

A typological study of vowel interactions in Basque

Varun deCastro-Arrazola
Leiden University & Meertens Institute
Varun.deCastro-Arrazola@meertens.knaw.nl

Edoardo Cavarani
Meertens Institute
Edoardo.Cavarani@meertens.knaw.nl

Kathrin Linke
Meertens Institute & Leiden University
Kathrin.Linke@meertens.knaw.nl

Francesc Torres-Tamarit
UMR 7023 (CNRS/Paris 8)
FrancescJosep.Torres@gmail.com

Received: 28-04-15
Accepted: 21-07-15

Abstract

The aim of this paper is to check the factorial typology for a set of phonological constraints on vowel interactions in Basque against corpus data (Hualde and Gaminde 1998, *Euskararen Herri Hizkeren Atlas*, ‘The Basque Dialectological Atlas’) with the help of OT-Help 2.0 (Staub *et al.* 2010), a specialized software that calculates factorial typologies. The formal analysis developed to account for different patterns of vowel interactions in Basque, including those patterns displaying phonological opacity, implements Element Theory (Bacley 2011) in Turbidity Theory (Goldrick 2001, Van Oostendorp 2008). The proposed analysis has the virtue of predicting all attested patterns of a specific type of vowel interactions in Basque and excluding the unattested patterns.

Keywords: Basque; vowel interactions; factorial typology; opacity; Optimality Theory; Element Theory; Turbidity Theory.

Table of Contents

<p>1. Introduction</p> <p>2. Overview of the data</p> <p>3. Theoretical framework</p> <p>4. Factorial typology</p> <p>5. Opaque interaction</p>	<p>6. Underived patterns</p> <p>7. Discussion</p> <p>8. Conclusions</p> <p>9. References</p> <p>10. Appendix</p>
---	--

1. Introduction¹

In Basque, an uninflected NP like *gizon* ‘man’ takes the suffix /-a/ in order to create the singular absolutive DP *gizona* ‘the man’. When the stem ends in a consonant, the suffix takes the form [a] in all varieties of Basque. When the stem ends in a vowel (e.g. *neska* ‘girl’), however, a number of patterns show up (e.g. *neskia*, *neskie*, *neski*, *neska* ‘the girl’).²

Vowel interactions in Basque inflectional morphology had been profusely used as examples of extrinsic rule ordering in the literature of classic generative phonology although some empirical facts had often been misunderstood. In this context, Hualde and Gaminde published a thorough description of patterns of vowel interactions found in numerous dialects of Basque in 1998.³ However, still up to this day, the same misrepresented facts are still being reported for Basque in more recent studies (Kawahara 2002, Moreton and Smolensky 2002, Baković 2011).

Fortunately, the recent publication of the *Euskararen Herri Hizkeren Atlas* (‘The Basque Dialectological Atlas’; Aurrekoetxea and Videgain 2013), together with Hualde and Gaminde’s (1998) pivotal study, gives a fairly reliable

¹ The authors would like to thank the participants of the PhonoLAM meeting at the Meertens Institute, especially Marc van Oostendorp, and two anonymous reviewers for insightful comments on an earlier draft of this paper.

² Orthographic forms will be used throughout the text as a shorthand. Basque exhibits a standard five-vowel inventory, except for some Eastern varieties, which also include the high front rounded vowel /y/. Regarding consonants, their phonetic correspondence is fairly transparent, but note that <z, s, x> map to [ʒ, ʃ, ʃ], respectively.

³ A supposedly attested Western Basque dialect characterized by a synchronic chain shift mapping /a - a/ onto [ea] and /e - a/ onto [ie] was analyzed within Optimality Theory in Kirchner (1996), although such a system is unattested. However, no analyses on the attested synchronic chain shift mapping /e - a/ onto [ia] and /i - a/ onto [ie] is found in the theoretical literature, as far as we know. This paper aims to fill this gap. It is not our purpose, however, to deal with cases of low vowel assimilation that are blocked in particular morphological environments (/ur-a/ → [ure] ‘the water’, but /ur-a-k/ → [urak] ‘the waters’ cf. *[urek]) (see Hualde 1989, Orgun 1996, Inkelas 2000, Łubowicz 2002).

picture of the variation found in vowel interactions in the whole Basque speaking area.

The abstract models posited by formal linguists to explain observed patterns in language must both be tested on solid empirical grounds and aim at explanatory adequacy. With the help of those corpus data, this paper aims at developing a formal analysis of vowel interactions in Basque and also assessing its predictive power. We will show that the proposed analysis accounts for all the attested patterns of a specific type of vowel interactions in dialects of Basque and at the same time excludes the unattested patterns.⁴ Furthermore, we will show that an analysis based on Element Theory (Bacley 2011) and Turbidity Theory (Goldrick 2001, Van Oostendorp 2008), a version of Optimality Theory (Prince and Smolensky 1993/2004) that assumes containment, helps accounting for a counter-feeding opaque interaction between two phonological processes involved in vowel interactions.

Notice that, because of the formal properties just introduced, this paper represents an improvement with respect to the most recent and best-informed approach to Basque vowel alternations published so far, namely Hualde (1999). Indeed, his analysis, which crucially gives away with traditional generative devices such as rules, constraints and underlying representations and resorts instead to correspondences between surface forms, lacks the formal strictness necessary to constrain the observed alternations and to distinguish therefore between the attested/attestable versus non-attestable patterns. As discussed in section 7, this translates in a loss of explanatory adequacy: every pattern could be possible/learned, its absence being no more than an accident of history.

The outline of the paper is as follows. Section 2 illustrates the data on vowel interactions in Basque. Section 3 describes the theoretical framework adopted. Section 4 presents the factorial typology that follows from the adopted analysis. Section 5 discusses two patterns showing phonological opacity and proposes a solution. Section 6 briefly explains why unattested patterns are in fact unpredicted in our system. Section 7 discusses the approach to vowel interactions in Basque by Hualde (1999). Section 8 concludes the paper.

2. Overview of the data

2.1. Data sources

There are two main reliable and comprehensive sources of data for the dialectological variation of vowel interactions in Basque: HG (Hualde and Gaminde 1998) and EHHA⁵ (Aurrekoetxea and Videgain 2013). The data in HG were collected mainly through fieldwork conducted by Iñaki Gaminde; it contains vowel alternations for about 50 varieties of Basque. The data in EHHA were collected through fieldwork by a team of linguists coordinated by *Euskaltzaindia* ('the Royal Academy of the Basque Language'); it contains vowel alternations for 145 locations, covering virtually all the Basque-speaking area.

⁴ The type of vowel interactions considered involve only element changing operations, but not second vowel deletion or consonant epenthesis (see 2).

⁵ This is the acronym for *Euskararen Herri Hizkeren Atlas*.

In the present article we have combined both sources into a single dataset containing 195 data points (Appendix A and B). Most of these refer to dialects covering a single municipality, but some refer to smaller or bigger entities. There is some degree of overlap between the sources. In some cases, a data point in HG corresponds to a whole region that is decomposed in EHHA (i.e. Zeanuri and Igorre are towns within Arratia Valley). In other cases, both sources provide diverging descriptions for the same variety; out of the 22 such cases of overlap, 8 provide inconsistent data. We have opted to combine the data from both sources in our dataset (see the appendixes).

2.2. Constraints

As already mentioned, stems ending in a vowel show a variety of alternations in the singular definite absolutive. These alternations have been analysed by means of a number of synchronic or diachronic processes. Dialects differ in the number of processes that play a role, and the way these processes interact. This paper focuses on the three processes summarized in (1).

(1) Phonological processes

- a. Low vowel raising (/a - a/ → [ea]): It affects stems ending in the low vowel /a/. When the singular definite absolutive suffix /-a/ is added, this produces the alternation *alaba* / *alabea* ‘(the) daughter’.
- b. Mid vowel raising (/e - a/ → [ia]): It affects stems ending in the mid vowels /e/ or /o/. When the suffix /-a/ is added, this produces alternations such as *seme* / *semia* ‘(the) son’, or *baso* / *basua* ‘(the) forest’. The process is also triggered when a suffix starting with a different vowel is added, such as the proximate plural determiner: *seme* / *semiok*. Finally, cross-dialectal data provides evidence that the process is also active within morphemes: *beor* (Zugarramurdi dialect), *bior* (Deba dialect) ‘mare’ (Aurrekoetxea and Videgain 2011).
- c. Low vowel assimilation (/i - a/ → [ie]): It affects the suffix /-a/ when added to a stem ending in the high vowels /i/ or /u/: *idi* / *idie* ‘(the) ox’, *buru* / *burue* ‘(the) head’. As with mid vowel raising, cross-dialectal evidence indicates that the process also operates within morphemes: *biar* (Sara dialect), *bier* (Mungia dialect) ‘tomorrow’ (Aurrekoetxea and Videgain 2010).

In the rest of the paper, the back vowels /o, u/ will not be discussed. Stems ending in these vowels behave roughly as those ending in the front counterparts /e, i/. Low vowel assimilation, when active, is consistently triggered by both /i, u/. However, asymmetries concerning mid vowels do exist. In Larrauri, for instance, mid vowel raising affects /e/ but not /o/: *abade* / *abadia* ‘(the) priest’, but *asto* / *astoa* ‘(the) donkey’.

A number of additional processes can also affect the phonological contexts just described; they are briefly illustrated in (2), though not taken into account in the analysis that follows. Given that the emphasis of the paper lies on vowel quality, forms with epenthesis or gliding ((2a), (2c)) have been treated as though these processes were absent: a form like *idixa* is treated as equivalent to *idia*, and

astwa as equivalent to *astua*. The reader is referred to the paper by Hualde and Gaminde (1998) for more details.

(2) *Other phonological processes*

- a. Consonant epenthesis: *idi* / *idixa* ‘(the) ox’
- b. Second vowel deletion: *seme* / *semi* ‘(the) son’
- c. Gliding: *asto* / *astwa* ‘(the) donkey’
- d. Fronting of /u/: *esku* / *eskia* ‘(the) hand’

2.3. *Corpus annotation*

The dataset we use has been simplified in a number of ways in order to focus on the phenomenon we are interested in, namely, vowel alternations. Each variety is described using three labels. These labels capture the phonetic outcome for stems ending in /a, e, i/ when uttered in the singular definite absolutive form (i.e. when the suffix /-a/ is added). Except for the cases in which second vowel deletion is active, the label comprises two vowels, the first one corresponding to the stem-final vowel, and the second one corresponding to the suffix. Hence, a variety like Orozko, where the stems *alaba* ‘daughter’, *seme* ‘son’, *erri* ‘village’ are inflected as *alabea*, *semea*, *errie*, would be given the set of labels [ea], [ea], [ie].

It should be stressed that these labels constitute schematic representations of the phonetic output. The main simplifications are the following. Epenthetic consonants intervening between the two vowels were omitted. The first vowel loses its syllabicity in certain dialects, but the distinction is ignored here. Finally, the label [aa] virtually always corresponds to a single short [a], but we have chosen to represent it with a two-letter label for the sake of uniformity with the other labels.

2.4. *Attested and unattested patterns*

If we consider the three processes described above, stems ending in /a, e, i/ can take one of 4, 3, and 2 forms respectively. This derives from the fact that process 1 (/a - a/ → [ea]) can feed process 2 (/e - a/ → [ia]), which in turn can feed process 3 (/i - a/ → [ie]). When inflected, the stem *alaba* ‘daughter’ takes the form *alaba*⁶ in dialects in which no process is active, *alabea* in dialects in which low vowel raising is active, *alabia* in dialects in which both low vowel raising and mid vowel raising are active, and *alabie* if all three processes operate. Similarly, *seme* ‘son’ can be inflected as *semea*, *semia*, *semie*; and *idi* ‘ox’ can be inflected as *idia* or *idie*.

Pattern ID	/a - a/	/e - a/	/i - a/	Frequency
3	[aa]	[ia]	[ia]	70
24	[ie]	[ie]	[ie]	23
1	[aa]	[ea]	[ia]	18
2	[aa]	[ea]	[ie]	18
15	[ia]	[ia]	[ia]	15

⁶ Adding the suffix *-a* to the stem *alaba* would create two adjacent *a*'s, or [aa]. However, this is phonetically realised as a single [a].

8	[ea]	[ea]	[ie]	12
16	[ia]	[ia]	[ie]	5
4	[aa]	[ia]	[ie]	4
6	[aa]	[ie]	[ie]	4
7	[ea]	[ea]	[ia]	1
14	[ia]	[ea]	[ie]	1
18	[ia]	[ie]	[ie]	1
20	[ie]	[ea]	[ie]	1

Table 1. Logical patterns that are attested in the dataset and their observed frequencies in the Basque dialect area

From a logical point of view, these nine (4 + 3 + 2) forms can be combined in 24 ($4 \times 3 \times 2$) different patterns (corresponding to 24 potential dialects). In our dataset, 13 of these patterns are attested (table 1).

The observed absolute frequency of each pattern can be used as a proxy for the robustness of their description. Patterns that have been described in one variety are more prone to transcription inconsistencies than a pattern attested in 70 locations. However, it should also be pointed out that not all dialects listed in the dataset have the same status. Namely, some of them might suffer from under-representation because they refer to a whole set of locations which can potentially be partitioned into independent data points. Nonetheless, this cannot be argued for three of the four singletons (14, 18, 20), since they describe varieties spoken in individual towns (less than 10 km² each). Pattern 7 is attested in Literary Bizkaian, a variety no longer spoken but with reliable written evidence. In this case it is harder to estimate the geographical extension it had as a spoken variety.

For the sake of completeness, table 2 displays the remaining attested patterns that do not conform to any of the 24 logical patterns presented before. The reason they do not show any of these 24 patterns is that second vowel deletion has applied, or that they show a vowel not present in the standard five-vowel system. The patterns in table 2 represent roughly ten percent of the dataset. They will not be discussed in the present paper, but will be the object of future work.

Pattern ID	/a - a/	/e - a/	/i - a/	Frequency
33	[e]	[e]	[i]	4
26	[aa]	[e]	[i]	3
29	[aa]	[eœ]	[ie]	3
35	[i]	[i]	[ie]	3
28	[aa]	[e]	[ie]	2
31	[aa]	[i]	[i]	2
25	[aa]	[ea]	[i]	1
27	[aa]	[e]	[ia]	1
30	[aa]	[ia]	[i]	1
32	[aa]	[iə]	[ia]	1
34	[ia]	[ia]	[i]	1

Table 2. Patterns that depart from the 24 basic combinations

3. Theoretical framework

The main objectives of this paper are to account for the attested systems of vowel interaction in dialects of Basque and at the same time to exclude the unattested patterns.

Before presenting the OT analyses of Basque vowel interactions observed in different sets of dialects, we first define the representational principles assumed in the analyses and then formalize the proposed set of constraints.

3.1. Representations

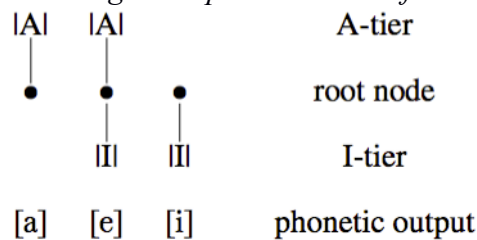
To simplify the analyses, we will exclude back vowels for now. Therefore, only the interactions between [a], [e] and [i] will be considered.⁷

We follow Element Theory (Backley 2011) in assuming that the primitives of phonological segments are a set of elements characterized by being *privative*, that is, present or absent from the phonological representation, and by being *autonomously interpretable*, meaning that at each point in the derivation, elements must be phonetically fully interpretable (see Harris and Lindsey 1995 for a more detailed explanation).

In order to represent the set of front vowels in a five-vowel system like the one that Basque displays, the two elements |A| and |I| are sufficient. The element |A| is (universally) interpreted as the low vowel [a], the element |I| is (universally) interpreted as the front high vowel [i], and the front mid vowel [e] is the result of combining the two primitive elements |A| and |I|. We assume the standard autosegmental principle that each element occupies its own tier. The phonological representation of front vowels in Basque in terms of elements is given in (3).⁸

⁷ As previously noted, [e] and [o], and [i] and [u], display the same phonological behavior in most dialects. In other dialects, however, there is an asymmetrical relation between the set of front vowels and the set of back vowels. If a certain phonological process targets a back vowel, then it always targets the corresponding front vowel. However, the opposite situation is not always true for some dialects: a phonological process can target a front vowel without necessarily targeting the corresponding back vowel. It is not the purpose of this paper to offer an explanation of these cases and we leave this issue for future research.

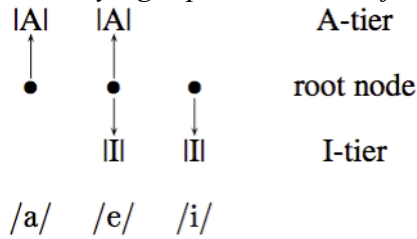
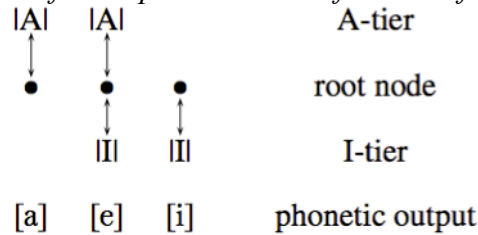
⁸ Using the primitive element C instead of |I| and the primitive element V instead of |A|, as in Dependency Phonology (Van der Hulst 2005), is also a possibility. Using C and V as primitive elements would in fact explain why /u/, and not only /i/, also triggers low vowel assimilation. This is so because the element C correlates with highness without making reference to place distinctions. We have opted to use the elements |A| and |I| in this study since our analysis is restricted to those vowel interactions affecting the set of front vowels.

(3) *Phonological representations of the set of Basque front vowels*

Apart from using elements for the representation of segments, the constraint-based analysis developed in this paper assumes Turbidity Theory (Goldrick 2001, Van Oostendorp 2008). Turbidity Theory is based on containment (Prince and Smolensky 1993/2004), i.e. on a particular approach to the theory of faithfulness that assumes an input-output relationship whereby the former is contained in the latter: “no element may be literally removed from the input form. The input is thus contained in every candidate form” (Prince and Smolensky 1993/2004:88).⁹ Turbidity Theory implements containment in the following way: there are two distinct types of relations between phonological elements and root nodes. On the one hand, underlying (or projection) relations express the lexical affiliation between phonological elements and root nodes.¹⁰ Graphically, this relation is depicted by an arrow pointing from the root node to the element. These arrows are always present in surface representations because of containment, i.e. they can never be deleted. On the other hand, surface (or pronunciation) relations express the phonetic realization of phonological elements. Graphically, this relation is depicted by an arrow pointing from the element to the root node. This is illustrated in (4) for underlying representations and in (5) for surface representations.

⁹ Notice that within this approach the constraints only evaluate output representations. As a consequence, faithfulness and markedness constraints are extensionally analogous: “containment effect is to make it possible to state all constraints on the output, without reference to the input-output relation [...]. Containment means, for example, that segmental deletion phenomena involve underparsing [viz underpronunciation of] a segment of the input [...] rather than outright replacement of a segment by \emptyset ” (McCarthy and Prince 1993: 88). The correspondence constraints DEP and MAX are hence substituted by FILL (“syllable positions are filled with segmental material”) and PARSE (“every phonological element needs to be parsed in the prosodic structure”), respectively (Van Oostendorp 2007), or, as shown in (6), by PROJECT and PRONOUNCE. For expository ease, though, we keep on referring to faithfulness and markedness constraints.

¹⁰ We depart from the notational conventions used in Turbidity Theory analyses to enhance the readability for readers unfamiliar with this theory.

(4) *Underlying representations of the set of Basque front vowels*(5) *Surface representations of the set of Basque front vowels*

In a nutshell, hence, projection lines are considered to be part of the lexical representation of a morpheme. The projection lines, in turn, cannot be altered by Gen because of the Consistency of Exponence (McCarthy and Prince 1994): “Gen can neither insert nor delete lexical material, hence it simply cannot change projection lines, but it can freely manipulate pronunciation lines” (Van Oostendorp 2008: 137).

3.2. *Constraints*

In Turbidity Theory, we need a family of constraints that ensure underlying projection relations to be realized as surface pronunciation relations, on the one hand, and surface pronunciation relations to correspond to underlying projection relations, on the other hand. We will call these constraints $\text{PRONOUNCE}(E)$ and $\text{PROJECT}(E)$, respectively. Their definitions are given in (6).¹¹

(6) *Constraints in Turbidity Theory*

- a. $\text{PROJECT}(E)$ ($\text{PROJ}(E)$): Assign a violation mark for every pronounced element E that does not correspond to any projection of E .
- b. $\text{PRONOUNCE}(E)$ ($\text{PRON}(E)$): Assign a violation mark for every projected element E that does not correspond to any pronunciation of E .

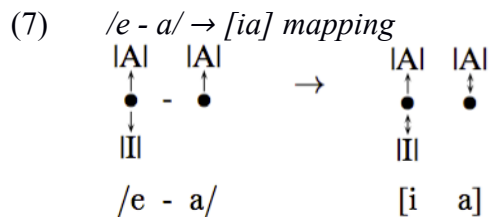
In Turbidity Theory, elements can be inserted. Inserting an element means inserting a surface pronunciation relation between the element E and the root node r with which it associates. In this case, the constraint $\text{PROJECT}(E)$ is violated. On the contrary, underlying elements cannot be deleted in Turbidity Theory due to containment. This means that an underlying projection relation between an element E and a root node r remains intact in the output representation. ‘Deletion’ in Turbidity Theory is in fact underpronunciation, that is, the absence of a surface

¹¹ In Van Oostendorp (2008), these constraints are called RECIPROCITY constraints. This change in nomenclature makes no changes.

pronunciation relation between an element E and the root node r that projects it. In this latter case, the constraint that is violated is $\text{PRONOUNCE}(E)$.

3.3. Input-output mappings

The phonological process mid vowel raising maps a stem-final underlying /e/ onto a surface [i] when preceding the absolutive singular /a/. This process is formalized as underpronunciation of an underlying element |A|. The input-output mapping /e/ → [i] is represented in (7); only the element |I|, but not the element |A|, is pronounced in the output representation.

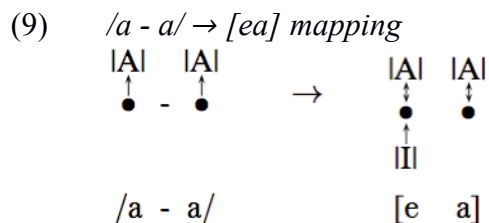


This unfaithful mapping is triggered by an OCP constraint against the pronunciation of pairs of adjacent |A| elements, $\text{OCP}(|A|)$. This constraint is formulated in (8).

- (8) $\text{OCP}(|A|)$
Assign a violation mark for every pair of adjacent root nodes that pronounce an |A| element.

The markedness constraint $\text{OCP}(|A|)$ outranks the constraint $\text{PRONOUNCE}(|A|)$. This constraint ranking produces the input-output mapping in (7).¹²

Basque also exhibits the phonological process of low vowel raising, which maps a stem-final underlying /a/ onto a surface [e] when preceding the absolutive singular /a/. In this case, low vowel raising is the result of pronouncing an element |I| that is not projected. The input-output mapping /a/ → [e] is represented in (9), in which the projected element |A| is pronounced together with an inserted element |I|.



¹² Why the constraint $\text{OCP}(|A|)$ is satisfied by means of underpronunciation of the element belonging to the root instead of underpronunciation of the element belonging to the suffix could be derived by splitting the constraint $\text{PRONOUNCE}(|A|)$ into two morphologically-sensitive constraints $\text{PRONOUNCE}(|A|)_{\text{Root}}$ and $\text{PRONOUNCE}(|A|)_{\text{Suffix}}$, respectively. The ranking for Basque could thus be $\text{PRONOUNCE}(|A|)_{\text{Suffix}} \gg \text{OCP}(|A|) \gg \text{PRONOUNCE}(|A|)_{\text{Root}}$.

We propose that the pronunciation of a non-projected element |I| is due to the satisfaction of a different OCP constraint. Instead of applying at one specific autosegmental tier, like OCP(|A|), this OCP constraint applies at the level of the root node, OCP(root). This constraint is defined in (10).

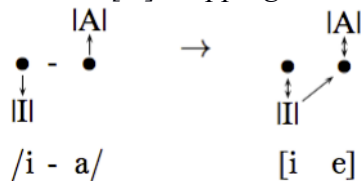
(10) *OCP(root)*

Assign a violation mark for every pair of adjacent root nodes that pronounce the same set of elements.

The output representation in (9) violates the local constraint OCP(|A|), because there are two adjacent pronounced |A| elements. However, this representation satisfies OCP(root) because the two adjacent root nodes, although sharing the pronunciation of a subset of elements, do not pronounce the same full set of segments. The pronunciation of the element |I| adds a violation of the constraint PROJECT(|I|).

The last phonological process considered here is low vowel assimilation, by which an underlying /a/ belonging to the absolutive singular morpheme maps onto a surface [e] when following an i-final root. The input-output mapping /a/ → [e] is represented in (11), in which the element |I| belonging to the root is doubly pronounced by its own root node and by the suffixal root node.¹³

(11) */i - a/ → [ie] mapping*



This phonological process of assimilation is interpreted here as being triggered by a markedness constraint SPREAD(|I|), which demands pronouncing the element |I| by a neighboring root node. This constraint is formulated in (12).

(12) *SPREAD(|I|)*

Assign a violation mark for every pronounced element |I| that does not spread (i.e. that is not pronounced by a neighboring root node).

¹³ As will be shown in the next section, the assimilation-triggering vowel can be underlying or derived depending on the dialect. In (11), low vowel assimilation is represented as triggered by underlying /i/. Notice, also, that the element undergoing the spreading process in (11) is the only element making up the relevant segment. This is not the case, though, in the forms represented in (7) and (9), in which |I| combines with |A| (either in the input or in the output form). As shown in section 5, the sensitivity of the spreading process to the uniqueness of a (segment's) element is formalized by a more restricted version of the constraint in (12), which asks for the spreading of |I| only in case |I| is the only element that is linked to the relevant root node.

(13).¹⁴ The constraint SPREAD(|I|) dominates the constraint OCP(|I|) defined in

- (13) *OCP(|I|)*
Assign a violation mark for every pair of adjacent root nodes that pronounce an |I| element.

To sum up, the constraint rankings for the three phonological processes are given in table 3.

Phonological processes	Mappings	Constraint rankings
low vowel raising	/a - a/ → [ea]	OCP(root) >> PROJ(I)
mid vowel raising	/e - a/ → [ia]	OCP(A) >> PRON(A)
low vowel assimilation	/i - a/ → [ie]	SPREAD(I) >> OCP(I)

Table 3. Constraint rankings for the three considered phonological processes

In the next section, we analyze all (sets of) dialects in terms of those constraints, and show how the factorial typology of the constraints accounts for the attested patterns of vowel interaction and discards the unattested patterns. Following this, the two dialects showing counter-feeding opacity between mid vowel raising and low vowel assimilation will be discussed.

4. Factorial typology

We used OT-Help 2.0 (Staub *et al.* 2010) to calculate the factorial typology for the 6 constraints given in table 3 with respect to the three relevant inputs /a - a/, /e - a/ and /i - a/. For each input, we only considered [aa], [ea], [ia] and [ie] as possible output candidates since each of them corresponds to an actual surface form.¹⁵

Having four potential candidates for each of the three inputs results in 64 (4³) possible sets of input-output mappings, i.e. 64 potential dialects.¹⁶ However,

¹⁴ The constraint SPREAD(|I|) could be defined in other ways. It is not the purpose of this paper, however, to develop a theory of assimilation or vowel harmony in Turbidity Theory. For the purpose of this paper, this constraint does the necessary job.

¹⁵ We excluded potential output candidates like [ee] and [ii] from the candidate set. Recall that this type of candidates are never pronounced as a hiatus but always undergo fusion as in /a - a/ → [a]. Although [ee] and [ii] are found in some dialects, we do not consider them in this typological study. Furthermore, all three considered phonological processes show a clear directionality effect. In the formulation of the constraints, we abstract away from the directionality options. Therefore, output candidates that would be the result of the inverse directionality effect (such as **alabae* vs *alabea*) are also excluded.

¹⁶ In section 2, we reported 24 logically possible dialects. However, this number already takes forward-feeding into account, i.e. any phonological process that applies to an underlying representation also applies to the identical derived representation. For instance, mid vowel raising not only enforces /e - a/ → [ia] but

OT-Help 2.0 can only generate constraint rankings for 8 out of the 64 sets of input-output mappings. This result is expected since the 6 constraints interact pairwise in 3 independent constraint rankings (3^2 ; cf. table 3). The generated 8 grammars coincide with 8 attested dialects,¹⁷ all of which show transparent interactions between the constraints. The rankings for each of these transparent dialects are given in table 4. The present constraint set accounts for more than 90 percent of the considered data. Section 5 extends the analysis to two additional patterns involving opaque relations.

Pattern ID	/a - a/	/e - a/	/i - a/	Constraint ranking	Frequency
1	[aa]	[ea]	[ia]	PROJ(I) >> OCP(root) PRON(A) >> OCP(A) OCP(I) >> SPREAD(I)	18
24	[ie]	[ie]	[ie]	OCP(root) >> PROJ(I) OCP(A) >> PRON(A) SPREAD(I) >> OCP(I)	23
6	[aa]	[ie]	[ie]	PROJ(I) >> OCP(root) OCP(A) >> PRON(A) SPREAD(I) >> OCP(I)	4
3	[aa]	[ia]	[ia]	PROJ(I) >> OCP(root) OCP(A) >> PRON(A) OCP(I) >> SPREAD(I)	70
15	[ia]	[ia]	[ia]	OCP(root) >> PROJ(I) OCP(A) >> PRON(A) OCP(I) >> SPREAD(I)	15
7	[ea]	[ea]	[ia]	OCP(root) >> PROJ(I) PRON(A) >> OCP(A) OCP(I) >> SPREAD(I)	1
8	[ea]	[ea]	[ie]	OCP(root) >> PROJ(I) PRON(A) >> OCP(A) SPREAD(I) >> OCP(I)	12
2	[aa]	[ea]	[ie]	PROJ(I) >> OCP(root) PRON(A) >> OCP(A) SPREAD(I) >> OCP(I)	18

Table 4. Factorial typology for attested transparent patterns

Pattern 1 shows the faithful input-output mappings, also corresponding with standard Basque. This is obtained by ranking all faithfulness constraints

also applies to derived [ea] from underlying /a - a/. So, an input /e - a/ will never be mapped onto [aa]. For the factorial typology, however, we adopted the most conservative view and inserted all four possible output candidates for each of the three phonological processes (therefore, 64 possible dialects). The observed forward-feeding is a result of the calculated grammars.

¹⁷ Sometimes we use the term dialect and sometimes the term pattern. When we use the term dialect, we mean sets of dialects displaying the same patterning of vowel alternations (cf. table 1). We use dialects as a shorthand.

above all markedness constraints. The constraint OCP(|I|), although being a markedness constraint, is responsible for blocking spreading of the element |I|. In what follows, we will refer descriptively to OCP(|I|) as a faithfulness constraint, in the sense of being a blocker constraint.¹⁸

Pattern 24 is the least marked pattern, meaning that all markedness constraints outrank all ‘faithfulness’ constraints. Therefore, all outputs take the form [ie]. Pattern 6 resembles pattern 24 except for the ranking of PROJECT(|I|) and OCP(root), the former dominating the latter. This ranking causes the input /a - a/ to map faithfully onto the output [aa].

Both patterns 3 and 15 lack spreading of the element |I| due to ranking OCP(|I|) above SPREAD(|I|). However, only pattern 3 faithfully maps the input /a - a/ onto the output [aa], as in pattern 6. One interesting aspect of the constraint ranking given for pattern 3 is that the constraint OCP(|A|) dominates the constraint PRONOUNCE(|A|). This ranking is responsible for mapping the input /e - a/ onto the output [ia]. However, the output [aa] violates the top-ranked constraint OCP(|A|). The only way to satisfy the constraint OCP(|A|) is leaving the element |A| unpronounced. However, this strategy would result in an empty, element-less root node. A candidate containing such a root node could be fought against by positing a markedness constraint such as HAVE-ELEMENT, demanding root nodes to be specified. Given the fact that in all Basque dialects HAVE-ELEMENT is undominated, we have not included it in our constraint set.¹⁹

The patterns 7 and 8 share the ranking in which the constraint OCP(root) dominates the constraint PROJECT(|I|). This is why the output [aa] derived from the input /a - a/ is discarded in favor of the output [ea]. The constraint OCP(|A|) is dominated by PRONOUNCE(|A|), thereby enforcing the input /e - a/ to be faithfully mapped onto the output [ea]. These two patterns only differ with respect to the absence *versus* the presence of spreading of the element |I|. Finally, pattern 2 resembles pattern 8 except for the ranking of the constraints PROJECT(|I|) and OCP(root), the former dominating the latter, which prevents the insertion of the element |I| in outputs that derive from the input /a - a/.

Finally, consider again pattern 8. This pattern is characterized by spreading the element |I| (/i - a/ → [ie]) and allowing the pronunciation of two adjacent elements |A| ([ea]). The reason why the element |I| does not spread in those outputs derived from the input /e - a/ is due to the constraint OCP(root). So, the element |I| does not spread in the output derived from the input /e - a/ because that would create an OCP(root)-violating output [ee]. This is an interesting aspect of our analysis: OCP(root), an independently motivated constraint that accounts for the mapping /a - a/ → [ea], is enough to block the spreading of the element |I| in that case. This situation only applies in those dialects in which the pronunciation of the element |A| is mandatory (PRONOUNCE(|A|) >> OCP(|A|)).

¹⁸ Cf. footnote 8 on the markedness/faithfulness (non-)distinction characterizing containment-based OT approaches.

¹⁹ OT-Help 2.0 does not allow for specifying constraints as undominated constraints. Therefore, the undominated constraint HAVE-ELEMENT and the potentially targeted output candidates containing empty root nodes had to be excluded for the typological study.

In the next section, we consider the two attested patterns that show a counter-feeding opaque interaction between mid vowel raising and low vowel assimilation.

5. Opaque interaction

Two attested dialects in Basque show a counter-feeding opaque interaction between mid vowel raising and low vowel assimilation. The input-output mappings of those two dialects are given in table 5.

Pattern ID	/a - a/	/e - a/	/i - a/	Frequency
16	[ia]	[ia]	[ie]	5
4	[aa]	[ia]	[ie]	4

Table 5. Attested opaque dialects

In these two opaque dialects, the input /e - a/ maps onto the output [ia], and the input /i - a/ maps onto the output [ie]; however, the input /e - a/ never maps onto the output [ie]. This is a classic case of a synchronic chain-shift.²⁰ Blocking the spreading of the element |I| in outputs that derive from underlying /e - a/ cannot be attributed to the activity of OCP(root), as in dialect 8, because in these two opaque dialects the element |A| is left unpronounced (OCP(|A|) >> PRONOUNCE(|A|), which causes underlying /e/ to map onto surface [i]).

If /i - a/ maps onto [ie], it means that SPREAD(|I|) dominates OCP(|I|). The output [ia] derived from /e - a/ violates top-ranked SPREAD(|I|). For this reason, the output [ia] can never be the optimal candidate with the constraint set presented so far.

In our proposal, we make use of two ingredients: containment and privativity. Due to containment, the underlying composition of root nodes is always present in output representations. On the other hand, privativity, as defined in Element Theory, implicitly relies on set theory (see Breit 2013 for an implementation).

We understand that the projected elements of a root node form a set of elements. For instance, the segment /e/ is formally defined by the set { |A|, |I| }. Therefore, constraints make reference to these sets of elements projected by root nodes.

We further propose that constraints may include conditions on the identity between the set of elements projected by a root node and a given set of elements.

We apply this idea to solve the opacity problem mentioned above. In patterns 16 and 4, the element |I| only spreads when it is the exclusive element projected by the root node, but not when it co-occurs with another projected element, as |A|. In other words, only the faithfully derived segment [i] (/i/ → [i])

²⁰ The same holds for the input /a - a/ of dialect 16, which surfaces as [ia]. Also in this case, [ia] behaves differently from /i - a/. For this input-output mapping, though, constraint (14b) is not necessary to account for the opaque pattern. As shown in table 6, the relevant constraint ranking for the /a - a/ → [ia] mapping is OCP(root) >> PROJ(|I|).

triggers low vowel assimilation, but not the unfaithfully derived segment [i] (/e/ → [i]).

In order to derive the opaque candidates, we propose to split the constraint SPREAD(|I|) into two constraints standing in a stringency relation: a more stringent constraint SPREAD(|I|) (already defined in (12) and repeated in (14a)) and a less stringent constraint SPREAD(|I|)', defined in (14b).

- (14) *Spread(|I|) constraint in stringency relation*
- a. SPREAD(|I|): Assign a violation mark for every pronounced element I that does not spread.
 - b. SPREAD(|I|)': Let the set S of projected elements by a root node be identical to the set $\{|I|\}$, i.e. $\{|I|\} \subseteq S \wedge S \subseteq \{|I|\}$. Assign a violation mark for every pronounced element $|I| \in S$ that does not spread.

The more stringent constraint SPREAD(|I|) assigns a violation mark for any element |I| that does not spread irrespective of whether it is the only element projected by the root node or not. That is, it assigns a violation mark for any output [ia] derived from either /e - a/ or /i - a/. However, the less stringent constraint SPREAD(|I|)' assigns a violation mark for any element |I| that does not spread if and only if this element |I| is identical to the full set of elements projected by the root node. That is, it assigns a violation mark for any output [ia] derived from /i - a/, but not when it derives from /e - a/.

The whole constraint rankings for patterns 16 and 4 appear in table 6.

Pattern ID	/a - a/	/e - a/	/i - a/	Constraint ranking	Frequency
16	[ia]	[ia]	[ie]	OCP(root) >> PROJ(I) OCP(A) >> PRON(A) SPREAD(I)' >> OCP(I) >> SPREAD(I)	5
4	[aa]	[ia]	[ie]	PROJ(I) >> OCP(root) OCP(A) >> PRON(A) SPREAD(I)' >> OCP(I) >> SPREAD(I)	4

Table 6. Factorial typology for attested opaque dialects

We have shown that assuming standard representations in Element Theory together with a set of basic operations and constraints correctly accounts for the full set of observed patterns of vowel interactions in the Basque speaking area and discards the unattested patterns. Making use of Turbidity Theory combined with Element Theory has also been shown to be advantageous in accounting for the two opaque dialects.

6. Underived patterns

The factorial typology of constraint rankings based on the possible input-output mappings revealed that, once the constraint we need for the opaque patterns (14b)

is introduced in the grammar, 10 out of the formally 64 logical patterns can be derived. Crucially, these patterns coincide with the 10 attested dialect types. Out of the remaining 54 patterns, 40 patterns are excluded due to a lack of what we called ‘forward-feeding’ (cf. footnote 15). The remaining 14 logical patterns cannot be derived in our system based on the constraints given in either table 3 or table 6. From these 14 underived patterns, 11 are unattested (table 7).

Pattern ID	/a - a/	/e - a/	/i - a/	Frequency
5	[aa]	[ie]	[ia]	0
9	[ea]	[ia]	[ia]	0
10	[ea]	[ia]	[ie]	0
11	[ea]	[ie]	[ia]	0
12	[ea]	[ie]	[ie]	0
13	[ia]	[ea]	[ia]	0
17	[ia]	[ie]	[ia]	0
19	[ie]	[ea]	[ia]	0
21	[ie]	[ia]	[ia]	0
22	[ie]	[ia]	[ie]	0
23	[ie]	[ie]	[ia]	0

Table 7. Logical patterns which are unattested in the dataset

From the perspective of our parallel constraint-based analysis, the above patterns are discarded because they simply require contradictory constraint rankings, and therefore a consistent grammar cannot be obtained for these data. For instance, consider dialects 9, 10, 11 and 12 in table 7. These patterns require the final vowel of e-stems to map onto [i]. In our analysis, this is the result of ranking the constraint OCP(|A|) above the constraint PRONOUNCE(|A|). At the same time, these patterns map the final vowel of a-stems onto [e] because of the activity of OCP(root). However, the latter mapping also requires the ranking PRONOUNCE(|A|) above OCP(|A|), which is contradictory to the other facts of the language. The other underived patterns involve other ranking paradoxes, which is why they are excluded from the factorial typology.

However, there are 3 dialects that our theory does not predict for the same reasons exposed above, but that are seemingly attested (last 3 dialects in table 1, repeated in table 8).

Pattern ID	/a - a/	/e - a/	/i - a/	Frequency
14	[ia]	[ea]	[ie]	1
18	[ia]	[ie]	[ie]	1
20	[ie]	[ea]	[ie]	1

Table 8. Attested but underived patterns

So far we have taken the vowel alternations from a single morphological context: the last vowel of the noun root plus the definitive absolutive suffix /-a/. The maps in EHHA, however, include information for a number of other suffixes which also begin with /a/. In table 9 we include the interactions between the stem-final vowel and suffix-initial vowel (four different suffixes) in the three non-

predicted dialects: Zaratamo (14), Etxeberri (20), and Gizaburuaga (18). All four suffixes presented here begin with /a/, and can be represented as follows: *-a* (absolute), *-ak* (ergative), *-ari* (dative), and *-aren* (genitive). Overall, variation can be observed in each paradigm.

Dialect	Context	/a - a/	Dialect	Context	/e - a/
14	absolute sing	ia, ie	14	absolute sing	ea
14	ergative sing	ia	14	ergative sing	ea
14	dative sing	ie	14	dative sing	ea
14	genitive sing	aa, ie	14	genitive sing	ie
18	absolute sing	ia	18	absolute sing	ie
18	ergative sing	i	18	ergative sing	ie
18	dative sing	ia	18	dative sing	ie
18	genitive sing	ie	18	genitive sing	i
20	absolute sing	ie	20	absolute sing	ea
20	ergative sing	ie	20	ergative sing	ea
20	dative sing	ie	20	dative sing	ea
20	genitive sing	ie	20	genitive sing	ie

Table 9. Interaction between the last vowel of a-final stems (left) and e-final stems (right) and the first vowel of four a-initial suffixes, as attested in three dialects

We consider that not predicting these 3 dialects does not overthrow our typological study. We understand that excluding them from our factorial typology is in fact desirable. First, the inconsistencies found for each system suggest that it might be the case that informants of these dialects produce outcomes derived from two different systems due to interdialectal contact or due to an unstable situation of on-going language change. Second, these dialects can be considered marginal since each of them is only attested once. Third, they are not substantially different from all the other underived patterns that are in fact unattested.

In the next section, we will discuss Hualde's (1999) approach to Basque vowel alternations. In the light of the analysis of the Basque data presented in this paper, we will conclude that there is no substantial reason to reject generative phonology as a competence model characterized as a system of mappings between (at least) two levels of representation.

7. Discussion

As hinted at in section 1, an analysis of Basque vowel interactions has already been proposed by Hualde (1999). However, he neither accounts for the counter-feeding opaque interaction just explained, nor does he develop an analysis that is formal enough to constrain the attested variation. Worried by the alleged lack of psychological plausibility of the intermediate representations characterizing the generative approaches that resort to rule ordering, Hualde's (1999) approach resorts to a mechanism whereby the phonological knowledge that Basque speakers have regarding vowel alternations is defined in terms of correspondences

between surface forms that share the same stem (similarly to what is proposed, for instance, by Bybee 1994, 2001). As a result, he disposes any intermediate representation together with all underlying representations. This is because, he claims,

“postulating abstract underlying representations results in both incorrect predictions and contradictory analyses on the Basque alternations [that] are also perfectly compatible with a much simpler approach where unobservable entities are not posited and only surface-to-surface correspondences are employed” (Hualde 1999:33-35).

However, the formal devices he proposes are not explicitly developed, nor exploited to formalize the variation characterizing the vowel interactions under concern. As a consequence, his approach lacks restrictiveness and predictive power. Indeed, there seems to be no way to constrain what language users can learn about their phonological system, and therefore no way to constrain the phonological patterns a language can display: in principle, every possible pattern could be learnable, and the absence of a given pattern from the attested typology seems no more than an accident of history. An approach such as the one developed in this paper, instead, allows for clear predictions regarding the attestability of a given system. Indeed, as shown in sections 4 and 6, from 64 logical systems we only predict 8 systems, or 10 once we introduce the specific opacity-solving constraint. None of the remaining 54 systems are reliably attested.

Notice that resorting to correspondences between surface forms and, crucially, the absence of the underlying/surface representation distinction, apparently eliminates the opacity issue, as well as the alleged lack of naturalness of rules or constraints postulated by ‘traditional’ generative approaches. For instance, when discussing the opaque case of the Ultzama dialect (Pattern ID 4), Hualde claims that it

“is certainly possible to write a rule (or a constraint in an optimality theory approach) that will produce the gliding of mid vowels but not of high vowels in this context” (*seme* / *semja* ‘(the) son’ cf. *mendi* / *mendie* ‘(the) mountain’).²¹

However, “such rule or constraint would be completely unexplanatory”. This is because

“the situation is not a natural one. The expected situation is that if mid vowels glide, high vowels should also glide. Our hypothesis should be that the sound changes that gave rise to these synchronic patterns were natural, phonetically grounded ones, and that the naturalness of the changes was later masked by further developments” (Hualde 1999:40).

²¹ Notice that in the Ultzama dialect, variation can be observed in the application of the mid vowel raising process. When the absolutive marker *-a* is affixed to an *e*-final stem such as *seme*, the stem-final vowel can be raised, as in *semja*, or not, as in *semea*. As a consequence of this alternation, the Ultzama dialect seems to oscillate between a transparent and an opaque system. The gliding process applies to both outcomes.

In fact, this is a problem only if we assume that synchronic phonological (computational and representational) properties must be phonetically-grounded. As claimed by Hualde (1999) himself, it could have been like this at a given historical stage, but then a given rule could have “gone crazy by aging” (Scheer in press). In other words, a grammar can generate input-output mappings that are as arbitrary as the ones described by Hualde’s correspondences. Later on, Hualde (1999) hypothesizes that

“when gliding was acquired by the Ultzama dialect it did not affect forms like *mendie* because at the time there was an intervening consonant between the two vowels [...]. In fact, not far from Ultzama, in Lizarraga, we find forms such as *semia*, *mendiye* [...], *otsua*, *eskube*. We would be justified in assuming that Ultzama had similar forms more-or-less recently and that subsequently there has been a change *-iye* > *-ie*, *-ube* > *-ue*, by which the once epenthetic consonants were lost. As a historical change, gliding applied in a natural fashion. The present-day alternations, however, lack naturalness. But this lack of naturalness does not make the alternations less learnable” (Hualde 1999:41).

Nor does it make the alternations less phonologically determined, we would like to add. Furthermore, notice that utilizing containment allows for the presence, in the underlying representation, of precisely the segment that interferes with the gliding process (or at least an underspecified consonantal slot). Moreover, Hualde (1999) lacks any definition of the representational properties of segments in general and vowels in particular. Applying containment and Element Theory, instead, allows us to maintain the naturalness that the system of correspondences proposed by Hualde (1999) lacks. In fact, the Ultzama case seems exactly to be a case in which resorting to underlying representations could improve (the naturalness of) the analysis. Furthermore, as shown in section 5, the formal model we propose allows for an account of counter-feeding opacity characterizing the alternations of this dialect.

As a conclusion, let us recall the following passage: “the burden of proof must lie with those trying to defend the existence of underlying representations and derivations” (Hualde 1999:35). With these words, Hualde threw down the gauntlet to generative phonologists. We hope that, with our paper, we have taken it up.

8. Conclusion

We have proposed an analysis of vowel alternations in Basque couched within Element Theory and Turbidity Theory. The proposed set of constraints predicts all attested patterns of a specific type of vowel interactions and excludes the unattested patterns. We have further shown that Element Theory, together with containment, allows us to make reference to identity conditions between the set of projected (underlying) elements of a root node and a given set of elements. This device has the potential to solve the counter-feeding opaque interaction between vowel raising and low vowel assimilation. Referring to identity conditions of this kind for the treatment of phonological opacity deserves future research.

This paper has made an integrated use of formal theories, corpus data and computational tools to study the typology of vowel interactions in Basque. We hope that the approach taken in this paper also contributes to a more general discussion of methodological aspects in the study of (micro-)variation at the level of phonological analysis.

9. References

- Aurrekoetxea, Gotzon and Xarles Videgain (eds.). 2010. *Euskararen Herri Hizkeren Atlas 2. Lexikoa*. Bilbo: Euskaltzaindia.
- Aurrekoetxea, Gotzon and Xarles Videgain (eds.). 2011. *Euskararen Herri Hizkeren Atlas 3. Lexikoa*. Bilbo: Euskaltzaindia.
- Aurrekoetxea, Gotzon and Xarles Videgain (eds.). 2013. *Euskararen Herri Hizkeren Atlas 5. Izen morfologia*. Bilbo: Euskaltzaindia.
- Backley, Phillip. 2011. *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.
- Baković, Eric. 2011. "Opacity and ordering". In J. Goldsmith, J. Riggle and A. Yu (eds.), *The Handbook of Phonological Theory*, 40–67. Malden, MA: Wiley-Blackwell.
<http://dx.doi.org/10.1002/9781444343069.ch2>
- Breit, Florian. 2013. Formal aspects of Element Theory. MA thesis, University College London.
- Bybee, Joan. 1994. "A view of phonology from a cognitive and functional perspective". *Cognitive Linguistics* 5(4): 285–305.
<http://dx.doi.org/10.1515/cogl.1994.5.4.285>
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge, UK: Cambridge University Press.
<http://dx.doi.org/10.1017/cbo9780511612886>
- Goldrick, Matthew. 2001. "Turbid output representations and the unity of opacity". In M. Hirotani, A. Coetzee, N. Hall and J. Y. Kim (eds.), *Proceedings of the North East Linguistic Society*, 30: 231–245. Amherst, MA: GLSA.
- Harris, John and Geoff Lindsey. 1995. "The elements of phonological representation". In J. Durand and F. Katamba (eds.), *Frontiers of phonology: Atoms, structures, derivations*, 34–79. Essex: Longman.
- Hualde, José I. 1989. "The strict cycle condition and noncyclic rules". *Linguistic Inquiry* 20(4): 675–680.
- Hualde, José I. and Iñaki Gaminde. 1998. "Vowel interaction in Basque: A nearly exhaustive catalogue". *Studies in the Linguistic Sciences* 28(1): 41–77.
- Hualde, José I. 1999. "Vowel interaction and related phenomena in Basque and the nature of morphophonological knowledge". *Cognitive Linguistics* 10(1): 33–56.
<http://dx.doi.org/10.1515/cogl.1999.002>
- van der Hulst, Harry. 2005. "The molecular structure of phonological segments". In P. Carry, J. Durand and C. Ewen (eds.), *Headhood, elements, specification and contrastivity*, 193–234. Amsterdam: John Benjamins.
<http://dx.doi.org/10.1075/cilt.259.13hul>

- Inkelas, Sharon. 2000. "Phonotactic blocking through structural immunity". In B. Stiebels and D. Wunderlich (eds.), *Lexicon in focus*, 7–40. Berlin: Akademie Verlag.
<http://dx.doi.org/10.1515/9783050073712-002>
- Kawahara, Shigeto. 2002. Similarity among variants: Output-variant correspondence. Undergraduate thesis, International Christian University, Tokyo.
- Kirchner, Robert. 1996. "Synchronic chain shift in Optimality Theory". *Linguistic Inquiry* 27(2): 340–350.
<http://dx.doi.org/10.1002/9780470756171.ch21>
- Łubowicz, Anna. 2002. "Derived environments effects in Optimality Theory". *Lingua* 112: 243–280.
[http://dx.doi.org/10.1016/s0024-3841\(01\)00043-2](http://dx.doi.org/10.1016/s0024-3841(01)00043-2)
- McCarthy, John J. and Alan Prince. 1993. "Generalized alignment". In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1993*, 79–153. Springer Netherlands.
http://dx.doi.org/10.1007/978-94-017-3712-8_4
- McCarthy, John and Alan Prince. 1994. "The emergence of the unmarked: Optimality in Prosodic Morphology". In M. González (ed.), *Proceedings of the North East Linguistic Society*, 24: 333–379. Amherst, MA: GLSA.
- Moreton, Elliott and Paul Smolensky. 2002. "Typological consequences of local constraint conjunction". In L. Mikkelsen and C. Potts (eds.), *WCCFL*, 21: 306–319. Cambridge, MA: Cascadilla Press.
- van Oostendorp, Marc. 2007. "Derived environment effects and consistency of exponence". In S. Blaho, P. Bye and M. Krämer (eds.), *Freedom of analysis?*, 95: 123–148. Berlin: Walter de Gruyter.
<http://dx.doi.org/10.1515/9783110198591.123>
- van Oostendorp, Marc. 2008. "Incomplete Devoicing in Formal Phonology". *Lingua* 118: 127–142.
<http://dx.doi.org/10.1016/j.lingua.2007.09.009>
- Orgun, Cemil Orhan. 1996. Sign-based morphology and phonology: With special attention to Optimality Theory. Ph.D. thesis, University of California Berkeley.
- Prince, Alan and Paul Smolensky. 1993. "Optimality Theory: Constraint interaction in generative grammar". Ms., Rutgers University and University of Colorado, Boulder. Published 2004, Malden, MA and Oxford: Wiley-Blackwell.
- Scheer, Tobias. In press. "Crazy rules, regularity and naturalness". In J. Salmons and P. Honeybone (eds.), *The Handbook of Historical Phonology*. Oxford: Oxford University Press.
- Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy and Joe Pater. 2010. *OT-Help 2.0*. Software package. Amherst, MA: University of Massachusetts Amherst.

10. Appendix

A. Attested combinations (ordered by pattern)

Dialect	Source	/a-a/	/e-a/	/i-a/	Pattern ID
ahetze	ehha	aa	ea	ia	1
ainhoa	hg	aa	ea	ia	1
arrangoitze	ehha	aa	ea	ia	1
goizueta	hg	aa	ea	ia	1
abaurregaina	ehha	aa	ea	ia	1
urepel	hg	aa	ea	ia	1
irisarri	ehha	aa	ea	ia	1
aezkoaValley	hg	aa	ea	ia	1
luzaide	ehha	aa	ea	ia	1
bidarraia	ehha	aa	ea	ia	1
aldude	ehha	aa	ea	ia	1
sara	ehha	aa	ea	ia	1
baigorri	ehha	aa	ea	ia	1
tolosa	ehha	aa	ea	ia	1
zugarramurdi	ehha	aa	ea	ia	1
itsasu	ehha	aa	ea	ia	1
sara	hg	aa	ea	ia	1
arbizu	ehha	aa	ea	ia	1
zilbeti	ehha	aa	ea	ie	2
beasain	ehha	aa	ea	ie	2
alegia	hg	aa	ea	ie	2
ikaztegieta	ehha	aa	ea	ie	2
mezkiritz	ehha	aa	ea	ie	2
alkotz	ehha	aa	ea	ie	2
elduain	ehha	aa	ea	ie	2
zegama	hg	aa	ea	ie	2
etxaleku	hg	aa	ea	ie	2
lizartza	hg	aa	ea	ie	2
gaintza	ehha	aa	ea	ie	2
leitza	ehha	aa	ea	ie	2
orexa	ehha	aa	ea	ie	2
ataun	ehha	aa	ea	ie	2
amezketa	ehha	aa	ea	ie	2
etxaleku	ehha	aa	ea	ie	2
zegama	ehha	aa	ea	ie	2
ezkio	ehha	aa	ea	ie	2
beizama	ehha	aa	ia	ia	3
donostia	ehha	aa	ia	ia	3
larraul	hg	aa	ia	ia	3
arboti	ehha	aa	ia	ia	3
andoain	ehha	aa	ia	ia	3
sohüta	ehha	aa	ia	ia	3
erronkariValley	hg	aa	ia	ia	3

armendaritze	ehha	aa	ia	ia	3
sunbilla	ehha	aa	ia	ia	3
arnegi	ehha	aa	ia	ia	3
altzürükü	ehha	aa	ia	ia	3
behorlegi	ehha	aa	ia	ia	3
gamarte	ehha	aa	ia	ia	3
etxalar	ehha	aa	ia	ia	3
garrüze	ehha	aa	ia	ia	3
donibane	ehha	aa	ia	ia	3
aiherra	hg	aa	ia	ia	3
ürrüstoi	ehha	aa	ia	ia	3
arrueta	ehha	aa	ia	ia	3
beskoitze	ehha	aa	ia	ia	3
arbona	hg	aa	ia	ia	3
hazparne	ehha	aa	ia	ia	3
domintxaine	ehha	aa	ia	ia	3
donamaria	ehha	aa	ia	ia	3
ezterenzubi	ehha	aa	ia	ia	3
hondarribia	ehha	aa	ia	ia	3
dorrao	ehha	aa	ia	ia	3
lasarte	ehha	aa	ia	ia	3
oiartzun	ehha	aa	ia	ia	3
arroa	ehha	aa	ia	ia	3
bastida	ehha	aa	ia	ia	3
senpere	ehha	aa	ia	ia	3
jaurrieta	ehha	aa	ia	ia	3
usurbil	hg	aa	ia	ia	3
azkaine	ehha	aa	ia	ia	3
orio	hg	aa	ia	ia	3
larraine	ehha	aa	ia	ia	3
hernani	ehha	aa	ia	ia	3
uharte	ehha	aa	ia	ia	3
eskiula	ehha	aa	ia	ia	3
larzabale	ehha	aa	ia	ia	3
bardoze	ehha	aa	ia	ia	3
barkoxe	ehha	aa	ia	ia	3
izturitze	ehha	aa	ia	ia	3
arano	hg	aa	ia	ia	3
makea	ehha	aa	ia	ia	3
jutsi	ehha	aa	ia	ia	3
deba	ehha	aa	ia	ia	3
landibarre	ehha	aa	ia	ia	3
altzai	ehha	aa	ia	ia	3
bergara	ehha	aa	ia	ia	3
montori	ehha	aa	ia	ia	3
mugerre	ehha	aa	ia	ia	3
zarautz	hg	aa	ia	ia	3
urdiñarbe	ehha	aa	ia	ia	3

getaria	ehha	aa	ia	ia	3
beskoitze	hg	aa	ia	ia	3
aia	ehha	aa	ia	ia	3
hendaia	ehha	aa	ia	ia	3
urretxu	ehha	aa	ia	ia	3
zuberoaProvince	hg	aa	ia	ia	3
mendaro	ehha	aa	ia	ia	3
urketa	ehha	aa	ia	ia	3
zaraitzuValley	hg	aa	ia	ia	3
santagrazi	ehha	aa	ia	ia	3
pasaia	ehha	aa	ia	ia	3
zumaia	hg	aa	ia	ia	3
orio	ehha	aa	ia	ia	3
pagola	ehha	aa	ia	ia	3
uztaritze	ehha	aa	ia	ia	3
ultzama	hg	aa	ia	ie	4
baztanValley	hg	aa	ia	ie	4
eugi	ehha	aa	ia	ie	4
suarbe	ehha	aa	ia	ie	4
azpeitia	ehha	aa	ie	ie	6
lizarraga	hg	aa	ie	ie	6
azkoitia	ehha	aa	ie	ie	6
etxarri	hg	aa	ie	ie	6
literaryBizkaian	hg	ea	ea	ia	7
orozko	hg	ea	ea	ie	8
arratiaValley	hg	ea	ea	ie	8
orozko	ehha	ea	ea	ie	8
lemoa	ehha	ea	ea	ie	8
zeanuri	ehha	ea	ea	ie	8
sondika	ehha	ea	ea	ie	8
leioa	ehha	ea	ea	ie	8
zeberio	ehha	ea	ea	ie	8
dima	ehha	ea	ea	ie	8
zaldibia	hg	ea	ea	ie	8
legazpi	ehha	ea	ea	ie	8
legazpi	hg	ea	ea	ie	8
zaratamo	ehha	ia	ea	ie	14
leintz	ehha	ia	ia	ia	15
lekeitio	hg	ia	ia	ia	15
urdiain	hg	ia	ia	ia	15
antzuola	hg	ia	ia	ia	15
elorrío	hg	ia	ia	ia	15
oñati	hg	ia	ia	ia	15
elgoibar	ehha	ia	ia	ia	15
markina18thcent	hg	ia	ia	ia	15
mutriku	hg	ia	ia	ia	15
urdiain	ehha	ia	ia	ia	15
lekeitio	ehha	ia	ia	ia	15

eibar	ehha	ia	ia	ia	15
araotz	ehha	ia	ia	ia	15
oñati	ehha	ia	ia	ia	15
elorrío	ehha	ia	ia	ia	15
gamiz	ehha	ia	ia	ie	16
lezama	hg	ia	ia	ie	16
larrabetzu	ehha	ia	ia	ie	16
larrauri	hg	ia	ia	ie	16
mungia	ehha	ia	ia	ie	16
gizaburuaga	ehha	ia	ie	ie	18
etxebarri	ehha	ie	ea	ie	20
ibarruri	ehha	ie	ie	ie	24
bermeo	hg	ie	ie	ie	24
bolibar	ehha	ie	ie	ie	24
aramaio	ehha	ie	ie	ie	24
arrasate	ehha	ie	ie	ie	24
etxebarria	ehha	ie	ie	ie	24
berriz	ehha	ie	ie	ie	24
azkoitia	hg	ie	ie	ie	24
mañaria	ehha	ie	ie	ie	24
otxandio	ehha	ie	ie	ie	24
mendata	ehha	ie	ie	ie	24
aramaio	hg	ie	ie	ie	24
azpeitia	hg	ie	ie	ie	24
bakio	ehha	ie	ie	ie	24
arrazola	ehha	ie	ie	ie	24
busturia	ehha	ie	ie	ie	24
arrieta	ehha	ie	ie	ie	24
zollo	ehha	ie	ie	ie	24
zornotza	ehha	ie	ie	ie	24
arratzu	hg	ie	ie	ie	24
kortezubi	ehha	ie	ie	ie	24
errigoiti	ehha	ie	ie	ie	24
elantxobe	ehha	ie	ie	ie	24
ezkurra	ehha	aa	ea	i	25
etxarri	ehha	aa	e	i	26
beruete	hg	aa	e	i	26
oderitz	ehha	aa	e	i	26
goizueta	ehha	aa	e	ia	27
bermeo	ehha	aa	e	ie	28
beruete	ehha	aa	e	ie	28
aniz	ehha	aa	eœ	ie	29
lekaroz	ehha	aa	eœ	ie	29
erratzu	ehha	aa	eœ	ie	29
errezil	ehha	aa	ia	i	30
basaburua	hg	aa	i	i	31
igoa	ehha	aa	i	i	31
asteasu	ehha	aa	iə	ia	32

laukiz	ehha	e	e	i	33
getxo	hg	e	e	i	33
lemoiz	ehha	e	e	i	33
getxo	ehha	e	e	i	33
errezil	hg	ia	ia	i	34
ondarroa	ehha	i	i	ie	35
ondarroa	hg	i	i	ie	35
elantxobe	hg	i	i	ie	35

Table 10. Summary of all the combinations attested in EHHA and HG

B. Attested combinations (ordered by dialect)

Dialect	Source	a-a	e-a	i-a	PatternID
abaurregaina	ehha	aa	ea	ia	1
aezkoaValley	hg	aa	ea	ia	1
ahetze	ehha	aa	ea	ia	1
aia	ehha	aa	ia	ia	3
aiherra	hg	aa	ia	ia	3
ainhoa	hg	aa	ea	ia	1
aldude	ehha	aa	ea	ia	1
alegia	hg	aa	ea	ie	2
alkotz	ehha	aa	ea	ie	2
altzai	ehha	aa	ia	ia	3
altzürükü	ehha	aa	ia	ia	3
amezketa	ehha	aa	ea	ie	2
andoain	ehha	aa	ia	ia	3
aniz	ehha	aa	eœ	ie	29
antzuola	hg	ia	ia	ia	15
aramaio	ehha	ie	ie	ie	24
aramaio	hg	ie	ie	ie	24
arano	hg	aa	ia	ia	3
araotz	ehha	ia	ia	ia	15
arbizu	ehha	aa	ea	ia	1
arbona	hg	aa	ia	ia	3
arboti	ehha	aa	ia	ia	3
armendaritze	ehha	aa	ia	ia	3
arnegi	ehha	aa	ia	ia	3
arrangoitze	ehha	aa	ea	ia	1
arrasate	ehha	ie	ie	ie	24
arratiaValley	hg	ea	ea	ie	8
arratzu	hg	ie	ie	ie	24
arrazola	ehha	ie	ie	ie	24
arrieta	ehha	ie	ie	ie	24
arroa	ehha	aa	ia	ia	3
arrueta	ehha	aa	ia	ia	3
asteasu	ehha	aa	iə	ia	32
ataun	ehha	aa	ea	ie	2

azkaine	ehha	aa	ia	ia	3
azkoitia	ehha	aa	ie	ie	6
azkoitia	hg	ie	ie	ie	24
azpeitia	ehha	aa	ie	ie	6
azpeitia	hg	ie	ie	ie	24
baigorri	ehha	aa	ea	ia	1
bakio	ehha	ie	ie	ie	24
bardoze	ehha	aa	ia	ia	3
barkoxe	ehha	aa	ia	ia	3
basaburua	hg	aa	i	i	31
bastida	ehha	aa	ia	ia	3
baztanValley	hg	aa	ia	ie	4
beasain	ehha	aa	ea	ie	2
behorlegi	ehha	aa	ia	ia	3
beizama	ehha	aa	ia	ia	3
bergara	ehha	aa	ia	ia	3
bermeo	ehha	aa	e	ie	28
bermeo	hg	ie	ie	ie	24
berriz	ehha	ie	ie	ie	24
beruete	hg	aa	e	i	26
beruete	ehha	aa	e	ie	28
beskoitze	ehha	aa	ia	ia	3
beskoitze	hg	aa	ia	ia	3
bidarra	ehha	aa	ea	ia	1
bolibar	ehha	ie	ie	ie	24
busturia	ehha	ie	ie	ie	24
deba	ehha	aa	ia	ia	3
dima	ehha	ea	ea	ie	8
domintxaine	ehha	aa	ia	ia	3
donamaria	ehha	aa	ia	ia	3
donibane	ehha	aa	ia	ia	3
donostia	ehha	aa	ia	ia	3
dorrao	ehha	aa	ia	ia	3
eibar	ehha	ia	ia	ia	15
elantxobe	ehha	ie	ie	ie	24
elantxobe	hg	i	i	ie	35
elduain	ehha	aa	ea	ie	2
elgoibar	ehha	ia	ia	ia	15
elorrío	hg	ia	ia	ia	15
elorrío	ehha	ia	ia	ia	15
erratzu	ehha	aa	ee	ie	29
errezil	ehha	aa	ia	i	30
errezil	hg	ia	ia	i	34
errigoiti	ehha	ie	ie	ie	24
erronkariValley	hg	aa	ia	ia	3
eskiula	ehha	aa	ia	ia	3
etxalar	ehha	aa	ia	ia	3
etxaleku	hg	aa	ea	ie	2

etxaleku	ehha	aa	ea	ie	2
etxarri	ehha	aa	e	i	26
etxarri	hg	aa	ie	ie	6
etxebarri	ehha	ie	ea	ie	20
etxebarria	ehha	ie	ie	ie	24
eugi	ehha	aa	ia	ie	4
ezkio	ehha	aa	ea	ie	2
ezkurra	ehha	aa	ea	i	25
ezterenzubi	ehha	aa	ia	ia	3
gaintza	ehha	aa	ea	ie	2
gamarte	ehha	aa	ia	ia	3
gamiz	ehha	ia	ia	ie	16
garrüze	ehha	aa	ia	ia	3
getaria	ehha	aa	ia	ia	3
getxo	hg	e	e	i	33
getxo	ehha	e	e	i	33
gizaburuaga	ehha	ia	ie	ie	18
goizueta	hg	aa	ea	ia	1
goizueta	ehha	aa	e	ia	27
hazparne	ehha	aa	ia	ia	3
hendaia	ehha	aa	ia	ia	3
hernani	ehha	aa	ia	ia	3
hondarribia	ehha	aa	ia	ia	3
ibarruri	ehha	ie	ie	ie	24
igoa	ehha	aa	i	i	31
ikaztegieta	ehha	aa	ea	ie	2
irisarri	ehha	aa	ea	ia	1
itsasu	ehha	aa	ea	ia	1
izturitze	ehha	aa	ia	ia	3
jaurrieta	ehha	aa	ia	ia	3
jutsi	ehha	aa	ia	ia	3
kortezubi	ehha	ie	ie	ie	24
landibarre	ehha	aa	ia	ia	3
larrabetzu	ehha	ia	ia	ie	16
larraine	ehha	aa	ia	ia	3
larraul	hg	aa	ia	ia	3
larrauri	hg	ia	ia	ie	16
larzabale	ehha	aa	ia	ia	3
lasarte	ehha	aa	ia	ia	3
laukiz	ehha	e	e	i	33
legazpi	ehha	ea	ea	ie	8
legazpi	hg	ea	ea	ie	8
leintz	ehha	ia	ia	ia	15
leioa	ehha	ea	ea	ie	8
leitza	ehha	aa	ea	ie	2
lekaroz	ehha	aa	eœ	ie	29
lekeitio	hg	ia	ia	ia	15
lekeitio	ehha	ia	ia	ia	15

lemoa	ehha	ea	ea	ie	8
lemoiz	ehha	e	e	i	33
lezama	hg	ia	ia	ie	16
literaryBizkaian	hg	ea	ea	ia	7
lizarraga	hg	aa	ie	ie	6
lizartza	hg	aa	ea	ie	2
luzaide	ehha	aa	ea	ia	1
makea	ehha	aa	ia	ia	3
mañaria	ehha	ie	ie	ie	24
markina18thcent	hg	ia	ia	ia	15
mendaro	ehha	aa	ia	ia	3
mendata	ehha	ie	ie	ie	24
mezkiritz	ehha	aa	ea	ie	2
montori	ehha	aa	ia	ia	3
mugerre	ehha	aa	ia	ia	3
mungia	ehha	ia	ia	ie	16
mutriku	hg	ia	ia	ia	15
oderitz	ehha	aa	e	i	26
oiartzun	ehha	aa	ia	ia	3
oñati	hg	ia	ia	ia	15
oñati	ehha	ia	ia	ia	15
ondarroa	ehha	i	i	ie	35
ondarroa	hg	i	i	ie	35
orexa	ehha	aa	ea	ie	2
orio	hg	aa	ia	ia	3
orio	ehha	aa	ia	ia	3
orozko	hg	ea	ea	ie	8
orozko	ehha	ea	ea	ie	8
otxandio	ehha	ie	ie	ie	24
pagola	ehha	aa	ia	ia	3
pasaia	ehha	aa	ia	ia	3
santagrazi	ehha	aa	ia	ia	3
sara	ehha	aa	ea	ia	1
sara	hg	aa	ea	ia	1
senpere	ehha	aa	ia	ia	3
sohüta	ehha	aa	ia	ia	3
sondika	ehha	ea	ea	ie	8
suarbe	ehha	aa	ia	ie	4
sunbilla	ehha	aa	ia	ia	3
tolosa	ehha	aa	ea	ia	1
uharte	ehha	aa	ia	ia	3
ultzama	hg	aa	ia	ie	4
urdiain	hg	ia	ia	ia	15
urdiain	ehha	ia	ia	ia	15
urdiñarbe	ehha	aa	ia	ia	3
urepel	hg	aa	ea	ia	1
urketa	ehha	aa	ia	ia	3
urretxu	ehha	aa	ia	ia	3

ürüstoi	ehha	aa	ia	ia	3
usurbil	hg	aa	ia	ia	3
uztaritze	ehha	aa	ia	ia	3
zaldibia	hg	ea	ea	ie	8
zaraitzuValley	hg	aa	ia	ia	3
zaratamo	ehha	ia	ea	ie	14
zarautz	hg	aa	ia	ia	3
zeanuri	ehha	ea	ea	ie	8
zeberio	ehha	ea	ea	ie	8
zegama	hg	aa	ea	ie	2
zegama	ehha	aa	ea	ie	2
zilbeti	ehha	aa	ea	ie	2
zollo	ehha	ie	ie	ie	24
zornotza	ehha	ie	ie	ie	24
zuberoaProvince	hg	aa	ia	ia	3
zugarramurdi	ehha	aa	ea	ia	1
zumaia	hg	aa	ia	ia	3

Table 11. Summary of all the combinations attested in EHHA and HG