catalan

d. *vagò+restauránt*

  wagon+restaurant

  'dining-car'

Additionally, since PF does not have access to the compound's category head (*n*), it is puzzling how (13a) assures the correct distribution of class markers in BP N-N compounds (e.g., *peix-e espad-a* lit. fish+sword 'sword fish'; *ministr-o chef-e* lit. minister+chief 'chief minister'). An account as (13a) also fails to differentiate cases where the first root does not receive a class marker, such as in N-N stem-based compounds:[2/3]

(18) BP

a. *lul-o-fobia*

  Lula-LE-phobia

  'aversion to Lula'

b. *cervej-o-chato*

  beer-LE-snob

  'beer snob'

As a second context, let us explore BP compositional RootCs. If Romance RootCs were made up of two uncategorized roots, then we would find additional problems to account for some interpretive effects at LF, along with the PF issues identified for the first context. First, this analysis would not be able to explain why we find different grammatical relations holding between the constituent members of nominal RootCs, such as attributive (e.g., BP. *trem-bala* lit. train+bullet 'bullet train') and coordination relations (e.g., BP. *sofá-cama* lit. sofa+bed 'daybed'), see also (15). Second, this analysis does not elucidate, in a principled way, why in some nominal RootCs only one root (viz., the non-head noun) is interpreted idiosyncratically, as in (19):

(19) BP

a. *bolsa+sanduíche*

  financial.aid+sandwich

  'financial aid/scholarship to graduate students visit a foreign university'

b. *banana+maçã*

banana+apple

'Latundan banana'

c. *cão-salsicha*

dog+sausage

'dachshund'

Now let us consider non-compositionality in other compound types. If non-compositionality in the two-root domain is associated with the absence of category heads, as (13a) implies, how then could we explain the correct distribution of nominal class markers and verbal theme vowels in non-compositional V-N compounds, such as those in (20)?

(20) BP

a. *bat-e$_v$ pap-o$_n$*

hit-ThV.3p.PRT+goiter-CM

'chat'

b. *pux-a$_v$ sac-o$_n$*

pull-ThV.3p.PRT+ball-CM

'yes-man'

Finally, let us keep assuming that Romance RootCs are made up of two bare roots merged without undergoing categorization. Then, let us admit that the nominal category head (n) is visible at both interfaces. In this scenario, we would be able to account for the correct distribution of class markers in Romance N-N RootCs (e.g., *peix-e espad-a* lit. fish+sword 'sword fish'). However, in the case of nominal N-A RootCs, it is hard to determine to which root a nominal class marker will be attached, and to which root gender agreement will be attached, as illustrated with the examples in (21).

(21) BP

a. *pã-o*   *dur-o*

bread-CM   hard-MASC

'iron-fisted; mean'

b. *águ-a*   *viv-a*

water-CM   alive-FEM

'jellyfish'

These four contexts suggest that an analysis for Romance RootCs along the lines of Bauke's (2013, 2014, 2016) proposal cannot account for a set of morphological and morpho-phonological facts. Bearing this in mind, we claim that non-compositionality in compounding should be dissociated

from categorization. We suggest that even non-compositional RootCs must have their roots independently categorized. In the next sub-section, we will point out that lexical integrity effects provide evidence for postulating a third category head in compounding, which turns both categorized roots into a single SO for the purposes of movement and binding, and elucidates cases of categorial exocentricity.

2.2 Lexical integrity effects in compounding

Zhang (2007) claims that (i) categorial exocentricity —i.e., compounds where the constituent in the head position does not impose its categorial features on the whole construction (Scalise, Fábregas, & Forza, 2009, p. 58)—, illustrated in (22), (ii) the disappearance of subcategorization —i.e., cases where the subcategorization of a verb has not been satisfied—, in (23), and the impossibility of movement (24) and pronominalization (25) of a single constituent member, serve as evidence that Chinese RootCs result from the merger of two uncategorized roots.

(22)   Chinese (Zhang, 2007, p. 172)
    a. *zhe*   *zhang*   *zhuozi*   *de*   *da-xiao*   (A+A → N)
       this   CL   table   MOD   big-small
       'the size of this table'

    b. *yi*   *ge*   *hen*   *bao-shou*   *de*   *ren*   (V+V → N)
       one   CL   very   keep-defend   MOD   person
       'a very conservative person'

(23)   Chinese (Zhang, 2007, p. 174)
    a. *Ta*   *mai-le*   *shu/*mai*
       he   buy-PRF   book/sell
       'he bought books'
    b. *yi*   *zhuang*   *mai-mai*
       one   CL   buy-sell
       'a transaction of trade'

(24)   Chinese (Zhang, 2007, p. 176)
    a. *Tamen*   *yixiang*   *fu-ze*
       they   always   carry-duty
       'they are always responsible'
    b. **Tamen* *yixiang* *lian* *ze*   *dou*   *fu*
       they   always   even   duty   also   carry
       Intended: 'they are always even responsible'

(25) Chinese (Zhang, 2007, p. 177)

Ta xian na-le   yi   ba cha₁-hu ranhou ba ta₁   dao-ru   beizi-li
he first take-PRF one CL tea-pot then   BA it   pour-in cup-in
Intended: 'he first took a tea-pot, and then poured the tea into a cup'

As an alternative to Zhang's (2007) proposal, we suggest that these phenomena may in fact be indicating that the unifying characteristic of compounds is the presence of a category domain on the top of two or more categorized roots, which is responsible for turning one or more syntactic elements (e.g., categorized roots, phrasal constituents) into a single SO (see Nóbrega, 2014, 2015). Following Nóbrega & Panagiotidis (2020, p. 230), we incorporate this assumption in the syntactic definition of compounds presented in (26).

(26) *Compounds within syntax*
Compounds are phrasal structures with two or more categorized roots combined in a specific grammatical relation —viz. subordination, attribution, or coordination—, which are further categorized by a category head, *n, v* or *a*.

According to (26), compounds should be analyzed as the by-product of the recategorization of an endocentric syntactic structure. As a consequence, the overall category of the compound may —in some cases, such as those in (22)— differ from the category of its internal constituent members. Furthermore, the nominal status of deverbal compounds would inhibit the subcategorization frames of their internal verbs, which would explain (23). Finally, the category domain on the top of this compound structure would require both categorized roots to be moved together, thus no element can be moved out of the compound in isolation, as observed in (24). It also inhibits reference to some of the compound's roots by using an anaphoric device, as in (25). Other empirical facts motivating the assumption of a category domain in compounding are: (i) the addition of a subcategorization frame in verbal N-V and synthetic compounds, as in (27) and (28); (ii) parasynthetic compounds (i.e., when two roots form a (non-existent) compound with a derivational suffix; see Bisetto & Melloni, 2008, 2010), and (iii) the addition of inflectional markers distinct from those expected for the compound members when used as an autonomous word, as observed in Modern Greek and Slavic compounds (Nespor & Ralli, 1996; Ralli, 2008, 2009; Ralli & Karasimos, 2009); see Nóbrega (2014) for a detailed description.
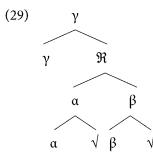
(27) Catalan (Padrosa-Trias, 2007, p. 95)
a. *cama+trencar*
   leg+break
   'to break the leg(s)'

b. *El   caçador   va   <u>camatrencar</u>   l'ocell.*
the   hunter   went   leg-break   the-bird
'the hunter broke the bird's leg(s)'

(28)   English

Tonight I'm <u>babysitting</u> my two-month-old nephew for my sister in law.

The generic structure for a RootC, including the categorial domain alluded in (26), is thus as follows:

(29)   γ

γ   ℜ

α   β   In which α, β, γ are category-defining heads
and ℜ stands for the grammatical relations of
α   √ β   √   subordination, attribution and coordination.

Based on the empirical facts discussed so far, we may admit that categorization is not an optional derivational step, especially in non-compositional domains. In addition to the empirical issues explored, merging two uncategorized roots leads to a set of theoretical drawbacks, most of them associated with their feature-less nature (Panagiotidis, 2011, 2014, 2015, 2020). First, by postulating that two free acategorial roots create a well-formed SO, we would necessarily have to admit that feature-less items are able to project (Acquaviva, 2009). Second, compounds made up of two free acategorial roots would be inherently headless. Roots are 'weak' syntactic elements, since they remain feature-less due to the 'No Tampering Condition' (Chomsky, 2008, 2015); thus, the output of two uncategorized roots would in principle induce formal crashing at the interfaces, since no immediate head can be identified. With this in mind, we will now discuss how the derivation of RootCs in BP is structured, and how a non-compositional meaning can be assigned to both roots.

3. Deriving and interpreting BP nominal RootCs

We have concluded thus far that roots are never interpreted independently. Their meaning is negotiated when they are dispatched to the interfaces, right after being categorized. Second, roots are defective syntactic objects, since they are content-less and feature-less primitives (Arad, 2003, 2005; Acquaviva & Panagiotidis, 2012; Harley, 2014; Panagiotidis, 2011, 2014, 2015, 2020, a.o.). Therefore, roots must be merged with a category head; otherwise they cannot be semantically and phonologically interpreted (Panagiotidis, 2014, 2015). Third, non-compositional interpretation in compounding emerges when both roots are shipped together to the interpretive interfaces, allowing this combination to be

assigned an idiosyncratic meaning (Nóbrega & Panagiotidis, 2020, p. 229). In the compound structure in (29), both roots are merged to a categorizing head as their complements, as shown in (30).
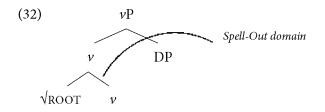
(30)   a. [$_v$ *v* √ROOT]   b. [$_n$ *n* √ROOT]

Nevertheless, as indicated by Nóbrega and Panagiotidis (2020, p. 231-233), once category heads are phase heads, both roots will be spelled-out separately, precluding the assignment of a single idiosyncratic meaning to the compound. It is not consensual, however, that roots are complements of category heads. Marantz (2013), in particular, shows that contextual allomorphy requires a root to be concatenated as an adjunct to a category head (specifically *v*). Irregular past-tense morphology in English, for instance, is sensitive to the past-tense feature of T. The root √TEACH, in (31), has to be realized as /tɔ/ in the environment of √TEACH, and the past-tense feature has to be realized as /t/ in the environment of √TEACH.

(31)   *English past tense*
       √TEACH + *v*(Voice) + past = taught
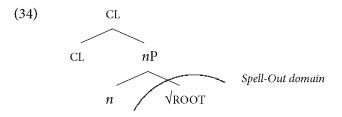
(Marantz, 2013, p. 98)

(31) thus describes a locality issue: if *v* is a phase head, then the root and T are on different sides of a phase boundary. In order to allow for a root to be in the same Spell-Out domain as *v* (or Voice), Marantz argues that each root is adjoined to the verbal category head. Consequently, the root will not be in the Spell-Out domain of *v*.

(32)

As a result, *v*(+Voice) does not interfere with Tense serving as the context for the vocabulary item (VI) at the root, since all of these heads will be spelled-out at the same time, in the complement domain of C. Additional evidence and possible extensions are cases of contextual allosemy in Murinypata, a language of North-West Australia. In languages with noun classifiers, distinct classifiers can be used with the same noun to specify its meaning. In Murinypata, the choice of the noun classifier will govern the meaning assigned to the root, as illustrated with the noun *kamarl* 'eye' in (33):

(33)   Murinypata (Walsh, 1976, p. 275)

    a. *nanth*   *kamarl*

        CL:GENERIC   eye

        'eye/face'

    b. *kura*   *kamarl*

        CL:ACQUATIC   eye

        'water-hole'

    c. *kardu*   *kamarl*

        CL:HUMAN   eye

        'sweetheart'

    d. *mi*   *kamarl*

        CL:VEGETABLE   eye

        'seed'

These examples instantiate a case of contextual allosemy, where the meaning of the root is sensitive to the noun classifier attached above its nominal category head. Thus, for LF to assign a meaning to the root √KAMARL 'eye', both the root and the classifier node (CL) have to be in the same Spell-Out domain, otherwise such an outer morphology cannot influence the root's meaning, as schematically illustrated in (34). If we admit that the root is adjoined to its category head, then *n* will not interfere with CL serving as the context for the VI at the root, since all of these heads will be spelled-out at the same time, in the complement domain of D.

(34)

Nóbrega and Panagiotidis (2020) developed an analysis along these lines to explain root categorization in compounding. Expanding Marantz's (2013) adjunction proposal, the authors suggest that roots are externally pair-merged with category heads, as in (35). This assumption would assure that: (i) each root will be in a local domain with a category head, and that (ii) both roots will be in the same Spell-Out domain, allowing the assignment of an idiosyncratic meaning to the compound.[4/5]

(35)

    <√ROOT, *x*>   <√ROOT, *x*>   in which *x* stands for *n*, *v* or *a*.

Following Nóbrega and Panagiotidis (2020, p. 232), we also admit that the grammatical relations connecting the members of a compound (ℜ) are derived by the nature of the operation Merge applying to combine their categorized roots,

whether set-Merge (i.e., {...}) of pair-Merge (i.e., <...>) (Chomsky, 2000, 2004), according to the distribution in (36):

(36)   *Grammatical relations internal to compounds*

    a. **Subordination**

       Head-complement relation in which a constituent α is set-Merged with β.

       (e.g., {<$\sqrt{}$$\alpha$, x>, <$\sqrt{}$ β, x>}

    b. **Attribution**

       Adjunction relation in which the non-head constituent β is pair-Merged to its head α.

       (e.g., <<$\sqrt{}$$\alpha$, x>, <$\sqrt{}$ β, x>>)
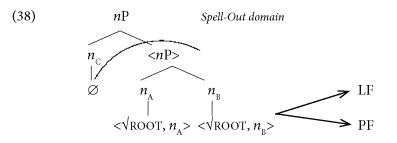
    c. **Coordination**

       Conjunctive relation in which two categorially identical constituents, α and β, are connected by means of a functional coordination head, giving rise to a Boolean phrase &P.

Furthermore, we admit that the syntactic derivation of a compound would follow the same derivational steps of complex specifiers or complex adjuncts (following the technical implementation envisaged in Nunes & Uriagereka, 2000; Nunes, 2012; Piggott & Travis, 2013). Thus, considering the Numeration in (37a), the computational system would independently select a nominal category head (n) and a root (37b-c), and subsequently it would externally pair-merge them together (37d). A second root would be externally pair-merged with a nominal category head as a second root syntactic object (37e-g). Finally, both SOs would be merged to each other following one of the specifications in (36). The derivational step in (37h) indicates that both roots were concatenated in an attributive relation.

(37) a. $N = \{n_{A1}, n_{B1}, \sqrt{A_1}, \sqrt{B_1}\}$

     b. $N' = \{n_{A1}, n_{B1}, \sqrt{A_0}, \sqrt{B_1}\}$

        $K = n_A$

     c. $N'' = \{n_{A0}, n_{B1}, \sqrt{A_0}, \sqrt{B_1}\}$

        $K = n_A$

        $L = \sqrt{A}$

     d. $N'' = \{n_{A0}, n_{B1}, \sqrt{A_0}, \sqrt{B_1}\}$

        $M = <\sqrt{A}, n_A>$

     e. $N''' = \{n_{A0}, n_{B0}, \sqrt{A_0}, \sqrt{B_0}\}$

        $M = <\sqrt{A}, n_A>$

        $O = n_B$

     f. $N''' = \{n_{A0}, n_{B0}, \sqrt{A_0}, \sqrt{B_0}\}$

        $M = <\sqrt{A}, n_A>$

        $O = n_B$

        $P = \sqrt{B}$

g. $N''' = \{n_{A0}, n_{B0}, \sqrt{A_0}, \sqrt{B_0}\}$

  $M = <\sqrt{A}, n_A>$

  $Q = <\sqrt{B}, n_B>$

h. $N'''' = \{n_{A0}, n_{B0}, \sqrt{A_0}, \sqrt{B_0}\}$

  $U = <<\sqrt{B}, n_B>, <\sqrt{A}, n_A>>$

After being concatenated, both categorized roots (i.e., roots in a strict local domain with a category head) have a third category head set-merged on the top of them. At this moment, this category head determines a phase head and triggers the Spell-Out of its complement —since it was set-merged to the structure—, allowing both categorized roots to be dispatched to the interfaces together. In (38), we describe the generic structure of a nominal RootC with an attributive interpretation (e.g., *trem-bala* lit. train+bullet 'bullet train'; *aula-debate* class+debate 'debate class', and the compounds listed in (14)), and indicate the Spell-Out domain.

(38)

Now to account for the assignment of a single idiosyncratic reading to two categorized roots, we claim that the interpretation of a nominal RootC is determined at LF, through instructions associated with roots and the syntactic structure they are in, following Harley's (2014) proposal. LF instructions assign interpretations to roots, taking into account the overall syntactic environment in which they are inserted. Thus, the syntactic context —and, more importantly, the categorial environment— is extremely relevant to determine the roots' meaning. Harley (2014), following Acquaviva (2009), assumes that roots are phonologically abstract and semantically vacuous linguistic primitives, and that a root terminal node is solely individuated by an alphanumeric index. For example, an arbitrary English root, such as $\sqrt{77}$ —associated to the phonological matrix /θrow/—, can be assigned multiple interpretations at LF, depending on its syntactic environment, as illustrated with the LF instructions in (39).

(39) <u>LF instructions</u> (Harley, 2014, p. 244)

  $\sqrt{77}$ ←→   'vomit'/ $[_v [[\_\_]_\sqrt{} [up]_p]]_{vP}$

        ←→   'a light blanket'/ $[_n [\_\_]_\sqrt{}]$

        {...other meanings in other contexts...}

        ←→   'throw'/ elsewhere

Following this rationale, a non-compositional RootC, such as (BP) *samba-canção* lit. samba+song 'boxers', will have its idiosyncratic meaning listed as part of

the LF instructions of both roots. For instance, $\sqrt{243}$ "SAMB-", in the context of root $\sqrt{38}$ "CANÇ-", will be assigned the meaning 'boxers'. The same meaning is codified as part of the LF instructions of the root $\sqrt{38}$ "CANÇ-" in the context of root $\sqrt{243}$ "SAMB-", as described on the right-hand portion of the LF instructions in (40).[6]

(40)   LF instructions for *samba-canção* 'boxers'

    a. $\sqrt{243}$ $\leftrightarrow$ "boxers"/ $[[n[\underline{\quad}]_{\sqrt{}}] [\sqrt{38}]n]_{nP}$

    b. $\sqrt{38}$ $\leftrightarrow$ "boxers"/ $[[n[\sqrt{243}]_{\sqrt{}}] [\underline{\quad}]n]_{nP}$

An account along these lines can also straightforwardly explain why some RootCs have just one of their constituent members interpreted idiosyncratically, such as those in (19), repeated below as (41).

(41)   BP

    a. *bolsa-sanduíche*

       financial.aid+sandwich

       'financial aid to graduate students visit a foreign university'

    b. *banana-maçã*

       banana+apple

       'Latundan banana'

    c. *cão-salsicha*

       dog+sausage

       'dachshund'

In these compounds, only the non-head noun —viz., the second member— is assigned a drifted interpretation, while the compound's head generally receives an elsewhere interpretation. This is illustrated with the compound *bolsa-sanduíche* in (42).

(42)   LF instructions for *bolsa-sanduíche* 'scholarship to visit a foreign institution'

    a. $\sqrt{59}$ $\leftrightarrow$ "financial aid; scholarship"

    b. $\sqrt{540}$ $\leftrightarrow$ "visiting period in a foreign institution"/$[[n[\sqrt{59}]_{\sqrt{}}] [\underline{\quad}]n]_{nP}$ "sandwich"/ elsewhere

To conclude, a short note on weak compositionality. Based on what has been discussed, we hypothesize that weak compositionality, as observed in (10b) – (10d), is an interpretive effect that arises when the compound is interpreted compositionally. This suggests that weak-compositionality is not necessarily dependent of a lexicalized meaning codified as part of an LF instruction, as is the case of the non-compositional RootCs in (40) and (42). Furthermore, weak compositionality seems to be restricted to nominal attributive compounds, and although their meaning may vary substantially, such variation is restricted by the attributive reading (Scalise & Vogel, 2010). We assume for now that weak compositionality is context-dependent, and may be regulated by pragmatic factors.

4. Final remarks

In this article, we argued that categorization is not an optional derivation step. Contrarily to approaches that resort to a delayed categorization to account for non-compositionality and lexical integrity effects in root compounding, we argued that non-compositionality is dissociated from categorization, and that lexical integrity effects are a reflex of a categorial domain established on the top of a complex structure (which generally comprises two categorized roots). We also argued that non-compositionality and weak compositionality should be analyzed as different LF phenomena. The former arises when LF instructions determine particular interpretations to a syntactic structure, while the latter is context-dependent, presumably pragmatically determined.

Notes

1. In the coordination reading, order variation is possible, *pastor-deputado* and *deputado-pastor* (Bauer & Tarasova, 2013; Arcodia, Grandi, & Wälchli, 2010), while in the attributive reading —when the speaker assigns prominence to one of the compound members— order variation is not allowed; for this reason, the interpretation in (15i) cannot be extracted from the compound *deputado-pastor*.

2. The vowel -*o*- connecting both roots is a linking element (LE), which is inserted for phonotactic reasons when the first root ends in a consonant, and the second root begins with a consonant (Nóbrega, 2013, 2014; Scher & Nóbrega, 2014). This LE does not need to be inserted in cases where this consonant cluster is not formed, e.g. *psic-análise* lit. psych+analysis 'psychotherapy' and *hidr-elétrica* 'lit. hidr+electric 'hydroelectric'.

3. In fact, (13a) is not capable of differentiating word-based from stem-based nominal compounds, commonly found in some Romance languages. Since Romance nouns are generally linked to a class marker, we can admit that stem-based and word-based compounds are, in these languages, two different modes of externalizing RootCs.

4. One consequence of assuming that roots are externally pair-merged with category heads is that they may not be in fact "categorized" (Alexiadou & Lohndal, 2017, p. 220). In light of this, we restate the canonical notion of categorization (*Categorization Assumption*; see Embick & Marantz, 2008) as follows: "A root must be merged in a strict local domain with a category-assigning head. Categorially non-individuated roots are not legitimate LF and PF objects, inducing formal crashing at the interfaces."

5. Since roots are adjoined to category heads, they could be seen as "optional." Thus, it would be expected to find syntactic structures with grammatical features, but no roots. Possible examples are "this is here", "it is here" (Emonds, 1985 *apud* Panagiotidis, 2011, 2014).

6. Borer (2014, p. 355-356) points out that this type of double marking, in cases of non-compositionality, would be a drawback. We do not see, however, how listing the same meaning as part of the LF instructions of both roots would interfere in a significant way in the compound's interpretation.

References

Acquaviva, P. (2009) Roots and Lexicality in Distributed Morphology. In A. Galani, D. Redinger, & N. Yeo. (Eds.) *York Papers in Linguistics Vol. 2* (pp. 1-21). York: University of York.

Acquaviva, P., & Panagiotidis, P. (2012). Lexical decomposition meets conceptual atomism. *Lingue e Linguaggio* 11(2), 165-180.

Alexiadou, A., & Lohndal, T. (2017). The structural configurations of root categorization. In L. Bauke, & A. Blümel. (Eds.). *Labels and Roots* (pp. 203-232). Berlim: de Gruyter.

Arad, M. (2003). Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory* 21, 737-778.

Arad, M. (2005). *Roots and Patterns:* Hebrew Morphosyntax. Amsterdam: Sprinter.

Arcodia, G. F., Grandi, N., & Wälchli, B. (2010). Coordination in compounding. In S. Scalise, & I. Vogel (Eds.). *Cross-disciplinary issues in compounding* (pp. 177-197). Amsterdam: John Benjamins.

Bauer, L., & Tarasova, E. (2013). The meaning link in nominal compounds. *SKASE Journal of Theoretical Linguistics* 10(3), 2-18.

Bauke, L. (2013). A Phase-Theoretic Analysis of Root Compounds (from a Cross-linguistic Perspective). In C. Rhys, P. Iosad, & A. Henry. (Eds.) *Microvariation, Minority Languages, Minimalism and Meaning: Proceedings of the Irish Network in Formal Linguistics* (pp. 150-170). Cambridge: Cambridge Scholars Publishing.

Bauke, L. (2014). *Symmetry breaking in syntax and the lexicon.* Amsterdam: John Benjamins.

Bauke, L. (2016). Parametric Variation in Nominal Root Compounding. In L. Eguren, O. Férnandez, & A. Mendikoetxea (Eds.) *Rethinking Parameters* (pp. 203-235). Oxford: Oxford University Press.

Bisetto, A., & Melloni, C. (2008). Parasynthetic Compounds. *Lingue e Linguaggio* 7(2), 233-260.

Bisetto, A., & Melloni, C. (2010). Parasynthetic compounds: Data and theory. In S. Scalise, & I. Vogel. (Eds.) *Cross-disciplinary issues in compounding* (pp. 199-217). Amsterdam: John Benjamins.

Borer, H. (2013). *Taking Form: Structuring Sense, Vol. III.* Oxford: Oxford University Press.

Borer, H. (2014). Wherefore roots? *Theoretical Linguistics* 40(3/4), 343-359.

Chomsky, N. (2000). Minimalist inquiries: the framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.) *Step by Step: Essays on Minimalist Syntax in Honor to Howard Lasnik* (pp. 89-155). Cambridge, MA: The MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti. (Ed.) *Structure and Beyond* (pp. 104-131). Oxford: Oxford University Press.

Chomsky, N. (2008). On Phases. In R. Freidin, C. P. Othero, & M. L. Zubizarreta (Eds.) *Foundational Issues in Linguistic Theory* (pp. 133-166). Cambridge, MA: MIT Press.

Chomsky, N. (2015). Problems of projection: Extensions. In E. Di Domenico, C. Hamann, & S. Matteini (Eds.) *Structures, strategies and beyond: Studies in honour of Adriana Belletti* (pp. 3-16). Amsterdam: John Benjamins.

De Belder, M. (2017). The root and nothing but the root: Primary compounds in Dutch. *Syntax* 20(2), 138-169.

Delfitto, D., Fábregas, A., & Melloni, C. (2011). Compounding at the interfaces. In: *Proceedings of the 39th Meeting of the North East Linguistic Society* (Amherst, MA), 253–268.

Di Sciullo, A. M. (2005). Decomposing compounds. *SKASE Journal of Theoretical Linguistics* 2–3, 14-33.

Di Sciullo, A. M., & Williams, E. (1987). *On the definition of word.* Cambridge, MA: MIT Press.

Embick, D., & Marantz, A. (2008). Architecture and Blocking. *Linguistic Inquiry* 39, 1-53.

Emonds, J. (1985). *A Unified Theory of Syntactic Category*. Dordrecht: Foris Publications.

Harley, H. (2014). On the identity of roots. *Theoretical Linguistics* 40(3/4), 225-276.

Marantz, A. (1995). "Cat" as a phrasal idiom: consequences of late insertion in Distributed Morphology, Ms., MIT.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2), 201-225.

Marantz, A. (2001). *Words and Things*. Handout, MIT.

Marantz, A. (2008). Phases and words. In S. Choe (Ed.) *Phases in the Theory of Grammar*. Seoul: Dong In Publisher.

Marantz, A. (2013). Locality domains for contextual allomorphy across the interfaces. In O. Matushansky, & A. Marantz (Eds.) *Distributed Morphology Today: Morphemes for Morris Halle* (pp. 95-115). Cambridge, MA: MIT Press.

Marvin, T. (2002). *Topics in the stress and syntax of words*. Doctoral Dissertation, MIT.

Marvin, T. (2013). Is word structure relevant for stress assignment? In O. Matushansky, & A. Marantz (Eds.) *Distributed Morphology Today: Morphemes for Morris Halle* (pp. 79-93). Cambridge, MA: MIT Press.

Nespor, M. (1999). Word Stress. In H. van der Hulst (Eds.) *Word Prosodic Systems in the Languages of Europe* (pp. 1-115). Berlim, New York: Mouton de Gruyter.

Nespor, M., & Ralli, A. (1996). Morphology-Phonology Interface: Phonological domains in Greek Compounds. *The Linguistic Review* 13, 357-382.

Nóbrega, V. A. (2013). A morfossintaxe da composição neoclássica. *Estudos linguísticos: Textos selecionados /Abralin – 2013* (pp. 274-289). João Pessoa: Ideia.

Nóbrega, V. A. (2014). *Tópicos em Composição: Estrutura, formação e acento*. Master's thesis. São Paulo. Universidade de São Paulo.

Nóbrega, V. A. (2015). Composição na Morfologia Distribuída: Dos universais à variação. *ReVEL* 13, 87-119.

Nóbrega, V. A., & Miyagawa, N. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the Integration Hypothesis. *Frontiers in Psychology* 6:271. *doi: 10.3389/fpsyg.2015.00271*

Nóbrega, V. A., & Panagiotidis, P. (2020). Headedness and exocentric compounding. *Word Structure* 13(2), 211-249.

Nunes, J. (2012). Sideward movements: Triggers, timing, and outputs. In M. Uribe-Etxebarria, V. Valmala (Eds.) *Ways of Structure Building* (pp. 114-142). Oxford: Oxford University Press.

Nunes, J., & Uriagereka, J. (2000). Cyclicity and extraction domains. *Syntax*, v. 3, n. 1, 20-43.

Padrosa-Trias, S. (2007). Catalan verbal compounds and the syntax-morphology competition. In F. Montermini, G. Boyé, & N. Hathout (Eds.) *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse* (pp. 91-107). Somerville, MA: Cascadilla Proceedings Project.

Panagiotidis, P. (2011). Categorial features and categorizers. *The Linguistic Review* 28, 325-346.

Panagiotidis, P. (2014). A minimalist approach to roots. In P. Kosta, L. Schürcks, S. Franks, & T. Radeva-Bork (Eds.) *Minimalism and beyond: Radicalizing the interfaces* (pp. 287-303). Amsterdam: John Benjamins.

Panagiotidis, P. (2015). *Categorial Features: A generative theory of word class categories.* Cambridge, UK: Cambridge University Press.

Panagiotidis, P. (2020). On the nature of roots: Content, form, identification. *Evolutionary Linguistics Theory* 2(1), 56-83.

Pigott, G., & Travis, L. (2013). Adjuncts within words and complex heads. In R. Folli, C. Sevdali, & R. Truswell (Eds.). *Syntax and its limits* (pp. 157-174). Oxford: Oxford University Press.

Pirelli, V. (2002). Per un superamento della dicotomia lessico-grammatica. Aspeti di composizionalità 'debole' nel linguaggio. In E. Ferrario, & V. Pulcini (Eds.) *La lessicografia bilingue tra presente e avvenire* (pp. 187-203). Vercelli, Italy: Edizioni Mercurio.

Ralli, A. (2008). Compound markers and parametric variation. *Language Typology and Universals* 61(1), 19-38.

Ralli, A. (2009). I.E. Hellenic: Modern Greek. In R. Lieber, & P. Štekauer. (Eds.) *The Oxford Handbook of Compounding* (pp. 453-463). Oxford: Oxford University Press.

Ralli, A., & Karasimos, A. (2009). The Bare-Stem Constraint in Greek Compound Formation. *Gengo Kenkyu* 135, 29-48.

Roeper, T., Snyder, W., & Hiramatsu, K. (2002). Learnability in a Minimalist Framework: Root Compounds, Merger and the Syntax-Morphology Interface. In I. Lasser (Ed.) *The Process of Language Acquisition* (pp. 25-35). Frankfurt: Peter Lang.

Roeper, T., & Snyder, W. (2005). Language Learnability and the Forms of Recursion. In: A-M. Di Sciullo (Ed.) *UG and External Systems* (pp. 155-169). Amsterdam: John Benjamins.

Scalise, S., Fábregas, A., & Forza, F. (2009). Exocentricity in compounding. *Gengo Kenkyu* 135(1), 49-84.

Scalise, S., & Vogel, I. (2010). Why compounding? In S. Scalise, & I. Vogel. (Eds.) *Cross-disciplinary issues in compounding* (pp. 1-18). Amsterdam: John Benjamins.

Scher, A. P., & Nóbrega, V. A. (2014). Unifying neoclassical and stem-based compounds: A non-lexicalist approach. *Revista LinguíStica* 10, 74-98.

Snyder, W. (1995). *Language Acquisition and Language Variation: The Role of Morphology.* Doctoral dissertation. MIT. Cambridge, MA.

Snyder, W. (2001). On the Nature of Syntactic Variation: Evidence from Complex Predicates and Word-Formation. *Language* 77, 324-342.

Walsh, M. (1976). *The Murinypata language of North-west Australia.* PhD dissertation. Australian National University.

Zhang, N. N. (2007). Root merger in Chinese compounds. *Studia Linguistica* 61(2), 170-184.

Zwitserlood, I. (2008). Morphology below the level of the sign "frozen" forms and classifier predicates. In J. Quer. (Ed.) *Proceedings of the 8th Conference on Theoretical issues in Sign-language Research* (pp. 251-272). Hamburg: Signum Verlag.