

VOWEL REDUCTION AND VOWEL HARMONY IN EASTERN CATALAN LOANWORD PHONOLOGY¹

TERESA CABRÉ

Universitat Autònoma de Barcelona

The aim of this paper is to account for the phonological adaptation of loanwords in Eastern Catalan. As the phonology of these new words deviates from that of the native Catalan vocabulary set (with a certain amount of variation among speakers), the new phonetic features would seem to be borrowed from Spanish. We suggest that a new phonology has emerged whose purpose is to identify loans among the lexicon, the most striking element of this phonology being a harmony effect on stressed mid vowels in the presence of post-tonic [+ATR] mid vowels. The existence of unstressed [+ATR] mid vowels [e, o] in Eastern Catalan has been previously interpreted as lexical exceptions to vowel reduction (Fabra 1912 and Mascaró 2002, among others). However, the phonetic variation in the new lexicon is analyzed here as being fully consistent with Catalan phonology within the theory of lexical strata (Itô & Mester 1999).

keywords: loanwords, vowel reduction, vowel harmony, lexicon strata, Catalan

1. Introduction

This work examines the strategies that Catalan has developed for the incorporation of loans and offers an optimality-theoretical account of the adaptation process of these new words in Eastern Catalan. As the result of being a minority language, Catalan has historically borrowed a certain number of words from neighboring languages, particularly Spanish. This borrowing process has accelerated on a massive scale in the last 100 years, with

¹ This work was partially presented at the *XIVth International Colloquium on Catalan Language and Literature* (Budapest, September 2006) and at *PaPI 2007* (Phonetics and Phonology in Iberia, Braga). I would like to thank the audience at these gatherings as well as Maria Ohannesian, Pilar Prieto, Donka Steriade, Jaume Solà and especially Michael Kenstowicz, for their useful observations and comments. Thanks are also due to the informants who patiently answered my questions about a long list of loans. I am grateful to the reviewers for their suggestions and advice, which has allowed me to improve certain aspects of this paper. This research is funded by grants 2005 SGR 00753 from the Generalitat de Catalunya and HUM2006-01758/FILO and HUM2006-13295/FILO from the Ministerio de Educación y Ciencia-FEDER.

increasing numbers of loans coming from languages other than Spanish, especially English, and the consequence has been the emergence of a new phonology.

These more recent loanwords show striking phonetic differences relative to the native vocabulary in terms of the vowel system of Eastern Catalan. Most notably, speakers of this Catalan variety identify new words through the pronunciation of unstressed mid vowels. Unlike what happens with native words, in loans, unstressed mid vowels never reduce to schwa (from [-post]) or [u] (from [+lab]), instead surfacing as [+ATR], e.g. *C[o]p[e]nhagu[e]*, *N[e]pal*, *V[e]rsall[e]s*, *t[e]mpura* ‘Japanese dish’, *C[o]lgat[e]* ‘brand of toothpaste’, *jud[o]*, *G[e]stap[o]*, etc. In addition, a vowel harmony effect arises: mid vowels in stressed position are pronounced as close mid vowels when in the presence of post-tonic close mid vowels, e.g. *pesto* ['pesto], *Boston* ['boston], *gueto* ['geto] ‘ghetto’, *Oslo* ['oslo], *Volvo* ['bolβo], *Tere* ['tere] ‘Teresa’, *secre* ['sekre] ‘secretary’, *cole* ['kole] ‘school’, etc. This paper analyzes why reduction fails to occur in such loanwords and examines the scope of and variation in the vowel harmony phenomenon. Our main goal is to provide a unified account of the different stages in the adaptation process from within the theory of lexical strata (Itô & Mester 1999). If we accept a hierarchy of foreignness, we can account for the phonology of Catalan vocabulary as a consistent whole, notwithstanding the phonetic variation observed.

Interest in loanword phonology has increased in the last fifty years mainly because of its relation to linguistic change and language acquisition. All languages borrow the vocabulary they need from neighbors and develop adaptation strategies in different ways. Nevertheless, the new vocabulary frequently preserves some phonological (and morphological) aspects that distinguish it from the native vocabulary.

When one language (L1) borrows words from another (L2), there occurs a conflict between trying to preserve the phonological information of L2 on the one hand and trying to satisfy the phonological conditions of L1 on the other. Speakers of L1 tend to keep the more salient characteristics of L2 words while modifying less prominent segments to bring them closer to L1 (Steriade 2001, Kenstowicz 2001). In other words, speakers distinguish loans by pronouncing them differently from how they would pronounce if they were native words, and speakers do this more or less consciously.

When linguistic borrowing occurs on a very large scale (as tends to happen with minority languages such as Catalan), the L1 may adopt elements of the L2's phonology. In this process of incorporation the intrinsic of L1 system undergoes some extrinsic intrusion and change. These modifications are consistent with natural and attested processes and conditions in many languages. It is therefore possible to attribute these conditions to Universal Grammar (UG), with the implication that adult speakers can always resort to them (Kenstowicz 2001).

The particular situation of Catalan under the influence of Spanish has meant that it adapts any sounds in the foreign vocabulary to its phonology in the same way that it has done for Spanish, at least apparently. For example, modern Catalan has assimilated the Spanish segments [x] and [θ], and uses them in all borrowings of any language which contains similar segments. This is in contrast with the situation before the 20th century, when, as far as we know, Spanish loans were completely assimilated to Catalan phonology,² with more or less the same being true of all other foreign vocabulary (Bruguera 1985).

The assimilation of L2 phonological segments raises a question: if loanwords exhibit a different phonology from the native vocabulary, how many phonologies (i.e. grammars) do we have and how many lexicons correspond to each language? Variation among speakers adds another level of difficulty to this already complex map. However, it may be that the adaptation of loans sometimes allows us to discover certain regularities or default elements that are hidden in native vocabulary because of faithfulness to the input.

Our task in this paper is to explain this kind of phonological variation within a language's lexicon and, if possible, describe the factors or processes that can phonologically differentiate loanwords from native vocabulary. Specifically, the aim of this paper is to account for the vowel phonology of loanwords in Eastern Catalan, including the stressed and unstressed vowel systems. The unstressed system differs from the native system in its resolution of mid vowels: since unstressed mid vowels in loans never reduce to schwa or [u] as they would in native vocabulary. Instead, they surface as

² Formerly, these two segments from Spanish were always rendered in Catalan as [k] and [s] respectively, and vowels were also adapted to Catalan phonology: ojo! Sp. [ˈoxo], Cat. [ˈɔku] 'eye' (meaning 'Be careful!'), Rodriguez Sp. [roˈðriyeθ], Cat. [ruˈðriyəs] 'Spanish surname'.

[+ATR], as in the loans *C[o]p[e]nhagu[e]*, *Ibs[e]n*, *bàsqu[e]t* ‘basketball’, *t[e]mpura* ‘Japanese dish’, *c[o]ns[o]mé* ‘consommé’, *B[o]mbai* ‘Bombay’, *Xil[e]* ‘Chile’, etc. The stressed system in loans follows the general tendency of the language to open mid vowels, except for those words with a close mid vowel in post-tonic position. In these cases the [+ATR] value of the post-tonic vowel spreads left onto the stressed mid vowel, thus triggering a vowel harmony process which is completely new in Catalan phonology, as in *Penèlope* [pe'nelope], *UNESCO* [u'nesko], *Beethoven* [be'toβen], *Volvo* ['bolβo], *Irene* [i'rene], *polièster* [po'ljester] ‘polyester’, *profe* ['profe] ‘teacher’, etc. These phenomena are subject to a certain amount of inter-speaker variation, not only in blocking vowel reduction but also in spreading [+ATR] to the stressed vowel and in the scope of this vowel harmony process. Finally, we adapt Ito & Mester’s (1999) model according to which lexical items are organized in a core-periphery structure. A lexicon divided in strata provides us a feasible explanation for the different subphonologies and variation found in Catalan loans, as well as the different stages of loan adaptation.

This paper is organized as follows. Section 2 presents the extended vowel system of Eastern Catalan that includes the native and non-native vocabulary. Section 3 focuses on the earliest attested words showing blockage of vowel reduction in unstressed position and the subsequent expansion of the phenomenon. Section 4 shows the preference for [–ATR] mid vowels in stressed position in loans. Section 5 describes the vowel harmony process and accounts for it by means of constraint ranking. Finally, section 6 links the dialectal variation found to stages in the nativization process and adapts the lexicon strata theory to the three main varieties of non-native words we found in Eastern Catalan.

2. The Catalan vowel system

The phonology of Eastern Catalan loanwords shows striking differences from the phonology of native vocabulary. All these differences have been traditionally attributed to the influence of Spanish (Fabra 1912, Recasens 1993, etc.) due to the fact that vowel reduction to schwa or labial high vowel is blocked when mid vowels appear, yielding outcomes with close mid vowels, as in the Spanish unstressed system: *sin[e] di[e]*, *v[e]det* ‘starlet’, *Pinotx[o]* ‘Pinocchio’, *Nix[o]n*. Mascaró (2002:110) points out that exceptions to

vowel reduction may appear in a great variety of lexical items without any specific phonological context. In addition, some stressed mid vowels do not follow the distribution of Catalan native vocabulary, but instead are pronounced as close mid vowels as well, which once again appears to be consistent with the Spanish system: *gu[é]t[o]*, *c[ó]ll[e]*, *Ir[é]n[e]*, *N[é]st[o]r*, *B[e]th[ó]v[e]n*, *B[o]st[o]n* (Cabr  2002).

Concerning stressed vowels, studies have been carried out in this area but none of them make any reference to the harmonic process. Badia (1968, 1970) and Pi-Mallarach (1997) merely emphasize the generalized tendency of Catalan to open all stressed mid vowels in loanword adaptation. Fabra (1906: 22) refers only to front mid vowels: “Les noms emprunt s poss dant un *e* tonique, sont prononc s par r gle g n ral avec un *e* ouvert”. In fact, many borrowings —whether from Spanish or any other languages— involving mid vowels in either stressed or unstressed position have been adapted to the Catalan native system, as is shown in (1).

- | | | |
|-----|-----------------------------|----------------------------|
| (1) | robo [r βu] ‘theft’ | s tano [s t nu] ‘basement’ |
| | bulto [bu tu] ‘lump’ | enterro [  teru] ‘burial’ |
| | regalo [r  alu] ‘gift’ | pago [pa u] ‘payment’ |
| | desaig e [d zaj w ] ‘drain’ | laringe [l rin  ] ‘larynx’ |
| | merci [mersi] ‘thanks’ | futbol [fub l] ‘football’ |

In addition, certain Spanish proper names such as *Francisco*, *Fernando*, *Celestino* or *Jacinto* have been not only completely assimilated phonologically but also used as bases for the traditional Catalan truncation process (*Cisco* [sisku], *Nando* [nan du], *Tino* [tinu], *Cinto* [sin tu]) (Cabr  1993). Therefore, it is not that Eastern Catalan simply borrows and uses the Spanish vowel system for loans, but rather that Eastern Catalan has introduced certain changes to its phonological system that yield words that sound closer to Spanish. A more detailed examination of the phonology of loans shows that the Eastern Catalan extended system for loans is quite removed from the five-vowel Spanish system, as we can see in (2).

(2)	Stressed system	Native unstressed system
	i u	i u
	e o	
	ɛ ɔ	ə
	a	
	Unstressed system for loans	Spanish stressed and unstressed system
	i u	i u
	e o	e o
	ə	a
	(a)	

The stressed vowel system is common to both native vocabulary and loanwords, because we can find both [+ATR] and [-ATR] mid vowels, but their distribution within each sublexicon is not exactly the same. The tendency to have [-ATR] stressed mid vowels in adapted loans, as the authors referred to above (Fabra, Badia and Pi-Mallarach) state, follows the general pattern of Catalan. If we look at the stress position in native words, we see a clear predominance of [-ATR] among front vowels in non-derived stems. The preference for [-ATR] stressed mid vowels is also evident in suffixed forms: more than twenty stressed suffixes contain a [-ATR] mid vowel (Mascaró 1985).³ In addition, all pre-stressed suffixes (all of them with a high vowel) and all the so-called suffixed forms (such as /ləg/ or /fon/ in *fonòleg* [fu'nələk] 'phonologist' or *gramòfon* [grə'məfun] 'gramophone') show the same pattern. In (3) we offer some examples that illustrate this generalization.

(3)	bleda ['bleðə] 'chard'	hora ['ɔrə] 'hour'
	metro ['mɛtru] 'subway'	lloro ['lɔru] 'parrot'
	col·legi [ku'lɛzi] 'school'	petroli [pə'trɔli] 'oil'
	café [kə'fɛ] 'coffee'	història [is'tɔrjə] 'history'
	còmode ['kɔmuðə] 'comfortable'	desè [də'zɛ] 'tenth'
	ventijol [bən̩ti'ʒɔl] 'light wind'	cavalleresc [kəβələ'rɛsk] 'chivalric'

³ The most common suffixes with [+ATR] mid vowel are / 'er/ > ['e], /'ment/ > ['men], /'on/ > ['o], /'or/ > ['o], /'oz/ > ['os]: *mentider* 'liar' (from *mentida* 'lie'), *sofriment* 'suffering' (from *sofrir* 'to suffer'), *petitó* 'small, dim.' (from *petit* 'small'), *negror* 'blackness' (from *negre* 'black'), *plujós* 'rainy' (from *pluja* 'rain').

unglot [uŋ'glɔt] 'hoof'
 francès [frən'ses] 'French'
 adulteri [əðul'teri] 'adultery'

verdet [bər'det] 'moss'
 nadalenc [nəðə'lɛŋ] 'relative to Christmas'
 harmònic [ər'mònik] 'harmonic'

The emergence of [-ATR] mid vowels in loans depends on the context: partial reduction of mid vowels after stress (that is [+ATR] mid vowels) generally blocks [-ATR] mid vowels in stressed position, whereas stressed mid vowels surface as [-ATR] when a post-stressed high vowel or a schwa occurs, as in the preceding examples from native vocabulary (3).

Catalan native vocabulary exhibits what Crosswhite (2004) calls a sonority-driven unstressed system, which is presented in (4). Seven stressed vowels reduce to three in unstressed position following the vocalic prominence scale (Crosswhite 2001) according to which [i, u, ə] are the lesser prominent vowels. Consequently, they are the most suitable to appear in unstressed position in a prominent-driven neutralization system, which is what Catalan has.

(4)

	- back		+ back		
+ high	i		u	↑ ↑	
- high	e	→	ə	o	↑ ↑
- low	ɛ	↘		ɔ	↑ ↑
+ low			a		
			- lab		+ lab

+ ATR
- ATR

In loanword phonology, the unstressed system is extended relative to that of native words. Schwa always comes from *a*, but we also find instances where this segment is not reduced. Vowel reduction involves the ATR feature, with the rest of the feature values remaining unchanged. The extended system is shown in (5):

(5)

	- back		+ back	
+ high	i		u	
- high	e ↑		ə ↑	+ ATR
- low	ε ↑		ɔ ↑	- ATR
+ low			a	
			- lab	+ lab

We may regard the extended system as following a sonority-driven reduction (Crosswhite 2004) since low and low mid vowels change to [+ATR], with only the lesser prominent vowels arising in unstressed position, according to the vocalic prominence scale shown in (6).

(6) Vocalic prominence scale (after Crosswhite 2001): a > ε, ɔ > e, o > i, u > ə.

3. Vowel reduction blockage

We have some evidence of an extended unstressed system from the beginning of the 20th century. At that time, Catalan grammarians noted the non-reduced pronunciation in post-tonic position in some learned words in the Barcelona area, as in *bas*[e] ‘base’, *catàstrof*[e] ‘catastrophe’, *class*[e] ‘class’, *fas*[e] ‘stage’, *fras*[e] ‘sentence’, *pest*[e] ‘plague’, *superfici*[e] ‘surface’, etc. (Fabra 1912). This phenomenon is consistent with the behavior of another group of learned words which is made up of compounds built with two (or more) Greek or Latin elements or learned variants and created as scientific or technical terminology, even though some of them may later move into the colloquial vocabulary (Mascaró 1985). They differ from other compounds because of the inflectional element *-o* that appears suffixed at the end of each non-final element in many compounds.⁴ Speakers are aware of this fact and often apply the extended unstressed vowel system to at least the first component. The result is a phonetic realization with unstressed [+ATR] mid vowels. It is important to stress that this pronunciation is not extended to the isolated colloquial

⁴ We can add to this group some recent words with the monosyllabic unstressed prefix *co-* ‘with’ such as *codirector* [koðirək'to] ‘codirector’ and *coeditar* [koəði'ta] ‘copublish’. In all items in the native vocabulary with the same prefix, vowel reduction occurs, e.g. *col·laborar* [kuləβu'ra] ‘collaborate’, *confiar* [kunjfi'a] ‘trust’.

elements: compare, for example, *politico-social* [politikosu'sjal] ‘sociopolitical’ or *sociocultural* [sosjoku'tu'ral] ‘sociocultural’ with *polític* [pu'litik] ‘politician’ or *social* [su'sjal] ‘social’ In (7) we show some examples of this lexical class:

- (7) antropomòrfic [əntropo'mərfik] ‘anthropomorphic’
 monolingüe [mono'liŋgwe] ‘monolingual’
 oto-rino-laringòleg [oto,ri noləriŋ'gələk] ‘ear, nose and throat specialist’
 politico-social [politikosu'sjal] ‘sociopolitical’
 sociocultural [sosjoku'tu'ral] ‘sociocultural’
 morfo-sintàctic [morfosiŋ'taktik] ‘morphosyntactic’

The non-reduced vowel system can arise in unstressed position in a great variety of lexical items, such as in Latin or Greek expressions (*sin[e] di[e]*, *èpsil[o]n*), in learned words (*cosm[o]s* ‘cosmos’, *[e]ff[e]mèrid[e]* ‘event’), in classical-language-origin compounds and technical terms (*[o]l[eo]ducte* ‘pipeline’, *p[o]lític[o]-social* ‘sociopolitical’), in certain Catalan proper names (*S[o]là*, *Viv[e]s*), in acronyms (*UNESC[o]*, *B[e]N[e]LUX*), in some non-traditional truncated forms (*prof[e]* ‘teacher’, *dem[o]* ‘demonstration’), and in common and proper names borrowed from any foreign language (*B[e]thov[e]n*, *D[o]st[o]ievsky*, *s[o]pran[o]*, *B[o]d[e]laire*), etc. Some of these words also exhibit close mid vowels in stressed position, as we will see below: *c[ó]sm[o]s*, *UN[é]SC[o]*, *pr[ó]ff[e]*, *d[é]m[o]*, *B[e]th[ó]v[e]n*. We must bear in mind that a large degree of variation among speakers can be observed in both processes, but in general the more recent an item, the more patent the process.

Leaving aside the idiolectal and dialectal variation we can find in all these types of words, the important thing about them is that they all have something in common in that they somehow stand out in the way speakers treat them. That is, speakers do not regard these words as belonging to the native vocabulary because they resort to a phonological process that never occurs in native words.

In contrast to older borrowings, new words are not generally common words introduced through oral transmission but instead pertain to more formal registers and tend to be introduced through the written form. Speakers have to read them and incorporate

them into scientific or technical speech. So the phonology of the language that lends a word is overridden by the graphic form in which this word is transmitted.

In Eastern Catalan orthographical *a* is usually pronounced as [ə] (closer to [a] in the area of Barcelona) both in pre-tonic and in post-tonic position, e.g. *Chanel, canapé, in fraganti, Jaguar, Karamazov, Berna, Armani, or Philadelphia*, and even in words with unstressed mid vowels that are not reduced like *alel·luia* [əle'lujə], or *Antígona* [ən̩'tiɣonə]. And at the present time Eastern Catalan speakers apparently perceive schwa to be closer to [a] than [e ε]. This correspondence illustrates the psychological (and phonetic) distance between [ə] and the low vowel [a] on the one hand or the mid vowels [e, ε] on the other. We have evidence for this from the early stages of writing, in which children tend to systematically represent all schwas by the letter *a*. In addition, many adult speakers pronounce [e] or [ə] depending on the written form, associating *e* with [e] and *a* with [ə]. Often in Catalan the same proper name can have different spellings, and this can determine whether the pronunciation shows vowel reduction or not: *Queralt* [ke'ral] / *Caralt* [kə'ral] and *Torres* ['tores] / *Torras* ['torəs].

We also find other instances of schwa in neologisms. Initial epenthesis in foreign words with a specific initial cluster is obligatory in Catalan. Thus, words such as *Fnac* ‘a chain store’ *slogan, sketch, speaker, Slazenger* or *slalom* are unfailingly pronounced with an initial schwa. Note that here schwa occurs even though no vowel is transcribed.

When mid vowels appear in written form, [e o] are the general pronunciations. (8) provides some examples of words in which unstressed mid vowels are not reduced as they would be in native vocabulary but surface instead as [+ATR] mid vowels. To facilitate understanding, only mid vowels are transcribed:

- | | | |
|-----|---|--|
| (8) | agr[o]turisme ‘agrotourism’
cin[e] ‘cinema’
v[e]det ‘starlet’
p[e]r[o]né ‘fibula’
hipèrb[o]l[e] ‘hyperbole’
[e]sp[e]rant[o] ‘Esperanto’
Dick[e]ns
Nix[o]n
F[e]rrari | franc[o]prussia ‘Franco-Prussian’
íd[e]m ‘Latinism, the same’
c[o]llage ‘collage’
c[o]ns[o]mé ‘consommé’
síndr[o]m[e] ‘syndrome’
brav[o] ‘well done’
lbs[e]n
Byr[o]n
P[e]trarca ‘Petrarch’ |
|-----|---|--|

Regarding the realization of unstressed mid vowels, it must be noted that [e o] (+ATR) are unmarked relative to [ɛ ɔ] (–ATR). As we saw in the table in (5) above, [ɛ ɔ] are blocked in unstressed position, thus changing the ATR value. The phonetic distinction of five vowels based on the written forms *a, e, o, u, i* has the goal of emphasizing loans by improving the equidistant positions of vowels in the vowel trapezoid, thereby making them clearer, due to the fact that there is an unambiguous correspondence between the written and phonetic forms. This can be interpreted as a faithfulness constraint that arises in loanword phonology. This constraint must be indexed to the specific stratum because the native vocabulary remains unaffected. The following hierarchy in (9) accounts for the extended vowel system in unstressed positions for a subset of the lexicon.

(9) * ɛ ɔ >> L–Faith (e o) >> * ɛ ɔ

The constraints * ɛ ɔ and * ɛ ɔ force vowel reduction of mid vowels and are active when native vocabulary is used. When one of the so-called non-native words occurs, the indexed faithfulness constraint applies, thus blocking complete reduction of mid vowels. We can also rewrite this ranking to show how the ATR feature values are crucial in this process. In (10) the new ranking is shown and the tableau in (11) exemplifies this ranking. The written form acts like the underlying form.

(10) *[-Stress, –ATR] >> L–Faith (e o) >> *[-Stress, +ATR]

(11)

síndrome	*[-Stress, –ATR]	L–Faith (e o)	*[-Stress, +ATR]
síndr[ɔ]m[ɛ]	*!*		
síndr[u]m[ɔ]		*!*	
☞ síndr[o]m[e]			**

The ranking presented in (10) can be considered an example of what Pater (2004) calls exceptional blocking by faithfulness, that is, when a general process of the language is

blocked in specific words. In this case, the general process is that mid vowels reduce to [ə] or [u] and the blocking yields the exceptional pattern seen in loans.⁵

4. The emergence of the unmarked: the preference for [-ATR] stressed vowels

In contrast to what was claimed by the grammarians referred to above, our data show that the tendency for Eastern Catalan to open mid vowels in stressed position in foreign adaptations has been reduced to certain specific contexts (Cabré 2006). We find open mid vowels in borrowed oxytone words ending in a consonant or glide, regardless of the feature values of the segment. Some examples of this are provided in (12).

(12)	minu[ɛ]t	ved[ɛ]t	Budap[ɛ]st
	sk[ɛ]tch	estr[ɛ]s	Queb[ɛ]c
	Ever[ɛ]st	Flaub[ɛ]rt	p[ɔ]nx
	Nova Y[ɔ]rk	cowb[ɔ]y	ad h[ɔ]c
	esn[ɔ]b	Reps[ɔ]l	sh[ɔ]rts

If we compare these outcomes with those of the most productive stressed suffixes that yield oxytone words, we see that the adaptation of these loans follows the general tendency of the language because mid vowels generally surface [-ATR] in native words too, as in *verd[ɛ]t* ‘mold’, *ungl[ɔ]t* ‘hoof’, *puj[ɔ]l* ‘mountain’, *portugu[ɛ]s* ‘Portuguese’, *nadal[ɛ]nc* ‘relative to Christmas’, *cavaller[ɛ]sc* ‘chivalric’. This pronunciation is generalized independently of whether preceding syllables undergo vowel reduction or not. Though there is some variation, pre-tonic mid vowels generally surface as closed [+ATR] and stressed mid vowels surface as open [-ATR]: *Quebec* [keβék], *Flaubert* [floβért] or [floβér], *Repsol* [repsól], *Tolstoj* [tolstój] (at least this is the most common pronunciation in the area of Barcelona).

⁵ Catalan exhibits other examples of this typology. For example, final deletion of /n/ after a stressed vowel in nominal forms is active in native words but not in loans: compare *camí*, [kə'mi], *camins* [kə'mins] ‘path, sg/pl’ with *caiman* [kəj'man], *caimans* [kəj'mans] ‘caiman, sg/pl’.

The preference for open mid vowels in stressed position is also evident in penult and antipenult stress words, but only when *a*, *i* or *u* appears in post-tonic position, as shown in (13).

(13)	Calif[ɔ]rnia	T[ɛ]xas	B[ɛ]rna ‘Berne’
	Vars[ɔ]via ‘Warsaw’	t[ɔ]fu	Z[ɛ]us
	[ɛ]du ‘Edward’	V[ɛ]nus	lato s[ɛ]nsu
	b[ɔ]dy	Art[ɛ]mis	[ɛ]mi ‘Emily’
	Tr[ɔ]tsky	Dostoi[ɛ]vsky	a posteri[ɔ]ri
	in extr[ɛ]mis	urbi et [ɔ]rbi	vox p[ɔ]puli

The vowel pattern of the language clearly asserts itself in this case, above all when final high vowels are involved. All stressed mid vowels from native words ending in *i* or *u* in the written form are pronounced as [–ATR] (e.g. *inn[ɔ]cu* ‘harmless’, *ing[ɛ]nu* ‘candid’, *micr[ɔ]bi* ‘microbe’, *sil[ɛ]nci* ‘silence’, *petr[ɔ]li* ‘petrol’, *pr[ɛ]mi* ‘prize’, *dim[ɔ]ni* ‘devil’, *mist[ɛ]ri* ‘mystery’, etc).⁶

We can interpret this [+Stress, –ATR] preference as the emergence of the unmarked (McCarthy & Prince 1994) since open vowels are phonetically longer and acoustically more noticeable (Maddieson 1997) than closed ones because of their greater sonority. This is also favored by the unstressed mid vowels so that they surface as [+ATR], that is, [–Stress, +ATR] is phonetically shorter because of its lower sonority. On the basis of this generalization, we propose the constraint *[+Stress, +ATR] in order to guarantee the preference for open mid vowels in stressed position in either loanwords or native vocabulary.

(14) *[+Stress, +ATR]: Close mid vowels are not allowed in stressed position.

The examples in (12) and (13) show that the ATR value of stressed mid vowels is independent of the presence of partial reduced vowels. The faithfulness constraint that blocks native vowel reduction is ranked above *[+Stress, +ATR] because it is specifically indexed for loans (Faith (e o)–L >> *[+Stress, +ATR]).

⁶ It is important to point out that words ending in unstressed and orthographic ‘-u’ are practically non-existent in the native Catalan vocabulary. On the other hand, they are abundant in borrowings from Latin.

Nevertheless, when stressed mid vowels occur in final position —mostly in gallicisms— the tendency is to surface as [+ATR], as is shown in (15).

- | | | | |
|------|-------------|-----------|-----------|
| (15) | beb[e] | consom[e] | neglig[e] |
| | jaqu[e] | pur[e] | cup[e] |
| | frivolit[e] | clix[e] | pat[e] |
| | ximpanz[e] | quingu[e] | moar[e] |

These examples contrast with the following adapted loans: *t*[ɛ] ‘tea’, *caf*[ɛ] ‘coffee’, *obo*[ɛ] ‘oboe’, *comit*[ɛ] ‘committee’. Relative to the native vocabulary, such words must be regarded as morphological exceptions because the lack of final /n/ in the underlying form prevents them from undergoing any suffixation process. If we compare their morphological productivity with that of final-stressed native words, the contrast is even more evident: *morè* [mu're], *morens* [mu'rens] ‘brown, sg. and pl.’, *morenet* [murə'net] ‘little brown’, *morenor* [murə'no] ‘brownness’, etc. versus *clixé* [kli'ʃe] and *clixés* [kli'ʃes] ‘cliché sg. and pl.’). We must also bear in mind that the native process of final /r/ deletion in stressed syllables is not active (with a few exceptions) in loanword phonology: *plumi*[er] ‘pencil case’, *necess*[er] ‘toilet bag’ and *premi*[er] ‘prime minister’, while *Molière* and *Baudelaire* are pronounced with final [ɛr]. Even though the context calls for a [–ATR] mid vowel, the stressed vowel in native words is [+ATR]. One would expect the pronunciation to follow the pattern we find in the stressed suffix /–er/ with final /r/ deletion (one of the most productive suffixes in the language: *fuster* [fus'te] ‘woodworker’, *llimoner* [ʎimu'ne] ‘lemon tree’, *sabater* [səβə'te] ‘shoemaker’). Nevertheless, loans never delete final /r/.

We therefore propose the markedness constraint L*[–ATR]# for loans ending in a mid vowel:

- (16) L*[–ATR]#: Open mid vowels in stressed final position are not allowed in loans.

This markedness constraint indexed to loans accounts for what Pater (2004) denotes an exceptional triggering process, that is, a process that only applies to exceptional forms. Because of its exceptionality this constraint is ranked above and dominates all other conditions, as shown in (17).

- (17) L*[-ATR]# >> *[-Stress, -ATR] >> L-Faith (e o) >> * [+Stress, +ATR] >> *[-Stress, +ATR]

We illustrate in (18) how this ranking accounts for the surface form of loans such as *consomé* or *Dostoievsky*:

(18)

Dostoievsky	L*[-ATR]#	L-Faith (e o)	* [+Stress,+ATR]	*[-Stress,+ATR]
☞ d[o]st[o]i['ɛ]vsk[i]				**
d[u]st[u]i['ɛ]vsk[i]		*!*		
d[o]st[o]i['e]vsk[i]			*!	**
d[u]st[u]i['e]vsk[i]		*!*	*	
consomé				
☞ c[o]ns[o]m['e]			*	**
c[o]ns[o]m['ɛ]	*!			**
c[u]ns[u]m['e]		*!*	*	
c[u]ns[u]m['ɛ]	*!	**		

5. Vowel harmony

When post-tonic mid vowels surface as [+ATR], stressed mid vowels undergo a process of vowel harmony: the otherwise expected outcome with [-ATR] is blocked and open mid vowels are raised to close when followed by a close mid vowel (Cabré 2006, Bonet et al. 2007).). Some examples provided in (19) show the harmonic process.

- | | |
|------------------------------------|-------------------------------|
| (19) pesto ['pesto] ‘pesto sauce’ | Boston ['boston] |
| gueto ['geto] ‘ghetto’ | Oslo ['oslo] |
| Volvo ['bolʋo] | Tere ['tere] |
| secre ['sekre] ‘secretary’ | grosso modo [grosɔ'moðo] |
| cole ['kole] ‘school’ | jefe ['xefe] ‘boss’ |
| Irene [i'rene] | Penélope [pe'nelope] |
| UNESCO [u'nesko], | Beethoven [be'toβen] |
| Hölderlin ['xolderlin] | Essex ['eseks] |
| polièster [po'ljester] ‘polyester’ | Euterpe [ew'terpe] |
| Montse ['monse] | forense [fo'rense] ‘forensic’ |
| pòquer ['poker] ‘poker’ | cosmos ['kosmos] |
| Pinotxo [pi'notʃo] ‘Pinocchio’ | Nèstor ['nestor] |

Tòquio ['tokjo] 'Tokyo'
Jespersen ['ʒespersen]
Opel ['opel]

Mendelsohn ['mendelson]
Apolo [ə'polo] 'Apollo'
folklore [fol'klore]

In the examples above, we note that the [+ATR] value has spread from the post-tonic vowel to the stressed mid vowel. The scope of this harmonic spread is one trochee, the basic foot of Catalan (Cabré 1993). If we focus on proparoxytones, the harmonic process arises when a stressed vowel is followed by a close mid vowel. When the medial post-tonic vowel is a schwa or high, the general outcome is an absence of harmony (e.g. *Sòcrates* ['sòkrətes], *Hèrcules* ['èrkules]). Nevertheless, it is important to bear in mind that this process presents a great amount of variation across speakers and dialects. For example, data gathered from Eastern Catalan speakers showed that some of them did not apply the harmonic process to words such as *r[ɛ]qui[e]m* 'requiem', *r[ɛ]c[o]rd* 'record', *N[ɔ]bel*, *S[ɛ]n[e]ca*, *C[ɔ]rc[e]ga*, or *G[ɔ]lg[o]ta*. Others extended the process outside the trochee and pronounced the words *Sòcrates* ['sokrates] and *Hèrcules* ['erkules]. At any rate, the more distant (or foreign) a word is felt to be, the more likely it is that the harmony process will apply; by the same token, the closer (or nativized) a word is felt to be, the more likely it is that vowel reduction will occur.

The vowel harmony process presented here shows some parallels with the vowel harmony in Central Italian described by Stefano Canalis (2009, this volume). In both cases the trigger for the harmony is an unstressed vowel and the target is also to the left of the trigger. This indicates that the direction of the harmonic process goes from right to left, exactly the opposite of the best-known harmony systems such as in Turkish or Finnish. Conversely, there is a crucial difference between the two processes, because whereas in the Italian dialects the target is likewise a weak vowel, in Eastern Catalan the target is the stressed vowel of the word, the processes thus entail distinct analyses. Moreover, the prosodic context of the Catalan process contrasts with the general harmony cases in which the trigger is a prominent position and the target a weak position.

Indeed, the fact that the trigger for vowel harmony is the post-tonic [+ATR] vowel, an element in a weak position, seems to disagree with analyses based on the perceptual prominence of strong positions, which is certainly what the stressed syllable is. According

to Walker (2005), the ATR feature of the post-tonic vowel associates with the preceding stressed vowel in order to become more perceptible. Weak position (the post-tonic vowel) “is licensed by association to strong position” (the stressed vowel) in the scope of a trochee. In other words, the ATR feature is attracted to the stress position, yielding ATR harmony. This phenomenon has a particularly pronounced effect on loans and generates a prominent difference between them and the native vocabulary. In order to account for this fact we propose the following constraint of positional markedness:

(20) License [ATR post-tonic, σ']: ATR value spreads to the preceding stressed vowel.

License [ATR post-tonic, σ'] dominates *[+Stress, +ATR]. This markedness constraint is indexed to a subset of words we have labeled as loans that is in turn a subset of learned words. As L–Faith (e o) is indexed to learned words, License [ATR post-tonic, σ'] must dominate L–Faith (e o) and *[+Stress, +ATR] too. The ranking in (21) accounts for the vowel harmony process, the most salient feature of Catalan loanword phonology.

(21) [ATR post-tonic, σ'], L*[-ATR]# >> *[-Stress, -ATR] >> L–Faith (e o) >> *[+Stress, +ATR] >> *[-Stress, +ATR]

Tableau (22) illustrates one example of vowel harmony in Eastern Catalan. The constraints L*[-ATR]# and *[-Stress, -ATR] are not relevant for the process. *Pinotxo* ‘Pinocchio’ may be interpreted by some Catalan speakers as a native name or it has simply become completely nativized, thus surfacing as *pin[ɔ]tx[u]*. Nevertheless, this pronunciation never arises when the word is indexed as a loan.

(22)

Pinotxo	[ATR post-tonic, σ']	L–Faith (e o)	*[+Str, +ATR]	*[-Str, +ATR]
☞ pin[o]tx[o]			*	*
pin[o]tx[u]		*!	*	
pin[ɔ]tx[o]	*!			*
pin[ɔ]tx[u]		*!		

Even though adjacency inside the trochee is needed for the general process, we find some loans—in greater or lesser quantity depending on the speaker—with the same effects of vowel harmony yet without the requirements of adjacency in the foot. The mere presence of an unstressed [+ATR] mid vowel inside the word is enough to trigger unbounded vowel harmony. In this case, the scope of the harmony process is the whole word. Some examples are *Otegi* [o'teyi], *Sòcrates* ['sokrates], *Khomeini* [xo'mejni], and *vedet* [be'ðet]. In sum, much work is needed to clarify the differences among dialects and speakers.

6. Variation and stages of nativization

As we have noted, there is a certain amount of variation in the application of all these processes. For example, unstressed close mid vowels in older borrowings are quite common in the northeast area (from Barcelona to Girona), but less so in the center and the south of the dialect region where Eastern Catalan is spoken. Moreover, we also have found disagreement within the same dialect, often across different generations. This variation may or may not involve the presence of vowel harmony when the segmental context allows it: *euro* ['ewro] / ['ewru], *Nobel* ['noβel] / ['noβəl], *Irene* [i'rene] / [i'renə]. Within the Barcelona area, it is also possible to hear some words with unstressed [+ATR] vowels pronounced without evidence of the vowel harmony process: *Sèneca* ['senekə], *Còrcega* ['kørseyə], *Nobel* ['noβel], *rèquiem* ['rəkjem], *rècord* ['rəkər]. As far as we know, this possibility is blocked in the rest of the linguistic domain when mid vowels occur in stressed positions.

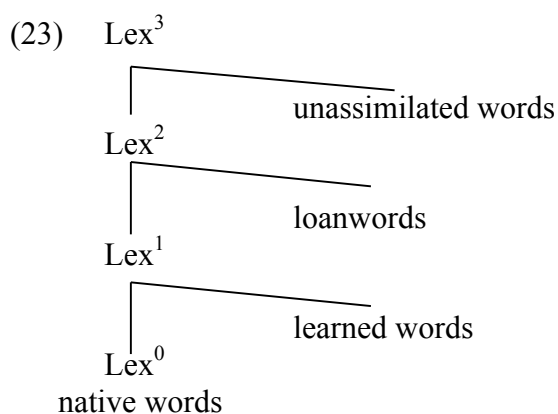
The age of a speaker also seems to play an important role: in general younger people are more innovative than older people and therefore apply the different processes across a broader range of words. In addition, vowel harmony applied to the whole word is quite widespread in the younger generation, where one hears *vedet* [be'ðet], *Khomeini* [xo'mejni], *Toyota* [to'jota], *Minnesota* [mine'sota], etc.

If we accept the notion of different strata in the lexicon, it is possible to explain these different phonologies and the variation found in their application. As Ito & Mester (1999: 64) say, “Lexical items do not come neatly packaged into groups labeled [+/-foreign]; rather, different degrees of nativization among foreign words are commonplace.

Instead of the partitioning into parallel and disjoint [+foreign] and [-foreign] sublexica, we have a hierarchy of foreignness, with exceptions to one rule always being exceptions to another rule, but not vice versa.”

In summarizing the differences found in the vocabulary at hand that we have labeled as *new words*, it is possible to define three groups according to the three main processes we have offered.⁷ The first group includes those new words (learned words, classical language compounds, etc.) whose main characteristic is to fail mid vowel reduction; the second group corresponds to loanwords or any new word that undergoes a harmony process inside the trochee; the third one includes the newest vocabulary (unassimilated words), and shows unbounded harmony when no reduced mid vowels are present in the word. Native vocabulary consistently exhibits none of these processes.

By taking up the proposal of Itô & Mester, we regard lexical items as “organized in terms of an overall core-periphery structure” and adapt their model to the Catalan sublexica: Lex⁰ – native vocabulary; Lex¹ – learned words; Lex² – loanwords; Lex³ – unassimilated words. Considering the variation found in the data, a word belongs to one or another level according to how many processes it has undergone, which will depend on the speaker and dialect. Moreover, we must bear in mind that some items historically labeled as borrowings (such as *enterro* [əŋˈteru] ‘burial’) have a core behavior, that is, they are completely nativized, whereas other items that are historically native (such as *literature* [literəˈturə] ‘literature’) are still peripheral for some speakers because of the pronunciation of the front mid vowel. The structure of the Catalan sublexica is represented in (23):



⁷ It is not completely clear what set words with stressed final [+ATR] mid vowel such as *consomé* belong.

Native words constitute the core of the lexicon because they fulfill all constraints. As Ito & Mester (1999: 65) emphasize, “moving outwards from the core, we encounter items that violate more and more constraints until we encounter, at the periphery, items fulfilling only a small subset of the constraints. These constraints are truly fundamental in the sense that they define the basic syllable canons and other central aspects of the language”. Thus, our data clearly show that initial epenthesis is one of the basic and unavoidable phonological processes of Catalan; “native” vowel reduction, on the other hand, is not such a central process.

We show in (24) the relation between the different classes of lexical items discussed above and the processes active in the Catalan lexicon. The basic syllable structure is represented here by the initial vowel epenthesis but includes other phonological processes (such as sonority or articulation place assimilation) that we have not discussed.

(24)

EPENTHESIS (INEP)	Initial vowel epenthesis
NOEXTENDEDREDUCT (NO-EXRED)	No extended vowel reduction
NOFOOTHARMONY (NO-FTHARM)	No vowel harmony in the foot domain
NOWORDHARMONY (NO-WDHARM)	No vowel harmony in the word domain

Except for initial vowel epenthesis, processes applying to a specific sublexicon must be labeled in negative terms. Native vocabulary items must fulfill all core requirements, so it would violate them in positive terms because of their peripheral status. The table in (25) clearly shows the hierarchical relations between the different strata in the Eastern Catalan phonological lexicon.

(25)

	INEP	NO- WDHARM	NO- FTHARM	NO-EXRED
Native words	√	√	√	√
Learned words	√	√	√	violated
Loanwords	√	√	violated	violated
Unassimilated	√	violated	violated	violated

The dialectal and idiolectal variation is interpreted in terms of item place, that is, each dialect –even each speaker– can vary in the consideration of a particular word in its lexicon. As Itô & Mester (1999: 70) say, “the less nativized an item is, the more it disobeys lexical constraints, i.e. the more it falls outside of various constraint domains and is located towards the periphery of the lexical space.”

The stratification of the Catalan lexicon presented above accounts for the gradual application of the constraints to different subsets of words. A grammar with indexed faithfulness constraints interspersed between markedness constraints solves not only the problem of apparent different grammars but also its learnability in an Optimality Theory account (Pater 2005).

7. Conclusion

Summarizing our study, we can say that it offers a feasible interpretation of the amalgam of data at hand, although considerable work is still needed to clarify dialectal differences. On the one hand, the exceptions to vowel reduction in learned words are interpreted in terms of faithfulness to written form in order to make them stand out, by improving the equidistance of positions in the vowel trapezoid. On the other hand, Catalan loanwords show two opposite effects with respect to stressed vowels: the emergence of the unmarked generalizes the [–ATR] value on stressed mid vowels in some specific positions, whereas the [+ATR] value harmonizes the stressed mid vowel in the presence of a post-tonic [+ATR] mid vowel. A word that undergoes this vowel harmony becomes more perceptible, so that the scope of the process is the stress foot. The lexical strata framework accounts for the apparent random facts and idiosyncrasies of non-native words in a very intuitive way. The stratification of the lexicon into domains is clearly evident in Eastern Catalan, but the variation found in the data must be interpreted as reflecting the idiolectal location in one of these domains.

References

- Badia i Margarit, Antoni M. 1968. “Predominio de las vocales abiertas e y o en el catalán de Barcelona”. *Sons i fonemes de la llengua catalana*. 1988. 209-214. Barcelona: Universitat de Barcelona
- 1970. “Les vocals tòniques e i o en el català de Barcelona”. *Sons i fonemes de la llengua catalana*. 1988. 141-190. Barcelona: Universitat de Barcelona.
- Bonet, Eulàlia, Maria-Rosa Lloret & Joan Mascaró. 2007. “Domain and directionality in Catalan ATR harmony”. Paper presented at the workshop “Harmony in the languages of the Mediterranean” *OCP* 4, 18-21 January, Rhodes.
- Bruguera, Jordi. 1985. *Història del lèxic català*. Barcelona: Enciclopèdia Catalana.
- Cabré, Teresa. 1993. *Estructura gramatical i lexicó: el mot mínim català*. PhD dissertation. Universitat Autònoma de Barcelona.
- 2002. “Altres mecanismes de formació de mots”. *Gramàtica del català contemporani*, ed. by Joan Solà et alii, vol. 1, 889-932. Barcelona: Empúries.
- 2006. Els sistema vocàlic del català central i l’adaptació dels manlleus. Paper presented at *XIV Col·loqui internacional de llengua i literatura catalanes (AILLC)*. 4-9 September, Budapest.
- Canalis, Stefano. 2008. “Post-tonic vowel harmony in some dialects of Central Italy: the role of prosodic structure, contrast and consonants”. In this volume.
- Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. London: Routledge.
- 2004. “Vowel reduction”. *Phonetically based Phonology*, ed. by Bruce Hayes et alii, 191-231. Cambridge: Cambridge University Press.
- Fabra, Pompeu. 1906. “Les E toniques du Catalan”. *Révue Hispanique*, XV. 9-23.
- 1912. *Gramàtica de la llengua catalana*. Barcelona: L’Avenç.
- Itô, Junko & Armin Mester. 1999. “The Structure of Phonological Lexicon”. *A handbook of Japanese Linguistics*, ed. by Natsuko Tsujimura, 62-100. Oxford: Blackwell.
- Kenstowicz, Michael. 2001. *The role of perception in loanword phonology*, Ms.
- 2005. *Loanword Phonology*. Lectures at LSA Institute, July. Cambridge, MIT.
- McCarthy, John & Alan Prince. 1994. “The emergence of the unmarked: Optimality in prosodic morphology”. *Proceedings of the North East Linguistic Society*, ed. by M. González, 333-379. Amherst, MA: GLSA Publications.

- Maddieson, Ian. 1997. "Phonetic Universals". *The Handbook of Phonetic Sciences*, ed. by W. Hardcastle & J. Laver, 619-39. Oxford: Blackwell.
- Mascaró, Joan. 1985. *Morfologia*. Barcelona: Enciclopèdia Catalana.
- . 2002. "El sistema vocàlic. Reducció vacàlica". *Gramàtica del català contemporani*, ed. by Joan Solà et alii, vol 1, 89-123. Barcelona: Empúries.
- Pater, Joe. 2004. "Exceptions in Optimality Theory: Typology and Learnability". Paper presented at the conference on *Redefining Elicitation: Novel Data in Phonological Theory*, New York University.
- Pater, Joe. 2005. "Learning a Stratified Grammar". *Proceedings of the 29th Boston University Conference on Language Development*, ed. by Alejna Brugos et alii, 482-492. Somerville, MA: Cascadilla Press.
- Pi-Mallarach, Josep. 1997. *L'obertura de les vocals mitjanes en una mostra de neologismes del català*. PhD dissertation. Universitat Autònoma de Barcelona.
- Recasens, Daniel. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.
- Steriade, Donca. 2001. "Directional asymmetries in place assimilation: a perceptual account". *The role of speech perception in Phonology*, ed. by E. Hume and K. Johnson, 119-150. San Diego: San Diego Academic Press.
- Walker, Rachel. 2005. "Weak triggers in vowel harmony". *Natural language and Linguistic Theory*, 23. 917-989.