Morphological Exceptions to Vowel Reduction in Central Catalan and the Problem of the Missing Base*

Joan Mascaró

Universitat Autònoma de Barcelona. Centre de Lingüística Teòrica
joan.mascaro@uab.cat

Abstract

A certain class of Central Catalan compounds characterized by a first component that lacks a related output word is discussed and analyzed in connection with Vowel Reduction and Destressing. The first component of these compounds contains a vowel [a, ɛ, ɔ] that undergoes Destressing before the stressed vowel of the second component, but does not reduce. This causes an opacity problem because the generalization that there are no unstressed [a, ɛ, ɔ] is not surface-true for these cases. An analysis in the framework of parallel OT with output-to-output constraints is examined in detail and the function that returns the base of the relevant constituent in the candidates being evaluated is made precise. It is shown that such an analysis is not feasible. After showing that compounds have internal constituent structure even under noncompositional semantics, a Stratal OT analysis is presented that can handle such cases.

Keywords: compounds; Central Catalan; Vowel Reduction; Destressing; opacity; output-to-output constraints; Stratal OT; compositionality

Resum. *Les excepcions morfològiques a la reducció vocàlica i el problema de la base absent*

S'examina, en relació amb la reducció vocàlica i la desaccentuació, una classe de compostos en català central caracteritzats per un primer component que no està relacionat amb cap paraula en l'*output*. El primer component d'aquests compostos conté una vocal [a, ɛ, ɔ] que sofreix desaccentuació davant de la vocal accentuada del segon component, però que no es redueix. Això causa un problema d'opacitat ja que la generalització segons la qual no hi ha [a, ɛ, ɔ] àtones no és certa superficialment per a aquests casos. S'examina en detall una anàlisi en el marc de la TO paral·lela i es defineix de forma precisa la funció que assigna la base al constituent rellevant dels candidats avaluats. Es mostra que aquesta anàlisi no és satisfactòria. Després de demostrar que els compostos poden tenir estructura de constituents interna encara que no tinguin semàntica composicional, es presenta una anàlisi en el marc de la TO amb estrats que dóna compte d'aquests casos.

Paraules clau: compostos; català central; reducció vocàlica; desaccentuació; opacitat; restricció d'*output* a *output*; TO amb estrats; composicionalitat

Table of Contents

1. Introduction

The existence of analogical relations and groups of related words (paradigms) has been recognized for a long time. As put in general terms by Hermann Paul (1968 [1880]: 106) a long time ago, "individual words attract each other in the soul, and as a result a set of bigger or smaller groups arise."[1] The analysis of the role of paradigms in optimality theory has brought forward a range of different mechanisms to account for paradigm uniformity effects, most significantly the Base-Priority model of output-output correspondence, the theory of Optimal Paradigms, and cyclic evaluation in Stratal OT. When we face a given phenomenon, we may ask whether it is best analyzable, or analyzable at all, within a given model. Depending on the answers we get for different phenomena, we can either conclude that some of the models should be rejected or that linguistic theory must incorporate all or several of them in order to account for the facts. In this paper I analyze a set of compounds in Central Catalan with the intention to make some progress in answering such questions.

It has been noticed that in Catalan the stem of derivatives, the first component of compounds, and stressed prefixes lose their stress because they are followed by another stress. However, they differ in that the destressed vowel of derivatives reduces, but in the case of compounds and prefixed words the destressed vowel does not reduce. Before we examine such cases, some background in vowel reduction in Central Catalan is necessary (Mascaró 1976, 2002, Wheeler 2005).

Central Catalan has seven underlying vowels, /a, ε, ɔ, e, o, i, u/, and the derived vowel [ə]. Vowel reduction is governed by the initial, approximate generalizations in (1):

(1) a. All and only a, ε, ɔ, e, o, i, u appear in stressed position.

b. All and only i, u, ə appear in unstressed position.

In the case of alternations, vowel reduction follows the mappings a, ε, e → ə, and ɔ, o → u, as shown in (2a,b) below. (2a) shows stressed stems and (2b) the same stems destressed by a following stressed suffix, and (2c) presents cases of

---

1. "… attrahieren sich die einzelnen Wörter in der Seele, und es entstehen dadurch eine Menge grösserer oder kleinerer Gruppen."

nonalternating unstressed vowels. They all follow the generalizations in (1). The examples in (2d) will be discussed directly.

(2) a.   b.   c.   d.

    p[á]ra   p[ə]r[ɛ́]m   frar[ə]   fr[á]s[e]

    's/he stops'   'we stop'   'monk'   'sentence'

    tr[ɛ́]nta   tr[ə]nt[ɛ́]   bíg[ə]m   [í]t[e]m

    'thirty'   'thirtieth'   'bigamist'   'item'

    r[é]nta   r[ə]nt[é]s   [ə]l[ə]ment   [e]v[e]r[ɛ́]st

    's/he washes'   's/he washes-pres.subj' 'element'   'Everest'

    [ɔ́]bre   [u]brir[á]   tràng[u]l   pl[á]nct[o]n

    's/he opens'   's/he will open'   'turmoil'   'plankton'

    c[ó]ntra   c[u]ntr[á]ri   b[u]coi   b[o]ns[á]i

    'against'   'contrary'   'cask'   'bonsai'

    A set of cases that have been termed contextual, lexical, and morphological exceptions to vowel reduction in the literature disobey the generalizations in (1). In this paper we will be interested in *morphological exceptions* and will ignore *contextual exceptions* (Mascaró 1976: §1.6, 2002: 107-110, Wheeler 2005: 61-70), since they are not relevant to the central discussion. I will refer incidentally to *lexical exceptions* later; they affect a numerous set of lexical items in which some vowels appear as [e] or [o] even if unstressed. Unstressed [a], [ɛ], [ɔ], however, never appear as unstressed vowels in lexical exceptions.[2] The corresponding generalization (3a) is exemplified in (2d) above; (3b) is discussed directly.

(3) a. *Lexical exceptions.* A set of lexically marked items have underlying /e/, /o/ which do not undergo vowel reduction, even if unstressed.

    b. *Morphological exceptions.* In compounds and some prefixed words a destressed vowel in the first component does not undergo vowel reduction.

    Let us examine *morphological exceptions*. Beyond the (simple) word level, several processes of phrase phonology rearrange the stress contour of the concatenated words in different ways, and the generalizations in (1) are rendered opaque (Mascaró 2002: 93-95). The sentence *Empenyeu ara!* 'Push-pl now!' can be pronounced among other configurations with the stress pattern [ə̀mpəɲɛw áɾə], although the word stress pattern of the constituent words is [əmpəɲéw] and [áɾə]. The generalization (1a) is not obeyed because [ɛ] in the first word has undergone destressing but is not reduced, and (1b) is not obeyed because the initial [ə] has received stress. I will concentrate my attention on one of these processes, namely

---

2.   In northern subdialects [a] can appear also in lexical exceptions.

destressing in the first component of compound structures,[3] but there is a wide range of processes that have the same opacity effects.[4]

The different effects of a stressed vowel followed by another stressed vowel can be observed in the examples in (4a-b) repeated from (2a-b). (4a) shows a stressed vowel that is not followed by another stress. In (4b) the same root of (4a) is followed by a derivational or inflectional stressed suffix that causes destressing and vowel reduction. The examples in (4c) are compounds; the stress in the second component of the compound (4c) causes destressing but no vowel reduction in the first component. Finally, in (4d) the same word in (4a) is followed by another word within a phrase; in this case the stress weakens, but does not disappear, and the vowel does not reduce.[5]

(4)   a.   b.   c.   d.

p[á]ra   p[ə]r[έ]m   p[a]racaig[ú]des   p[à]ra caig[ú]des
's/he stops'   'we stop'   'parachute'   'stops falls'

tr[έ]nta   tr[ə]nt[έ]   tr[ε]nta-c[í]nc   tr[ὲ]nta c[í]ncs
'thirty'   'thirtieth'   'thirty-five'   'thirty fives'

r[é]nta   r[ə]nt[é]s   r[e]ntapl[á]ts   r[ὲ]nta pl[á]ts
's/he washes'   's/he washes-pres.subj' 'dishwasher'   's/he washes dishes'

[ɔ́]bre   [u]brir[á]   [ɔ]brell[á]unes   [ɔ́]bre ll[á]unes
's/he opens'   's/he will open'   'can opener'   's/he opens cans'

c[ó]ntra   c[u]ntr[á]ri   c[o]ntracultur[á]l   c[ò]ntra cult[ú]res
'against'   'contrary'   'countercultural'   'against cultures'

(5) shows the structures of the examples in (4b-d), illustrated with one of them, and (6) states the corresponding generalizations.

(5)   a. *Derivative*   b. *Compound*   c. *Phrase*

[$_{Wd}$[$_{St}$trənt] [$_{Sf}$έ]]   [$_{Wd}$[$_{Wd}$[$_{St}$trεnt-ə]] [$_{Wd}$[$_{St}$síŋ]]]   [$_{NP}$[$_{Wd}$[$_{St}$trὲnt-ə]] [$_{Wd}$[$_{St}$síŋ] s]]
'thirtieth'   'thirty-five'   'thirty fives'

---

3. I will use the term *compound structure* to refer to both compounds and prefixed words.
4. There is a tradition, mainly in the more prescriptively-oriented literature (but also elsewhere, Oliva 1992, Wheeler 2005), that considers that the first element of compound structures retains a secondary stress. This assumption derives from a confusion between word stress and emphatic and rhythmic stress. Those authors who have based their conclusions on experimental results (Mascaró 1976, Prieto 2003, Nadeu 2016) assume destressing, as does Recasens (1993). Destressing also follows from the transcriptions of accurate traditional historical linguists (e.g. Moll 1931, Coromines 1989-1997).
5. Destressing of the first constituent of compounds also obtains in Spanish (Hualde 2007).

(6) a. *(Simple) word stress.* The stressed vowel of a root/stem of a derivative or inflected form followed by a stressed suffix loses its stress and reduces.

   b. *Sentence stress.* A word followed by another word in a phrase usually weakens its stress but does not lose it. The vowel does not reduce.[6]

   c. *Underapplication of vowel reduction in compound structure.* The stressed vowel of the first component of a compound structure loses its stress but does not reduce.

2. Compound structures: stress, vowel reduction and missing bases

Let us examine in more detail the morphology and phonology of morphological exceptions. Focusing on their first component, we might distinguish three types of compounds:

(7) a. $[[L_O] [C]]$ L_O is a lexical element with an independent output

   parə-kəjɣúðəs (4c)

   b. $[[L_B] [C]]$ L_B is a lexical element with no independent output

   nɛu-klásik (12)

   c. $[[I] [C]]$ I not a lexical element (it is an *isolate*)   bɛt-əkí (9)

   In (7a) the first constituent of *parə-kəjɣúðəs* is related to the lexical element *párə* 's/he stops', whereas *nɛu* in *nɛu-klásik* in (7b) is a lexical element but is a bound form that never appears as the output *[nɛ́u]. In the case of *bɛt-əkí* the first constituent *bɛt* cannot be successfully related to any independent lexical item, word or affix. I will use the term *isolate* to refer to such constituents, and I will refer to constituents that lack a related output (both (7b) and (7c)), the structures that are of interest in this paper, as *constituents that lack a base* or *constituents with a missing base*.

   Compound structures with a first component missing a base can have several origins. Some of them originate through word formation processes that use as a first component a lexical item that is not a word, as is the case with reduplicative compounds, prefixes and many neoclassical compounds (§§ 3.2, 3.4). In the case of borrowings (§ 3.3) the compound structure in the source language can be preserved with loss of the original semantic compositionality. Finally, linguistic change can result in loss of semantic compositionality with preservation of constituent structure; since this case is of specific interest I will consider it briefly below. What is important to notice is that, no matter what the origin, all these cases result in the same structure: a compound structure that contains two constituents, the first being unrelated to an independent output and an exception to vowel reduction. The examples presented in this paper correspond to my own speech, but even if there

---

6. Under different circumstances, in particular under stress clash, the first stress can disappear, as in the third syllable of the sentence *Empenyeu ára* cited above.

is individual variation with regard to particular words, for all speakers the effects described are robust.

In productive compounds that show transparent compositional meaning like *para-sol* 'parasol', literally 'stops sun', and *parallamps* 'lightning rod', lit. 'stops lightning', the speaker implicitly knows that they share a common first constituent, and for *para-sol* and *gira-sol* 'sunflower' that they share a common second constituent.[7] Some compounds are not compositional semantically and keep their constituent structure, e.g. [[matə][pərén]], a toxic mushroom of the family Boletus, lit. 'kills relative'. In other cases noncompositionality is accompanied by loss of constituent structure, usually through historical change. For a word like *passa-port* [pəsəpórt] 'passport' (see 8a), the speaker can only discover the fact that [pəsə] is (historically) related to the form [pásə] of the verb *passar* 'to pass' through the orthographic form and rational deduction.[8] This process of lexicalization of productive compounds can result in the loss of their original internal constituent structure, the two components becoming a monomorphemic root. As should be expected, in such cases, since the compound structure has been lost, all the vowels within the original first component reduce. Some examples are presented in (8a), followed in (8b) by their form in the reconstructed, previous, nonlexicalized stage; (8c) shows a proper synchronic compound with a first component coinciding with the one in (8b):

(8) a. *Lexicalized*   b. *Previous form*   c. *Compound with same 1st component*

[pərájɣwə]   *[[paɾ][ájɣwə]]   [[paɾə][ʎáms]]
'umbrella'   'stops water'   'lightning rod', lit. 'stops lightning'

[pəsəpórt]   *[[pasə][pórt]]   [[pasə][puɾé]]
'passport'   'passes gate'   'food mill', lit. 'passes purée'

[bərsəmblán]   *[[ber][səmblán]]   [[beɾə][mén]]
 'likely, credible'   'true looking'   'truly', lit. 'true-suffix'

[kəpfikát]   *[[kap][fikát]]   [[kab][ʒiɾát]]
'concerned'   'head put'   'upside-down', lit. 'head turned'

[fərukəríl]   *[[fɛru][kəríl]]   [[fɛru][məŋnətízmə]]
'train'   'iron track'   'ferromagnetism', lit. 'iron magnetism'

[mələmén]   *[[malə][mén]]   [[mal][pərlát]]
'badly'   '(in) bad mind'   'foulmouthed', lit. 'bad spoken'

[bunumíə]   *[[bɔn][umíə]]   [[bɔnə][βəntúɾə]]
'conviviality'   'good manhood'   'fortune telling', lit. 'good venture'

7. It is easy to elicit such relatedness by different means. Speakers can also make up expressions built on *gira-sol*, lit. 'turns-flower', like *Gira més que un gira-sol* 'it turns more than a sunflower', or *El fa girar més el sol que un gira-sol* 'it is turned around more by the sun than a sunflower'.
8. Although most Catalan compounds are not hyphenated, from now on their components will be always separated by a hyphen in order to make them apparent.

But crucially, in other cases of lexicalization, in fact in the majority of cases, the compound undergoes semantic drift but the constituent structure is retained and the two components do not become a monomorphemic root. Let us examine a specific case in some detail. The noun meaning 'stone', derived from Latin *pĕtrām*, is part of many place names. The regular evolution of the internal cluster is TR>dɾ after a stressed vowel, TR>ɾ after an unstressed vowel (Coromines 1971: 183-188),[9] thus we get *pĕtrām>péðɾə*. When the stress of the preceding vowel disappears as in compounds (also in proclitic prepositions: *rĕtrō>reɾə* 'behind') the result of TR is [ɾ]. In compounds formed at the stage when the phonetic change took place, we got *pĕtram>péðɾə* for the simple word, but when this word is the first component of a compound, as in *Pera-tallada*, lit. 'cut stone', the evolution is *pĕtrām-talleātām>peɾə-təʎáðə*. At some point after this stage the lexical relation *péðɾə-péɾə* is lost; at later stages transparent place name compounds of the same class are formed again on *péðɾə*, e.g. *peðɾə-fórkə Pedra-forca*, lit. 'fork stone'. What is relevant here is that at the time that the result of Latin *pĕtram* in the noun in isolation and in the first position of the compound diverged, this did not force vowel reduction of the unstressed vowel, i.e. it is not the case that /[peɾə-təʎáðə]/ became *[pəɾətəʎáðə]. We could possibly argue, for present day Catalan, that /[peɾətəʎáðə]/ has lost its compound structure, and that its syllable-initial /e/ is marked as a lexical exception to vowel reduction (see (2d) above, and discussion), as in [soprán-o] 'soprano', [deskáɾt-es] 'Descartes', [memoɾándum] 'memorandum' (Mascaró 1976, 2002, 2015). As argued in the next section, this explanation is impossible for cases with unstressed [a], [ɛ], [ɔ], which can never be exceptions to vowel reduction. But it does not hold even in cases with [e], [o] like [peɾətəʎáðə], because lexical exceptions to vowel reduction are recent (they probably arose in the 19th century), and it is necessary to assume a stage after TR>ɾ in which there were no exceptions to vowel reduction and in which *Pera-tallada* had to have the compound structure /[[peɾə]-[təʎáðə]]/. The obvious explanation for the lack of vowel reduction is that the word did not lose its compound structure and that it remained /[[peɾə]-[təʎáðə]]/, even though /[peɾə]/ was unrelated at this point to any other lexical item. This compound structure has survived in several place names (*Pera-fita, Pera-fort, Pera-tallada*) for more than ten centuries.[10]

Summing up, we can distinguish two classes of [[C][C']] compounds, those like *trenta-cinc* in (8c), for which the first component [C], *trenta-* (4a) can be identified as an independent synchronic lexical element, the numeral *trenta*, and those like *Pera* in *Pera-tallada*, for which we can only identify the first component [C] as a word through diachronic analysis. Following current practice, we will refer to this independent lexical element related to a given constituent [C] (or [C']) of a compound structure as the *base* of [C] (or [C']). I will use the term *isolate* to refer to the member of a compound structure with a missing base, like *Pera* in *Pera-tallada*. Thus the preposition *contra* is the base of *contra-* in *contra-cultural* (4c), but the

---

9. We also get ɾ when a vowel between T and R was lost, i.e. VTVR>Vr: *mātĕr>máɾə* 'mother'.
10. There is loss of constituent structure and reduction in the case of *Peralada* [pəɾəláðə]<*pĕtrām-lātam*.

first components in *anti-cultural* 'anticultural' does not have a base and is hence an isolate. In the next section I will examine in turn different sets of cases of word with a first component that lacks a base that does not undergo vowel reduction. In order to exclude a possible analysis based on lexical exceptionality, which, as already said, affects only [e] and [o], all the examples will be cases of unreduced [a], [ɛ], [ɔ] in their first component.

3. Different kinds of compounds with a first member lacking a base

3.1. Proper and common nouns

One group of isolates with unreduced [a], [ɛ], [ɔ] contains place or person names and some common nouns. Examples are displayed in (9).

(9)	*1st component, etymon*	*Etymon glosses*	*Meaning*[11]
palə-mós	pălūm > Ø	marsh - liquid	*town name*
palə-furʒéʎ	pălātĭūm > pəláw	palace - *pers. name*	*town name*
bɛt-əkí	vĭdēte > bəʒéw	see-2pl.imp - here	'here it is'
kamə-miʎə	*Greek* chamai-mēlon	on the ground - apple	'chamomile'
kɔβrə-ʎít	kɔ́βre (Old Cat.) > kuβréʃ	covers - bed	'bed cover'
kɔβrə-kálzə	kɔ́βre (Old Cat.) > kuβréʃ	covers - chalice	'chalice cover'
ʃɛrkə-βíns	sérkə > ʃérkə	seeks - wine-pl.	*family name*
bɛni-kásim	*Arabic* banī	sons - *pers. name*	*town name*
taɣə-mənén	—	*unknown*	*mountain name*
tɛri-káβɾəs	—	*unknown*	*family name*
parə-sɛtə-mɔ́l	—	*novel creation*	'paracetamol'

In *palə-mós* the Latin etymon of [palə], *pălūm* 'marsh', has left no simple word descendants and as a result the first component has become an isolate; the first component in *palə-furʒéʎ*, from Lat. *pălātĭūm*, has resulted in *pəláw*, to which *palə* cannot be related anymore. For both cases, there is a noun *pálə* 'shovel' which cannot be put in any regular compounding relation to *mós* 'bite', nor to *furʒéʎ*, which is not an independent word. In *bɛt-əkí* the original first component, *vĭdēte* (regularly >*bəʒéw*), has been reduced diachronically to *bɛt* which no longer can be related synchronically to *bəʒéw* or to any existing word.[12] In the case of *kamə-miʎə* the

---

11. Orthographic forms: *Palamós, Palafrugell, vet aquí, camamilla, cobrellit, cobrecalze, Xercavins, Benicàssim, Tagamanent, Terricabres*, *paracetamol*.
12. For many speakers it is *bɛt-əkít* with the second component also an isolate.

original Greek compound either kept its complex character through Latin or was reinterpreted as a compound after a monomorphemic phase. The next two examples contain the apparent verbal form *kɔβrə*. In Old Catalan the verb *cobrir* belonged to conjugation IIIb and 3sg.pres.ind. had the form *kɔ́βɾe*. Later, when the verb changed to conjugation IIIa, *kɔβrə* became an isolate, because the 3sg.pres.ind. form had become *kubɾ-ɛ́ʃ*. Today speakers might establish a regular compound relation of VN compounds with regular compositional semantics for items like *cobre-llits* ('instrument used to cover beds'). But if the base of *kɔβrə* were *kubɾɛ́ʃ*, we would predict *kuβɾə-ʎít*. There is another verb *cobrar* 'to cash, to retrieve' with 3sg.pres. ind. *kɔ́βrə* which could be identified as the correspondent, but then we would get the wrong semantics, i.e. *'bed casher/retriever' instead of 'bed cover'. In *ʃɛrkə-βíns* the sporadic change s>ʃ has rendered the relation to the base opaque. In the following three cases the Arabic or pre-Roman origin of the first component makes the identification of a base impossible. Finally, the last example *parə-sɛtə-mɔ́l* is a case of two nonfinal isolate components with destressed and unreduced vowels.

3.2. Reduplicative compounds

Another source of isolates is found in reduplicative expressive compounds. The members of a group of such compounds consist of a reduplicated monosyllabic or disyllabic component, in many cases of onomatopeic origin. Thus if speakers want to imitate the sound of a boiling syrup they can invent the imitation form or inter-jection [blɔbblɔp] which can be pronounced with varying degrees of stress on its syllables. But if it is turned into a noun as in the sentence *Quan sentis el blop-blop, para* 'when you hear the blop-blop, stop' it will appear invariably as [blɔbblɔ́p] with final stress and unreduced unstressed vowel in the first component. Here are some actual common cases (Cabré 1993, 2002; Riera-Eures 2002).[13]

(10) nyeu-nyeu	ɲɛwɲéw	'hypocritical talk' (N)
mec-mec	mɛgmɛ́k	'horn blast' (N)
nyam-nyam	ɲamɲám	'eating' (N)
tau-tau	tawtáw	'on equal terms' (Adv.)
xano-xano	ʃanuʃánu	'(walking) slowly' (Adv.)
poti-poti	pɔtipɔ́ti	'disorder, chaos' (N)
taf-taf	taftáf	'sound of an engine' (N)
(fer la) gara-gara	garəɣárə	'to try to please through flattering' (N) (lit. 'to make the *gara-gara*')

---

13. We can also assign to this class a set of borrowings with identical structure: *Bora-Bora*, *Baden-Baden*, *pai-pai*, *tse-tse*, *beri-beri*, *yo-yo*.

Of course we also find cases in which the first component of these reduplicative compounds is not an isolate: krɛk-krɛ́k 'repeated cracking sound' (N), krɛ́k 'cracking sound'; mɛwmɛ́w 'repeated meow' (N), mɛ́w 'meow'. But in the examples in (10) the first component is always an isolate, because there is no independent word it can be related to.

3.3. Borrowings

Borrowings are another source of isolates. I show some instances in (11). Some of these examples present, in addition to the nonreduced destressed vowels [a], [ɛ], [ɔ], the nonreduced vowels [e], [o] that correspond to lexical exceptions (see §1), as in [rɔkefélər] and [lɛjdmotíf]:

(11) McDonalds　　mag-dónəls　　Eng.

　　　playboy　　　plɛj-βój　　　Eng.

　　　Pepsi-Cola　　pɛpsi-kólə　　Eng.

　　　Quasimodo　　kwasi-móðo　　Lat.

　　　leitmotiv　　　lɛjd-motíf　　Ger.

　　　Tel Aviv　　　tɛl-əβíp　　　Heb.

　　　Vietcong　　　bjek-kóŋ　　　Viet.

　　　outsider　　　ɔwt-sájðər　　Eng.

　　　boy scout　　　bɔj-əskút　　Eng.

　　　Rockefeller　　rɔke-félər　　Eng.

The reasons for attributing compound structure to borrowings is diverse. In some cases the orthographic form might have suggested compound constituent structure (Tel Aviv), in others the speakers who introduced the borrowing might have known the source language, where the word was a compound (playboy, leitmotiv) or a phrase (Quasimodo).[14]

3.4. Neoclassical compounds and prefixes

Many cases of compounds with a missing base are found in neoclassical compounds. In this case the first component is not an isolate, because it is a lexical element, but one which has no independent output, (7b) in the previous classification. The relevant examples displayed in (12) show an unstressed, unreduced vowel in the first component. Here and in (13) I omit the gloss if the English form is sufficiently close to the Catalan one.

---

14. Quasimodo derives from the Gregorian Introit *Quasi modo geniti infantes* 'in the way of newborn babies', based on 1 Peter 2:2.

(12) [a]nglo-franc[ɛ́]s   'Anglo-French'   [ɔ]vi-f[ó]rme

    m[ɛ]ta-llengu[á]tge   'metalanguage'   c[ɔ]rtico-ester[ɔ́]ide

    l[a]bio-dent[á]l   [ɛ]cto-pl[á]sma

    h[ɛ]li-oc[ɛ́]ntric   p[a]ra-norm[á]l

    p[ɛ]tro-d[ɔ́]lar   t[ɛ]tra-pl[ɛ́]gia

    [ɔ]rto-tipograf[í]a   c[a]rdio-vascul[á]r

    p[ɔ]li-traumat[í]sme   n[ɛ]o-cl[á]ssic

The other source of words with compound constituent structure and a first component which is a lexical element with no independent output is the class of stressed prefixes.[15] Notice that since prefixes are, by definition, not independent words, it is impossible to establish a precise border between clear prefixes like *pre-* in pr[ɛ]-nat[á]l in (13) and "root" isolates like *cardi(o)-* in c[a]rdio-vascul[á]r in (12). Since this distinction does not bear on the issues we are dealing with, the problem will not be addressed here. Some examples of isolates traditionally described as prefixes are shown in (13):

(13) [a]nti-c[ɔ́]s   'antibody'   ps[ɛ]udo-probl[ɛ́]ma

    [a]rxi-satisf[é]t   'supersatisfied'   pr[ɛ]-nat[á]l

    p[ɔ]li-sil·l[á]bic   p[ɔ]st-operat[ɔ́]ri

    tr[a]ns-sexu[á]l   s[ɛ]mi-form[á]l

4. Isolates and complex structure

Missing bases are bound morphemes like the examples in §3.4, or isolates like the examples in §3.1-3.3 and are crucial to the arguments I present in §5-7. Since isolates are peculiar structures, in this section I will examine the evidence in favor of their existence. Containing an isolate implies having a complex structure [$_Z$ [I] [Y]], where I is an isolate. Therefore we must rule out the possibility that Z has a flat structure ([$_Z$I⌢Y], X⌢Y a monomorphemic sequence) and give evidence in favor of the internal constituent structure. A possible argument against the complex structure [$_Z$ [I] [Y]] is that it violates strict compositionality: the meaning of [$_Z$ [I] [Y]] cannot be a function of the meaning of its parts because I is not a lexical element and therefore it has no meaning. An expression like [$_Z$ [$_ı$bɛt] [$_{Adv}$əkí]] 'here it is' should be monomorphemic. But there is extensive evidence against strict compositionality (Jackendoff 1997, 2010b, Jackendoff

---

15. There are also unstressed prefixes, like *a-*, *des-*: *ə-nurmál* 'abnormal', *dəz-órjðrə* 'disorder'. In stressed prefixes the existence of underlying stress is justified by the lack of reduction. In neoclassic compounds it can be justified also, in some cases, by the existence of surface stress in the case of components that can appear both in first and in second position: *m[ɔ]rfo-gènesi, antropo-m[ɔ́]rf; d[a]ctilo-forme, mono-d[á]ctil*.

& Audring to appear), which can be summarized in the generalization in (14a), briefly exemplified in (14b-e):

(14) a.   There are linguistic expressions that have complex structure and noncompositional meaning.

    b.   *Idioms*. Eng. *kick the bucket*; Cat. *prendre el pèl* 'to fool somebody', lit. 'to take somebody's hair'; Cat. *dinyar-la* 'to die' (**dinyar* , obj. clitic *la* nonreferential)

    c.   *Inflected forms*. pluralia tantum, Eng. *jean-s*; Cat. *pantalon-s* 'trousers'

    d.   *Derivatives*. Eng. *prob-able* (cf. *prob-abil-ity*); Cat. *lubr-i[k]* 'lubricious' (cf. *lubr-i[s]-itat* 'lubricity)

    e.   *Compounds*. Eng. *bull's eye*; Cat. *mata-parent* 'Boletus satanas', literally 'kills relative'

In (14b) the lack of semantic compositionality is obvious, and the need for several constituents also, because the verbs can inflect. The pluralia tantum in (14c) contain a plural morpheme because it triggers agreement and because the morpheme shows the typical plural allomorphy (*jean[z], pant[s], breech[ɪz]*). For Catalan, the sequence *pantalon-s* cannot be monomorphemic because in this case when referring to a set of trousers it would have a plural **pantalonsos*, as in monomorphemic *descans - descans-os* 'rest/rests'. In (14d) there must be a second component, the suffix *–ic*, as shown by its peculiar allomorphy.

In the case of compounds (14e) similar arguments apply. If [Y] in [$_Z$[I][Y]] is a constituent, then given that [$_Z$] is a proper bracketing of the terminal structure it dominates, [I] must also be a constituent. The second component Y is a constituent because it is identifiable as a lexical element, as in the examples in (15a), repeated from (9-13); '—' marks isolates:

(15)   *Components' glosses*   *Gloss*

    a.   Pera-tallada   —   'cut-part.fem-sg'   Town name
       vet-aquí   —   'here'   'here it is'
       Pepsi-Cola   —   'cola drink'
       ecto-plasma   —   'plasm'   'ectoplasm'
       Pala-folls   —   crazy-pl   Town name

    b.   poti-poti   —   —   'disorder, chaos' (N)

    c   Taga-manent   —   —   Mountain name
       cama-milla   —   —   'camomile'
       ɔwt-sájðər   —   —   'outsider'
       teri-yaki   —   —   'teriyaki'
       paɾə-sɛtə-mɔ́l   —   —   —   'paracetamol'

In some cases there is additional evidence for an analysis of the first component as an isolate. In the town name *Pala-folls* the second member is identified with the masculine plural form *foll-s* of the adjective *foll* 'crazy'. If the sequence *foll-s* were monomorphemic we would expect, like in the case of *pantalon-s* discussed above, the plural **Pala-folls-os*, as in *descans - descans-os* 'rest/rests', instead of *Pala-foll-s*. Notice that the semantic plural of fully compositional compounds with a plural second component is also identical to the semantic singular: *un espanta-ocell-s* 'one scarecrow', *dos espanta-ocell-s* two scarecrows'. The identifiability of the second component of *Pala-foll-s* is confirmed by the derived demonym *pala-foll-enc*. In the case of reduplicative compounds (15b), examined in §3.2, the evidence is furnished by the reduplication process itself. The examples in (15c), in which both components are isolates ([$_Z$ [I][I]]) constitute a special case. The only evidence for complex structure, at least for most cases, is the lack of vowel reduction in the first component. If we derive complex structure from lack of reduction we cannot predict the latter from the former. Therefore these cases do not constitute solid evidence for the arguments put forth in §5-8. The only plausible way to account for their phonological behavior, however, is to assume complex structure.[16]

5. Parallel and cyclic analyses

In a parallel framework like classical OT underapplication of vowel reduction causes an opacity problem, because given a noun like *para-caigudes* 'parachute', lit. 'stops falls', there is no way of forcing reduction in the destressed /a/ in the derivative *kəjg-úð-ə-s* without forcing it also in **pəɾ-ə*. In /paɾ-ə-kajg-ud-ə-s/ the constraints responsible for stress placement will favor *páɾ-ə-kájg-úð-ə-s*, but those responsible for destressing will favor elimination of all nonfinal stresses. Vowel reduction will then force both nonfinal vowels to reduce, yielding **pəɾ-ə-kəjg-úð-ə-s,* no matter how the constraints are ordered, as shown in (16). In (16) Final Stress stands for the set of constraints that assign stress to one of the last three syllables of a word, **ă,ĕ,ɔ̆,ĕ,ŏ* for the constraints that force vowel reduction, and **[$_{Wd}$ ...ˈV...ˈV...]* for constraints that favor candidates with a single rightmost stress in the simple or the compound word; '□', the necessity operator, marks the needed, acceptable candidate:

16. In some other cases there might be some kind of "weak identification" of one of the components when there is repeated partial coincidence, as in *Viet-nam*, *Viet-minh*, *Viet-cong*; *MacDonald*, *MacArthur*, etc.; *Amsterdam*, *Rotterdam*. Köhnlein (2015) gives extensive evidence for complex structure in Dutch place names with noncompositional semantics like *Wagening-en*, *Loos-drecht* and *Amster-dam*. He also shows that such structures are widespread in Germanic and in many languages from different families.

(16) *Regular compounds, parallel analysis*

paɾ-ə-kajg-ud-ə-s	Final Stress	*[_Wd ...ˈV...ˈV...]	*ă,ĕ,ɔ̆,ĕ,ŏ
paɾ-ə-kajg-ud-ə-s	*!		**
páɾ-ə-kəjɣ-úð-ə-s		*!	
☐ paɾ-ə-kəjɣ-úð-ə-s			*!
☞ pəɾ-ə-kəjɣ-úð-ə-s			

Harmonic Serialism would face the same problems, since constraint ordering is fixed. The obvious solution to this problem in parallel OT is transderivational output to output correspondence (OO-correspondence), which establishes a correspondence relation between two elements in different derivations (Kenstowicz 1996, Benua 1997, Kager 1999, Downing 2005, Downing et al. 2005). In the case of a compound structure with two constituents, [[C][C']], the first constituent C can be put into morphological correspondence (notated here as $\Re_M$) with some base, a word or a constituent C^B, more specifically with the output C'^B_O of C^B: $C\Re_M C'^B_O$. Vowel reduction in [C] underapplies because, given $C\Re_M C'^B_O$, [C] (more precisely, C's candidates) must be faithful to C'^B_O with respect to some specific properties. In our example the vowel features of the output of the /a/ of /paɾ-ə/ in /paɾ-ə-kajg-ud-ə-s/ must be faithful to those in the output of the independent verbal form [páɾ-ə] 'stop-3sg.pres.ind':

(17) *Regular compounds, parallel analysis with OO*

[_N[_Vpaɾ-ə][_Nkajg-ud-ə-s]] *Base*: [_Vpáɾ-ə]	OO-Id(V-features)	*[_Wd ...ˈV...ˈV...]	*ă,ĕ,ɔ̆,ĕ,ŏ
a. ☞ [_N[_Vpaɾ-ə][_Nkajg-úd-ə-s]]			*
b. [_N[_Vpáɾ-ə][_Nkájg-úd-ə-s]]		*!*	
c. [_N[_Vpáɾ-ə][_Nkəjg-úd-ə-s]]		*!	
d. [_N[_Vpəɾ-ə][_Nkəjg-úd-ə-s]]	*!		

In nonparallel Stratal OT (Kiparsky 2000, in press; Bermúdez-Otero 2003, 2011, in preparation), the idea that the first /a/ in *para-caigudes* does not reduce because *para-* is an independent element of a specific sort can be captured by different mechanisms, namely constituent structure and stratal ordering. Since *para* and *caigudes* are independent lexical elements, they must be cyclic constituents in the compound [_N[_Vpaɾ-ə][_Nkəjɣ-úð-ə-s]]. The constituents *para* and *caigudes* are processed first, in the word cycle, and are assigned stress. The stressed /a/ in [_Vpáɾ-ə] is not reduced, but the unstressed /a/ in [_Nkajɣ-úð-ə-s] undergoes reduction. The next cycle is the phrase cycle, where destressing can reapply because there is a stress to the right of [_Vpáɾ-ə], namely the ú in [_Nkajɣ-úð-ə-s]. But vowel reduction does not apply in this cycle, because it is a word-level process, i.e. it does not

apply in compound structures and beyond. In Lexical Phonology this would mean that the rule is not present in the phrasal cycles. In Stratal OT the corresponding mechanism is a different constraint ordering in the phrasal cycle, which will prevent the candidate with reduction from surfacing. The derivation proceeds as in (18a-b).

(18) *Regular compounds, stratal analysis*

 a. Word stratum $*[_{Wd} ...'V...'V...] \gg$ Final Stress*, Id(V-features); ă,ĕ,ɔ̆,ĕ,ŏ $\gg$ Id(V-features)

	paɾ-ə	$*[_{Wd} ...'V...'V...]$	Final Stress	*ă,ĕ,ɔ̆,ĕ,ŏ	Id(V-features)
☞	páɾ-ə				
	paɾ-ə		1W	1W	
	pə́ɾ-ə				1W

	kajg-ud-ə-s	$*[_{Wd} ...'V...'V...]$	Final Stress	*ă,ĕ,ɔ̆,ĕ,ŏ	Id(V-features)
☞	kəjg-úd-ə-s				1
	kajg-ud-ə-s		1W	1W	L
	kájg-úd-ə-s	1W			L
	kajg-úd-ə-s			1W	L

 b. Phrase stratum *[Wd ...'V...'V...], Id(V-features) $\gg$ *ă,ĕ,ɔ̆,ĕ,ŏ

	páɾ-ə-kəjg-úd-ə-s	$*[_{Wd} ...'V...'V...]$	Id(V-features)	*ă,ĕ,ɔ̆,ĕ,ŏ
☞	paɾ-ə-kəjɣ-úð-ə-s			1
	páɾ-ə-kəjɣ-úð-ə-s	1W		L
	pəɾ-ə-kəjɣ-úð-ə-s		1W	L

Both analyses, parallel and stratal, make similar predictions for compositional compounds like $[_N[_V\text{paɾ-ə}][_N\text{kəjɣ-úð-ə-s}]]$. For the OO-correspondence analysis compositionality implies the existence of an independent lexical form, for the stratal analysis it implies cyclic structure. Predictions differ, however, when there is internal constituent structure but there is a first constituent that is not an independent element, or is an element that lacks an independent output.[17]

6. The identification of the base

It follows from the previous considerations that a crucial step in the OO-correspondence analysis, and in general in the analysis of paradigm effects, is the determination of the independent output form with which the candidate stands in correspondence, its base, and the determination of the class of elements affected

---

17. Similar considerations are formulated by Trommer (2013: §2.5).

by the constraint. This is what Bachrach & Nevins (2008b) call *the asymmetry question*, i.e. "why […] do identity effects go from some members of the paradigm towards others, and not vice versa?" and the *inclusion question*, "what is the set of relevant forms that learners put together", i.e. what constitutes a relevant paradigm for identity effects. In many cases it is evident what the base of a candidate is, but in other cases it is not so clear.[18] Therefore a more explicit and careful formulation of the candidate-base relation is in order. I will follow the standard assumption that the elements standing in correspondence are words or clitic groups, or constituents they contain; in what follows the term *constituent of a word W* has to be understood not as denoting a proper subconstituent, but as possibly referring also to the word itself.

The correspondence relation that is crucial for the OO-constraints is established between a constituent C of an input word W and a constituent C_O^B of a morphologically related output word W^B, its base; as indicated above, C and C_O^B may also coincide with the whole words W and W_O^B, respectively. For *para-caigudes* in (17), W is the input noun /[$_N$[$_V$paɾ-ə][$_N$kajg-ud-ə-s]]/ and W^B is the verbal from /[$_V$paɾ-ə]/; the relevant constituents are the first constituent of the input noun, [$_V$páɾ-ə] and **[$_V$páɾ-ə]**, the output of the independent verbal form /[$_V$paɾ-ə]/ (henceforth, bases are set in boldface whenever they are not easy to distinguish from their correspondent).

This correspondence relation, $C\Re_M C_O^B$ is mediated by a morphological process M that relates the corresponding words: for any specific OO-constraint we should identify the base C_O^B of any input constituent C. We can define this identification as the function $\Im$ whose domain is the set of triples <C, W, M>, with C a constituent in an input word W, and M a morphological process, and whose range is the set of constituents in output words plus the null set. We can also view $\Im$ as the product of two functions $\Im_{LEX}$, which returns the lexical base constituent C^B, and $\Im_{GenEval}$, which applies to C^B and returns C_O^B, the output of C^B. In (19) $\Im$ is illustrated with the identification of the base of the first component in *para-caigudes;* M_{Cpd} stands for the set of relevant compounding and prefixation processes.

(19) *Identification of the base for* /[$_V$paɾ-ə]/ *in* /[$_N$[Vpaɾ-ə][$_N$kajg-ud-ə-s]]/

    a. $\Im_{LEX}$(/[$_V$paɾ-ə]/, /[$_N$[$_V$paɾ-ə][$_N$kajg-ud-ə-s]]/, M_{Cpd})= **/[$_V$páɾ-ə]/**, and

    b. $\Im_{GenEval}$ (**/[$_V$páɾ-ə]/**)=[$_V$páɾ-ə]; hence

    c. $\Im$(/[$_V$paɾ-ə]/, /[$_N$[$_V$paɾ-ə][$_N$kajg-ud-ə-s]]/, M_{Cpd})= **[$_V$páɾ-ə]**

We can now formulate an OO-constraint that requires identity of feature values between vowels in any input C and its base C'^B_O whenever they are related through the process M_{Cpd}:

---

18. "The asymmetry question and the inclusion question illustrate the need for a rigorous formalization of the principles governing the formation of the paradigms and 'mini-paradigms' used in invoking identity effects." (Bachrach & Nevins 2008b: 7). Bermúdez-Otero (2011: 12, 29) notes also that "The implementation of this [OO-correspondence] solution poses a number of nontrivial technical challenges, such as motivating the selection of the surface base […]; transderivational theories face other questions […]: what expressions can qualify as surface bases, and how are they selected? should OO-identity be symmetrical, base-prioritizing, or both?" See also Trommer (2006) and Downing et al. (2005).

(20) OO-IDENT(V-features, M_{Cpd}) Let C be a constituent in the input word W, C_O^B a constituent in the output word W_O^B, and M_{Cpd} a compounding or prefixation process such that $\Im(C, W, M_{Cpd})=C_O^B$. Then for any pair of vowels V, V' standing in correspondence, V in C and V' in C_O^B, assign a violation mark for every feature that does not have the same value in V and V'.

When the function $\Im$ returns Ø no output base is identified. This happens under different circumstances. First, if the constituent C is not an independent lexical element, as in the case of [taɣə], which exists only in [[taɣə][mənén]] (9), or [nɛw] which only appears in [[nɛw][nέw]] (10), no morphological process applies. Since $\Im_{LEX}$ cannot return any lexical constituent, $\Im$ will return Ø. In other cases there can be regular morphological relatedness, but not the one determined by the morphological process M of the function $\Im(C, W, M)$; consider the derivative *parada*, [N[Vpəɾ-á][Affð-ə]] 'stop', derived from the input /[N[Vpaɾ-a][Affd-ə]]/ and formed on the same base as the first constituent of *paracaigudes*, the verb *parar* 'to stop'. Here the first constituent /[Vpaɾ-a]/ contains the same root that appears in *paracaigudes*, but, unlike [N[Vpaɾ-ə][Nkəjɣ-úð-ə-s]] it does not show OO effects, since the vowel reduces to [ə]. The reason is that the morphological process M relating /[paɾ-a]/ to the verb *parar* is not M_{Cpd}, but a different, derivational process, hence $\Im$ cannot identify in this case any form for (/[paɾ-a]/, /[N[Vpaɾ-ə][Nkəjɣ-úð-ə-s]]/, M_{Cpd}). A more interesting case of $\Im$ returning Ø obtains in the case of the compound structures discussed in §3.4. Consider $\Im$ as the product $\Im_{LEX} \times \Im_{GenEval}$ (18). When the first constituent C is a bound lexical element, $\Im_{LEX}(C, W, M_{Cpd})$ will return a lexical element L, but $\Im_{GenEval}(L)=$Ø because bound elements have no independent output. In a case like [[karð j-u][βəskulár]] in (12), there exists a related lexical element /kardi/, that appears in *cardíac, cardiologia, miocardi, taquicàrdia*, etc., and $\Im_{LEX}$ will identify it. But it is always bound, and therefore $\Im_{GenEval}$ will not be able to identify an independent output *[kárði], or inflected *[kárð j-u], *[kárð j-ə], etc. The same happens with prefixes: [pɾɛ] has the same specific meaning in *prenatal, premolar, prevocàlic, preadolescent*, etc. but it never surfaces independently. In both cases there is no asymmetrical process that can derive one of the elements in the paradigm from the other. In the case of neoclassical compounds one might argue for some cases that the base is a constituent in a derivative, e.g. that *genito- in g[ɛ] nitounirari* derives from *genit-al*, or *antero- in ant[e]roposterior* from *anter-ior*, or *org[a]no- in organoplàstia* from *òrgan*. But then we would get the wrong results because the base in the derivatives is destressed and has a reduced [ə].

The identification function $\Im$ should be made precise in other important ways which I cannot examine in detail here. For instance, when the morphological process is not asymmetrical, as might be the case sometimes in word formation, and is typical of inflection, the asymmetry has to be built into $\Im$ itself, which will have to be able to select a single element according to some criterion (see Albright 2002, 2005, 2008, 2010, Steriade 2008, for evidence in favor of asymmetrical OO relations affecting inflection). Steriade & Yanovich (2015) have also shown that OO identity constraints must have in some cases an existential interpretation: a candidate of a derivative satisfies the OO constraint if there exists *some* member in the

paradigm of the base such that the properties in the candidate and in this member mentioned by the constraint match.

In order to make the morphological base of the correspondence relation more precise, I will assume an analysis of lexical representation that follows proposals in Jackendoff (2010b) and Jackendoff & Audring (to appear), an elaboration of Jackendoff (1975)—see also Bermúdez-Otero (2012). Complex words are fully specified in the lexicon and morphological schemas (similar to the redundancy rules in Jackendoff 1975) express lexical morphological regularities and relate specific lexical entries. I will not formulate the morphological processes/schemas here; for specific proposals, see the references cited. In any case, for a compound like *obre-llaunes* 'can opener', lit. 'opens cans' (4c) a morphological compounding process will relate the compound structure (21a-c) to the individual words (21d-f) and (21g-i).

(21) *obrellaunes* 'can opener'　　　　　*llaunes* 'cans', *obre* 'opens'

　　a.　$[\text{OBJECT}^{\alpha}; [\text{OPEN}_1(\text{INDEF}, \text{CAN}_2, \alpha)]]_3$　d.　$[\text{PLURAL}_4(\text{CAN}_5)]_6$

　　b.　$[_N [_{V3sg} X]_1[_{N, pl} Y]_2]_3$　　e.　$[_{N, pl} N_5 - af_4]_6$

　　c.　$[/\text{ɔbɾə}/_1 /\text{ʎawnə-s}/_2]_3$　　f.　$[/\text{ʎawnə}_5\text{-s}_4/]_6$

　　　　　　　　　　　　　　　　　　g.　$[\text{OPEN}_7; \text{PRES}_8]_9$

　　　　　　　　　　　　　　　　　　h.　$[_V [_{V3sg,pres.ind.8} V_7]_9$

　　　　　　　　　　　　　　　　　　i.　$[/\text{ɔbɾə}/_{7,8}]_9$

The entry *obrellaunes* consists of semantic (21a), morphosyntactic (21b) and phonological (21c) information, related by coindexation. Semantically it refers to an object OBJECT$^{\alpha}$ that is related to the predicate OPEN which takes three arguments, an agent, a theme and an instrument; the OBJECT$^{\alpha}$ refers to the third argument, the instrument α. Morphosyntactically, it has the structure (21b) with a verbal and a nominal constituent. Coindexation establishes associations among parts of these structures: the meaning OPEN$_1$ is linked to the morphosyntactic constituent $[_{V3sg} X]_1$ and to the phonological representation /ɔbɾə/$_1$; CAN$_2$ is linked to the constituent $[_{N, pl} Y]_2$ and to /ʎawn-ə-s/$_2$. The compounding process M$_{Cpd}$ relates (21a-c) to the pair (21d-f), (21g-i). In particular it relates /ɔbɾ-ə/ in (21c) to /ɔbɾ-ə/$_9$ in (21i), and since the output of /ɔbɾ-ə/$_9$ is [ɔβɾ-ə], ℑ can identify the base of /ɔbɾ-ə/$_1$ in (21c) for the process M$_{Cpd}$ as **[ɔ́βɾ-ə]** in (21i). Consider now the cases in which ℑ returns Ø, illustrated in (22-23) with *cobre-llit* and *Taga-manent*. Since relatedness is marked by coindexation, the structure of the lexical entries allows for different kinds of unrelatedness, expressed by lack of coindexation:

(22) a.　$[\text{BED COVER}]_3$　　　　　　d.　$[\text{BED}_4]_5$

　　b.　$[_V [X]_1[_N Y]_2]_3$　　　　e.　$[_N N_4]_5$

　　c.　$[/\text{kɔβɾə}/_1 /\text{ʎit}/_2]_3$　　　f.　$[/\text{ʎit}/_4]_5$

(23) a. [MOUNT TAGAMANENT]$_3$

    b. [$_N$ [X]$_1$ [Y]$_2$]$_3$

    c. [/tagə/$_1$ /manen/$_2$]$_3$

In (22a-c) the common noun meaning 'bed cover' (8) is connected through the subindex 3 to the global morphosyntactic and phonological structures, but the morphosyntactic and phonological subconstituents are unrelated to any semantic substructure. In the entry (22a-c) the second constituents in (22b,c) might be related through a morphological schema or process to the word (22d-f), but the first constituent cannot relate to any existing word. In (23a-c), representing the place name *Taga-manent* (8), the morphosyntactic and phonological subconstituents are not related to any existing, independent lexical entry. The lack of coindexation has as a consequence the impossibility for $\Im$ to identify a base; the morphological process M_{Cpd} cannot relate /kɔβɾə/$_1$, or /tagə/$_1$ to any base, hence $\Im$(/kɔβɾə/$_1$, [/kɔβɾə/$_1$ /ʎit/$_2$]$_3$, M_{Cpd})=Ø and $\Im$(/tagə/$_1$, [/tagə/$_1$ /manen/$_2$]$_3$, M_{Cpd})=Ø.[19]

7. The problem of the missing base

We now get to the crucial question. What happens when the $\Im$ function returns Ø? Obviously, if the identification of a constituent fails and an output base is missing, an OO constraint like (20) is trivially satisfied, since there is no C_O^B such that $\Im$(C, W, M_{Cpd})=C_O^B. An important prediction follows: there should be no misapplication (underapplication or overapplication) effects in this case, whereas we will find them when $\Im$ returns some C_O^B. One instance of "missing bases" of this kind has been analyzed in detail in Trommer (2006, 2013) and is discussed in Bermúdez-Otero (2011). In Albanian nonactive verb forms like [foɾ.mó.hem] 'form-nonact.1sg' should have final stress, but they retain the stress that appears in the active forms, as in [foɾ.mój] 'form-act.1sg'. There are however deponent verbs like the verb 'to regret', whose paradigm lacks active forms. Since the active base cannot be identified, the prediction is that in the case of a missing base the nonactive forms should show regular, not misapplied stress. This prediction is not borne out; stress falls on the stem also in these cases: [pen.dó.hem], *[pen.do.hém]. Bermúdez-Otero (2011) analyzes in detail two other cases, Quito Spanish /s/-voicing and lenition of linking and intrusive [ɹ] in nonrhotic English dialects.[20]

In the case of Catalan compound structures, we get similar predictions for an OO analysis. Thus for the compounds in (24a), repeated from (4), for which $\Im$ returns an output word or a word constituent, underapplication of vowel reduction is correctly predicted. For those in (24b), repeated from (11), (12), $\Im_{LEX}$ returns a

---

19. One possibility is that there is also an allomorph /kɔβɾ/, in addition to /kubr/, that would be selected in these particular instances. This would allow I_{Lexa} to select it, but since /kɔβɾ/ never has an independent output, $\Im_{GenEval}$ would return Ø.

20. Bermúdez-Otero (forthcoming) presents three more cases, schwa epenthesis in Itelmen intransitive verbs, failure of gliding of stem-final prevocalic /i/ in Bothoa Breton verbs and debuccalization of word-final prevocalic /s/ in Northern Chilean dialects of Spanish.

lexical element (bound root, prefix), but $\mathfrak{I}_{GenEval}$ cannot return an output, hence $\mathfrak{I}$ gives Ø and we incorrectly predict the reduced vowel. In (24c) $\mathfrak{I}_{LEX}$ cannot identify a base, hence $\mathfrak{I}$=Ø, and again we predict the wrong vowel. Therefore, in general, for missing output bases an OO analysis makes the wrong predictions.

(24) *Phonetic form* $\mathfrak{I}_{LEX}$ $\mathfrak{I}_{GenEval}$ *Predicted vowel*

   a. ɔβrə-ʎáwnəs $\mathfrak{I}_{LEX}$=/ɔbɾ/ $\mathfrak{I}_{GenEval}$(/ɔbɾ/)=[ɔ́βɾə]   [ɔ]

      trɛntə-síŋ $\mathfrak{I}_{LEX}$=/trɛntə/ $\mathfrak{I}_{GenEval}$(/trɛntə/)=[trɛ́ntə]   [ɛ]

   b. laβju-ðəntál $\mathfrak{I}_{LEX}$=/laβju/ $\mathfrak{I}_{GenEval}$(/laβju/)=Ø   *[ə]

      pɔst-upərətɔ́ri $\mathfrak{I}_{LEX}$=/pɔst/ $\mathfrak{I}_{GenEval}$(/pɔst/)=Ø   *[u]

   c. kɔβrə-ʎít $\mathfrak{I}_{LEX}$=Ø $\mathfrak{I}_{GenEval}$(Ø)=Ø   *[u]

      perə-təʎáðə $\mathfrak{I}_{LEX}$=Ø/ $\mathfrak{I}_{GenEval}$(Ø)=Ø   *[ə]

      bɔj-əskút $\mathfrak{I}_{LEX}$=Ø/ $\mathfrak{I}_{GenEval}$(Ø)=Ø   *[u]

There is one class of compounds, however, that yields to an alternative account under a parallel analysis. In reduplicatives (§3.2) like *nyeu-nyeu* (10) we could attribute the lack of reduction to the reduplicative process itself, which requires phonological identity between the two components. A constraint requiring partial identity, including identity in vowel features would penalize the candidate with reduction in the first component ([[ɲəw][ɲɛ́w]]) and favor [[ɲɛw][ɲɛ́w]].[21]

The predictions of an OO analysis can be illustrated with [$_A$[$_N$laβj-u][$_A$[$_N$ðənt] ál]] 'labiodental' whose second member, the derivative [$_A$[$_N$ðənt]ál], like our previous example [$_N$kəjg-úð-ə-s] in [$_N$[$_V$paɾ-ə][$_N$kəjg-úð-ə-s]], has undergone regular vowel reduction (cf. [$_N$dén] 'tooth'), but the vowel in [$_N$laβj-u] should not reduce. In [$_N$[$_V$paɾ-ə][$_N$kəjg-úð-ə-s]], [$_V$paɾ-ə] has a base, and the candidate [$_V$pəɾ-ə] with reduced [ə] in the first constituent (17d) above violates OO-IDENT(V-features) because there is an output **[$_V$páɾ-ə]**. But in (25) below the first constituent of the candidate (25d) [$_A$[$_N$ləβj-u][$_A$[$_N$ðənt]ál]] has a missing base, because it lacks an output base; therefore the OO constraint is trivially satisfied by all candidates and (25d) is the unwanted winner.[22]

21. I owe this observation to Eulàlia Bonet.
22. The first constituent of *labiodental* can of course be related to the first constituent of other words like *labiovelar*, but we do not know the output of *labiovelar* unless we find a base, and we run into a vicious circle. One could entertain the idea that, since /labio/ is a lexical element, it is evaluated, the optimal candidate being the null parse (McCarthy & Wolf 2010). The OO constraint would then pick as the base the next most harmonic candidate. There are obvious empirical and theoretical difficulties in such an analysis. In any case, it cannot be extended to isolates.

(25) *Missing base, parallel analysis*

$\Im([_N labj\text{-}o], M_{Cpd})=\emptyset$

$[_A[_N labj\text{-}o][_A[_N dent]al]]$ *Base*: *none*	OO-IDENT(V-features)	*$[_{Wd} ...\text{ˈV}...\text{ˈV}...]$	*ă,ɛ̆,ɔ̆,ĕ,ŏ
a. □ $[_A[_N laβbj\text{-}u][_A[_N ðənt]ál]]$			*!
b. $[_A[_N láβbj\text{-}u][_A[_N ðént]ál]]$		*!*	
c. $[_A[_N láβbj\text{-}u][_A[_N ðənt]ál]]$		*!	
d. ☞ $[_A[_N ləβbj\text{-}u][_A[_N ðənt]ál]]$			

8. The stratal analysis

A stratal analysis does not rely on the existence of an independent word [lábj-o], [pérə] or [kɔ́βrə], or one that contains these constituents; the existence of the constituent in the compound's lexical representation is sufficient to predict underapplication. Since both first members of a compound with a base (*para-caigudes*) and first members missing a base (*cobre-llit*) are constituents, underapplication effects apply equally to both. Given the constituent structure $[_A[_N labj\text{-}o][_A[_N dent]al]]$, $[_N labj\text{-}o]$ and $[_A[_N dent]al]$ are computed at the word stratum, and the corresponding outputs are $[lábj\text{-}u]_N$ and $[[dənt]_N ál]_A$ (26a). At this level, FINAL STRESS, *ă,ɛ̆,ɔ̆,ĕ,ŏ, and *$[_{Wd} ...\text{ˈV}...\text{ˈV}...]$ must dominate ID(V-features), as shown in (26a). This causes reduction of all destressed vowels. At the phrase level (26b), the input $[[lábj\text{-}u]_N[[dənt]_N ál]_A]$ is submitted to evaluation by a different constraint ordering, with ID(V-features) and *$[_{Wd} ...\text{ˈV}...\text{ˈV}...] \gg$ *ă,ɛ̆,ɔ̆,ĕ,ŏ; this prevents phrase-destressed vowels from reducing.

(26) *Missing base, stratal analysis*

a. Word stratum $[_{Wd} ...\text{ˈV}...\text{ˈV}...] \gg$ FINAL STRESS, *ă,ɛ̆,ɔ̆,ĕ,ŏ $\gg$ *ID(V-features)

$[_N labj\text{-}o]$	*$[_{Wd} ...\text{ˈV}...\text{ˈV}...]$	FINAL STRESS	*ă,ɛ̆,ɔ̆,ĕ,ŏ	ID(V-features)
☞ $[_N lábj\text{-}u]$				1
$[_N labj\text{-}o]$		1W	2W	L
$[_N lábj\text{-}o]$			1W	L

$[_A[_N dent]al]$	*$[_{Wd} ...\text{ˈV}...\text{ˈV}...]$	FINAL STRESS	*ă,ɛ̆,ɔ̆,ĕ,ŏ	ID(V-features)
☞ $[_A[_N dənt]ál]$				1
$[_A[_N dént]ál]$	1W			L
$[_A[_N dént]əl]$		1W		1
$[_A[_N dent]ál]$			1W	L

b. Phrase stratum *[Wd ...ˈV...ˈV...], Id(V-features) ≫ *ă,ĕ,ŏ,ĕ,ŏ

	*[$_{Wd}$ ...ˈV...ˈV...]	ID(V-features)	*ă,ĕ,ŏ,ĕ,ŏ
[$_A$[$_N$lábj-u][$_A$[$_N$dənt]ál]]			
☞ [$_A$[$_N$laβj-u][$_A$[$_N$ðənt]ál]]			1
[$_A$[$_N$láβj-u][$_A$[$_N$ðənt]ál]]	1W		L
[$_A$[$_N$ləβj-u][$_A$[$_N$ðənt]ál]]		1W	L

In neoclassical compounds like *labiodental* and in the case of prefixes a base can be identified which is not an output. In the case of proper and some common nouns in (9), reduplicated forms (10) and borrowings (11), usually no base can be identified. For these cases we have to assume a constituent structure that is not justfied by the existence of a base. The lexical configurations proposed at the end of §6 allow for such structures, i.e. for the existence of morphosyntactic constituent structure which is not matched by strictly compositional semantics. Thus for the place name *Tagamanent* in (23) coindexation relates the constituents in [/tagə/$_1$ /manen/$_2$]$_3$ to the morphosyntactic constituents in [$_N$ [X]$_1$ [Y]$_2$]$_3$, deriving /[[tagə][manen]]/, although there is a single global meaning 'mount tagamanent', the internal constituent not being coindexed with individual semantic representations.

9. Conclusions

I have shown that there is a large set of compound structures that share two properties: a first component with a missing base and underapplication of vowel reduction. Lack of vowel reduction in compounds with an identifiable base for the first component yields to both transderivational output-to-output and cyclic analyses. The compound structures with a first component missing a base cannot be handled through OO-constraints, but they can be derived cyclically under the reasonable assumption that lexical items can lack semantic compositionality while still retaining morphosyntactic constituent structure. At the same time, however, there are analogical effects that seem difficult to derive in a stratal analysis, without OO constraints. This is clearly the case whenever ℑ can identify an output as a base, but this base is not morphologically contained in the derived form, as in many cases of bases in inflectional paradigms (Steriade 2007, Steriade & Yanovich 2015, Albright 2002 and subsequent work). Thus in Romanian (Steriade 2008) the derivative *stíndʒ-íst* 'leftist' appears with the palatalized consonant of the plural *stíndʒ-i̯* 'left(hand)-pl.', not with the underlying velar that shows up in the singular *stíŋg-ʌ*. Obviously, the derivative does not contain morphologically the plural, i.e. *[[N,pl stíndʒ-i̯] íst]. In other words, some cases derive from analogic effects caused by the influence of words on words, but others derive from the way lexical representations are structured in terms of constituent structure and from cyclic effects. Whether some unification of these two mechanisms is possible must be left for future research.

References

Albright, Adam. 2002. *The identification of bases in morphological paradigms*. Ph.D. dissertation, UCLA.

Albright, Adam. 2005. The morphological basis of paradigm leveling. In Downing et al. (2005a), 17-43.

Albright, Adam. 2008. Inflectional paradigms have bases too: evidence from Yiddish. In Bachrach & Nevins (2008a), 271-312.

Albright, Adam. 2010. Base-driven leveling in Yiddish verb paradigms. *Natural Language & Linguistic Theory* 28.3: 475-537.

Bachrach, Asaf & Nevins, Andrew (eds.). 2008a. *Inflectional Identity*. Oxford: Oxford University Press.

Bachrach, Asaf & Nevins, Andrew. 2008b. Introduction: approaching inflectional identity. In Bachrach & Nevins (2008a), 1-28.

Benua, Laura. 1997. *Transderivational identity: phonological relations between words*. Ph.D. dissertation, University of Massachusetts, Amherst, MA. Available at ROA-259-0498, Rutgers Optimality Archive, <http://roa.rutgers.edu>.

Bermúdez-Otero, Ricardo. 2003. The acquisition of phonological opacity. In Spenader, Jennifer, Eriksson, Anders & Dahl, Östen (eds.). *Variation within Optimality Theory: Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory'*, 25-36. Stockholm: Department of Linguistics, Stockholm University. Expanded version (2003) available at ROA-593-0403, Rutgers Optimality Archive, <http://roa.rutgers.edu>.

Bermúdez-Otero, Ricardo. 2011. Cyclicity. In van Oostendorp, Marc, Ewen, Colin, Hume, Elizabeth & Rice, Keren (eds.). *The Blackwell Companion to Phonology*, vol. 4, 2019-2048. Malden, MA: Wiley-Blackwell.

Bermúdez-Otero, Ricardo. 2012. The architecture of grammar and the division of labour in exponence. In Trommer, Jochen (ed.). *The morphology and phonology of exponence*, 8-83. Oxford: Oxford University Press.

Bermúdez-Otero, Ricardo. Forthcoming. Stratal Phonology. In S.J. Hannahs & Anna R. K. Bosch (eds). *The Routledge handbook of phonological theory*. Abingdon: Routledge.

Cabré, Teresa. 1993. *Estructura gramatical i lexicó: el mot mínim català*. Ph.D. dissertation, Universitat Autònoma de Barcelona.

Cabré, Teresa. 2002. Altres sistemes de formació de mots. In Solà et al. (2002), vol. I, 891-932.

Coromines, Joan. 1971. *Lleures i converses d'un filòleg*. Barcelona: Club Editor.

Coromines, Joan. 1989-1997. *Onomasticon Cataloniae*. Barcelona: Curial.

Downing, Laura J., Hall, T. Alan & Raffelsiefen, Renate (eds.). 2005a. *Paradigms in Phonological Theory*. Oxford: Oxford University Press.

Downing, Laura J., Hall, T. Alan & Raffelsiefen, Renate. 2005b. Introduction: The Role of Paradigms in Phonological Theory. In Downing et al. (2005a), 1-16.

Hualde, José-Ignacio. 2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5.2/6.1: 59-89.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51: 639-71. Reprinted in Ray Jackendoff (2010a), 40-84.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2010a. *Meaning and the Lexicon*. Oxford: Oxford University Press.

Jackendoff, Ray. 2010b. The ecology of English noun-noun compounds. In Jackendoff (2010a), 413-45. (An expanded version of Compounding in the Parallel Architecture and Conceptual Semantics, in Rochelle Lieber & Pavol Štekauer (eds.). *The Oxford Handbook of Compounding*, 105-28. Oxford: Oxford University Press, 2009.

Jackendoff, Ray, & Audring, Jenny. To appear. Morphology in the Parallel Architecture. In Audring, Jenny & Masini, Francesca (eds.). *The Oxford Handbook of Morphological Theory*. Oxford: Oxford University Press.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In Durand Jacques & Laks, Bernard (eds.). *Currents trends in phonology: models and methods*, vol. 1, 363-393. Salford: European Studies Research Institute, University of Salford.

Kiparsky, Paul. 2000. Opacity and cyclicity. In Ritter, Nancy A. (ed.). *A review of Optimality Theory. The Linguistic Review* (Special issue), 17.2-4: 351-67.

Kiparsky, Paul. In press. *Paradigms and Opacity*. Vol. I. Stanford: CSLI.

Köhnlein, Björn. 2015. The morphological structure of complex place names: the case of Dutch. *Journal of Comparative German Linguistics* 18: 183-212.

Mascaró, Joan. 1976. *Catalan Phonology and the Phonological Cycle*. Ph.D. dissertation, MIT: Published by Indiana University Linguistics Club, Bloomington, 1978. Expanded Catalan translation: *La fonologia catalana i el cicle fonològic*. Bellaterra: Sèrie Lingüística, UAB, 1983.

Mascaró, Joan. 2002. El sistema vocàlic. Reducció vocàlica. In Solà et al. (2002), vol. I, 89-123.

Mascaró, Joan. 2015: Regularitat i excepcions en fonologia: les reduccions vocàliques. In Lloret, Maria-Rosa, Pons-Moll, Clàudia & Bosch-Roura, Eva (eds.). *Clàssics d'ahir i d'avui en la gramàtica del català*, 59-69. Barcelona: Universitat de Barcelona.

McCarthy, John J. & Wolf, Matthew. 2010. Less than zero: Correspondence and the null output. In Blaho, Sylvia & Rice, Curt (eds.). *Modeling Ungrammaticality in Optimality Theory*, 17-66. London: Equinox.

Moll, Francesc de B. 1931. Estudi fonètic i lexical del dialecte de Ciutadella. In *Miscelànea Filològica dedicada a don Antonio Ma. Alcover*. Palma, 397-460. Reproduced in *Randa* 8, 1979: 5-48.

Nadeu, Marianna. 2016. Phonetic and phonological vowel reduction in Central Catalan. *Journal of the International Phonetic Association* 46.1: 33-60.

Oliva, Salvador. 1992. *La mètrica i el ritme de la prosa*. Barcelona: Quaderns Crema.

Paul, Hermann. 1968 [1880]. *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer.

Prieto, Pilar. 2003. Correlats acústics de l'accent secundari en català. *Estudios de Fonética Experimental* 12: 107-142.

Recasens, Daniel. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.

Riera-Eures, Manel. 2002. *Diccionari d'onomatopeies i mots de creació expressiva: les paraules transparents de la llengua catalana*. Barcelona: Edicions 62.

Solà, Joan, Lloret, Maria-Rosa, Mascaró, Joan & Pérez Saldanya, Manel (eds.). 2002. *Gramàtica del català contemporani*. Barcelona: Empúries.

Steriade, Donca. 2008. A pseudo-cyclic effect in the Romanian declension. In Bachrach & Nevins (2008), 313-361.

Steriade, Donca & Yanovich, Igor. 2015. Accentual allomorphs in East Slavic: an argument for inflection dependence. In Bonet, Eulàlia, Lloret, Maria-Rosa & Mascaró, Joan (eds.). *Understanding Allomorphy*, 254-314. Shefield and Bristol: Equinox.

Trommer, Jochen. 2006. Stress uniformity in Albanian: morphological arguments for cyclicity. Paper given at the workshop *Approaches to Phonological Opacity, 29th GLOW Colloquium,* Barcelona.

Trommer, Jochen. 2013. Stress uniformity in Albanian: morphological arguments for cyclicity. *Linguistic Inquiry* 44.1: 109-143.

Wheeler, Max W. 2005. *The phonology of Catalan*. Oxford: Oxford University Press.