

The free-ride procedure to morphophonemic learning is correct. Some evidence from Catalan vowel epenthesis¹

Clàudia Pons-Moll (UB) & Maria-Rosa Lloret (UB)

claudia.pons@ub.edu, mrosa.lloret@ub.edu

*Tutte le strade portano a Roma
(All roads lead to Rome)*

Simplified syllabus

1. Uncertainty in URs
2. UR Uncertainty within Optimality Theory
3. Goals and arguments
4. The challenge and the potential solution
5. Getting close to the evidence. Underapplication of vowel reduction in Majorcan Catalan
6. Into the evidence. The interaction between word-initial epenthesis and underapplication of vowel reduction
7. The Free-Ride Learning Algorithm applied to Majorcan Catalan
8. Remaining formal and empirical issues
9. Selected references

1. Uncertainty in URs and vowel epenthesis in Catalan

The structure present in URs is not challenging when (at the same time) productive morphophonemic alternations shed light on it, and when there is a clear and systematic phonological condition in the language that justifies the differences between the UR that has to be established and the corresponding surface representation. The challenge, or the uncertainty, appears in those cases in which such morphophonemic alternations do not exist, when they are not fully productive or transparent, and also when alternative interpretations are possible.

This is the case of the URs corresponding to words beginning with <esC>- in Catalan (*escriure* ‘to write’, *estona* ‘while’, *espot* ‘spot’), for which there is a small amount of morphological evidence for treating the initial vowel as underlying or as epenthetic (Wheeler 2005: sec. 8.1-8.2).

2. UR uncertainty within Optimality Theory

Within Optimality Theory (OT), there are **three main hypotheses regarding the nature of URs** and their process of acquisition, construction, storage, and access when no dynamic morphophonemic

alternations are available. Two of these hypotheses (which are not always mutually exclusive but to some extent complementary) appear in the foundational work of OT (Prince and Smolensky 1993 / 2004):

2.1. Richness of the Base (RotB). According to the RotB hypothesis (“all inputs are possible in all languages”, Prince and Smolensky 1993 / 2004: 225), the analyst must project all possible URs (which must be understood as theoretical constructs with no psychological reality) for every surface form. The grammar, that is, the constraint hierarchy (EVAL), is ultimately responsible for selecting the actual surface form in a given language, no matter which UR is taken. The base is, by hypothesis, rich in that it is not subject to language-particular restrictions. In accordance with this thesis, there can be no morpheme structure constraints, stipulated underspecification or similar devices that preemptively constrain possible inputs of the grammar (McCarthy 2002).

2.2. Lexicon Optimization (LO). However, in the process of storage and access to URs, a principle called LO is assumed to be at play (Prince and Smolensky 1993 / 2004: 205, 225), according to which, when there is no morphophonemic evidence bearing on the choice of URs, phonological representations are stored identically to their surface form. This principle, which echoes Kiparsky’s (1968) anti-abstractionist “Alternation Condition”, leads to a direct economization of input-output mappings, given that the map from underlying to surface representations is accomplished more faithfully. The base that learners access is by hypothesis the one that is identical to the corresponding surface form.

Additionally, McCarthy (2005a) develops an additional mechanism:

2.3. Free-Ride approach to Morphophonemic Learning (FRML). McCarthy (2005a: 19) argues that “[w]hen alternation data tell the learner that some surface [B]s are derived from underlying /A/s, the learner will under certain conditions generalize by deriving all [B]s, even nonalternating ones, from /A/s”. Given this observation, “[a]n adequate learning theory must [...] incorporate a procedure that allows nonalternating [B]s to take a ‘free ride’ on the /A/ → [B] unfaithful map”. The FRML approach is **enlightening** not only for the cases where **morphophonemic alternations are unhelpful or nonexistent** but also for the process of language acquisition itself: it is a matter of fact that learners have **limited experience** (especially at early stages of acquisition) and that hence are often unaware of the morphophonemic alternations that would allow them to construct URs.

3. Goals and arguments

The purpose of this talk is twofold. On the empirical side, we aim at contributing to the body of knowledge about the UR of words beginning with <esC>- in Catalan. On the theoretical side, and relying on the empirical results obtained, we provide arguments that support the FRML.

¹ For fruitful discussion on this subject we are grateful to R. Bermúdez-Otero, J. Jiménez, J. J. McCarthy, A. Nevins, M. van Oostendorp, D. Steriade, F. Torres-Tamarit, and M. W. Wheeler. A later version of this work has been published in *Linguistic Inquiry*, 47.1. (Winter 2016). Previous versions of the same issue have been presented at the Meertens Institute / Royal Academy of Arts and Sciences (Leiden-Amsterdam Phonology Circle, Amsterdam, 2012), at the 10th Old World Conference in Phonology (Istanbul, 2013), and at the University College London (London, 2014). This research has been supported by the projects FFI2010-22181-C03-02 (UB) and RecerCaixa 2011 (UB).

On the basis of the casuistry arising from the interaction between word-initial vowel epenthesis in Catalan and particular cases of underapplication of vowel reduction in Majorcan Catalan, we provide empirical arguments, *lying inside the grammar*, for the underlying absence of the initial vowel in words beginning with <esC>-. On the basis of these results, we show how Majorcan Catalan learners take a “free ride” in the process of constructing the UR of nonalternating forms by generalizing the pattern—and the subsequent input-output mapping—observed in cases with transparent morphophonemic alternations and a parallel syllabification problem.

4. The challenge and the potential solution

4.1 The challenge. Epenthesis and lack of (fully productive) morphophonemic alternations

In Catalan, vowel epenthesis has often been invoked to explain the presence of a vowel (typically [ə] in Eastern dialects and [e] in Western dialects) in those situations where its absence would entail the occurrence of a structure defying some kind of syllabic constraint (see Mascaró 1976, Wheeler 1975, DeCesaris 1987, Lloret 2002, among others). In few of these cases, though, the postulation of epenthesis is fully legitimate by truly productive morphophonemic alternations, so that other interpretations of the vowel are available (Wheeler 2005).

Word-initial epenthesis has typically been adduced in words like *escriure* ‘to write’, *estona* ‘while’ and *eslògan* ‘slogan’ as a strategy to avoid word-initial sC- clusters.

(1) Cases *escriure*, *esperar*, and the like

(1a) Realizations

<i>escriure</i>	[ə]scriure	‘to write’
<i>esperar</i>	[ə]sperar	‘to wait’

(1b) Prefixed forms **without** the initial vowel

<i>inscriure</i>	in[∅]scriure	‘to register’
<i>descriure</i>	de[∅]scriure	‘to describe’
<i>subscriure</i>	subs[∅]scriure	‘to subscribe’

<i>prosperar</i>	pro[∅]sperar	‘to prosper’
<i>exasperar</i>	exa[∅]sperar	‘to exasperate’

- Morphophonemic alternations [ə] ~ [∅]

BUT

- These alternations are not fully productive.
- The morphological compositionality of the forms in *ib* is not transparent, while it is in the forms of *ic* (just below).

(1c) Other prefixed forms **with** the initial vowel

<i>reescriure</i>	re[ə]scriure	‘to rewrite’
<i>sobreescriure</i>	sobre[ə]scriure	‘to overwrite’
<i>desesperar</i>	des[ə]sperar	‘to despair’
<i>inesperat</i>	in[ə]sperat	‘unexpected’

- No morphophonemic alternation [ə] ~ [∅]

AND

- This lack of alternation is fully productive.
- The morphological compositionality of the forms in *ic* is transparent, unlike the one seen in (1b)

(2) Cases *estona*, *escala*, and the like

(2a) Realizations

<i>estona</i>	[ə]stona	‘while’
<i>escala</i>	[ə]scala	‘scale’
<i>estructura</i>	[ə]structura	‘structure’
<i>especial</i>	[ə]special	‘special’
<i>especific</i>	[ə]specific	‘specific’
<i>estereotip</i>	[ə]stereotip	‘stereotype’
<i>esport</i>	[ə]sport	‘sport’

(2b) Inexistence of prefixed forms **without** the initial vowel

.....
.....

(2c) Prefixed forms **always with** the initial vowel

<i>superestructura</i>	super[ə]structura	‘superstructure’
<i>infraestructura</i>	infra[ə]structura	‘infrastructure’
<i>superespecial</i>	super[ə]special	‘super especial’

<i>inespecífic</i>	<i>in[ə]specific</i>	‘unspecific’
<i>subestereotip</i>	<i>sub[ə]stereotip</i>	‘substereotype’
<i>Interesport</i>	<i>Inter[ə]sport</i>	‘commercial name’
<i>antiesportiu</i>	<i>anti[ə]sportiu</i>	‘unsporting’
<i>poliesportiu</i>	<i>poli[ə]sportiu</i>	‘sports center’

- Absolute lack of the morphophonemic alternation [ə] ~ [∅].
- This lack of alternations is fully productive.
- The morphological compositionality of the forms in *2c* is transparent.

(3) Cases *espot*, *eslògan*, and the like (loanwords)

(3a) Realizations

<i>espot</i>	<i>[ə]spot</i>	‘spot’
<i>eslògan</i>	<i>[ə]slògan</i>	‘slogan’
<i>esteps</i>	<i>[ə]steps</i>	‘steps’

Other cases: [ə]*Skype*, [ə]*spinning*, [ə]*Spotify*, [ə]*slang* (not included in the dictionary and not adapted to the Catalan orthography).

(3b) Inexistence of prefixed forms **without** the initial vowel

.....
.....

(3c) Prefixed forms **with** the initial vowel

<i>superespot</i>	<i>super[ə]spot</i>	‘superspot’
<i>antislògan</i>	<i>anti[ə]slogan</i>	‘antislogan’

- Absolute lack of the morphophonemic alternation [ə] ~ [∅].
- This lack of alternations is fully productive.
- The morphological compositionality of the forms in *3c* is transparent.

Interim remarks:

- There is unsubstantial empirical evidence to consider the first vowel in words starting in <esC>- (like *escriure*, *esperar*, *estructura*, *esport*, *estona*, *eslògan*) as epenthetic.
- The underlying representations proposed are frequently not fully legitimated empirically.
- The elements the analyst counts with in order to determine underlying representations are very often blurred.

- One has to inevitably resort to theoretical speculations or, at best, extrapolation of the data.

- *Rebus sic stantibus*, the learner has inexistent (or, in any case, very poor) empirical evidence to project the mapping /∅/ → [ə].

4.2 The potential solution. Epenthesis and fully productive morphophonemic alternations

Unlike word-initial epenthesis, word-internal epenthesis in certain verbal forms is claimed to be uncontroversial (4a), because there are fully transparent morphophonemic alternations, both within the paradigm of the same verbs (4b) and with respect to parallel inflected forms of verbs that belong to the same class but whose URs do not give rise to potential syllabic problems (4c), and because there exist regular phonological conditions that drive the insertion of the epenthetic vowel (Lloret 2002, Pons-Moll 2005, Wheeler 2005).

- | | | | |
|--------|-----------------------------|-----------------------------------|------------------------------|
| (4) a. | <i>temgré / temgria</i> | <i>tem[ə]-ré / tem[ə]-ria</i> | ‘(I) will / would be afraid’ |
| | <i>creixgré / creixgria</i> | <i>creix[ə]-ré / creix[ə]-ria</i> | ‘(I) will / would grow up’ |
| b. | <i>temo / temia</i> | <i>tem[∅]-o / tem[∅]-ia</i> | ‘(I) am / was afraid’ |
| | <i>creixo / creixia</i> | <i>creix[∅]-o / creix[∅]-ia</i> | ‘(I) grow / grew up’ |
| c. | <i>rompré / rompria</i> | <i>romp[∅]-ré / romp[∅]-ria</i> | ‘(I) will / would break’ |
| | <i>beuré / beuria</i> | <i>beu[∅]-ré / beu[∅]-ria</i> | ‘(I) will / would drink’ |

Interim remarks:

- In the word-internal cases presented in (4a) the alternation /∅/ ~ [ə] is dynamic, and hence the /∅/ → [ə] map is fully legitimated.
- The key question that we address here is whether the learner makes use of these word-internal alternations (and the subsequent fully legitimated unfaithful /∅/ → [ə] map) to project the same mapping in the word-initial cases, which do not display transparent and productive alternations.

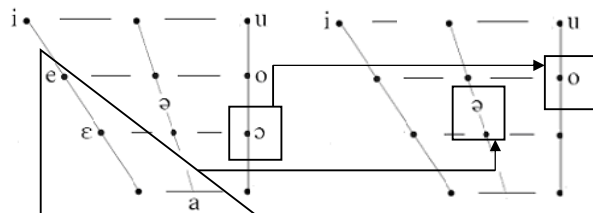
- As we will next show, the patterns derived from the interaction between vowel epenthesis and underapplication of vowel reduction in Majorcan Catalan demonstrate that this is the case.

5. Getting Close to the Evidence. Underapplication of vowel reduction in Majorcan Catalan

5.1. Normal application of vowel reduction in Majorcan Catalan

(5a) Process of vowel reduction in Majorcan Catalan

a. Stressed vowel system b. Unstressed vowel system



(5b) Morphophonemic alternations driven by vowel reduction

a. Stressed position

c[á]sa 'coffee'
caf[é] 'coffee'
carr[é]r 'street'
cont[é]st '(I) answer'
tol[é]r '(I) tolerate'

b. Unstressed position

c[ə]s[ə]ta 'house dim.'
caf[ə]t[ə]t 'coffee dim.'
carr[ə]r[ə] 'street dim.'
cont[ə]st[ə]m '(we) answer'
tol[ə]r[ə]m '(we) tolerate'

5.2. Underapplication of vowel reduction in Majorcan Catalan

The process of vowel reduction of /e/ and /ε/ underapplies under the following circumstances (see, among others, Veny 1962, Bibiloni 1998, Mascaró 2002, 2005; Pons-Moll 2012a, b, 2013):

a) IN PRODUCTIVE DERIVED FORMS with an unstressed vowel which alternates with a stressed mid front vowel ([é] / [é]) in the stem of the primitive:

(6) Derived forms

festassa [e] 'party *augm.*' ~ *festa* [é] 'party'
celet [e] 'sky *dim.*' ~ *cel* [é] 'sky'

b) IN VERBAL FORMS (I conjugation) with an unstressed vowel which alternates with a stressed close mid front vowel ([é]) in another verbal form of the same inflectional paradigm:

(7) Verbal forms

pegar [e] 'to hit' ~ *pega* [é] '(he/she) hits'
quedam [e] 'we stay' ~ *queda* [é] '(he/she) stays'

c) IN LEARNED WORDS AND LOANWORDS with an unstressed *e* generally preceded by a labial consonant:

(8) Loanwords and learned words

vermut [e] 'vermouth' *benigne* [e] 'benign'
vedet [e] 'star' *fetixisme* [e] 'fetishism'
el Pentàgon [e] *penicil·lina* [e] 'penicillin'

As shown in previous studies (Pons-Moll 2011a, 2012, 2013), the lack of vowel reduction affects, indeed, the unstressed *e* in the mentioned situations, **but only *d*** when it is located **in the initial syllable of the stem**:

- This is why we find normal application of vowel reduction in forms like the ones in (9) [see also (5b)], which meet the conditions in *a*, *b* and *c*.

(9) Derived forms, verbal forms and loanwords and learned words

paperet 'paper *dim.*', *castellet* 'castle *dim.*' [ə] Cf. *a* and (6)
contestam 'we answer', *tolerau* 'we tolerate' [ə] Cf. *b* and (7)
amenitzar 'to liven up', *preferent* 'preferable' [ə] Cf. *c* and (8)

[Note again that in any of these cases the vowel is not located in the initial syllable of the stem.]

- This is why we find underapplication of vowel reduction of the second vowel and not of the first (belonging to the prefix) in prefixed forms like the ones in (10).

(10) Prefixed forms

renetet [ə] [e] 'great grandson *dim.*'
recremar [ə] [e] 'to burn again'
efeminat [ə] [e] 'effeminate'

(11) Cf. Bases for the prefixed forms in (10)

netet [e] 'grandson *dim.*'
cremar [e] 'to burn'
feminisme [e] 'feminism'

Analysis of underapplication of vowel reduction in Pons-Moll (2011a, 2012, 2013)

- Productive derivation (cases in 6, with vowel alternations, repeated here)

festassa [e] ‘party *augm.*’ ~ *fiesta* [é] ‘party’

celet [e] ‘sky *dim.*’ ~ *cel* [é] ‘sky’

(12) → Output-output faithfulness constraints (TCT, Benua 1995, 1997) relativized according to the position of the vowel within the stem (Pons-Moll 2011a, 2013) [there is also a formalization of the “productivity effect” that we omit here]

- Verbal inflection (cases in 7, with vowel alternations, repeated here)

pegar [e] ‘to hit’ ~ *pega* [é] ‘(he/she) hits’

quedam [e] ‘we stay’ ~ *queda* [é] ‘(he/she) stays’

(13) → Output-output faithfulness constraints (OP, McCarthy 2005b) relativized according to the position of the vowel within the stem (Pons-Moll 2011a, 2013)

- Learned and loanwords (cases in c, without vowel alternations)

vermut [e] ‘vermouth’

benigne [e] ‘benign’

vedet [e] ‘star’

fetixisme [e] ‘fetishism’

el Pentàgon [e]

penicil·lina [e] ‘penicillin’

(14) → Contextual markedness constraint relativized according to the position of the vowel within the stem, only active in the productive phonology [loanwords and learned words; cf. *menorquí* [ə] ‘Minorcan’ (Pons-Moll 2011a, 2012, 2013)]

[For the details about the formalization, see Pons-Moll (2011a [derivation and inflection], 2012 [loanwords], 2013 [both])]

Crucial and of interest...

- We find underapplication of vowel reduction of the second vowel in words like:

(15)

Estevet [ə]st[e]vet ‘Stephen *dim.*’

esquemet [ə]squ[e]met ‘scheme *dim.*’

esperau [ə]sp[e]rau ‘(you) wait’

estenem [ə]st[e]nem ‘(we) tend’

<i>especial</i>	[ə]sp[e]cial	‘especial’
<i>especialista</i>	[ə]sp[e]cialista	‘specialist’
<i>especialitat</i>	[ə]sp[e]cialitat	‘speciality’

It is important to note that the first vowel never alternates with a stressed vowel nor with [Ø], and that the second vowel can show vowel alternations or not; see § 6.

6. Into the evidence. The interaction between word-initial epenthesis and underapplication of vowel reduction

6.1. The facts

(16) Productive derived forms and verbal forms with initial schwa + sC

[ə]st[é]ve ‘Stephen’ [ə]st[e]vet ‘Stephen *dim.*’

[ə]squ[é]ma ‘Scheme’ [ə]squ[e]met ‘scheme *dim.*’

[ə]sp[é]ra ‘(s/he) waits’ [ə]sp[e]r[á]u ‘(you) wait’

[ə]st[é]n ‘(s/he) tends’ [ə]st[e]nem ‘(we) tend’

(17) Loanwords and learned words with initial schwa + sC

[ə]sp[e]cial ‘especial’ [ə]sp[e]cífic ‘specific’

[ə]sp[e]cialista ‘specialist’ [ə]sp[e]cificador ‘specifier’

[ə]sp[e]cialitat ‘speciality’ [ə]sp[e]cificitat ‘specificity’

[ə]sp[e]cialitzar ‘specialize’ [ə]sp[e]rmatozou ‘spermatozoon’

[ə]sp[e]cialment ‘especially’ [ə]st[e]roide ‘asteroid’

[ə]st[e]reotip ‘stereotype’ [ə]st[e]roides ‘steroids’

6.2. Empirical Consequences

- Since the generalization for underapplication of VR is that it only occurs when the vowel is located in the initial syllable of the stem, we now have external and independent evidence that lies inside the grammar to assert that the initial vowel of words beginning with [ə]sC- is not part of the stem and thus has to be considered epenthetic.
- If this vowel were part of the stem, the second vowel would be located in a noninitial syllable of the stem and thus should not be affected by underapplication of VR (but should reduce as it does in the noninitial-syllable cases mentioned above).
- The initial vowel of [ə]sC- words behaves indeed as “invisible” to the output-output positional faithfulness constraints alleged to account for underapplication of VR in derivation (cf. *f[e]stassa* ~ *f[é]sta*, and also *[ə]st[e]vet* ~ *[ə]st[é]ve*) and verbal inflection (cf. *p[e]gau* ~ *p[é]ga*, and

also [ə]sp[e]rau ~ [ə]sp[é]ra), and it is unaffected by the contextual markedness constraint against a schwa in the initial syllable of the stem alleged to account for underapplication of VR in loans and learned words (cf. *b[e]nigne*, *f[e]titxisme*, *v[e]det*, and also [ə]sp[e]cial).

- This is clear positive evidence for the epenthetic nature of the initial vowel, realized as [ə] in the dialect under study. If this was not the case, the second vowel would not be affected by these constraints, because it would occupy a position other than the initial one within the stem.

6.3. Theoretical Consequences

The data analyzed are relevant to test the three approaches developed within OT to account for the nature of URs and their process of acquisition, learning, and construction (see section 2).

- Under the **RotB** hypothesis and guided simply by morphophonemic alternations, the potential URs for words like *Estevet*, *esperau*, *especial* (without the alternation [ə] ~ [Ø]) should be /əsC/- and /ØsC/-.
- Obviously, though, if we assumed an UR with the vowel (that is, /əsC/-), we would obtain nonexistent forms with reduction of the second vowel to schwa.
- This is so because the output-output faithfulness constraints relativized according to the position of the vowel within the stem, adduced to explain cases like *f[e]stassa* or *p[e]gam*, or the contextual markedness constraint against a schwa in the initial syllable of the stem, adduced to explain cases such as *f[e]titxisme*, would not have any effect in the second vowel of words beginning with [əsC]-, because it would not be considered to be located in the initial syllable of the stem.
- Consequently, VR would erroneously be expected to apply in the second syllable of these words: **Est[ə]vet*, **esp[ə]rau*, **esp[ə]cial*.
- This does not counter the RotB hypothesis in a general way, but instead limits its application with respect to the data we are dealing with here.
- Under the **LO** hypothesis, the UR for words like *Estevet*, *esperau* or *especial* (without the [ə] ~ [Ø] alternation) should be /əsC/-, because it is the closest representation to the surface form.
- The same reasons adduced for the RotB hypothesis invalidate this approach. It is, in fact, impossible to derive *Est[e]vet*, *esp[e]rau*, *esp[e]cial* (without VR in the second surface syllable), if the initial vowel is already present in the UR, that is, if it is part of the stem.

After all, it seems that learners have taken a “free ride” with the unfaithful /Ø/ → [ə] map.

- In the absence of (fully productive) morphophonemic alternations that shed light on the URs of words with [əsC]- initials (and also given the fact that learners have limited experience, and hence are often unaware of the morphophonemic alternations that would allow them to discover URs), speakers generalize an unfaithful /Ø/ → [ə] map, which they deduce from the cases in which there are dynamic alternations (see section 4.2: *tem[Ø]o*, *perd[Ø]ré* ~ *tem[ə]ré*) and extend to nonalternating cases ([ə]stevet, [ə]sperau, [ə]special).
- Our proposal matches McCarthy’s (2005b) prediction according to which learners take the free-ride strategy in nonalternating forms under certain conditions *only*: when, by generalizing the unfaithful mapping, a consistent and a more restrictive grammar than the one obtained by an identity mapping (*à la LO*) is achieved. This is exactly the case we are dealing with here.
- First, the grammar obtained by generalizing the unfaithful /Ø/ → [ə] mapping to all cases—even to no alternating cases—is “**consistent**” with the primary data, because it homogeneously explains the absolute lack of words beginning with [sC]- in the language and it further limits the number of input-output mappings.
- Second, the grammar obtained is “**more restrictive**” because it has a higher “r-measure” (“The r-measure for a constraint hierarchy is determined by adding, for each faithfulness constraint in the hierarchy, the number of markedness constraints that dominate that faithfulness constraint”, Prince and Tesar 2004:252), and, as McCarthy (2005b:32) claims, a grammar with a higher r-measure is more restrictive because it grants “more power to markedness constraints”.
- In the alleged grammar for Majorcan Catalan, not only the markedness syllabic constraints SYLLABLECONTACT (that penalize syllabic contacts like **tem.ré*; see Wheeler 2005:83; Pons 2005, 2011b) and MINIMUM SONORITY DISTANCE (that avoid onset sequences with low sonority distances such as **te.mré*; see Wheeler 2005:255; Pons 2005) must be ranked above the faithfulness constraint DEP-IO (to justify vowel insertion), but also *sC- (penalizing words starting with a sibilant followed by a consonant), with the result that the grammar has a higher r-measure.

Summing up, once the identity /ə/ → [ə] map of the earliest stages is abandoned because the morphophonemic [Ø] ~ [ə] alternations are discovered and incorporated into the grammar, the learner is fully committed to the unfaithful /Ø/ → [ə] map, which applies to all cases with a parallel syllabic problem (following, hence, the orthodox, not contextually determined, “across-

the-board” free ride) driven by the constraint ranking $[[\text{SYLLABLECONTACT}, \text{MINIMUMSONORITYDISTANCE}, *sC- \gg \text{DEP-IO}]]$, with all markedness constraints relative to syllabic wellformedness outranking faithfulness.

7. The Free-Ride Learning Algorithm applied to Majorcan Catalan

The free-ride procedure for our case would roughly go as follows (we follow McCarthy 2005a very closely, here).

● **Phonotactic learning.** At the earliest stages of learning, there is presumably little or no awareness of morphological structure and no awareness of alternations, so only the phonotactics are being learned (on this assumption, see Hayes 2004; Pater and Tessier 2003; Prince and Tesar 2004). The phonotactic learner seeks a grammar that performs **identity maps** from the inferred lexicon to the observed surface forms.

7.1. Majorcan Catalan at onset of phonotactic learning initial stage
[At this stage all constraints are assumed to be unranked.]

(18) Initial stage: unranked constraints

		*UNSTR. a	*UNSTR. ε	*UNSTR. ə	*sC	DEP-IO	NO-CODA	IDENT(F)	MAX-IO
/əsc/ (escala)	[əsc]		1				1		
	[əC]		1				L		1 W
/asC/ (astre)	[ásC]						1		
	[áC]						L		1 W
/əsc/ (astral)	[əsc]		1				1		
	[asC]	1W	L				1	1 W	
	[əC]		1				L		1 W
/esC/ (estri)	[ésC]						1		
	[éC]						L		1 W
/əsc/ (estriet)	[əsc]		1				1		
	[esC]		1W	L			1	1W	
	[əC]		1				L		1 W

Comments on this tableau:

- In the case of Catalan, the ambient language contains words starting with $\#[\text{ə}]sC$ - (*escala*), so the only phonotactics the learner can learn are those related to consonants in coda position, or the ones related to the vowels (dis)allowed in unstressed position. The M and the F constraints we consider, then, are the ones correlated to these facts.
- We, though, include the constraint *sC, because it is going to be relevant in the following stages and, of course, because it is thought to be universal. For the same reasons, we include Dep-IO.
- Lexicon optimization is at play at this stage.

7.2. Finding the (phonotactically based) grammar via Biased Multirecursive constraint Demotion (BMCD) [Prince and Tesar 2004]

7.2.1. A constraint is rankable if it favors no losers among the candidates that have not yet been accounted for.

7.2.2. Once some constraint has been ranked, the tableau with this new ranking is then submitted to the same procedure, **recursively**.

7.2.3. The **bias** in BMCD favors grammars that are **more restrictive**. The bias says which constraints to rank first (=higher) when there are two or more constraints that favor no losers.

7.2.4. By preference, markedness constraints are ranked first / higher. If there are no rankable markedness constraints, the ranking preference performs goes to a faithfulness constraint.

7.2.5. When several faithfulness constraints are available for ranking, the algorithm chooses the one that, by accounting for certain candidates, allows a markedness constraint to be ranked on the next recursive pass.

(19) BMCD: Learning *Unstr. a, *Unstr. ε >> *Unstr. ə

		*UNSTR. a	*UNSTR. ε	*UNSTR. ə	*sC	DEP-IO	NO-CODA	IDENT(F)	MAX-IO
/əsc/ (escala)	[əsc]			1			1		
	[əC]			1			L		1 W
/asC/ (astre)	[ásC]						1		
	[áC]						L		1 W
/əsc/ (astral)	[əsc]			1			1		
	[asC]	1W		L			1	1 W	
	[əC]			1			L		1 W
/esC/ (estri)	[ésC]						1		
	[éC]						L		1 W
/əsc/ (estriet)	[əsc]			1			1		
	[esC]		1W	L			1	1W	
	[əC]			1			L		1 W

Comments on this support tableau:

- In this support tableau the learner learns that *UNSTR. a, *UNSTR. ε >> *UNSTR. ə.
- It is a matter of convention which ranking the learner learns first, this one or the one shown in (20) [MAX-IO >> NO-CODA]
- In fact, there is a previous stage (omitted here) with NO-CODA >> MAX-IO.

(20) BMCD: Learning MAX-IO >> NO-CODA

		*UNSTR. a	*UNSTR. ε	*sC	DEP-IO	*UNSTR. ə	MAX-IO	NO-CODA	IDENT(F)
/əsc/ (escala)	[əsc]					1		1	
	[əC]					1	1W	L	
/asC/ (astre)	[ásC]							1	
	[áC]						1W	L	
/əsc/ (astral)	[əsc]					1		1	
	[asC]	1W				L		1	1W
	[əC]					1	1W	L	

/esC/ (estri)	[ésC]						1	
	[éC]						1W	L
/əsC/ (estriet)	[əsC]					1		1
	[esC]		1W			L		1W
	[əC]					1	1W	L

(21) BMCD: Learning IDENT(F) >> *UNSTR. ə

		*UNSTR. a	*UNSTR. e	*sC	DEP-IO	IDENT(F)	*UNSTR. ə	MAX-IO	No-CODA
/əsC/ (escala)	[əsC]						1		1
	[əC]						1	1W	L
/asC/ (astre)	[ásC]							1W	L
	[áC]								1
/əsc/ (astral)	[əsC]						1		1
	[asC]	1W				1W	L		
	[əC]						1	1W	L
/esC/ (estri)	[ésC]								1
	[éC]							1W	L
/əsC/ (estriet)	[əsC]						1		1
	[esC]		1W			1W	L		1
	[əC]						1	1W	L



Majorcan Catalan at onset of phonotactic learning initial stage

● **Facing difficulties: The subset problem.** The grammar in (21) is consistent with the observed phonological system of Catalan, but it is also consistent with a proper superset of Catalan in which words starting with #sC- are allowed. We can infer this because, if this grammar is given the input /sC/, it yields the outputs *[sC] and [əsC] indistinctively). That is, as we only have positive evidence (words starting with #[ə]sC-), nothing prevents from selecting words starting with #sC- to be selected in case we have an UR /sC/, predicted by Richness of the Base. This grammar, then, is insufficiently restrictive, allowing things that Catalan does not allow. This is an instance of the Subset Problem (Angluin 1980; Baker 1979): learners presented with only positive evidence cannot proceed from a less restrictive grammar to a more restrictive one.

Solution: free-ride

● **Support tableau derived from phonotactic learning (see 21) + list of unfaithful mappings**

Assume that the learner has a support tableau τ_o (o for old) derived from previous learning [the one in (21)] and a list of L (alternation-supported) unfaithful maps in τ_o .

List of unfaithful maps: $\emptyset > ə$, supported by the alternations found in *tem \emptyset o*, *tem \emptyset ia*, *tem[ə]re*, *tem[ə]ria* and maybe (we cannot know) the controversial ones found in *in \emptyset scriure*, *des \emptyset criure*, [ə]scriure; *pro \emptyset sperar*, *exa \emptyset sperar*, [ə]sperar.

- The knowledge of $\emptyset > ə$ implies that the learner has previously discovered, after doing surgery (see below), the following ranking:

(22) Independent ranking: SYLLCONT, MINSONDIST >> DEP-IO

		SYLLCONT	MINSONDIST	DEP-IO
/tem \emptyset +re/	tə.mə.re			1
	təm.re	1W		L
/kreʃ \emptyset +ria/	tə.mre		1W	L
	krə.ʃə.ria			1
	krəʒ.ria	1W		L
	krə.ʃria		1W	L

● **Replace B,B mapping for A,B mappings via surgery**

Start with a member of L, the unfaithful map (A, B). Make a copy of τ_o called τ_n (n for new) and locate all (B, B) identity maps in τ_n . Perform «surgery» (Tesar *et al.* 2003) on τ_n , changing the underlying form of any morpheme that undergoes the (B, B) map, replacing it with a (A, B) map, and updating the faithfulness assessments accordingly.

- That is... τ_o (old support tableau in (21) + support tableau in (22); irrelevant constraints, like the ones related to vowel reduction, omitted here)

- From τ_o to τ_n via input surgery

(23) τ_o (before input surgery)

		SYLLCONT	MINSONDIST	*sC	DEP-IO	MAX-IO	No-CODA
/əsC/ (escala)	[əsC]						1
	[əC]					1W	L

[Still undefined]

(24) τ_n (after input surgery)

		SYLLCONT	MINSONDIST	*sC	DEP-IO	MAX-IO	No-CODA
/sC/ (escala)	[əsC]						1
	[əC]					1W	L



● Submitting τ_n to BMCD and comparison with τ_o

Now τ_n has the same outputs as τ_o but with a different lexicon. Submit τ_n to BMCD. If BMCD fails to find a grammar for τ_n , discard τ_n and go back to τ_o . If BMCD finds a grammar for τ_n , compare it with the grammar for τ_o . Of these two choices, the learner adopts the one that leads to a more restrictive grammar. In other words, the learner chooses the lexicon that is compatible with the subset language.

(25) τ_n (grammar after input surgery)

		SYLLCONT	MINSONDIST	*sC	DEP-IO	MAX-IO	NO-CODA
/sC/(escala)	[əC]				1		1
	[əC]				1	1W	L
	[sC]			1W	L		L

[Defined]

A grammar / ranking is found: *sC >> Dep-IO

[Note: This process continues with the other unfaithful maps in L and with all of the subsets of the unfaithful maps in L. If a free ride is warranted, then the support tableau at the end of the process will have different underlying representations than at the beginning of the process, and the grammar will be more restrictive.]

● **Competition between τ_n and τ_o .** If BMCD succeeds in finding a ranking for τ_n , a decision must be made: which hypothesis about the lexicon is superior, the one embodied by τ_n or the one embodied by τ_o ? We seek the lexicon that allows for a more restrictive grammar, thereby avoiding the Subset Problem. The choice between τ_n and τ_o therefore requires a metric for grammar restrictiveness, and one exists in the form of the «r-measure» of Prince and Tesar (2004). Grammars with higher r-measures are in general more restrictive, since they grant more power to markedness constraints. Prince and Tesar note that grammars with equal r-measures may also differ in restrictiveness, but this as-yet imperfect metric appears to be sufficient for present purposes. In short, τ_n beats τ_o if τ_n supports a grammar with a higher r-measure.

R-measure (Prince and Tesar 2004: 252)

The r-measure for a constraint hierarchy is determined by adding, for each faithfulness constraint in the hierarchy, the number of markedness constraints that dominate that faithfulness constraint.

In our case... τ_n is more restrictive than τ_o

- Not only because *sC >> DEP-IO
- BUT ALSO because SYLLCONT, MINSONDIST and *sC >> DEP-IO

8. Remaining formal issues

• About the morphological edges at the surface representations

A set of output-output positional faithfulness constraints that make reference to morphological edges (i.e. initial syllable of the stem) seem to be necessary to account for the data. This means that morphological edges are visible and active in the surface representations, supporting therefore the containment approach to faithfulness in OT.

• About the opaque character of alternations by prefixation and the UR of words starting with VCC-

Majorcan Catalan speakers alternate between realizations with [ə] and [e] in a very reduced set of cases such as *empr[ə/e]nyar* ‘to bother’, *engr[ə/e]ixar* ‘to gain weight’ or *al[ə/e]grar* ‘to make happy’ and sporadically show realizations with [e] when the vowel is not the initial of the stem (*ab[e]rració* and *am[e]ricà*).

This reveals the **opaque character** of morphophonological alternations by prefixation, in that it seems that speakers erroneously interpret these forms as prefixed (i.e. *a[e]rració*, *am[e]ricà*, ^{HISTPREF}*aix[ə/e]car*, ^{HISTPREF}*empr[ə/e]nyar*, ^{HISTPREF}*engr[ə/e]ixar*, ^{HISTPREF}*al[ə/e]grar*).² This reveals the **uncertainty** of speakers about the underlying character of the first vowel in the cases where a consonantal group follows (i.e. *empr[ə/e]nyar*, *engr[ə/e]ixar*). (Oostendorp and Bermúdez-Otero are skeptical about this explanation, though.)

• More empirical evidence in the same direction. Interaction between stress and the initial vowel of words starting with esC- in Catalan and Spanish.

(1) Epenthetic vowels cannot be stressed

(2) 1, 2, 3, 6 grammatical persons of the PI verbal forms are paroxytonic:

Catalan	Spanish
1 ju ro	1 ju ro
2 ju res	2 ju ras

² We mark historically prefixed words with HISTPREF.

3jura	3jura
6juren	6juran
(3) 1, 2, 3, 6 grammatical persons of verbs starting with VsC-, though, are oxytonic:	
Catalan	Spanish
1 estic	1 estoy
2 estàs	2 estás
3 està	3 está
6 estan	4 están
(4) No instances of words starting with <i>ésC-</i> in Spanish, except for: <i>éste</i> , <i>ésta</i> , etc.	

(See Harris 1970, for Spanish; Wheeler 2005, for Catalan.)

9. SELECTED REFERENCES

- Bibiloni, Gabriel. 1983. La llengua dels mallorquins. Anàlisi sociolingüística. Doctoral dissertation, University of Barcelona.
- Bibiloni, Gabriel. 1998. La *e* àtona a Mallorca. In *Estudis de llengua i literatura en honor de Joan Veny*. Vol. 2, ed. by Josep Massot i Muntaner, 533-539. Barcelona: Publicacions de l'Abadia de Montserrat.
- DeCesaris, Janet Ann. 1987. Epenthesis in Catalan. In *Studies in Romance languages: Proceedings of the 15th Linguistic Symposium on Romance Languages*, ed. by Carol Neidle and Rafael A. Núñez Cedeño, 79-91. Dordrecht: Foris.
- Lloret, Maria-Rosa. 2002. Estructura sil·làbica. In *Gramàtica del català contemporani*. Vol. 1, *Introducció. Fonètica i fonologia. Morfologia*, dir. by Joan Solà, Joan Mascaró, Maria-Rosa Lloret, and Manuel Pérez-Saldanya, 195-249. Barcelona: Empúries.
- Kiparsky, Paul. 1968. How abstract is phonology? Distributed by Indiana University Linguistics Club, Bloomington, IN. Published with some revisions as part 1 of Kiparsky (1973), 5-56. Reprinted in Kiparsky (1982), *Explanation in phonology*, 119-163. Dordrecht: Foris.
- Kiparsky, Paul. 1973. Phonological representations. In *Three dimensions of linguistic theory*, ed. by Osamu Fujimura, 1-136. Tokyo: TEC.
- Mascaró, Joan. 1976. Catalan phonology and the phonological cycle. Doctoral dissertation, MIT, Cambridge, MA.
- Mascaró, Joan. 2002. El sistema vocàlic. Reducció vocàlica. In *Gramàtica del català contemporani*. Vol. 1, *Introducció. Fonètica i fonologia. Morfologia*, dir. by Joan Solà, Joan Mascaró, Maria-Rosa Lloret, Manuel Pérez-Saldanya, 89-123. Barcelona: Empúries.
- Mascaró, Joan. 2005. Optimal Paradigms and lexical exceptions. Paper presented at the 2nd Old-World Conference in Phonology, University of Tromsø, 22-25 January.
- McCarthy, John J. 1981. The role of the evaluation metric in the acquisition of phonology. In *The logical problem of language acquisition*, ed. by C. L. Baker and John J. McCarthy, 218-248. Cambridge, MA: MIT Press.
- McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
- McCarthy, John J. 2005a. Taking a free ride in morphophonemic learning. *Catalan Journal of Linguistics* 4 (*Morphology in phonology*, ed. by Maria-Rosa Lloret and Jesús Jiménez):19-55.
- McCarthy, John J. 2005b. Optimal paradigms. In *Paradigms in phonological theory*, ed. by Laura J. Downing, T. Allan Hall, and Renate Raffelsiefen, 170-210. Oxford: Oxford University Press.
- Nevins, Andrew, and Bert Vaux. 2008. Underlying representations that do not minimize grammatical violations. In *Freedom of analysis*, ed. by Sylvia Blaho, Patrick Bye, and Martin Kramer, 35-61. Berlin: Mouton de Gruyter.
- Pons-Moll, Clàudia. 2005. Avall, que fa baixada. Els efectes de la Llei del Contacte sil·làbic en les llengües romàniques. *Revista de Linguística Occitana* 3:28-48. [Available at: <http://revistadoc.com/>.]
- Pons-Moll, Clàudia. 2011a. Underapplication of vowel reduction to schwa in Majorcan Catalan productive derivation and verbal inflection. In *Romance Languages and Linguistic Theory 2009*. Selected papers from 'Going Romance' Nice 2009, ed. by J. Berns, Haike Jakobs, and Tobias Scheer, 273-289. Amsterdam/Philadelphia: John Benjamins.
- Pons-Moll, Clàudia. 2011b. It is all downhill from here: the role of Syllable Contact in Romance languages. *Probus. International Journal of Latin and Romance Linguistics* 23.1:105-173.
- Pons-Moll, Clàudia. 2012. Loanword phonology, lexical exceptions, morphologically driven underapplication, and the nature of positionally biased constraints. *Catalan Journal of Linguistics* 11 (*On loanword phonology*, ed. by Michael J. Kenstowicz and Teresa Cabré):127-166.
- Pons-Moll, Clàudia. 2013. Underapplication of vowel reduction to schwa in Majorcan Catalan. Some evidence for the left syllable of the stem as a prominent position and for subparadigms. In *NELS 40: Proceedings of the Fortieth Annual Meeting of the North East Linguistic Society*. Vol. 2, ed. by Seda Kan, Claire Moore-Cantwell, and Robert Staubs, 121-135. Amherst: University of Massachusetts, Graduate Linguistics Student Association.
- Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In *Fixing priorities: Constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 245-291. Cambridge: Cambridge University Press.
- Veny, Joan. 1962. Notes phonétiques sur le parler de Campos (Majorque). *Boletim de Filologia* 20:334-335.
- Wheeler, Max W. 1975. Some rules in a generative phonology of modern Catalan. Doctoral dissertation, University of Oxford.
- Wheeler, Max W. 2005 *The phonology of Catalan*. Oxford: Oxford University Press.