External Allomorphy and Lexical Representation

Joan Mascaró

Many cases of allomorphic alternation are restricted to specific lexical items but at the same time show a regular phonological distribution. Standard approaches cannot deal with these cases because they must either resort to diacritic features or list regular phonological contexts as idiosyncratic. These problems can be overcome if we assume that allomorphs are lexically organized as a partially ordered set. If no ordering is established, allomorphic choice is determined by the phonology—in particular, by the emergence of the unmarked (TETU). In other cases, TETU effects are insufficient, and lexical ordering determines the preference for dominant allomorphs.

Keywords: allomorphy, emergence of the unmarked, external allomorphy, lexical representation

1 Introduction

In allomorphic variation in its strict sense, the variation in phonetic shape of a morpheme cannot be traced back to a single underlying form, and two or more underlying forms (the allomorphs) have to be posited. The conditions that determine the choice of one allomorph rather than another can be semantic, syntactic, morphological, or phonological, and in many cases they are highly idiosyncratic. Phonological conditioning appears in two very different forms. One type is irregular and is *internally* conditioned in the lexicon; it is illustrated in (1a) with the Tzeltal perfective suffix *-oh/ɛh*.[1] The other is regular; it is *externally* conditioned by phonology (hence the term *external allomorphy*). It can be illustrated by the Northwestern Catalan masculine singular definite article *lo/l*. In Tzeltal, *-oh* appears after monosyllabic stems and *-ɛh* appears after polysyllabic stems; in Catalan, *l* appears before vowels and *lo* appears before consonants. Allomorphic choice in Catalan is phonologically natural since it determines a less-marked CV syllable structure, whereas allomorphic choice in Tzeltal is arbitrary and does not improve the resulting structure in terms of (un)markedness.

[1] Data from Walsh Dickey 1996. The author argues convincingly that there is no possible natural phonological analysis.

(1) a. *Tzeltal*

      j-il-oh   'he has seen something'

      s-nuts-oh   'he has chased something'

      s-mak'lij-ɛh 'he has listened to something'

      s-tikun-ɛh   'he has sent something'

   b. *Northwestern Catalan*

      l ámo   'the owner'

      l uɾaká   'the hurricane'

      lo pá   'the bread'

      lo míte   'the myth'

Traditionally, allomorphic variation has been accounted for in two ways, both leading to unsatisfactory results, a fact already noted in the literature (see, e.g., Pullum and Zwicky 1988, Spencer 1991:127–129). One analysis is lexical: it assumes that each allomorph m_i is associated with a specific configuration C_i because the configuration is stipulated in the lexical entry (e.g., in English for the lexical item *to be*, the allomorph *am* is lexically listed as 1st sg., *is* is listed as 3rd sg., etc.). In the case of phonological conditioning, each allomorph should be associated with a phonological context. Such theories claim that external allomorphy is always idiosyncratic and that in a case like Northwestern Catalan the targeted CV structures are an accident. The other analysis is rule-based: for a case like Northwestern Catalan, it assumes a rule like $o \rightarrow \emptyset$ / _____ V, or its corresponding constraint-based analysis. But this rule is so restricted that it would apply to just one lexical item, the masculine singular definite article (see section 6 for detailed discussion of these analyses).

The solution to this dilemma rests on three basic ideas. First, lexical listing of contexts should be adopted only in the case of arbitrary, unmotivated instances like those in Tzeltal, where choice is not phonologically natural. Second, the linguistic generalization that the allomorph is chosen because it yields an unmarked structure should be incorporated into grammatical theory, since it rests on an extensive empirical base. In those cases in which allomorphic choice is external to the lexicon, an adequate theory should relegate to lexical listing only those aspects that are idiosyncratic; whatever is regular should be expressed in the grammar. To start with, the specific shape of the allomorph is idiosyncratic, and hence I propose that this is the only aspect that should be lexically specified; in other words, a phonological lexical representation can contain two or more allomorphs, which do not have contexts associated with them. A third idea is that mere lexical listing of allomorphs should be supplemented with partial ordering of the set, reflecting the relative markedness of the allomorphs. Unmarked allomorphs are preferred, although this preference can be violated to satisfy other phonological requirements.

Under standard assumptions, the claim that a lexical representation is a set of (unconditioned) allomorphs—that is, $\{m_1, m_2, \ldots, m_n\}$—would lead to free variation in the phonetic reflections of $m_1, m_2, \ldots, m_n$. A rule-based approach will simply interpret any $/m_i/$ as a phonetic reflex $[m_i']$. But Optimality Theory (OT), where phonological regularities are expressed through universal constraints, can force a choice between competing allomorphs. In the

remainder of this section, I will sketch and illustrate the type of analysis proposed; I will state it more formally in section 2.

When a set of markedness constraints is dominated by a set of conflicting faithfulness constraints, the former are inert. However, they can become active if faithfulness is suspended, as occurs in reduplication (the emergence of the unmarked; McCarthy and Prince 1994).

Faithfulness can also be suspended in allomorphy—in fact, it is a necessary consequence of (unconditioned) allomorphy, given fairly natural assumptions. This can be illustrated with a simple case, *h/u* selection in Moroccan Arabic. (The data are from Harrell 1962, and I follow the analysis in Mascaró 1996b.) The 3rd sg. masculine pronominal clitic is either [h] or [u]. [h] appears after vowels, and [u] appears after consonants. (2a) shows the noncliticized forms, (2b) the distribution of clitic allomorphs.

(2) a. xtˤa   'error'   b. xtˤa-h   'his error'
       ʃafu   'they saw'   ʃafu-h   'they saw him'
       mˤa   'with'   mˤa-h   'with him'
       ktab   'book'   ktab-u   'his book'
       ʃaf   'he saw'   ʃaf-u   'he saw him'
       menn   'from'   menn-u   'from him'
    c. iʃuf   'he sees'
       omm   'mother'
       aʒiw   'come-PL!'
       ka-ixroʒ   'he is going out'

ONSET and NO-CODA must be dominated by MAX and DEP, since the language allows codas and onsetless syllables freely, as shown in (2c). Assume that the underlying form of the clitic is the unordered pair /u, h/. GEN (/u, h/) contains, in addition to many other candidates, the maximally faithful identity maps [u], [h], which will trivially satisfy MAX and DEP. If we assume that ONSET ≫ NO-CODA, [u] or [h] will be chosen depending on how well each attaches to its host in terms of resulting syllable structure markedness (CV > CVC > V > VC). This is illustrated in tableaux (3) and (4).

(3)

/xtˤa-{h, u}/	MAX	DEP	ONSET	NO-CODA
☞ xtˤa-h				*
xtˤa.-u			*!	

(4)

/ktab-{h, u}/	MAX	DEP	ONSET	NO-CODA
ktab-h				*!
☞ kta.b-u				

2 Proposal

In this section, I formalize the theoretical proposal and discuss its main consequences.

Let M be a morpheme whose lexical representation contains allomorphs $/m_1, m_2, \ldots, m_n/$, and let C_F and C_M be conflicting faithfulness and markedness constraints, respectively.

(5) a. The set of allomorphs of a morpheme M $(m_1, m_2, \ldots, m_n)$ can be represented as a partially ordered set.
   b. For M $= /m_1, m_2, \ldots, m_n/$, GEN $(/m_1, m_2, \ldots, m_n/) =$ GEN (m_1) $\cup$ GEN (m_2) $\ldots \cup$ GEN (m_n). (Given a set of allomorphs, the candidate set is the collection of the individual candidate sets of each allomorph.)[2]
   c. Each candidate morph in b. stands in a correspondence relation to one of the underlying allomorphs (i.e., if $cand_i \in$ GEN $(/m_j/)$, then $cand_i \Re m_j$).
   d. Under input allomorphy, candidate faithfulness violations are computed with respect to the candidate's corresponding underlying allomorph.

Ordering in (5a) reflects preference for a given allomorph over its allomorphic competitors. The faithfulness constraint PRIORITY, to be discussed in sections 4–6, will favor dominant allomorphs. GEN will apply to m_1, the first allomorph, and the candidates generated will stand in correspondence to it. To the list of candidates to be evaluated we then add the result of applying GEN to m_2, and so on. Markedness constraints will apply as usual, and faithfulness constraints will ensure faithfulness of a candidate to its corresponding lexical allomorph.

The idea of allomorphic choice being guided by optimization of the output is of course not new. Although it has not received systematic attention, several authors have independently put forward important proposals along these lines. Mester (1994), for instance, in his study of the quantitative trochee in Latin, very convincingly analyzes two cases of allomorphy in terms of avoidance of prosodic trapping. Within OT, I have elsewhere (Mascaró 1996a,b) proposed allomorphic selection as TETU; and Kager (1996) adopts an analysis whereby phonological shape of morphs is introduced by constraints (e.g., GENITIVE = n); similar analyses can be found for French in Perlmutter 1998 and Tranel 1996.[3] The linguistic significance of such cases calls for a careful theoretical formulation, and also for a better determination of its empirical range, since the studies cited all deal with syllable structure and stress, that is, ONSET, NO-CODA, COMPLEX, PK-PROM, AL-R(Stem, Foot). In this article, I show that TETU in allomorphy has an extensive empirical basis, since it affects cliticization, derivation, and inflection[4] and since, moreover, many types of markedness constraints can be shown to be involved in external allomorphy. At the same time, I show that the lexical representation of allomorphs is a partially ordered set and that this

---

[2] This is similar to MI (multiple inputs) in Lapointe 1999:267.

[3] Both Perlmutter and Tranel treat *bel/beau* allomorphy in French as masculine/feminine competition. See Mascaró 1996a and Steriade 1999 for different analyses.

[4] There are indications that it also extends to words in prenominal position, as in French *bel/beau* allomorphy (*bel ami* vs. *beau mari*) and in Ribagorçan Catalan demonstrative selection (*esto llibre* 'this book' vs. *est (h)ome* 'this man'); see Mascaró 1996a,b.

ordering and a faithfulness constraint, PRIORITY, which enforces priority of dominant morphs, together make it possible to predict an important set of additional empirical cases.[5]

3 Exceptional Postnasal Voicing

I will first discuss a case that does not require allomorph ordering and PRIORITY. It is the emergence of postnasal voicing (*NÇ), which affects a wide range of bound forms in Basque. Postnasal voicing is a rather common phonological process (see Hayes and Stivers 1995, Hayes 1999, Pater 1999). Old Basque most probably had it as a general rule (Michelena 1965), as witnessed by early borrowings from Latin: *tempora* > *denbora* 'time', *uoluntate* > *borondate* 'will', *gente* > *jende* 'people', *incude* > *ingude* 'anvil'. Later the process was lost, but it was retained in some bound morphemes in derivation, inflection, and cliticization. The general situation in contemporary Basque, with no postnasal voicing internally, in either compounding, derivation, or inflection, is shown in (6) (contrasting nasal–voiced obstruent sequences are of course also possible, as in *mendi* 'mountain', *lan-gela* 'work-room', *zuen-gan* 'you-PL.INESS', etc.). (7)–(9) show the derivative and inflective morphemes that retain the process. Examples are in standard orthography (*np* is [mp], *nk* is [ŋk]; hyphens have been added to separate morphemes).[6]

(6) a. *Inside morphemes: No *NÇ*
    kanpo   'out'
    kontu   'count'
    Jainko   'God'

   b. *Compounds: No *NÇ*
    zezen-plaza   'bull-square (bull ring)'
    lan-talde   'work-group'
    Gabon-kanta   'Christmas-song'

   c. *Derivatives: No *NÇ*
    ilun   'dark'   ilun-ki   'darkly'
    ilun   'dark'   ilun-tasun   'darkness'
    zuzen   'right'   zuzen-keta,   'correction'
                      zuzen-pen

---

[5] As noted by Paster (to appear), if allomorphic choice is determined by optimization of the output, it should be sensitive to phonological elements in surface forms, not underlying forms. More specifically, in internal allomorphy, changes of underlying form do not affect allomorphic choice. In external allomorphy, the defining context cannot be underlying unless it is also found in the surface form. This is the case with most examples examined here and in the literature. If the defining context is found in the surface form but not in the underlying form, then allomorphic choice through optimization of the output becomes necessary. One such case is *en*/*l* article allomorphy before person names in Catalan (Mascaró 1996b). Here, initial epenthesis and optional gliding feed allomorphic choice, as shown by (underlyingly) consonant-initial *Stevens* and vowel-initial *Hualde*: [1 əstíβəns], *[ən stíβəns]; [1 uálde], [ən wálde], *[ən uálde], *[1 wálde]. Another example is Spanish *o*/*u*, examined at the end of section 3. Here diphthongization bleeds the *u* choice before *o*: *u [o]ler* 'or to sniff', *o [wé]le* (← /óle/) 'or she/he sniffs'.

[6] The process also applies after *l* in some cases, subject to dialectal variation. Extension to *l* seems natural, according to the results in Hayes 1999. I will center my attention on postnasal voicing only. Voicing of the clitic *ta* is more subject to dialectal variation.

Data for standard Basque are from De Rijk 1986, Euskaltzaindia 1990, Hualde 1991, and my own work with native-speaker consultants. For similar data in Bizkaian and in Baztan Basque, see Salaburu 1984 and Hualde, Elordieta, and Elordieta 1994. I am indebted to Pello Salaburu and Gorka Elordieta for empirical assistance.

    d. *Nominal inflection: No *NÇ*
       zein-taz   'which-INDEF.INSTR'
       hon-taz   'this-INSTR'

    e. *Verbal inflection: No *NÇ*
       gin-tuen   'she/he had (us)'
       nen-torren 'I went'

(7) **NÇ active in the derivational suffixes* -tar, -ta *and in the clitic* ta

    a. Eibar   'town name'   eibar-**t**ar   'Eibarese'
       Bilbo   'Bilbao'   bilbo-**t**ar   'Bilbaoese'
       Arizkun 'town name'   arizkun-**d**ar 'Arizkunese'
       Irun   'town name'   Irun-**d**ar   'Irunese'
       etorri   'to come'   etorri-**t**a   'come' (participial adjective)
       jan   'to eat'   jan-**d**a   'eaten' (participial adjective)

    b. i. ardorik   ez du edaten   [d]a
         wine.PART not has drink.PRES since
     ii. ardorik   ez du edango [t]a
         wine.PART not has drink.FUT since
         'since he doesn't drink/will not drink any wine'

(8) **NÇ active in some nominal inflective affixes:* -ko, -tik

		Place genitive	Ablative
Bilbo	'Bilbao'	Bilbo-**k**o	Bilbo-**t**ik
Irun	'town name'	Irun-**g**o	Irun-**d**ik
non	'where'	non-**g**o	non-**d**ik

(9) **NÇ active in some verbal inflective affixes: Participle markers* -ko, -tu[7]

    a. *Perfect*   *Future*
       etorri   etorri-**k**o 'come'
       hil   hil-**k**o   'kill'
       joan   joan-**g**o   'go'
       izan   izan-**g**o   'be'

    b. *Nominal*   *Verbal participle*
       gogor 'hard'   gogor-**t**u 'harden'
       argi   'light'   argi-**t**u   'clear up'
       ilun   'dark'   ilun-**d**u   'darken'
       gizon 'man'   gizon-**d**u 'become man'

The examples in (6) show that IDENT(voice) ≫ *NÇ: there is no general postnasal voicing in

---

[7] *-tu* is best analyzed as an inflectional suffix, not a derivational one. The nonfinite perfective, imperfective, and future participial forms are, for example, *gogor-tu*, *gogor-tzen*, *gogor-tu-ko*, respectively. *-tu* appears also in verbs that lack a derivative source: *sar-tu*, *sar-tzen*, *sar-tu-ko* 'to enter', *sar.

present-day Basque. But there is an exceptional class of lexical elements (*-tar*, *-ta*, *-ko-*GEN, *-tik*, *-ko-*FUT, *-tu*, *-ta*) that do undergo it, as shown in (7)–(9).

Postnasal voicing exhibits the two key features that are typical of external allomorphy: *lexical restriction* and *phonological naturalness*. The fact that it is restricted to a set of lexical elements excludes a general phonological treatment, since this would require lexically marking these morphemes as undergoing a ''minor'' rule or, in OT terms, as being subject to a set of additional indexed constraints. But it also shows phonological naturalness, since postnasal voicing is a natural phonological process. This excludes a totally idiosyncratic treatment, that is, listing the contexts of allomorphic choice in the lexicon, the analysis proposed by Piera (1985) and Hayes (1990) for similar cases. A purely lexical listing analysis (e.g., *dar* / [+nas] _____ , *tar* / elsewhere, etc.) would make equally natural the reverse lexical listing in which, instead of the natural outcomes *eibar-tar*, *arizkun-dar*, the results would be *eibar-dar, *arizkun-tar. The proposal in (5) predicts lexical restriction since listing of allomorphs is a prerogative of lexical items; it also predicts phonological naturalness because the choice of allomorph is not lexically idiosyncratic but an effect of the action of a universal constraint under TETU.

The analysis that derives from (5) holds that there is a special class of elements, *-tar*, *-ta*, *-ko*, *-tik*, *-ko*, *-tu*, *-ta*, that have two underlying allomorphic variants: /tar, dar/, /ta, da/, /ko, go/, /tik, dik/, /ko, go/, /tu, du/, /ta, da/. Their distribution follows from the existence of a universal constraint, *NC̥, that sets the preferences. Consider /tu, du/. The identity map candidates *tu*, *du* are both faithful with respect to IDENT(voice): they tie. Low-ranked *NC̥ resolves the tie by selecting candidates with postnasal voicing whenever the stem ends in a nasal, as shown in (10). When the stem ends in a nonnasal, *NC̥ is satisfied, and the tie is resolved through *VOICEDOBST, which favors the unmarked voiceless stop. Notice that low-ranked constraints must be adequately ordered—here, *NC̥ ≫ *VOICEDOBST, in order to get the right choice between *ilun-tu and *ilun-du*. Following (5c), I introduce correspondence indices for allomorphs and their correspondents in candidates.

(10) a.	/ilun-{tu_1, du_2}/	IDENT(voice)	*NC̥	*VOICEDOBST
	ilun-tu_1		*!	
	☞ ilun-du_2			*
b.	/argi-{tu_1, du_2}/			
	☞ argi-tu_1			
	argi-du_2			*!
c.	/ilun-ki_1/			
	☞ ilun-ki_1		*	
	ilun-gi_1	*!		*

(10c) shows that if there is a single allomorph in the lexical representation, as there is for the adverbial suffix *-ki* (6c), the faithfulness to voicing mandated by IDENT(voice) is not suspended and voicelessness is preserved even after nasals.

The Basque data are important because the lexical elements that induce postnasal voicing are randomly distributed among nominal inflection, verbal inflection, derivation, and cliticization. This makes an analysis in terms of different phonologies applying to different morphological levels infeasible, since it would lead to numerous ordering paradoxes, as shown by the examples in (11). (Root in boldface, voicing suffixes underlined, inflectional suffixes in italics, derivational suffixes in roman, clitic in small capitals. *-pe*, *-keria*, *-kari*, and *-koi(t)z* are derivational suffixes meaning 'under', 'negative quality or act', 'object, matter', 'multiple', respectively; the remaining suffixes have been glossed in (6)–(9).)

(11) **lan**-pe-*tu*-*ko*   'to overburden with work-FUT'
    kanpo-*ko*-tasun   'day school'
    baserri-tar-keria   'farmer's deed'
    egun-kari-*ko*   'newspaper's'
    ehun-koiz-*tu*   'to centuplicate'
    men-pe-*ko*-tasun-DA   'and subordination'

Before we turn to allomorph ordering, it might be worth reviewing the extent of the empirical evidence for external allomorphy and its typological variety. Most of the cases reported are related to syllable structure: Kager 1996 for Djabugay; Mascaró 1996a,b for French, English, Moroccan Arabic, and Catalan (three cases); Perlmutter 1998 and Tranel 1996 for French; Lapointe 1999 for Korean; and Rubach and Booij 2001 for Polish. Some cases are related to foot structure and stress: Kager 1996 for Estonian (two cases); Drachman, Kager, and Malikouti-Drachman 1996 for Modern Greek; Elías-Ulloa 2005 and González 2005 for Capanahua and other Panoan languages. There are also cases of tonal markedness and glottalization (Yip 2004, Zahao) and consonant-to-vowel assimilation (Rubach and Booij 2001, Polish) (see McCarthy 2002:183–184 for further references). In this article, I present detailed evidence for postnasal voicing (Basque, this section), consonantal assimilation (Catalan and Arabic, sections 4 and 5), consonantal Obligatory Contour Principle (OCP) phenomena (Catalan, section 5), and syllable structure again (Tzotzil, section 6).

Two other phenomena can be added to the list, namely, vowel OCP and vowel harmony. In Spanish, the conjunctions *y* 'and' and *o* 'or' have two allomorphs, /i/-/e/ and /o/-/u/, respectively. To avoid [ii] in one case and [uu] in the other, *e* appears before *i*, and *u* before *o*: *ironía y humor* but *humor e ironía* 'humor and irony'*; oraciones o palabras* but *palabras u oraciones* 'words or sentences'. Note that [ii] and [uu] do not otherwise trigger dissimilation: *m*[i i]*ironía* 'my irony', *s*[u u]*mor* 'her humor'. (For a more detailed analysis, see Bonet and Mascaró 2006.) A case triggered by vowel harmony is found in the Arabic 3rd person pronominal suffixes. They appear as *hu* (MASC.SG), *hum* (MASC.PL), *humaa* (DUAL), and *hunna* (FEM.PL) when the preceding vocoid is back, but as *hi*, *him*, *himaa*, and *hinna*, respectively, when they attach to a word ending in [i] or [j] (Wright 1991:101, 253).

(12) kalbu-hu   'his dog (NOM)'
     kalbi-hi   'his dog (GEN)'
     maʕa-hunna   'with them-FEM'
     ʔilaj-hinna   'to them-FEM'

4 Allomorph Ordering: Overassimilation in Baix Empordà Catalan

In the cases presented so far, and in previous OT studies as well, external allomorphy shows a peculiar property. From among the set of allomorph candidates, the constraints that are responsible for allomorphic distribution always single out one as most harmonic. But this need not be the case. Consider, for example, consonantal place assimilation. If AGREE(F) ranks above conflicting IDENT(F), the language will show generalized assimilation. If faithfulness outranks AGREE(F), only TETU effects may show up marginally. But in this case it is possible that some allomorphs, when bound to their hosts, will yield structures that are equally marked with respect to AGREE(F). Thus, for the three allomorphs /-pa/, /-ta/, /-ka/, AGREE(Place) will determine a unique choice after the hosts /sam-/, /san-/, /saŋ-/, and so on. But after vowel-final hosts like /sa-/, /si-/, the choice will remain indeterminate.

This poses a technical and an empirical problem. Empirically, the result that emerges from examining such cases of indeterminacy is that the language favors one of the allomorphs over the rest. In the next sections, I examine such cases and give a solution to the technical problem. The first case I will consider is infinitive overassimilation in one variety of Central Catalan. The Baix Empordà variety (henceforth BEC) exhibits two types of assimilation: the *normal assimilation* of Central Catalan and *overassimilation*, which is restricted to the infinitive marker /ɾ/. In normal assimilation, alveolar stops assimilate totally (alveolar nasals only in place) to a following consonant (13a), but liquids do not (13b) (Mascaró 1978, Wheeler 2005).

(13) *General assimilation*

a. /tn/, /tm/, /tl/, /t + obstruent/   b. /ɾn/, /ɾm/, /ɾl/, /ɾ + obstruent/

to[t], to[n n]adal	'all', 'all Christmas'	pe[r-n]adal	'for Christmas'
ga[t], ga[n n]egre	'cat', 'black cat'	ma[r n]egre	'Black Sea'
fe[t], fe[m m]olt	'done', 'done very'	per[r-m]olt	'for very'
ga[t], ga[m m]ort	'cat', 'dead cat'	ma[r m]ort	'Dead Sea'
to[t], to[l l]o bo	'all', 'all the good'	pe[r-l]o bo	'for the good'
to[t], to[p p]er tu	'all', 'all for you'	pe[r-t]u	'for you'

But the infinitive marker /ɾ/ behaves differently from other liquids, as we can see by putting it next to various pronominal clitics. The pronominal clitics are either vowel-initial (*u*, *i*) or consonant-initial (*mə*, *nə*, *li*, *lə*, *tə*, *sə*, etc.).[8]

In (14), I show the general form of the relevant enclitics (neuter *ho*, locative *hi*, 1st sg. *me*, 2nd sg. *te*, 3rd feminine accusative *la*, *les*, 3rd dative *li*, partitive *ne*, and 3rd reflexive *se*). In

---

[8] Other clitics (3rd /l/, /lzə/, /lzi/, 1st pl. /nzə/, 2nd pl. /wzə/) determine unacceptable clusters that cause *r*-deletion; *r* also deletes when no clitic follows.

(15), the same clitics follow the infinitive form of the verb /pɔz-a-ɾ/ '(to) put' (/a/ = conjugation marker, /ɾ/ = infinitive marker): when the clitic begins with a vowel, there is no change (15a); but when the clitic begins with a consonant, the infinitive /ɾ/ undergoes total assimilation (15b).[9]

(14) pose[m-u]   'let's put it!'   pose[m-i]   'let's put there!'
      fe[z-mə]   'make me!'   fe[s-tə]   'make yourself!'
      fe[z-lə]   'make it-FEM!'   fe[z-ləs]   'make them-FEM!'
      fe[z-li]   'make him/her-DAT!'   fe[z-nə]   'make some!'
      posi[n-sə]   'put-POL.PL yourself!'

(15) a. posa[ɾ-u]   'to put it'
          posa[ɾ-i]   'to put there'
      b. posa[m-mə]   'to put me'   posa[t-tə]   'to put you'
          posa[l-lə]   'to put it-FEM'   posa[l-ləs]   'to put them-FEM'
          posa[l-li]   'to put him/her-DAT' posa[n-nə]   'to put some'
          posa[s-sə]   'to put oneself'[10]

Any other /ɾ/ that is not the infinitive morpheme preceding a consonant does not undergo assimilation, even under the same prosodic conditions, that is, when the consonant cluster is separated by a clitic boundary. Compare *posa*[m-mə], *posa*[t-tə], *posa*[l-lə], *posa*[n-nə] in (15b) with the examples involving the clitic *per* in (13b): *per*[r-m]*olt*, *pe*[r-t]*u*, *pe*[r-l]*o bo*, *pe*[r-n]*adal*.

The regular data in (13) show that faithfulness to consonantal features must be dominated by the markedness constraints responsible for alveolar stop assimilation and nasal assimilation (AGREE/STOP for short), but the faithfulness constraint must dominate the more general constraint that requires total identity of any adjacent consonants (AGREE/CONSONANT): stops assimilate, liquids or any other types of consonant do not (Pons 2004, Wheeler 2005). The regressive character of assimilation is ensured by high-ranked IDENT-ONS(F) (not shown in the following tableaux). This general assimilatory pattern is illustrated in (16), where the phrase [əkɛm mar máj] 'this sea never' shows retention of underlying /ɾm/ and assimilation in /tm/ (/ɾ/ tensing in coda position is an independent process; I assume, for simplicity, that F in IDENT(F) does not include the feature that distinguishes [ɾ] from [r]).[11]

(16) Normal assimilation in Baix Empordà Catalan

/əkɛt mar maj/	AGREE/STOP	IDENT(F)	AGREE/C
əkɛ̀t mar máj	*!		**
☞ əkɛ̀m mar máj		*	*
əkɛ̀t mam máj	*!	*	*
əkɛ̀m mam máj		**!	

---

[9] Majorcan and Minorcan Catalan behave in a way quite similar to Baix Empordà.
[10] Before the clitic *se*, assimilation is usually followed by contraction of the *ss* sequence, in this case yielding *posa*[sə].
[11] If we allow AGREE/C to count disagreeing features, more distant pairs of consonants will incur more violations, but the results will be the same.

Overassimilation again exhibits the two key features of external allomorphy, lexical restriction and phonological naturalness, since it is restricted to the infinitive morpheme and total assimilation is a natural phonological process.

In the case under analysis, the infinitive morph will have the allomorphic variants /ɾ, n, l, t, s/.[12] Candidates identical to (or minimally differing from) these will comply with high-ranked faithfulness constraints, and the allomorph that best satisfies the low-ranked constraint AGREE/C will be chosen. This will be sufficient for all cases in which a pronominal clitic beginning with a consonant follows the infinitive, as illustrated for *posar-la* 'to put it-FEM' in tableau (17).

(17) Exceptional overassimilation in Baix Empordà Catalan. Infinitive = /ɾ, n, l, t, s/ (to be revised)

/pɔza-{ɾ$_1$, n$_2$, l$_3$, t$_4$, s$_5$}-lə/	IDENT(F)	AGREE/C
puzá-r$_1$-lə		*!
puzá-n$_2$-lə		*!
☞ puzá-l$_3$-lə		
puzá-t$_4$-lə		*!
puzá-s$_5$-lə		*!

If /pɔza-ɾ-lə/ were the only input, the nonidentical maps [puzá-n-lə], [puzá-l-lə], [puzá-t-lə], [puzá-s-lə] would all violate IDENT(F). But in (17) they are not GEN-modified versions of /pɔza-ɾ-lə/: each is in correspondence with one of the different inputs derived from the existence of allomorphy in the infinitive morpheme.

Nonassimilatory environments pose a problem, though—the problem of indeterminacy mentioned at the beginning of this section. Consider the cases in which the clitic starts with a vowel, as in (15a) *posa*[ɾ-u], *posa*[ɾ-i]. In the first case, the input is /pɔza-{ɾ, n, l, t, s}-u. AGREE/ C is vacuously satisfied by [puzá-ɾ-u], [puzá-n-u], [puzá-l-u], [puzá-t-u], [puzá-s-u], and we predict that there will be a multiple output or that, more probably, the candidate with the least marked consonant, [puzá-t-u], will be chosen. This indicates that the allomorph /ɾ/ has some priority over the rest of the allomorphs: it is the lexically unmarked allomorph. A morpheme should therefore not be considered an unordered set, but a partially ordered set of allomorphs, as stated in (5a). Outputs want to be faithful to this ordering by preferring the dominant, unmarked allomorph(s), all other things being equal. The faithfulness constraint PRIORITY will ensure, ranking permitting, the choice of the unmarked/dominant allomorph.

---

[12] It is unnecessary to include /m/ because—since AGREE/STOP dominates IDENT(F)—/n/ will yield [m] in *posa*[m-mə].

(18) *PRIORITY*

Respect lexical priority (ordering) of allomorphs.

Given an input containing allomorphs $m_1, m_2, \ldots, m_n$, and a candidate m_i', where m_i' is in correspondence with m_i, PRIORITY assigns as many violation marks as the depth of ordering between m_i and the highest dominating morph(s).[13]

Thus, if the set of allomorphs is $\{m_1 > m_2 > m_3\}$, candidate m_1' will obey PRIORITY, m_2' will incur one violation, and m_3' will incur two violations. In the present case, illustrated in (19a), we only need to assume that /ɾ/ dominates the rest (domination is represented with ''>''). We have an unmarked allomorph /ɾ/ and a set of (equally) marked allomorphs /n/, /l/, /t/, /s/. In (19b), I have repeated the input and two candidates from (17) for comparison; they show that for overassimilation cases, low-ranked PRIORITY has no effect.[14]

(19) Exceptional overassimilation in Baix Empordà Catalan. Infinitive = /ɾ > (n, l, t, s)/

a.	/pɔzá-{ɾ₁ > (n₂, l₃, t₄, s₅)}-u/	IDENT(F)	AGREE/C	PRIORITY
☞	puzá-ɾ₁-u			
	puzá-n₂-u			*!
	puzá-l₃-u			*!
	puzá-t₄-u			*!
	puzá-s₅-u			*!
b.	/pɔzá-{ɾ₁ > (n₂, l₃, t₄, s₅)}-lə/	IDENT(F)	AGREE/C	PRIORITY
	puzá-ɾ₁-lə		*!	
☞	puzá-l₃-lə			*

5 Arabic Solar Overassimilation

BEC infinitive overassimilation is not an isolated case. A well-known phenomenon, Arabic solar assimilation, comes quite close to it. In Modern Standard Arabic, the liquid /l/ maintains its features when followed by any other consonant. Only in the case of the so-called definite article

---

[13] When there are more than two ordered allomorphs (see section 5 for an example), PRIORITY appears to apply gradiently. This would run against McCarthy's (2003) proposal that constraints are categorical. Whether PRIORITY is gradient in McCarthy's sense depends on the definition of *locus* for faithfulness constraints, which is unclear. Assume that dominance relations in lexical representations are taken pairwise; for example, the lexical representation /Ø > u > ə/ in section 5 defines the separate dominance relations Ø > u, u > ə, and u > ə. If PRIORITY assigns a violation mark for each dominance relation that is not satisfied, then the constraint would apply categorically, under a reasonable extension of the definition of *locus* to constraints relating two representations.

[14] An obvious alternative to PRIORITY is allowing constraints to introduce morphs (INF = ɾ, etc.). This is an extremely powerful mechanism and, as Bonet (2004) notes, it has the unacceptable consequence of excluding allophonic variation of the morph, since it forces the phonetic form expressed in the constraint.

/l/, which is prefixed to a nominal, do we find total assimilation to a set of consonants, traditionally called solar consonants. The complement set of lunar consonants does not cause assimilation (the terms *solar* and *lunar* are in honor of *ʃams* 'sun', which begins with an assimilator consonant, and *qamar* 'moon', which begins with a neutral one). The situation is the same in Classical Arabic and many modern colloquial varieties. (20) shows nouns in isolation in the (bare) indefinite form and in the article-prefixed definite form; [ʔa] is epenthetic.

(20) *Arabic solar assimilation*

No assimilation before lunar			Solar assimilation		
Indefinite	Definite		Indefinite	Definite	
qamar	ʔal-qamar	'moon'	ʃams	ʔaʃ-ʃams	'sun'
ʤaziira	ʔal-ʤaziira	'island'	tˤaalib	ʔatˤ-tˤaalib	'teacher'
madiina	ʔal-madiina	'town'	raʔiis	ʔar-raʔiis	'chief'
jasiir	ʔal-jasiir	'easy'	zaʕfaraan	ʔaz-zaʕfaraan	'saffron'
ɣarb	ʔal-ɣarb	'Occident'	safariiya	ʔas-safariiya	'trip'

Solar assimilation has the same general properties as Catalan overassimilation. First, it is restricted to a single lexical morpheme (the definite article), and it does not apply outside this restricted domain. Thus, *l* does not assimilate to a following solar consonant in any of the many other prosodic contexts where this occurs: morpheme-internally (*maltˤa* 'Malta' *tˤtˤ*); when the contact between *l* and the solar consonant is produced by nonconcatenative inflection or derivation (*lisaan* 'tongue', *ʔalsina* *ss* 'tongue-(broken)PL', *lazima* 'adhere-3SG.MASC.PERF', *jalzamu* *zz* 'adhere-3SG.MASC.IMPERF'; in the derived form IV *ʔalzama* *zz* 'force-3SG.MASC.PERF'); in inflectional suffixation (*qatal-at* 'killed-3SG.FEM.PERF', *qatal-naa* *nn* 'killed-1PL.PERF'); in cliticization (*idʒʕal* 'put-IMP', *idʒʕal-nii* *nn* 'put-IMP-me'); or in phrasal phonology (*hal ʔanta* 'whether you', *hal ʃafajtu* *ʃʃ* 'whether I have cured').

Again, in an analysis based on a single underlying form /l/, we would need a specific phonology for a single morpheme, the definiteness marker. The second property of this process is that it only causes total assimilation to a set of consonants, namely, *t*, *θ*, *d*, *ð*, *r*, *l*, *n*, *z*, *s*, *sˤ*, *dˤ*, *tˤ*, *ðˤ*, and *ʃ*. The analysis is thus parallel to that of overassimilation in BEC. A markedness constraint, AGREE/C, is ranked below IDENT/C and therefore does not determine assimilation generally. The definiteness marker has as allomorphs the set of solar consonants, but one allomorph—/l/, the unmarked choice—is ordered before the rest and is the preferred choice in those cases in which AGREE/C has not set any preferences. Otherwise, AGREE/C chooses the allomorph that agrees with the nominal-initial consonant.[15]

---

[15] In cases of proliferation of underlying allomorphs, as in BEC and Arabic overassimilation, one might consider analyzing them as just two allomorphs: the fully specified dominating allomorph and an underspecified allomorph (e.g., /l > [+ cons]/ for BEC). Leaving aside problems regarding the status of underspecification in OT, this solution is impossible for Arabic, because the class of assimilatory triggers {t, θ, d, ð, r, l, n, z, s, sˤ, dˤ, tˤ, ðˤ, ʃ} is formed by coronals, but it does not contain the coronal ʤ and hence is not a natural class. See section 6 for discussion.

As noted by one reviewer, ordering of allomorphs as stated in (5a) allows partial ordering, but the two cases we have examined so far order just one allomorph before the rest. A more restrictive model would in fact allow only one allomorph to dominate the rest. But, as in other cases of markedness, even if this situation is common, depth of ordering can be greater than one. A totally ordered set of three allomorphs is found in Catalan gender allomorphy, analyzed in detail in Bonet, Lloret, and Mascaró 2007. The authors give independent justification for the relative markedness of the masculine gender marker allomorphs /∅ > u > ə/ in Catalan: for example, *gɔ́t-∅*, *gɔ́t-∅-s* (least marked) 'glass, glasses'; *áwt-u*, *áwt-u-s* (intermediate) 'car, cars'; *pár-ə*, *pár-ə-s* (most marked) 'father, fathers'. Sequences of sibilants cause an OCP effect, for which the usual repair strategies are deletion and ə-epenthesis. But in masculine plural nominals whose root ends in a sibilant, like /gos-s/, a gender vowel is preferred as a repair: ∅, the best allomorphic choice, is rejected because it fails to repair the *s-s* sequence. A gender vowel is preferred over deletion (21b) or ə-epenthesis (21e), because of MAX-C and AL-MM (''Align the right edge of morpheme X with the left edge of morpheme Y''), respectively. Since the OCP violation forces the choice of a vowel allomorph, either [u] or [ə] must be chosen. PRIORITY forces the choice of the first, less-marked option. Thus, in the singular *gós* 'dog', the dominant allomorph, ∅, wins because there is no OCP violation; but in the plural *gós-u-s* 'dogs' (*gós-∅-s*, *gós-ə-s*), the OCP-offending configuration *s-s* triggers the choice of the [u] allomorph.

(21)

/gos-{∅$_1$ > u$_2$ > ə$_3$}/-s	OCP-SIB	AL-MM	MAX-C	PRIORITY
a. gos-∅$_1$-s	*!			
b. gos-∅$_1$			*!	
c. ☞ gos-u$_2$-s				*
d. gos-ə$_3$-s				**!
e. gos-∅$_1$ə-s		*!		

6 Factorial Typology and Alternative Analyses

In this section, I examine the possible typological situations derived from the position of PRIORITY in the hierarchy, and I compare the present proposal with alternative analyses. Given a set of (possibly ordered) allomorphs, the constraint PRIORITY, and the set of relevant faithfulness and markedness constraints, the following ordering situations obtain:

(22) *Factorial typology*
   a. PRIORITY ≫ MARKEDNESS. Under this hypothetical ranking, the first-ordered allomorph would always be chosen; hence, no allomorphy will obtain.[16]

---

[16] A special case of PRIORITY not dominated by markedness is Bonet's (2004) analysis of ergative allomorph selection in Dyirbal.

b. FAITHFULNESS ≫ MARKEDNESS ≫ . . . ≫ PRIORITY. If PRIORITY is at the end of the hierarchy, it will be totally inactive, and ordering in the allomorph set makes no sense. (Moroccan Arabic, Basque (sections 1, 3), in which PRIORITY is not involved.)

c. FAITHFULNESS ≫ MARKEDNESS-1 ≫ PRIORITY ≫ MARKEDNESS-2. Selection is carried out first by MARKEDNESS-1, but undecided (tied) candidates are not further subject to MARKEDNESS-2; instead, PRIORITY selects the dominating, unmarked allomorph. (Catalan infinitive, Arabic solar overassimilations (sections 4, 5).)

d. MARKEDNESS ≫ FAITHFULNESS, PRIORITY. Dominating MARKEDNESS is now active (no TETU effects) and bans one allomorphic option in one context because it is ill formed in the language. It does not apply in other contexts, where PRIORITY decides.

Instances of (22d) have not been reported so far. I will deal briefly with two such cases, beginning with an allomorphy example presented by Kager (1996). In Djabugay, the genitive marker is *-n/-ŋun*, with *-n* appearing after vowels and *-ŋun* after consonants.

(23) a. gaŋal   gaŋal-ŋun   *gaŋal-n   'goanna lizard'
     b. guludu   *guludu-ŋun   guludu-n   'dove'

Complex codas are not tolerated; hence, COMPLEX ≫ MAX, DEP. *gaŋal-ŋun* (23a) is thus preferred because *gaŋal-n*, with an *ln* coda, is impossible. Both *guludu-ŋun* and *guludu-n* in (23b) are fine in terms of syllable structure, hence the need for PRIORITY to set the preference for the *-n* allomorph. But note that in the first case, (23a), the losing allomorph could set up a repair strategy, such as /gaŋal-n/ → [gaŋal-na], to circumvent COMPLEX. Under such circumstances, however, low-ranked, usually inactive MAX, DEP will become active, giving rise to an ''emergence of the faithful'' effect, as illustrated in (24): both *gaŋal-na* and *gaŋal-ŋun* obey COMPLEX, but the more faithful, nonepenthesized candidate is chosen.

(24) /gaŋal-{n$_1$>ŋun$_2$}/	COMPLEX	MAX, DEP	PRIORITY
ga.ŋal-n$_1$	*!		
ga.ŋal.-n$_1$a		*!	
☞ ga.ŋal.-ŋun$_2$			*
/guludu-{n$_1$>ŋun$_2$}/	COMPLEX	MAX, DEP	PRIORITY
☞ gu.lu.du-n$_1$			
gu.lu.du.-ŋun$_2$			*!

A parallel case is the 3rd person possessed form in Tzotzil (Haviland 1981), which has two allomorphs, *s-* and *j-*. They are prefixed to nominal roots, which are always consonant-initial; initial [ʔ], however, regularly deletes after prefixes.[17]

---

[17] A similar situation obtains in Tzeltal. See Smith n.d.:secs. 3.2.3, 3.2.6.

(25) a. na   'house'   s-na   *j-na   'his/her house'
      b. ʔosil   'land'   *s-osil   j-osil   'his/her land'

In (25a), a glide-initial j+C onset, as in *j-na, is impossible, but sibilant+C, as in s-na, is permitted; hence, the choice of s-na is dictated by the high-ranking markedness constraint prohibiting a j+C onset. However, in (25b) this constraint does not apply and both *s-osil and j-osil would be possible. Thus, if the lexical representation of the possessed form is /j > s/, PRIORITY will choose j-osil, but it will be rendered inactive in the case of s-na/*j-na by the higher-ranked markedness constraint.

Let us now consider possible alternatives to the theory based on allomorphs as a (possibly ordered) set plus PRIORITY. The obvious choice is exceptional marking through indexed constraints—in particular, constraints indexed for individual lexical items (the parallel to diacritic marks and minor rules in derivational models). In this model, a lexical element /L/ is idiosyncratically marked, $/L_D/$, and a low-ranked constraint C is "replicated" as C_D (it will apply only to candidates marked with D) and ranked higher in the hierarchy (Itô and Mester 1999, Pater 2000, 2004).[18] Since allowing free indexing and free reordering of indexed constraints is equivalent to allowing any lexical item or random class of lexical items to choose the phonology of any natural language, indexed constraints must be severely restricted. Some proposals, like restricting marking to natural classes (noun, verb, lexical strata like native and nonnative, etc.), cannot apply to external allomorphy, which involves an individual property of specific lexical elements, not a property of a large, natural class. Another proposal has been to restrict indexed constraints to faithfulness constraints, so as to take care of negative exceptions (underapplication) (see Itô and Mester 1999). Note that underapplication can be treated in a simple fashion in OT, but overapplication cannot. A high-ranking indexed faithfulness constraint F_D dictates faithfulness to the input and can successfully eliminate the set of excluded, "underapplied" candidates. On the other hand, a markedness constraint can prohibit a given structure but cannot dictate by itself what the right repair should be; hence, a much larger set of markedness constraints must be replicated, added to the hierarchy, and ordered accordingly. Thus, a language with postnasal voicing (*NC̥ ≫ IDENT(voice)) can mark lexical exceptions to the process with the diacritic D, for example, $/anta_D/$. The hierarchy IDENT(voice)$_D$ ≫ *NC̥ ≫ IDENT(voice) will render *NC̥ inactive for $/anta_D/$ and ensure the mapping $/anta_D/ \rightarrow$ [anta]. But for overapplication, as in Basque, a similar move—now replicating the markedness constraint—is normally insufficient. *NC̥$_D$ ≫ IDENT(voice) ≫ *NC̥ ensures that /ilun-ki/ fails to become [iʎuŋ-ki], but does not dictate the actual repair strategy (voicing vs. nasalization, epenthesis, deletion, etc.). Hence, in most cases several markedness constraints have to be replicated. The basic reason for preferring the present proposal to indexed constraints is therefore restrictiveness. Both mechanisms add complexity to the lexicon (more than one allomorph and diacritic marking, respectively), but in the present proposal the universal set of constraints is enriched with a single general constraint, PRIORITY,

---

[18] Other proposals, like indexing lexical items for specific constraint orderings, are equivalent for present purposes. See Pater 2004 for a good review of alternatives.

for all cases, whereas the alternative approach requires one or several indexed markedness constraints for each case of allomorphy.[19]

In cases like Basque and BEC, restrictiveness seems to be the only difference between analyses, but in other cases there are additional reasons to prefer the allomorphic analysis. Many cases of external allomorphy can be viewed as an intermediate stage in the historical process of lexicalization of phonological processes. Whenever the process is restricted to a natural prosodic context or a natural lexical class, a regular phonological analysis through indexed constraints might be justified, as in postnasal voicing in Japanese, which is restricted to the Yamato (native) vocabulary (Itô and Mester 1999). But sometimes it becomes restricted to a lexical item or a short random class of lexical items. Given the restrictiveness argument just presented, at this point an allomorphic analysis begins to be justified. At an early stage, phonological similarity among allomorphs, which is often found in external allomorphy, is a reflex of the origin of the phenomenon; at later stages, phonological similarity tends to diminish gradually.[20] In some cases there is further lexicalization, and the phonological process can become fully lexicalized: the alternations and/or the distributional context become unnatural and have to be idiosyncratically listed in the lexicon, as with the Tzeltal example in (1a). Many similar examples can be cited, like 2nd sg. present indefinite in Hungarian (Carstairs 1987) and plural in Armenian (Vaux 2003) (see other cases in Carstairs 1988, Lapointe 1999). In many of the attested cases of external allomorphy, the historical changes have been drastic enough to make the phonological similarity among allomorphs scarce; hence, a natural regular phonological analysis (i.e., using indexed markedness constraints) becomes impossible. This is the case of Moroccan Arabic *h/u* < Classical Arabic *hu* (see (2)), Modern Greek *imo/ma* and *ino/énjo* (Drachman, Kager, and Malikouti-Drachman 1996), Catalan *o/w* (Mascaró 1996a), and Tzeltal *h/k*, *s/y* (Maxwell 1998). And there are even cases in which the allomorphy does not originate historically from a single source. In Catalan, Latin *illum* and *dominum* have given rise to two different articles that have supplied allomorphs for the masculine personal article: [l] before a vowel and [ən] before a consonant (footnote 5, Mascaró 1996b). Another case that illustrates the effect of change is definite article allomorphy in the Pollença variety of Majorcan Catalan. The relative closeness of the historical allomorphs *lo* / _____ C and *l* / _____ V (the same situation as illustrated in (1b) for another variety of Catalan) has been blurred by successive phonological changes, *lo* > *l* > *əl* > *əw* > *u*, to give the actual distribution *u* / _____ C, *l* / _____ V: *u pá* 'the bread', *l ɔ́mo* 'the man'. A theory based on lexical allomorphs does not preclude phonological similarity among them, but it predicts that the opposite—phonological dissimilarity—will also arise, as all these examples show.

---

[19] As one reviewer notes, the theory predicts that, given two external allomorphic processes, there should be no contradictory ranking of PRIORITY with respect to other constraints needed. The cases of ordered allomorphs that I am familiar with rank PRIORITY with respect to different constraints and offer no evidence either way. For a similar position with respect to diacritic marking, see Kager, to appear.

[20] Phonological similarity among allomorphs is a general property of all kinds of allomorphy at first stages of lexicalization, not a peculiarity of external allomorphy. See, for instance, the internal (i.e., lexically irregular) allomorphs of Italian irregular verbs like *spegnere* 'to extinguish, turn off' /spɛɲ/, /spɛns/, /spɛn/, /spɛŋg/, or *dipingere* 'to paint' /dipiŋg/, /dipinʤ/, /dipins/, /dipin/.

In other cases, a stronger argument can be derived from the fact that the theory predicts that, given that there is a lexical list of allomorphs, changes in the list are possible and will lead to unnatural classes without affecting the status of the constraint set. Let us examine one such case in some detail. In Arabic solar assimilation, analyzed in section 5, if we eliminate one consonant from the list of allomorphs, we predict nonassimilation to that consonant, which is indeed what has happened historically. Coronal consonants are {t, θ, d, ð, r, l, n, z, ʤ, s, sˤ, dˤ, tˤ, ðˤ, ʃ}. But there is one in the set that does not cause assimilation, namely, ʤ: we find overassimilation of *l* to ʃ in /l-ʃams/ → [ʔaʃ-ʃams], and we would expect overassimilation of *l* to ʤ in /1-ʤaziira/ → *[ʔaʤ-ʤaziira], but the actual form is [ʔal-ʤaziira]. Since we can restrict assimilatory extension by not listing a /ʤ/ allomorph, this outcome is predictable. At the same time, we cannot find just any possible alternation, because we are restricted to (assimilatory) TETU effects. The origin of this state of affairs in Arabic is of course language change (Fleisch 1961, Moscati et al. 1964). The original Proto-Semitic system contained both noncoronal nontriggers /ɣ/ and /g/. When solar assimilation was established through a set of lexical allomorphs, this set did not contain /ɣ/ or /g/, and of course did not contain /ʤ/ either, since this consonant did not exist at that stage. Later, the sound shift /g/ > /gʲ/ > /ʤ/ took place; since there is no reference, given our analysis, to the class [+cor] in solar assimilation, and /ʤ/ was not in the set of lexical allomorphs, lack of assimilation before [ʤ] is correctly predicted. Note also that there is no problem with the sequence [(ʔa)ʤʤ], which is indeed very common: for example, [ʔaʤʤara] 'he hired out'.

Of course, a rule-based solution is possible, given a sufficiently abstract, *SPE*-type theory (Chomsky and Halle 1968) that mirrors the historical change. Assume that phonetic [ʤ] is /g/ and that there is a minor rule restricted to the article, assimilating /l/ to coronals. /l/ will fail to assimilate to [g], and a later, context-free rule will turn [g] into [ʤ]. This is in fact the proposal made in Brame 1970, 1972. But such extremely abstract analyses have been abandoned since the 1980s and are unavailable in OT. In some varieties of Arabic, including Maltese, the abstract approach fares even worse, because, as Brame himself notes, /g/ has been reintroduced into the inventory, and contrary to prediction, it does not parallel the fronting /g/: il-[ʤ]urnata 'the day', il-[g]werer 'the wars'. The same happens in the Sanʿaani variety of Yemen, where Classical Arabic uvulars have been lost, /q/ becoming /g/. Here again, the article does not assimilate to ʤ or to g: *algurʔaan* 'the Koran', *alʤajʃ* 'the army' (Watson 2002:216–225).

Under an indexed-constraint analysis, this forces an unnatural modification of the new replicated constraint. Since the set is not definable as a natural class, it must be listed (AGREE-t, θ, d, ð, r, 1, n, z, s, sˤ, dˤ, tˤ, ðˤ_D). Listing is unavoidable, but the lexicon is clearly the appropriate place to put a list, rather than the set of constraints.

7 Summary and Conclusion

A fairly large set of cases of allomorphy present an analytical paradox because they involve an idiosyncratic and a phonologically regular aspect at the same time. If they are treated lexically, a significant generalization is lost. If they are treated solely on the basis of regular phonology, phonological processes have to be encoded through diacritic features in lexical items. I have proposed a solution to this dilemma based on a minimal complication in the way certain bound

forms are represented (they consist of a set of partially ordered allomorphs) and the addition of the constraint PRIORITY, which favors allomorphs that are dominant in the lexical ordering. An important property of external allomorphy that deserves further research is that it seems to be restricted to bound forms, including clitics (but see footnote 4). The theory predicts emergence-of-the-unmarked (TETU) effects: once the ''regular'' phonology has selected a subset of the candidates, remaining allomorphic candidates will be selected according to the less-marked structure dictated by low-ranked constraints. In cases of underdetermination, PRIORITY will favor the dominant candidate (or candidates), the one that corresponds (or the ones that correspond) to higher-ranked allomorphs in the lexical representation. Active, higher-ranked markedness constraints can also indirectly determine allomorphic choice; in these cases, normally inert faithfulness becomes active, giving rise to an emergence-of-the-faithful effect. The proposal is supported by the fact that it affects inflection, derivation, and cliticization, and the fact that a wide range of constraint families is involved in the cases analyzed and cited.

References

Bonet, Eulàlia. 2004. Morph insertion and allomorphy in Optimality Theory. *International Journal of English Studies* 4(2):73–104.

Bonet, Eulàlia, Maria-Rosa Lloret, and Joan Mascaró. 2007. Lexical specifications and ordering of allomorphs: Two case studies. *Lingua* 117:903–927.

Bonet, Eulàlia, and Joan Mascaró. 2006. *U* u *o* e *y* o *e*. In *Cuadernos de Lingüística XIII 2006*, ed. by Antonio Fábregas, Natalia Curto, and José María Lahoz, 1–8. Madrid: Instituto Universitario Ortega y Gasset.

Brame, Michael K. 1970. Arabic phonology: Implications for phonological theory and historical Semitic. Doctoral dissertation, MIT, Cambridge, Mass.

Brame, Michael K. 1972. On the abstractness of phonology: Maltese ʕ. In *Contributions to generative phonology*, ed. by Michael K. Brame, 22–61. Austin: University of Texas Press.

Carstairs, Andrew. 1987. *Allomorphy in inflection*. London: Croom Helm.

Carstairs, Andrew. 1988. Some implications of phonologically conditioned suppletion. In *Yearbook of morphology 1988*, ed. by Geert Booij and Jaap van Marle, 67–94. Dordrecht: Foris. Reprinted in *Phonology: Critical concepts*, ed. by Charles Kreidler, 5:111–139. London: Routledge (2000).

Caspari, Carl Paul. 1859. *Grammatik der arabischen Sprache für akademische Vorlesungen*. Leipzig: C. L. Fritzsche.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Drachman, Gaberell, René Kager, and Angeliki Malikouti-Drachman. 1996. Greek allomorphy: An optimality account. In *Papers from the First Conference on Formal Approaches to South Slavic Languages*, ed. by Mila Dimitrova-Vulchanova and Lars Hellan, 345–361. University of Trondheim Working Papers in Linguistics 28. Trondheim: University of Trondheim, Department of Linguistics.

Elías-Ulloa, José. 2005. Variable syllable weight and quantity-insensitive allomorphy in Shipibo. In *NELS 35*, ed. by Leah Bateman and Cherlon Ussery, 171–186. Amherst: University of Massachusetts, GLSA.

Euskaltzaindia. 1990. *Euskal gramatika: Lehen urratsak*. Bilbao: Euskaltzaindia [Academy of the Basque Language].

Fleisch, Henri. 1961. *Traité de philologie arabe*. Beirut: Imprimerie Catholique.

González, Carolina. 2005. Phonologically-conditioned allomorphy in Panoan: Towards an analysis. In *Papers in phonology 6*, ed. by Jeffrey Heinz, Andrew Martin, and Katya Pertsova, 39–56. UCLA Working Papers in Linguistics 11. Los Angeles: UCLA, Department of Linguistics.

Harrell, Richard S. 1962. *A short reference grammar of Moroccan Arabic*. Washington, D.C.: Georgetown University Press.

Haviland, John B. 1981. *Sk'op Sotz'leb: El Tzotzil de San Lorenzo Zinacantán*. Mexico City: Universidad Nacional Autónoma de México. [English version available at http://www.zapata.org/Tzotzil.]

Hayes, Bruce. 1990. Precompiled phrasal phonology. In *The phonology-syntax connection*, ed. by Sharon Inkelas and Draga Zec, 85–108. Chicago: University of Chicago Press.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In *Functionalism and formalism in linguistics*. Vol. 1, *General papers*, ed. by Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick J. Newmeyer, and Kathleen M. Wheatley, 243–285. Amsterdam: John Benjamins.

Hayes, Bruce, and Tania Stivers. 1995. The phonetics of postnasal voicing. Ms., UCLA, Los Angeles, Calif.

Hualde, José Ignacio. 1991. *Basque phonology*. London: Routledge.

Hualde, José Ignacio, Gorka Elordieta, and Arantzazu Elordieta. 1994. *The Basque dialect of Lequeitio*. Bilbao: Euskal Herriko Unibertsitatea, and Donostia: Gipuzkoako Foru Aldundia.

Itô, Junko, and Armin Mester. 1999. The structure of the phonological lexicon. In *The handbook of Japanese linguistics*, ed. by Natsuko Tsujimura, 62–100. Oxford: Blackwell.

Kager, René. 1996. On affix allomorphy and syllable counting. In *Interfaces in phonology*, ed. by Ursula Kleinhenz, 155–171. Berlin: Akademie Verlag.

Kager, René. To appear. Lexical irregularity and the typology of contrast. In *The nature of the word: Essays in honor of Paul Kiparsky*, ed. by Kristin Hanson and Sharon Inkelas. Cambridge, Mass.: MIT Press.

Lapointe, Steven. 1999. Stem selection and OT. In *Yearbook of morphology 1999*, ed. by Geert Booij and Jaap van Marle, 263–297. Dordrecht: Kluwer.

Mascaró, Joan. 1978. *Catalan phonology and the phonological cycle*. Bloomington: Indiana University Linguistics Club.

Mascaró, Joan. 1996a. External allomorphy and contractions in Romance. *Probus* 8:181–205.

Mascaró, Joan. 1996b. External allomorphy as emergence of the unmarked. In *Current trends in phonology: Models and methods*, ed. by Jacques Durand and Bernard Laks, 2:473–483. Salford, Manchester: University of Salford, European Studies Research Institute. Reprinted in *Optimality Theory in phonology: A reader*, ed. by John McCarthy, 513–522. Oxford: Blackwell (2004).

Maxwell, Michael B. 1998. Two theories of morphology, one implementation. In *SIL electronic working papers 1998-001*. Available at http://www.sil.org/silewp/1998/001.

McCarthy, John. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John. 2003. OT constraints are categorical. *Phonology* 20:75–138.

McCarthy, John, and Alan Prince. 1994. The emergence of the unmarked: Optimality in Prosodic Morphology. In *NELS 24*, ed. by Mercè Gonzàlez, 333–379. Amherst: University of Massachusetts, GLSA.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12:1–61.

Michelena, Luis. 1965. *Fonética histórica vasca*. 3rd ed. San Sebastián: Publicaciones del Seminario Julio de Urquijo.

Moscati, Sabatino, Anton Spitaler, Edward Ullendorff, and Wolfram von Soden. 1964. *An introduction to the comparative grammar of the Semitic languages: Phonology and morphology*. Wiesbaden: Otto Harrassowitz.

Paster, Mary. To appear. Phonologically conditioned suppletive allomorphy: Cross-linguistic results and theoretical consequences. In *Understanding allomorphy: Perspectives from Optimality Theory*, ed. by Bernard Tranel. London: Equinox.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In *The prosody-morphology interface*, ed. by René Kager, Harry van der Hulst, and Wim Zonneveld, 310–343. Cambridge: Cambridge University Press. Edited version in *Optimality Theory in phonology: A reader*, ed. by John McCarthy, 271–289. Oxford: Blackwell (2004).

Pater, Joe. 2000. Nonuniformity in English stress: The role of ranked and lexically specific constraints. *Phonology* 17:237–274.

Pater, Joe. 2004. Exceptions in Optimality Theory: Typology and learnability. Paper presented at ''Redefining Elicitation: Novel Data in Phonological Theory,'' New York University, 9–11 April. Handout available at http://people.umass.edu/pater.

Perlmutter, David. 1998. Interfaces: Explanations of allomorphy and the architecture of grammars. In *Morphology and its relation to phonology and syntax*, ed. by Steven Lapointe, Diane Brentari, and Patrick M. Farrell, 307–338. Stanford, Calif.: CSLI Publications.

Piera, Carlos. 1985. On the representation of higher order complex words. In *Selected papers from the XIIIth Linguistic Symposium on Romance Languages*, ed. by Larry D. King and Catherine A. Male, 287–313. Amsterdam: John Benjamins.

Pons, Clàudia. 2004. Els contactes consonàntics en balear: Descripció i anàlisi. Doctoral dissertation, Universitat de Barcelona.

Pullum, Geoffrey K., and Arnold M. Zwicky. 1988. The syntax-phonology interface. In *Linguistics: The Cambridge survey*. Vol. 1, *Linguistic Theory: Foundations*, ed. by Frederick J. Newmeyer, 255–280. Cambridge: Cambridge University Press.

Rijk, Rudolf de. 1986. Euskara batua (Hoofdstuk 1–29). Ms., Universiteit van Leiden. Expanded English version, *Standard Basque*, is to appear, Cambridge, Mass.: MIT Press.

Rubach, Jerzy, and Geert E. Booij. 2001. Allomorphy in Optimality Theory: Polish iotation. *Language* 77: 26–60.

Salaburu, Pello. 1984. *Arau fonologikoak*. Bilbao: Argitarapen Zerbitzua EHU.

Smith, Joshua H. n.d. Manual de tzeltal de Manuel o El tzeltal como quien dice. Ms., El Instituto Lingüístico del Infierno, Mexico, and The Center for Concrete Linguistics, Santa Cruz, Calif. Available at http://www.tzeltal.org. The version cited is the English translation by Stuart P. Robinson, Manual of spoken Tzeltal, http://www.tzeltal.org.

Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Blackwell.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In *Formal perspectives in Romance linguistics: Selected papers from the 28th Linguistic Symposium on Romance Languages (LSRL XXVIII)*, ed. by Jean-Marc Authier, Barbara E. Bullock, and Lisa A. Reed, 243–270. Amsterdam: John Benjamins.

Tranel, Bernard. 1996. French liaison and elision revisited: A unified account within Optimality Theory. In *Aspects of Romance linguistics*, ed. by Claudia Parodi, Carlos Quicoli, Mario Saltarelli, and Maria Luisa Zubizarreta, 433–455. Washington, D.C.: Georgetown University Press.

Vaux, Bert. 2003. Syllabification in Armenian, Universal Grammar, and the lexicon. *Linguistic Inquiry* 34: 91–125.

Walsh Dickey, Laura. 1996. Syllable count and Tzeltal segmental allomorphy. In *Phonologica 1996: Syllables!?*, ed. by John R. Rennison and Klaus Kühnhammer, 323–334. The Hague: Holland Academic Graphics.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. New York: Oxford University Press.

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

Wright, William. 1991. *A grammar of the Arabic language*. Cambridge: Cambridge University Press. Reprint from the 3rd edition, 1896–1898, an enlarged and revised translation of Caspari 1859.

Yip, Moira. 2004. Phonological markedness and allomorph selection in Zahao. *Language and Linguistics* 5:969–1001.

Departament de Filologia Catalana
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain

joan.mascaro@uab.cat