Linguistic Variation in the Basque language and Education

Euskararen bariaazioa eta bariaazioaren irakaskuntza

Editoreak:
Gotzon Aurrekoetxea
Asier Romero
Aintzane Etxebarria
AURKIBIDEA

Aurrekoetxea, G., Romero, A. eta Etxebarria, A.:
Sarrera............................................................................................................................................ 2 - 3

Esteve-Gibert, N.:
The role of gesture and prosodic cues in the development of pragmatic meanings....................................................................................................................................... 4-14

De Pablo, I., Romero, A., Gaminde, I. eta Etxebarria, A.:
Keinuen eta bokalizazioen arteko harremanak komunikazio-funtzio goiztiaren funtzioaldeko euskara hizkuntzatik................................................................................................................. 15-28

Romero, A., Gaminde, I., Etxebarria, A. eta Sanz, A.:
Sinesgaiztasuna galderetan estrategia audiobisualen bitartez: prosodia eta keinuk................................................................................................. 29-38

Etxebarria, A., Gaminde, I., Romero, A. eta Sanz, A.:
Keinuaren eta azentuaren zerotzaren gauza euskaldun elebidunetan................................................................................................................................. 39-49

Garay, U., Gaminde, I., Etxebarria, A. eta Romero, A.:
Harridura adierazteko estrategia bisualak................................................................................... 50-58

Aurrekoetxea, G., Fernandez-Agirre, K., Ormaetxea, J.L. eta Rubio, J.:
Hizkuntza-bariozioa eta konbergentzia-prozesua: aldagai anizteko analisia euskararen kasuan........................................................................................................................... 59-76

Etchebest, X.:
Zubererako dardakari eta sudurkarien azterketa soziofonetikoa............................................. 77-80

Arcelus, A.:
Ahozko euskaran gertatzen diren arau fonologikoak, lexikoak eta postlexikalak.......................................... 81-92

Rada, L.:
Bizkaiko Gazteen Jarrerak Code-Switchingarekin.................................................................... 93-113

Videgain, X. eta Aurrekoetxea, G.:
Lehen kolpearen teoria eta EHHA...................................................................................................... 114-126

Masayuki, I.:
The ergative case in Lecumberry dialect.......................................................................................... 127-131

Aurrekoetxea, G., Iglesias, A., Santander, G. eta Usobiaga, I.:
'Diately' muga dialektalak zehazteko tresna.................................................................................. 132-144

Léonard, J. L., Heinsalu, E., Patriarca, M. eta Darlu, P.:
Modeling Regional Variation from EAS: complexity and communal aggregates.......................... 145-172

Ensunza, A., Garay, U., Gaminde, I. eta Romero, A.:
Bizkaiko gazte euskaldunen bai/ez galderen intonazioa................................................................ 173-182

Gaminde, I. eta Gandarias, L.:
Irizpide batzuk bariozioa aztertzeko intonazioan......................................................................... 183-205

Unamuno, L. eta Elordieta, G.:
Bariozioa Bizkaiko iparraldeko doinu-azentu sistemako berba azentugabez osatutako osagai prosodikoen eraketa.......................................................................................... 206-225

Aurrekoetxea, G, Arandia, A., Calvet, J. eta Camino, M.:
Oharmentzeko dialektologia euskaran: lehen urratsak..................................................................... 226-247
Abstract

The EAS data on auxiliary inflection in modern Basque dialects (elder population, aged over 40) is being processed from the standpoint of Complexity Theory (CT), i.e. comparing phylogenetic, ontogenetic and epigenetic results using several methods of computing interdialectal distance and algorithms (patristic distances versus Levenshtein distance). Moreover, a model for pondering morphophonological differentiation is proposed—the SFVL model (preserve Structure, Feed, Void & Loop)—, congruent with basic assumptions in current phonological and morphological theories, such as OT (Optimality Theory) in phonology and Word & Paradigm in morphology. Both cladistic and dialectometric results point at a threefold division between Eastern, Central and Western Basque, though the phylogenetic output from a sample of the EAS data clearly splits the central dialect into two components of the Eastern and the Western areas. Pondering innovations according to the SFVL model highlights the incidence of the road network on the diffusion of emerging norms, at least in the generational sample observed. Further results are expected applying similar CT methods on the younger group of speakers surveyed by the EAS project, in order to investigate to what extent these intricate patterns of divisions, tropisms and diffusion underwent change within the last decades, during which linguistic planning and the prestige of Euskara batua have played a role on Basque revitalization and structural change.

Key words: Complexity, morphophonology, cladistics, dialectometry, sociolinguistics.

1. Introduction: isoglosses & Complexity Theory

One of the basic flaws in many discussion about geolinguistic frontiers and aggregates stems from undermodeling isoglosses—whatever qualitative or quantitative efforts may be invested in the task—missing basic assumptions of what is called nowadays Complexity Theory (hence, CT). Among these prerequisites, autoorganization of aggregates out of flows of information, emergence of deterministic patterns instead of sharps constructs, and vicariance (multiplicity of standpoints, see Berthoz 2013) instead of one-sided explanations, provide a more realistic vision of dialect diversity within a language or a linguistic stock than current descriptions usually do. The fact that dialect frontiers are never definitive and always fluctuate should rather be considered as a heuristic premise than as a blind alley or an aporia. Geolinguistic patterns do indeed fluctuate, because they make up complex systems—so that they

11 See http://lipn.univ-paris13.fr/~duchamp/: Gérard H. E. Duchamp, initially trained by Marcel-Paul Schützenberger, has been a seminal and powerful "enzyme" in this endeavour to apply Complexity Theory to geolinguistics and sociolinguistics, since 2013.
should be processed as such, within the framework of Complexity Theory (Gribbin 2004; for applications to sociolinguistics and linguistic ecology, see Massip-Bonet & Bastardas-Boada 2013, Mufwene 2001, 2012). We will make here an attempt at multimodeling dynamic social aggregates through the Euskararen Atlas Sociolinguistikoa (EAS, 2010, see Unamuno, Ensunza, Iglesias & Ormaetxea 2012) data (questions V144-182: auxiliary inflection), surveying emerging dialect networks of older speakers from the standpoint of geosocial aggregates at various scales of magnitude through a $D$-normalizing analysis, according to a set of methods (scaled normalization of Levenshtein’s distance index, etc.) and cladistics processing of the 12 first variables of the EAS section on auxiliary inflection (see Table I in the appendix).

Indeed, in a CT approach, results are holistically interpreted from the standpoints of Deterministic Complexity (internal & external factors, hence DC), Algorithmic Complexity (identifying rules and controlling commands for data processing, hence AC) and Aggregate or Communal Complexity (social networks, diffusion, hence CC). The EAS answers have been documented for 94 locations (see list in Table II in the appendix).

AC will be understood here as multivariate algorithmic processing of the EDAK/EAS data and mapping proper on the one hand, providing cladistic phylograms according to patristic distances, and Levenshtein distances, and the use of declorative parameters defining inner diversity of the grammar of Basque dialects according to the SFVL model (diasystemic variables, see section 2 below) on the other hand. The latter will feed the former, as declarative variables will support ponderation and computation on raw data from the EAS database. CC, in turn, will apply to AC results: what kind of aggregates or communal agents eventually show up from cladistics and dialectometric analysis? Do geolinguistic aggregates vary according to town dialects (i.e. irradiating urban centers, or urban nodes at the crossroad between innovative centers) or according to rural networks as satellite dialects gravitating around leading centers, or to a combination of both (i.e. town & satellite dialects)? Or should we rather analyze diffusion processes as the result of the action of force fields, where many individual urban or rural aggregates compete, giving shape to intricate sets of substructures, in terms of communal aggregates? The relevance of grammar over the lexicon and phonology is well known, as far as Basque dialect classification is concerned. Auxiliary verbs have been considered as a powerful criterion, since the beginning of Basque dialectology (Luis-Lucien Bonaparte, Manuel Larramendi, Resurrección Maria de Azkue, etc.). Though, the data processed here provide only a glimpse at geosociolinguistic patterns: not only should multivariate analysis be implemented on each main component of the diasystem (i.e. phonology, morphology, syntax and the lexicon), but also on several subsystems, as nominal declension.

---

12 A trend of CT studies has also evolved in sociolinguistics, with a strong concern upon bilingual competition modeling, as Patriarca & al. 2012, Minett & Wang 2008, Heinsalu & al. 2013, etc.
13 A sample of 12 variables already provides robust results with the standard cladistics methods (see discussion on the size of the sample in Corvaglia & al. 2008). Though, we indeed plan to process the whole EAS aditza corpus in further research. About “standard” use of cladistics methods in dialectology, see Brun-trigaud & al. 2014, Gaillard-Corvaglia 2012.
14 This threefold patterning of DC, AC & CC has been elegantly developed by O’Sullivan, 2004 in a seminal contribution on CT assets for human geography.
Therefore, a CT approach based on auxiliary inflection is all the more strategic, as an attempt to develop appropriate methods from a more holistic point of view, in order to explore the diasytem and compare generational gaps through the EAS data (here, 94 localities and 3,816 tokens). This paper should be considered as the first step of a more extensive research. By no means it pretends to give a definitive assessment on dialect classification—though it significantly provides new and unexpected results, though congruent with former research and standard theories about Basque dialect areas (e.g. Zuazo 2013)\textsuperscript{16}.

2. Modeling Morpho(phono)logical Rules

As mentioned above, AC entails that we implement a declarative model in order to qualify variables for automatic processing of data. This model should be contrived uphill, in order to be able to account for the shape of results from the standpoint of dialectology. If data are processed trivially, it will give results, but these will still be questionable, as these results are under-informed, as if one would be comparing pears with oranges, or horses with chicken. If we want to obtain relevant results as far as DC and CC are concerned—i.e., about what may explain geolinguistic patterns and how they are reliable in terms of social determinism, we need to order facts and ponder every detail. Up to now, algorithms are not very good at doing that. Indeed, the Levenshtein algorithm can be run according to basic commands of ponderation, i.e. according to insertion or deletion of characters, and such ponderation can be monitored quite flexibly. We’ll use here this flexibility implementing our declarative model for morphophonological analysis (hence SFVL: preservation of Structure, Feed, Void and Loop). Instead, cladistics allows even a better balance of ponderation, inserted in the derivational graphs, as pondered items (or tests, see below) cumulate within chains of derived characters. This leads us to the SFVL model and its prerequisites—whose main labels are listed in (1) below. Table 1 shows some instances of these parameters in the rightmost column, such as PRESERVE X, FEED Stridency, FEED Stridency_Palatal, VOID Root_Nucleus, etc. Each parameter is associated to a type (A, B, C…), which can be added further to new descriptions (with the & symbol) as C & FEED Stridency in locality 134 (leftmost column). Local parameters enumerated in the rightmost column declare first the kind of process (S, F, V or L) in capital letters (PRESERVE, FEED, VOID or LOOP\textsuperscript{17}), followed by a short description of the process at stake—hinted at either through features, such as stridency or through syntagmatic operations such as inversion—and of its domain of application (root, affixes, Outer Coda, Left_Nucleus, Root CV, etc.), see tables 1 and 2.

\textsuperscript{16}Recent dialectometric research on Tseltal, a Western Mayan language of Chiapas based on the same kind of CT multivariate analysis as used here, suggests that the lexicon provides the most congruent results with “standard classifications”. Phonology shows similar results, though more intricate, whereas—interestingly enough—morphosyntax points at more interactive patterns, resorting to dynamic diversification but also contact. This entails that homoplasy tends to be stronger in the morphosyntactic component (MSC) of a diasytem, though underlying patterns of diversification may still be enhanced in MSC results (see especially Léonard & al. 2014 and Léonard & Polian 2009, 2013, 2014).

\textsuperscript{17}In generative phonology, FEED and VOID are often associated with other terms, such as BLEED (for subsequent rules in a morphophonological set of ordered rules or constraints), either for the study of cyclical processes in synchrony, or for the survey of phonological evolution. Indeed, we initially borrowed these terms from this framework, but we cling to a more trivial interpretation of these terms here, interpreting them rather as processes akin to licensing (FEED) and government (VOID), i.e. either filling or emptying inner structures or (morpho)phonological segments. For a much more sophisticated application of the “standard” generative model in (morpho)phonology, see Baković 2013.
Main operations according to the SFVL model

**Preserve STRUCTURE (S): conservation.** Initial state. In other words, nothing happens. Here, the basic forms will be those of standard Basque, which can be considered as both closer to the protoforms and as lemma for rooting the analysis.

**FEED (F): licensing** (i.e. enriching inner structure of phonological or morphological units), modification, constituent interaction, such as *palatalization*, *epenthesis*, etc.

**VOID (V): government** (i.e. weakening or deletion of elements, triggered by subsequent head units), dropping, e.g. *deletion.*

**LOOP (L): reanalysis, constituent inversion, e.g. *metathesis.*

As a model for the study of emerging norms, the SFVL grid relies on an adaptation of the *social cognitive model* of language change proposed by Hruschka & al. (2009), consisting of *form-function reanalysis, form-meaning mapping, iterated learning, replicators* and *exemplars.* At both ends of a polarity between conservation and change, preservation of Structures (the *S* parameter) resorts to trivial *exemplars,* whereas *Loop* (labelled here the *L* parameter) results from *form-function reanalysis.* Forms available in both extremes work as *exemplars,* and may change under the pressure of *iterated learning.* Forms spreading from a sociocognitive norm to another, through contact or imitation, are *replicators.* *Feed* and *Void* (*F & V*) modify substantially *exemplarity,* as feeding processes (*F* parameters) enrich exemplars, whereas voiding processes (*V* parameters) make them lighter. For instance, while *naiz* “I am” goes unmodified, items such as *naz* (type C, table 1), *niz* (type D) or *na* (type G) undergo processes of the *VOID* typ.e which have more or less strong consequences on the *form-meaning mapping* of these inflected forms – though, of course, these more or less dramatic morphophonological changes still remain at the brink of semantic shift or change, so that they do not endanger form-meaning mapping. To some extent, though mildly, it is also true of types *B* and *B’* (*nais* and *naix*), which show variation on the stridency of the coda fricative, challenging even less form-meaning mapping than the previous processes of the *VOID* type.

<table>
<thead>
<tr>
<th>EAS N°</th>
<th>Type</th>
<th>[na-iz]</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>330</td>
<td>A</td>
<td><em>naiz</em></td>
<td>PRESERVE X</td>
</tr>
<tr>
<td>118</td>
<td>B</td>
<td><em>nais</em></td>
<td>FEED Stridency</td>
</tr>
<tr>
<td>119</td>
<td>B’</td>
<td><em>naix</em></td>
<td>B &amp; FEED Stridency_Palatal</td>
</tr>
<tr>
<td>133</td>
<td>C</td>
<td><em>naz</em></td>
<td>VOID Root_Nucleus</td>
</tr>
<tr>
<td>703</td>
<td>D</td>
<td><em>niz</em></td>
<td>VOID Prefix_Nucleus</td>
</tr>
<tr>
<td>134</td>
<td>E</td>
<td><em>nas</em></td>
<td>C &amp; FEED Stridency</td>
</tr>
<tr>
<td>120</td>
<td>F</td>
<td><em>nai</em></td>
<td>VOID Outer_Coda</td>
</tr>
<tr>
<td>136</td>
<td>G</td>
<td><em>na</em></td>
<td>VOID VC Root</td>
</tr>
</tbody>
</table>

Table 1. Auxiliary “be”, 1 SG. *naiz: “I am”* as a sample of parameters according to the SFVL model.

Examples of *Loop* appear in table 2 (“be” in the 2PL cell of the paradigm: “you are”): *zarate* and *sarete* resort to a *S* pattern (no change, or insignificant), whereas *serate* (type B here) implies inversion of the vowel chain (called here *vocalic melody*), between the prefix and the lexical root of the auxiliary –amounting to a *LOOP.* Type C *zaete* implies a gap in the *form-meaning mapping,* described here as *VOID Root Onset,* as a result of intervocalic rhotic deletion (*zarate > zaete*). Interestingly enough, the reanalysis through *Inverse_Melody_Px-Root* looping can be iterated in other dialects, as
to be seen with type D zeate, which incorporates type C, and is therefore described as C & Inverse_Melody_Px-Root (zaete > seate).

In subsequent analysis (section 3) we will call test each of these answers enumerated from tables 1 & 2 (naiz, nais, naix…; zarete, zerate, zeate, zaete…), as these forms were elicited in face to face interaction in different localities, testing geo- and sociolinguistic variation in the Basque diasystem.

<table>
<thead>
<tr>
<th>EAS</th>
<th>[za-re-te] 5Ab.Pres-Lx-Pl</th>
<th>Type</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>zarete</td>
<td>A</td>
<td>PRESERVE X</td>
</tr>
<tr>
<td>202</td>
<td>sarete</td>
<td>A’</td>
<td>FEED Stridency</td>
</tr>
<tr>
<td>302</td>
<td>serate</td>
<td>B</td>
<td>LOOP Inverse_Melody_Px-Root (sarete &gt; serge)</td>
</tr>
<tr>
<td>306</td>
<td>zaete</td>
<td>C</td>
<td>VOID Root_Onset</td>
</tr>
<tr>
<td>310</td>
<td>zeate</td>
<td>D</td>
<td>C &amp; Inverse_Melody_Px-Root (zaete &gt; seate)</td>
</tr>
<tr>
<td>326</td>
<td>tzéate</td>
<td>D &amp; E</td>
<td>FEED Pfx_Initial_Onset</td>
</tr>
<tr>
<td>404</td>
<td>zaaté</td>
<td>F</td>
<td>FEED Left_Nucleus</td>
</tr>
<tr>
<td>408</td>
<td>zate</td>
<td>G</td>
<td>VOID Root_CV</td>
</tr>
<tr>
<td>301</td>
<td>sare</td>
<td>H</td>
<td>VOID CV_Sfx</td>
</tr>
<tr>
<td>402</td>
<td>zabie</td>
<td>J</td>
<td>FEED Labial_Epenthetic_Hiatus (*zarie &gt; *saie &gt; *sawie &gt; sabie)</td>
</tr>
<tr>
<td>506</td>
<td>zaizte</td>
<td>K</td>
<td>LOOP Reanalysis -iz-</td>
</tr>
<tr>
<td>503</td>
<td>ziezte</td>
<td>L</td>
<td>LOOP Inverted Contour_Pfx-Root (zaizte &gt; ziezte)</td>
</tr>
<tr>
<td>425</td>
<td>zizte</td>
<td>M</td>
<td>VOID &amp; LOOP Pfx_Nucleus</td>
</tr>
</tbody>
</table>

Table 2. Auxiliary “be”, 2 Pl. zarete: “you are” as a sample of parameters according to the SFVL model

Table 3 resumes the main correlates of the SFVL model, opposing two major trends in linguistic change: towards change or towards levelling. As these trends are cyclic and entail even inner cyclicity within paradigmatic changes and/or paradigmatic levelling, the relation between input (lexical form) and output (realizational form) alternate in the second row of the table. The chart takes over several points already enumerated in (1), but its main purpose is to correlate declarative factors (fourth row), structural consequences of change (fifth row), structural configurations and symmetry/asymmetry relationships between paradigms and subsystems (sixth row), and semiotic value or strength (last row).

<table>
<thead>
<tr>
<th>CHANGE</th>
<th>LEVELLING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input =&gt; Output</td>
<td>Output =&gt; Input</td>
</tr>
<tr>
<td>Input =&gt; Output</td>
<td>Output =&gt; Input</td>
</tr>
<tr>
<td>FEED</td>
<td>VOID</td>
</tr>
<tr>
<td>Modification</td>
<td>Neutralization</td>
</tr>
<tr>
<td>Interaction</td>
<td>Deleting</td>
</tr>
<tr>
<td>Pattern</td>
<td>Default</td>
</tr>
<tr>
<td>Markedness</td>
<td>Inner Skipping</td>
</tr>
</tbody>
</table>

Table 3. Structural change correlates according to the SFVL model
3. Mapping Patterns in Space: Topologies

3.1 Tools and data set
3.1.1 Geographical visualization and data set

Geographical representation can in general provide an immediate and intuitive picture of a certain situation allowing the direct visualization of a large amount of information which would be otherwise difficult to evaluate if presented e.g. in the form of tables.

In particular, it is a precious tool for highlighting the linguistic connections between different areas or populations, as shown for example by the compilation of word atlases.

We have used geographical visualization tools providing maps drawn with the help of GPS-based software, such as Google-Earth, in order to visualize automatically the data relative to the geographical locations where the tests were done, the different results of the tests, as well as the corresponding network structure between the different locations obtained from the statistical analysis of the results. Figure F1 illustrates the about one hundred locations (namely, 94) where tests were carried out by the EAS team in recent years.

![Figure F1. Geographical representation of the 94 locations where the AES tests were done.](image)

The data set available consists in the answers to a series of tests carried out in various Basque locations by the EAS team of the UPV (Universidad del País Vasco). Hence, tests will be labelled with the index \( k \), running from \( k = 1 \) to the total number of tests \( N_T \), while the locations will be labelled by indexes \( i \) (or \( j \)), running from \( i = 1 \) \( (j = 1) \) to the total number of locations \( i = N_L \) \((j = N_L)\). The answer to the \( k \)th test carried out in location \( i \) is represented as a certain string \( a_{ik}^k \) where \( k \) denotes the test and \( i \) the location where the test was carried out. All data obtained from the tests can be combined in a matrix of strings \( A \),

\[
A = \{ a_{ik}^k \}, k = 1,\ldots,N_T; i = 1,\ldots,N_L.
\]
Once the locations coordinates (latitudes and longitudes) are known, geographical tools can be used to draw word atlases directly from the data sets of the different tokens obtained as answers to tests. As an example, figures F2 and F3 visualize geographically the tokens $a_{i}^{1}$ and $a_{i}^{2}$ obtained in the tests $k = 1$ and $k = 2$ corresponding to IZAN $\rightarrow$ naiz “I am” and IZAN $\rightarrow$ zarete “You are”, respectively.

Figure F2. Geographical visualization of the results of test $k = 1$, i.e. the variants of the token IZAN $\rightarrow$ naiz « I am ». (file : naiz.png)

Figure F3. Geographical visualization of the results of test $k = 2$, i.e. of the variants of the token IZAN - zarete "you are". (file: zarete.png)
3.1.2 Levenshtein Distance between tokens

The Levenshtein distance $L(a, b)$ is a basic measure of the level of difference between two strings $a$ and $b$. Given two strings $a$ and $b$, for example $a = \textit{nais}$ and $b = \textit{nas}$, the Levenshtein distance $L(a, b)$ is defined as the minimum number of operations (represented by insertions, deletions, or editions) needed to turn $a$ to $b$ or vice versa. The Levenshtein distance $L(a, b)$ has the merit to have a simple definition and therefore to be straightforward to use. This same fact, however, may represent a limit when addressing more precise questions, as we discuss in the next sections. In fact, the Levenshtein distance is independent of

– the type of the actual operations used, whether insertions, deletions, or editions;
– the characters involved in the operation (e.g. if it is $a \rightarrow e$ or $a \rightarrow i$);
– the order in which operations are performed or the changed characters in the token.

As an example, the following scheme describes how the Levenshtein distances between the three tokens $\textit{nas}$, $\textit{nais}$, and $\textit{naiz}$ is computed:

- $\textit{naiz} \rightarrow \textit{nais}$ only one “z”→“s” edit $L(\textit{naiz}, \textit{nais}) = 1$
- $\textit{nais} \rightarrow \textit{nas}$ only one “i” delete $L(\textit{nais}, \textit{nas}) = 1$
- $\textit{naiz} \rightarrow \textit{nais} \rightarrow \textit{nas}$ one “z”→“s” edit and one “i” delete $L(\textit{naiz}, \textit{nas}) = 2$

In the following, the distances between two different tokens $a_i^k$ and $a_j^k$ representing the answers to the same test $k$ obtained in different locations $i$ and $j$ will be measured through the corresponding Levenshtein distances $d_{ij}^k = L(a_i^k, a_j^k)$.

3.2 Raw results obtained without SFVL method
3.2.1 Construction & Visualization of the Similarity Network (without SFVL method)

Visual representations of the connections between different units (the different dialects in this case) is based on the construction (through some algorithm discussed below) of the matrix of all the mutual distances between the units, $D = \{d_{ij}\}$, and on the graphical representation of the corresponding network topology. For instance, one can plot all the units $\{i\}$ as the nodes of a network embedded in the geographical background and draw all the links between units with different thickness inversely proportional to the distance, in order to emphasize stronger tights.

Alternatively, one can introduce a cutoff $d_{\text{MAX}}$ and only plot the links corresponding to distances $d_{ij} < d_{\text{MAX}}$.

For a more effective graphical representation here below we do both things. First, we construct the matrix $D = \{ d_{ij}\}$ of the average distances between different dialects. Then we plot the links between location pairs $i$ and $j$ only if $d_{ij} < d_{\text{MAX}}$ and with a thickness that is a decreasing function of $d_{ij}$.

Here below in figure F4 we present a simple example of geographical visualization in which the average distances $d_{ij}$ are obtained as arithmetic averages of the Levenshtein distances between the strings representing the answers to all the test $k = 1, \ldots, N_T$ obtained in all pairs of locations $i$ and $j$ in which the test $k$ was done. In this case the formula for the average distances used is the simple arithmetic average.
Representations of the unweighted dialect network are given in figure 4. At the top, a network corresponding to a lower intensity of differentiation ($d_{\text{MAX}} = 1.8$); at the bottom, a higher intensity of differentiation ($d_{\text{MAX}} = 2$) for the same date (AES tests, elders). The whole corpus of auxiliary inflection has been processed here, without any further intervention of the linguist than the trivial application of Levenshtein algorithm as explicated above. Even though, discrete and congruent patterns clearly emerge already at the $d_{\text{MAX}} = 1.8$ threshold of differentiation: a threefold division between a Western dialect (Bizkaian), a central dialect and an Eastern dialect (Navarrese-Lapurdian). Further subdialects or peripheral networks show up, as a southeastern Navarrese dialect (Zugarramurdi, Sunbilla, Lekaroz, Aniz), tightly embedded with the latter. Instead, the Zuberoan dialect appears either as a vertical strand (Urdinarbe-Larraine) or a as default zone (i.e. an empty space on the map, due to underscore as to the normalized $D$ threshold at stake). The Central dialect points at a complex internal structure, with subcomponents, as a core, in the Ezkio-Itsaso area, and compact embedded centers, such as the Getaria-Donostia-Hernani-Zestoa cluster on the North, and loosely embedded peripheries at the fringes, as Deba or Elgoibar on the west and Beruete or Etxarri in the South-east. The topology of $d_{\text{MAX}} = 1.8$ can be described mainly as *bulks* (the threefold division), *embedded patterns* (subdialects and strands) and *default areas*, i.e. constellations of unconnected varieties, such as Bajio, Laukiz or Mungia in the West, Hondarribia in the East, Urdiaín or Arbizu in the South and Sohúta or Luzaide in the East. Most of these default areas (or varieties) instead start their clustering process at further stages of the differentiation, as in the lower figure ($d_{\text{MAX}} = 2$).
Although the $d_{\text{MAX}} = 2$ results may seem intricate, the map provides a strong picture of communal clusters according to the *aditza* variables: i) it enhances the threefold division within the Basque dialectal network, and links the Easternmost dialect of Zuberoa (Souletin) with the Navarrese-Lapurdis dialect; ii) it highlights inner structures of each dialects through subclusters: the Westernmost dialect (Bizkaia) is not loose anymore: it gains in compactness, and it suggests a cyclic North-South division, leading from the valleys of the hinterland to the coast, and a *corniche* subdialect (Lekeitio-Bakio)\(^\text{18}\), as compared to more oblique stripes pointing at a South-west by North-east inclination in the central dialect; iii) it connects Zuberoa in the easternmost region with the bulk of Navarrese-Lapurdis; iv) it suggests some localities may work as crossroads for spreading innovations, as Bergara and Goizueta, which formerly appeared as default varieties in the former representation (threshold $d_{\text{MAX}} = 1.8$). The latter remark hints at a function of transition zones and cross-roads rather than isolates of the “default areas” skipped from networks with weaker threshold of differentiation. In other words, these representations point at *ontogenetic patterns* of diffusion – rather than *phylogenetic* –: they tell more about emerging clusters of innovations or about feature diffusion than about “classical” phylogenetic dialects, though being congruent with already established classifications. As such, they provide finer grain about geolinguistic areas and sociolinguistics networks, and the functions of these constructs.

A further question has to do with how can we measure better the impact of innovation, according to structural weight of the features at stake within the feature pool (Mufwene (2013: 322, 2001)? The Levenshtein algorithm being a routine, it

\(^{18}\) About the Corniche sub-dialect on the Bizkaian coast, see Léonard 2001.
unfortunately levels up the complexity of linguistic patterns. Could we contrive a better way to ponder variables and the results of the AES tests and implement this weight discrimination model into data proceeding? We will attempt at such a CT refinement through AC in the next section. Here, the SFVL model will be pondered and implemented into the Levenshtein algorithm, as a way of exploring further Algorithmic Complexity. We also expect from such a method to neutralize factors of bias, as the allocative responses to the AES inquiry, which apparently account for the star-like shape in the central part of the Navarrese-Lapurdin dialect in both maps in figure 4. Obviously, the allocative (HIKA and ZUKA) forms should be given a lower weight in order to control their incidence in shaping subnetworks. As we will soon see, a pondered version of the SFVL model turns out to be a solution to avoid this kind of bias.

3.2.2. A SFVL Levenshtein algorithm application: visualizing the Basque weighted network

A more general possibility to construct the dialect network is to start from an average Levenshtein distance between dialects in which each element of the average – i.e. each single token – is suitably weighted, depending on the type of change undergone. In this case the equation for the average distance between two assigned locations $i$ and $j$ should be generalized as

$$D_{i,j} = \frac{1}{\sum_{k=1}^{N_T} p_k d_{i,j}^k}$$

where the $p_k$ are the corresponding weights.

Table 4 shows the pondering grid used for the analysis taking into account structural weight. PRESERVE $X$, i.e. preservation of structure, from the lemmatic state, so to say, the reference dialect (Euskara batua), is given no weight. Positive changes run according to an increasing scale of values from F to L: Feed = 2, Void = 4, Loop L = 8. Three factors of bias are pondered according a neutralizing scale: Y for synthetic forms, as for variable V 147 (90310) edin [DADIN]: “We gave him some money so that he could go on holidays”, in 108 yoáteko, 109, joatêko, 120 dxúteko (synthetic forms) vs. 110 yoan deiten, 116 yoan daitén, etc. (analytical forms). In this case, analytical forms are analyzed according to the SFVL grid, while all forms elicited with the synthetic pattern (yoáteko, joatêko, dxúteko, etc.) are endowed an average neutral weight of 3. The EAS tests often gave ZUKA and HIKA forms as a result of the elicitation in some regions (especially Northern Basque). These specific paradigms strongly driven by pragmatics (as “vous” and “tu” agreement in French, Usted and tú in Spanish), induced a strong bias in the first results we got with Levenshtein distance. We therefore decided to give a 0.5 weight to these tests in order to neutralize this bias, which turned out quite efficient, comparing with prior results. As many tests give combined types, as already mentioned in tables 5.1 and 5.2, with combined processes, 119 naix type B’: B & FEED Stridency_Palatal, or 134 nas type E: C & FEED Stridency, the grid in table 4 is conceived in order to take into account the combination of processes (hinted at with the & symbol). A first set of values apply, in column WEIGHT (W), which may combine with further processes, as in the COMBINED WEIGHT (CW) column: values of the
former are divided by 2 in the later for the core positive processes (F, V, L). For S, Y, Z and H no further processes do apply.

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>ABBREVIATION</th>
<th>WEIGHT (W)</th>
<th>COMBINED WEIGHT (CW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRESERVE X</td>
<td>S</td>
<td>0</td>
<td>none</td>
</tr>
<tr>
<td>FEED</td>
<td>F</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>VOID</td>
<td>V</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>LOOP</td>
<td>L</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Synthetic</td>
<td>Y</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>ZUKA</td>
<td>Z</td>
<td>0.5</td>
<td>none</td>
</tr>
<tr>
<td>HIKA</td>
<td>H</td>
<td>0.5</td>
<td>none</td>
</tr>
</tbody>
</table>

Table 4. A SFVL grid for ponderating Levenshtein Distance

Tables 5.1 and 5.2 show a sample of this ponderation system applied to the same data as in tables 1 and 2 above, for naiz “I am” and zarete “you are”. This is a tentative application, though, as part of the richness of the declarative processes is levelled out by automatic computing of these values. Nevertheless, experience has proved that a first attempt according to an automatized procedure is necessary before trying a more fine-grained approach: the risk of ad hoc ponderation is as strong in both cases, but at least an automatic computation allows to check factors of bias more systematically. As the procedure of indexing the W and CW values is time consuming, we implemented this method on the 12 first variables, i.e. V144 (90010) izan [NAIZ]: “I am young”, V145 (90060) izan [ZARETE]: “You too are young”, V146 (90110) izan [GINEN]: “We too were young a long time ago”, V147 (90310) edin [DADIN]: “We gave him some money so that he could go on holidays”, V148(90360) edin [DAITEKE]: “Your friend can come at my place”, V149 (90410) izan [ZAIT]: “I like this meat”, V150 (90520) izan [ZITZAIDAN]: “Before, I used to like eating meat, too”, V151 (90590) izan [ZITZAIZKION]: “Before, my father used to like strawberries, too”, V152 (90700) edun [DUT]: “Today I didn’t eat much”, V153 (90730) edun [DU]: “He didn’t eat a lot”, V154 (90750) edun [DUZUE]: “You too, you didn’t eat a lot”, V155 (90760) edun [DUTE]: “My friends have eaten a sandwich”. A further reason to limit the pondered processing of Levenshtein distance on the data set to these 12 variables is that we could not process more variables for patristic distances either, in the span of time available for this research. For the sake of comparability of results from cladistics and dialectometry, we had to deal with a limited set of 12 variables, analyzing 1.167 tokens instead of 3.816, for the aditza corpus of the EAS survey. Although we could only check the validity of the SFVL cyclical weight on a quarter of the data set, it still proves encouraging, as suggests strong congruence between the unp pondered results and pondered results with Levenshtein distance. As to cladistics, the provisional output gets far better when applying the SFVL weighing method instead of automatic clustering of the whole data set.
Table 5.2 provides more examples of application of the SFVL grid, on a more complex variable (zarete “you are”), which entails more combinations of W and CW processes than Table 5.1. For instance, test sarete in locality 202 is given value 2 as a F expression (FEED Stridency), whereas serate in locality 302 receives a value string 8 + 1 = 9, for L (looping: LOOP Inverse_Melody_Px-Root) as a main process (W, valued 8), coupled with FEED as a secondary process (CW, valued 1). In 306, zaete rates 4 (VOID Root_Onset), whereas in 310 zeate inherits the LOOP Inverse_Melody_Px-Root, rated 4 as a secondary process, while it rates 4 as VOID Root_Onset, which is here the main positive process in the feature pool. With 326 tzéate, FEED Pfx_Initial_Onset stands as the main process (value: 2 as W) associated with LOOP melody inversion (value: 4 as CW) as in the previous set. In 404, zaaté is treated as a V & F string (4 + 2) (zarete > zaraté > zaate), as rhotic deletion (VOID) and low vowel propagation from the AGRS prefix (FEED) combine – although another interpretation could avoid the F rule, if a |za+ra+te| input were predicted, which entails the FEED process as CW rather than W in order to elude such a bias. In 408, zate implies VOID Root.CV (valued 4) and subsequent LOOP (valued 4 as CW), as meaning-form mapping is somewhat altered by elision of the auxiliary stem -ra-. In 301 and 110, sare and sarie are mainly characterized by V & L and F & L processes – so that the z- vs. s- stridency onset contrast turns out to be irrelevant here, according to a constraint of procedural parsimony. The last forms of the chart (402 zabie, 506 zaizte, 503 zietze, 425 zizte) strongly resort to various options of LOOP, combined with F or V, and even with further loop, as in the case of 503 zietze, from zeizte < zaizte.
### Table 5.2. The SFVL grid applied to zarete “you are”

<table>
<thead>
<tr>
<th>EAS</th>
<th>Parameters</th>
<th>Code</th>
<th>W</th>
<th>CW</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>za-re-te</td>
<td>PRESERVE X</td>
<td>S</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5Ab.Pres-Lx-Pl</td>
<td>FEED Stridency</td>
<td>F</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>zarete</td>
<td>LOOP Inverse_Melody_Px-Root</td>
<td>L &amp; F</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>202</td>
<td>sarete</td>
<td>VOID Root_Onset</td>
<td>V</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>302</td>
<td>zeate</td>
<td>LOOP Inverse_Melody_Px-Root</td>
<td>V &amp; L</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>306</td>
<td>zaete</td>
<td>FEED</td>
<td>F &amp; L</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>310</td>
<td>zaete</td>
<td>VOID Void_Px_Initial_Onset</td>
<td>V &amp; F</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>326</td>
<td>tzéate</td>
<td>FEED</td>
<td>F &amp; L</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>404</td>
<td>zaaté</td>
<td>FEED Void_Px_Initial_onset &amp; FEED Left_Nucleus</td>
<td>V &amp; F</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>408</td>
<td>zate</td>
<td>LOOP Reanalysis -iz-</td>
<td>L &amp; F</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>425</td>
<td>zisate</td>
<td>LOOP Void_Px_Initial_Onset &amp; LOOP Void_Px_Root</td>
<td>V &amp; L</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>503</td>
<td>zize</td>
<td>LOOP Void_Px_Initial_Onset &amp; LOOP Void_Px_Root</td>
<td>V &amp; L</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>506</td>
<td>zizte</td>
<td>LOOP Void_Px_Initial_Onset &amp; LOOP Void_Px_Root</td>
<td>V &amp; L</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 5 below shows representation of the weighted dialect network for the aditza variables (elders, 12 variables), pondered according to the SFVL model. Top: a lower intensity of differentiation ($d_{\text{MAX}} = 1.6$). Bottom: a higher intensity of differentiation ($d_{\text{MAX}} = 2$). The results of an application, albeit limited to 12 variables, of the SFVL model weight, proves fairly encouraging: not only do the three main clusters appear, within limits congruent to a broader application of the Levenshtein algorithm to the whole aditza corpus, but the HIKA and ZUKA forms do not distort any more inner patterns within the Eastern dialect. The inner structures of a peripheral dialect as Zuberoan have gained in consistency and compactness, and inner structures in main clusters are less intricate. Amphizones or transitional areas, or subdialects in the periphery appear more overtly too, as default areas: now the whole Bizkaian corniche, from Bakio to Deba at stage 1.6, and the westernmost part of the Bizkaian dialect appear as default zones, for both threshold of differentiation. Interestingly enough, the whole bulk of the s. c. Navarrese dialect (Sunbilla, Aniz, etc.) shows up as one of these default zones, pointing at transitionality rather than at singularity. Another interesting pattern is the link which in both topologies runs from the west of Legutiano to Donamaria in the South, hinting at a southern track for the spreading of innovations, from Southern Bizkaia to Southern Navarra. As those topologies tell more about flows of innovations or structural entropy (induced by the SFVL cumulated weights), i.e. ontogenetic trends, than about phylogenetic clusters, this southern track hints at relevant patterns to explore further. Nevertheless, these results are but a sample from 12 variables, and they will gain more relevance when being compared to phylogenetic results proper, from cladistics, from the same sample.
Figure 5. Representation of the weighted dialect network for the *aditza* variables. Top: a lower intensity of differentiation ($d_{\text{MAX}} = 1.6$). Bottom: a higher intensity of differentiation ($d_{\text{MAX}} = 2$).

We wish to add another methodological suggestion: in the analysis above (fig. 5) we fix the locations $i$ and $j$ and average over all the test $k$. One could check the complementary side, i.e. which (inflection form corresponding to a certain) test varies more within the Basque region, by averaging the distances over all location pairs (for
each fixed test $k$). It might turn out that some forms could either be more used or vary more than others. We leave this point for further research.

3.3. The phylogenetic standpoint: cladistics

The same steps will be applied to the phylogenetic approach through cladistics: first, we’ll base on the answers (or tests) to the 39 aditza questions (i.e. variables) of the EAS section dealing with auxiliary inflection (variable V144 to V182, see Table I in the appendix), in section 3.3.1; second, we’ll apply the cladistic method proper to a sample of the first 12 variables of the EAS aditza corpus for elder speakers, in section 3.3.2. The EAS answers have been documented for 94 locations (see Table II in the appendix).

3.3.1. Unpondered cladistic results

The results obtained with the whole sample of data (42 variables, see figure 5) are less reliable than those we’ll soon observe in the following section, from a fragment of the sample (12 variables): a cluster such as 336 (Hondarribia), 406 (Etxarri), 601 (Aldude) and 617 (Lazarbale) is both linguistically and geographically inconsistent. The central dialect mingles with too many other clusters, giving a confused picture of a dialect far less intricate, when looking at linguistic facts. Inconsistencies here point out at the risk of not monitoring data downhill with grammatical models uphill.
In the next section, we shall apply a more regular use of the PAUP algorithm to the same fragment of the data as previously (fig. 5), when applying the SFVL model to 12 variables of the corpus.

3.3.2. Cladistic processing according to the SFVL model and derivational graphs

In order to apply the cladistic procedure allowing to integrate hypotheses on the evolution of the morpho-phonological traits under scrutiny and to obtain clusters of the different locations, the regional variations of each of these traits were first represented in a graph tree where the changes from one form to another are oriented and weighted on a scale from 1 to 5, by giving heavy weights for rare or difficult transformations and light weights for easy or frequent transformations. As an example, figure 6 describes the graph of two of these 42 variables, naiz and zarete, as previously in tables 2, 3 and 5.1-2. Indeed, 12 graph-trees were constructed according to this method and then transformed into a binary coded (0/1) matrix without loss of information once the initial character state of the tree is specified (X in figure 6) and the changes constraint has been duly polarized. This factorisation process was performed with the Felsenstein’s factor program (http://evolution.genetics.washington.edu/phylip.html) – see the example of factorisation of naiz and zarete in figure 6.

The binary transformation of the 12 items leads to a vector of 260 binary characters for every 94 locations. Finally The 260x94 matrix is transformed to a parsimonious tree by constraining the changes from 0 to 1 to be irreversible and by locating the changes on branches using the delayed option (PAUP*,v4.0; Swofford, 2002) (Tree length = 1282 Consistency index (CI) = 0.44, Retention index (RI) = 0.7415 Rescaled consistency index (RC) = 0.3256).

Though similar as to the weight hierarchy of features of the S, F, V and L type, the range is smaller than for the Levenshtein Distance ponderation as presented in table 4 above. In other words, the basic hierarchy S << F << V << L still prevails, although intervals are made smaller for the sake of processing, for two reasons: i) derivational graphs as in figure 6 make hierarchical pairs as W & CW irrelevant, as characters are derived explicitly from one another, ii) each step of the string automatically cumulates previous weights, making necessary to lower the scale of values in order to preserve overall parsimony.
Figure 6. Graph-tree of the variables V144 and V145 (drawn with Pajek, Batagelj & Mrvar, 2011).
<table>
<thead>
<tr>
<th>Weight</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>4</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>naiz</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>náí</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>na</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nais</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nas</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>naix</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>niz</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nauk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nuk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>neok</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nuzu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nuzzû</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>nuxu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6. Factorisation of the graph-tree of V144 [naiz] (see figure 1), and weights of the changes.
Figure 7 shows the majority consensus tree obtained when keeping only branches found in at least 70% of the parsimonious trees. At least 5 main clades (or groups) can be distinguished. They are coded I to V and are localized on the map in figure 8.

Figure 7. Majority consensus tree showing only branches supported by at least 70% of the parsimonious trees (70%). Five different clades or groups, have been defined, I to V.
Map on figure 8 is strongly consistent with standard Basque dialect classification, as described in Zuazo (op.cit.): five main dialects, with Zuberoan embedded in the Navarrese-Lapurdian dialect. Moreover, the interesting point in this topology is the fact that the central dialect splits in two clusters, with a strong tropism either to the West (towards Western Gipuzkoan and Bizkaian) and the East (towards Navarrese proper or Navarrese-Lapurdian): Ataun and Hernani or Donostia cling to the former, whereas Tolosa, Azkoitia and Hondarribia bend to the latter. The Western dialect seems more intricate in its relation to these former clusters and to a peripheral subcomponent (Oñati, Bermeo, Ondarroa), making up a clade of its own.

In a few words, cladistics method properly applied to a fourth of the data set gives encouraging results, at phylogenetic level, with new patterns of intrication, complementary to the former results, provided by the Levenshtein algorithm: phylogenetic complexity highlights ontogenetic complexity, increasing deterministic and communal patterns (i.e. DC and CC), whereas ontogenetic complexity pointed at broad communal aggregates, tropisms, crossroads, transitions and singularities – i.e. patterns of diffusion and idiosyncrasies. Nevertheless, phylogenetic complexity processing can also give hints at diffusion patterns, as we’ll soon see.
Indeed, as most of the binary characters are not consistent, they can change more than once and on different branches. The numbers of changes occurring simultaneously within every main clades and occurring among two different clades are calculated (see figure 9). The state of the trait changes four times from the initial 0 to the derived 1, as the reversion from 1 to 0 is not allowed. The exchanges within and among groups G1 and G2 show up as dotted and dashed lines respectively. As the nodes cannot be dated, all exchanges between branches are taken into account, as if the structure of any clade would be a star-tree from their origin. The numbers of exchanges (borrowings, or *replicators*) are standardized by the maximum number of possible exchanges between branches within group and among groups (see fig. 9). These changes can be interpreted as possible borrowings. However, as the temporal hierarchy between nodes cannot be inferred, the tree structure has to be simplified to a star-tree. Moreover, as the number $n$ of terminals, and so the number of branches, are different from one clade to another, these numbers are standardized by the maximum number of possible different changes between branches. The results are shown figure 10. They can be carefully interpreted as borrowings or exchanges of linguistic traits between locations.

![Figure 9. Example of possible exchanges (or borrowings) within and among two main clades G1 and G2 for a binary trait.](image)

![Figure 10: normalized possible borrowing within and among groups I to V as defined in the tree figure 7](image)
The phenomena represented in figure 10 – especially the values of the chain 26, 21 and 23 below – do recall chain dialect models: features spread according to vicinity constraints, neighboring dialects being more liable to borrow from each other than far distant ones. Nevertheless, the discrepancy between zone III and V is striking. It corresponds to a political border (the French-Spanish frontier) and to a natural barrier (the Pyrenean mountains) except in the coastal region (the Pyrenean mountains). Nevertheless, the upper line of values (mostly 9 and 14, between clusters II-IV, II-III and I-III recall the diffusion dynamics already grasped by the former analysis, based on the Levenshtein distance – especially with the SFVL ponderation, where a southern track of diffusion could be glimpsed at. Clusters I and III obtained here from phylogenetic analysis happened to show up as cross-roads or transitional areas in the ontological approach – which could more properly be dubbed as the diffusion-typological approach. Grammars and phonological systems used in these areas differ so strongly from Standard Basque that we cannot either explain these interferences by mere standardization of the EAS elder speakers’ idiolects. All these facts converge towards a confirmation of a synthesis concerning phylogenetic diversification (dialect speciation) on the one hand, and typological diffusion on the other hand (feature pool) of the Basque dialect network from our CT approach. A model that we will now resume in the conclusion.
4. Conclusion & Prospects

The overall picture we can grasp from our CT approach, based on ontogenetic and phylogenetic processing of the EAS *aditza* data set for elder speakers points at a basic threefold division of Basque dialects, with subsequent subdivisions – such as peripheral singularities as Zuberoan and Western Bizkaian in the Western and the Eastern fringes. Between these blocks and sub-blocks intermediate “default areas” can be grasped, which appear more overtly from the first set of results (figures 4 and 5). They apparently played a very important role in the history of the language, as both sets of results confirm their incidence in propagating features between emerging blocks, or main dialect clusters, i.e. Bizkaian, Gipuzkoan, Navarrese-Lapurdin).

A clear phylogenetic split between Western (clades I to III) and Eastern Basque (clades IV to V) suggests that the central dialect has experienced for long a strong tropism either toward the Bizkaian or the Navarrese pool of features. Intermediate zones – or amphizones – may draw clear frontiers within the three main blocks, as far as diffusion routes are concerned, nevertheless, the central dialect is strongly driven on both fringes even in its core, towards these two peripheral feature pools, at opposite locations. Moreover, we could also identify *streams* or *nodes* conveying these trends: a southern route as a *stream*, and regional urban centers as Bergara or Hernani as *nodes*. Interestingly enough, our CT approach enhances the intricacy of embedded regional clusters, especially in zones corresponding to II and IV in the main cladistics results: Western Bizkaian and (Western) Navarrese, including the Eastern part of the Central dialect (Eastern Gipuzkoan). Last, but not least, this is only a tentative exploration of a subcomponent of Basque inflection – whole dimensions are still to explore from a CT standpoint, such as nominal inflection, phonological patterns, syntax and the lexicon. In tune with Berthoz’s vicariance concept, all these dimensions will highlight contrastive patterns of network embedding and phylogenetic clades. As pointed out in the introduction, *aditza* was used here because of its relevance in former attempts at classifying Basque dialects. But by no means one should rely in *aditza* only. Further explorations of other grammatical and lexical (sub)components will most probably unveil interesting asymmetries, as observed for other languages of high inflectional complexity, as mentioned above for Tseltal (Léonard & al. 2014).

We also want to suggest orientations for further developments of our CT approach, especially in the field of sociolinguistics and applied linguistics. Has Euskara Batua, the standard variety been lately the driving factor in the fabrics of geolinguistic variation in the younger generation of speakers, or on the contrary, has it been a weak epiphenomenon, unable to supersede the evolutive trends of *euskalkiak* (the traditional dialects)? Do geolinguistic aggregates vary according to *town dialects* (i.e. leading urban centers) or according to rural networks, or a combination of both (the town & satellite dialects)? What else can explain the diversity of geolinguistic patterns, from the standpoint of auxiliary verbs, which have been considered as a powerful criterion, since the beginning of Basque dialectology? Would other components of the grammar (or the lexicon) suggest different patterns and explanations? We hope to be able to answer these questions as soon as we’ll be able to process the second major EAS data set – the survey of younger speakers, aged under 40.

5. Bibliografía


Berthoz, Alain 2013, La vicariance. Le cerveau créateur de mondes, Paris, Odile Jacob.


169


Zuazo, Koldo 2013, The Dialects of Basque, Center for Basque Studies, University of Nevada, Reno.

Appendix
V144 [NAIZ]: Je suis jeune.
V145 [ZARETE]: Vous tous aussi vous êtes jeunes.
V146 [GINEN]: Nous aussi nous étions jeunes autrefois.
V147 [DADIN]: Nous lui avons donné de l'argent pour qu'il aille en vacances.
V148 [DAITEKE]: Ton ami il peut venir chez moi.
V149 [ZAIT]: La viande me plaît.
V150 [ZITZAIDAN]: Autrefois aussi la viande me plaisait.
V151 [ZITZAIZKION]: Autrefois (à mon père) les fraises LUI plaisaient
V152 [DUT]: Aujourd'hui j'ai mangé peu.
V153 [DU]: Il a peu mangé.
V154 [DUZUE]: Vous tous, vous avez peu mangé.
V155 [DUTE]: Mes amis ils ont mangé un sandwich.
V156 [DITU]: Peru (Pierre) il a mangé des prunes.
V157 [DITUZTE]: Eux ils mangent des anchois.
V158 [ZUEN]: Autrefois, Jon il buvait beaucoup trop.
V159 [ZENUTEN] Avant vous tous mangiez très peu.
V160 [ZITUEN]: Autrefois Andoni il mangeait beaucoup de noix.
V161 [DEZAN]: Nous lui avons donné de l'argent pour qu'il achète une voiture.
V162 [DITZAN]: Nous lui avons donné de l'argent pour qu'il achète des chocolatines.
V163 [DEZE]: Ton ami il peut le faire. Edo (il pourrait le faire)
V164 [DIO]: Il a vendu une vache au boucher.
V165 [DIZKIO]: Il a vendu des poulets à ton ami.
V166 [DIOGU]: Nous avons vendu une vache au boucher.
V167 [DIOTE]: Il ont vendu la maison à ton ami.
V168 [DIDAZU]: Tu m'as vendu un poulet.
V169 [DIEZAIKOE]: Andoni, il il peut lui porter le livre à Itziar. Edo Il pourrait lui porter le livre
V170 [ZOAZTE]: Où-c'est que vous allez /vous tous) si vite?
V171 [NINDOAN]: Moi j'allais à l'école quand c'est arrivé.
V172 [ZOAZ]: Partez vous d'ici et fichez -moi la paix.
V173 [ZAUDETE]: Vous êtes tous très contents
V174 [ZEGOEN]: Il était entrain de parler quand nous avons entendu le tonnerre. Iparraldean ezin da horrela edo
   " il dormait quand le tonnerre l'a réveillé = lo zagon.....
V175 [ZATOTO]: Venez tous vite à la maison tous.
V176 [ZABILTZATE]: Ces derniers temps vous étiez tous à travailler avec plaisir
V177 [DAUKATE]: Tes fils ils ont un bon prétexte pour arriver en retard (edo les Espagnols le tiennent prisonnier)
V178 [DAUZKA]: Andoni a vingt ans (L'Espagne les tient prisonniers)
V179 [ZEUKAN]: Andoni il aura vingt ans en finissant les études (mon fils L'Espagne le tenait prisonnier
V180 [DAKARTE]: Tes amis ils apportent toujours du vin pour dîner avec nous
V181 [DAKARTZA]: Ton ami il a apporté des bonnes nouvelles.
V182 [DIK]: dis donc, Mikael a bu de l'eau

TABLE I. The 39 variables (V144 to V182) of the EAS survey in the *aditza* section: auxiliary inflection.
TABLE II. The 94 locations for which the 39 variables of Table I are documented.