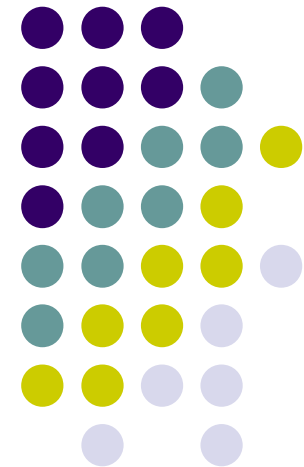


Analysis of the morphological variation using 'Diatech' tool

Gotzon Aurrekoetxea

University of the Basque Country
(UPV/EHU)





1. Basque Dialectology: some works

- Alvarez, J. L. & Aurrekoetxea, G., 1987, *Euskal dialektologiaren hastapenak [Handbook of the Basque dialectology]*, Bilbao: UEU.
- Martínez-Areta, M., 2013, “Basque dialects”, in M. Martínez-Areta (ed.), *Basque and proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*, Frankfurt and Main: Peter Lang, 31-87.
- Euskaltzaindia, 2010-2013, *Euskararen Herri Hizkeren Atlasa [Linguistic atlas of the Basque Language](EHHA)*, I-IV vol, Bilbao.

2. The Basque: an agglutinative language



[zazpi leiho]tatik
[seven windows] from
'from seven windows'



3. The inflexion of the Basque

- Grammatical cases:

- Absolutive (-Ø)
- Ergative (-k)
- Dative (-i)
- Partitive (-ik)

- Genitives

- genitive (-en)
- relational (-ko)

- Postpositions

- Locative cases:

- locative (-n) 'in'
- Ablative (-tik) 'from'
- Allative (-ra) 'to'
- Directional (-rantz) 'towards'
- Terminate (-raino) 'up to'

- Non locative cases:

- Commitative (-ekin) ('with')
- Benefactive (-rentzat) ('for')
- Instrumental (-z)
- prolative (-tzat)
- Cause (-gatik)

(Euskaltzaindia, 2003, *Euskal gramatika laburra: perpaus bakuna*
[Brief Grammar of the Basque], Bilbao: Euskaltzaindia.

For the names of the cases see Hualde & Ortiz de Urbina 2003)





4. The inflexion in the dialects

- Different suffices for the same inflexion case:

-areki(n)/-arekila(n) vs. -agaz ('with')

- Different phonological rules (PhRs): Dissimilation, assimilation, deletion, addition...

-o + -ak: -oak, -ook, -ok, -uak, -uek...





5. The inflexion in the EHHA project

- All inflexion cases
- Each case with words finished with different vowels and consonants
- Each word in indefinite, singular and plural forms
- 188 questions

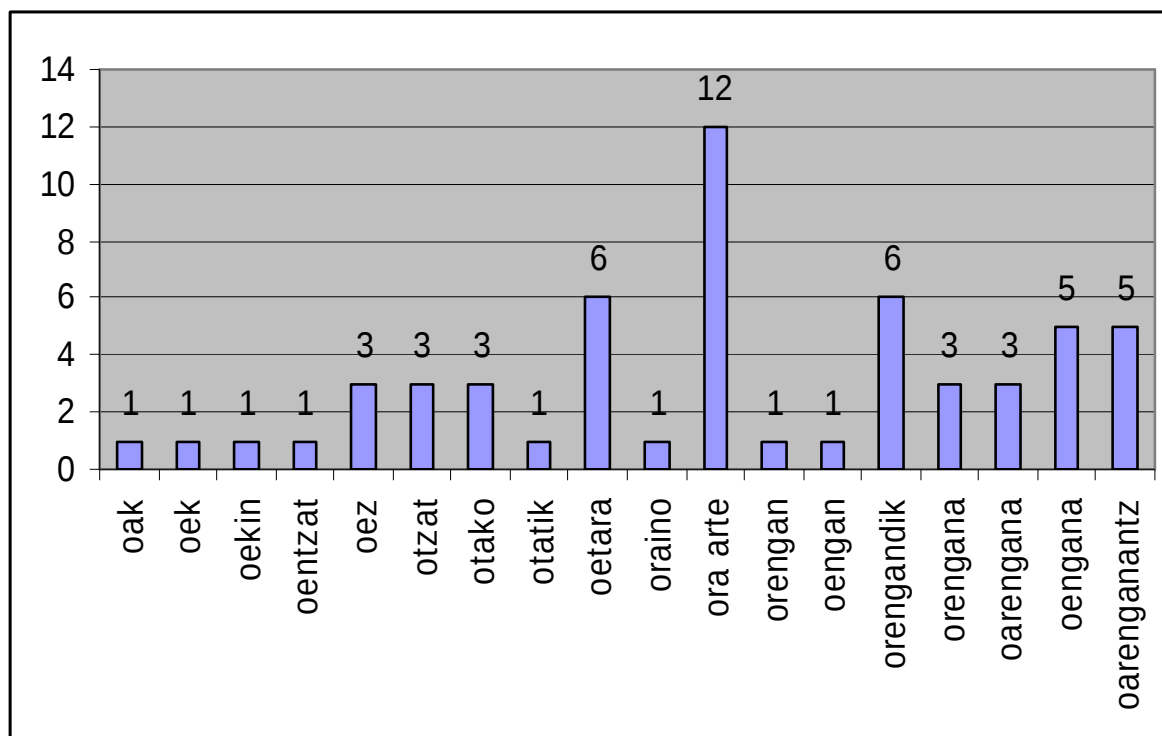


6. The data of the contribution

- Data from the EHHA-V
- 51 questions about the inflexion of the words finished by “-o” vowel
- Direct questions vs. Proposals
astuak vs. **astuek*
- Empty answers and multiple responses (MR)
- Responses and underlying representation



7. Empty answers



- 7.250 items
- 64 empty answers
- 0.88%
- From 51 cases in 18 empty answers

Fig. 1: Empty answers



8. Multiple Responses (MR)

- Different pronunciations of the same word or linguistic feature, when interviewers asked only one question to elicit information about this grammatical case: [aʃ'tuen] and [aʃ'tuan] (Deba). Also called 'polimorphisme' (Allières 1992).
- Different linguistic or grammatical forms, but with the same meaning, when the interviewer asked more than one question to elicit the same grammatical case: [aʃ'tokin] and [aʃ'toaʃ] (Berriz)
- Different linguistic forms given by different informants, but whereby the informants were also told, and they accepted, each others' different responses as possible utterances in that locality: [aʃ'toikin] and [aʃto'ekin] (Gamarte)
- Different linguistic forms, using proposals that were accepted by the informant: [aʃtoj'kin] and [aʃ'tojki'lan] (Uharte Garazi).|



9. MR (questions/localities)

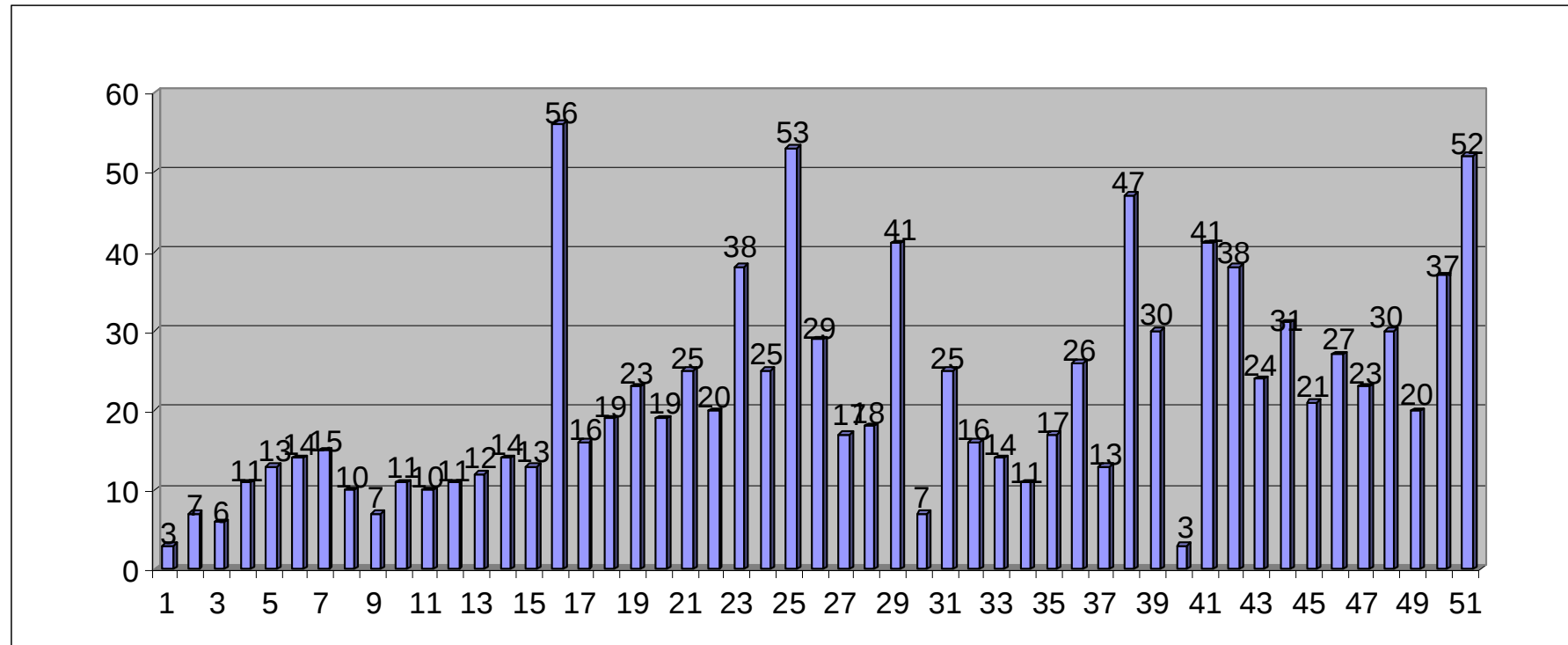


Fig. 2: Quantifications of MR in each question



10. The analysis of the data

a) Orthographic answers

astoak	>	-o + ak	>	<i>-oak</i>
'donkey' + det + abs. mark				
astok	>	-o + ak	>	<i>-ok</i>
astoog	>	-o + ak	>	<i>-oog</i>
astuak	>	-o + ak	>	<i>-uak</i>

...

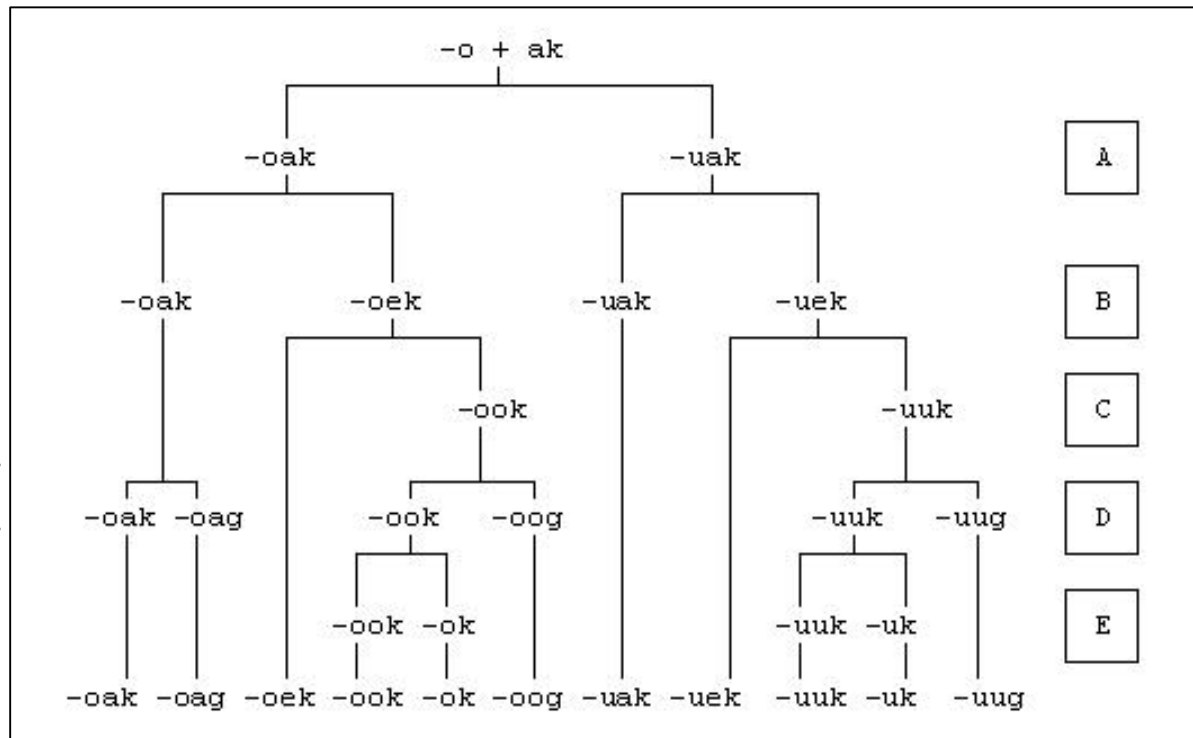
b) Underlying representations

-oak

c) Phonological rules



11. Hierarchical structure of PhR of “-o+ak” case



- A
- B
- C
- D
- E

A: Dissimilation rule
 B: Assimilation rule
 C: Assimilation rule
 D: Voiceless rule
 E: monoptongation rule

Fig 3: Hierarchical structure of the PhR



12. Linguistic distances in *Diatech*

Welcome Gotzon. Change password Management site Translations management Logout

DiaTech
Tool for doing dialectometry

Index Projects Help About us Account Users Languages English

Search answers Statistics Answers maps Manage database Manage project

Statistics: go-EHHA-D ?

Statistic Analysis

Analysis type: Synoptic map

Answer type: Orthographic answer

Secondary linguistic domain: Phonology

Linguistic domain: All domain

Statistic algorithm:

Linguistic distance: IPI, IRI, Levenshtein, IPD, IRD

Location:

Statistic operation: Distribution of similarity

Analyze Clear the cache

www.eudia.ehu.es/diatech



13. Analysis of the data: linguistic distance-1

a) Phonetic distance (Levenshtein unit)

(Heeringa 2004, Spruit, Heeringa & Nerbonne 2008...)

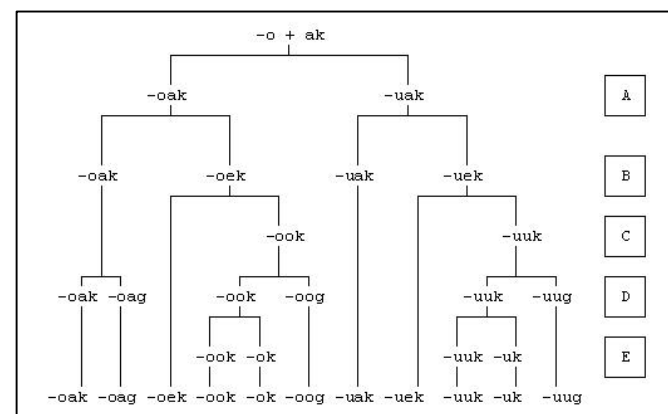
b) Phonological distance (RIV unit)

(Goebel 1981, 1992...)



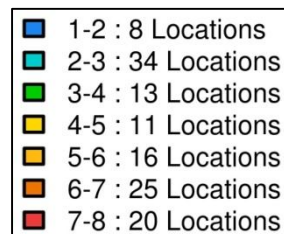
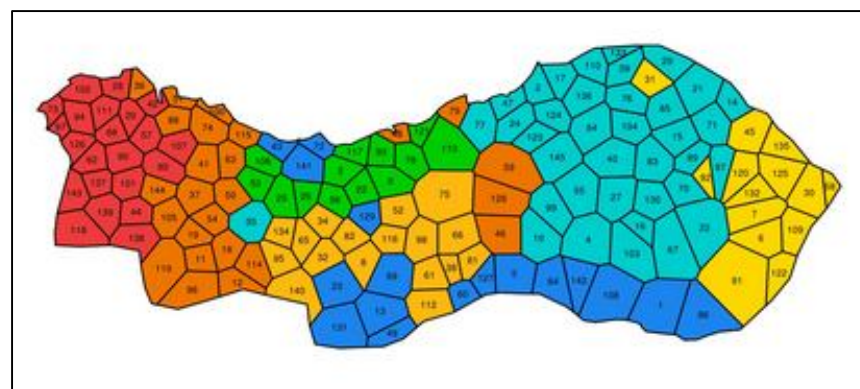
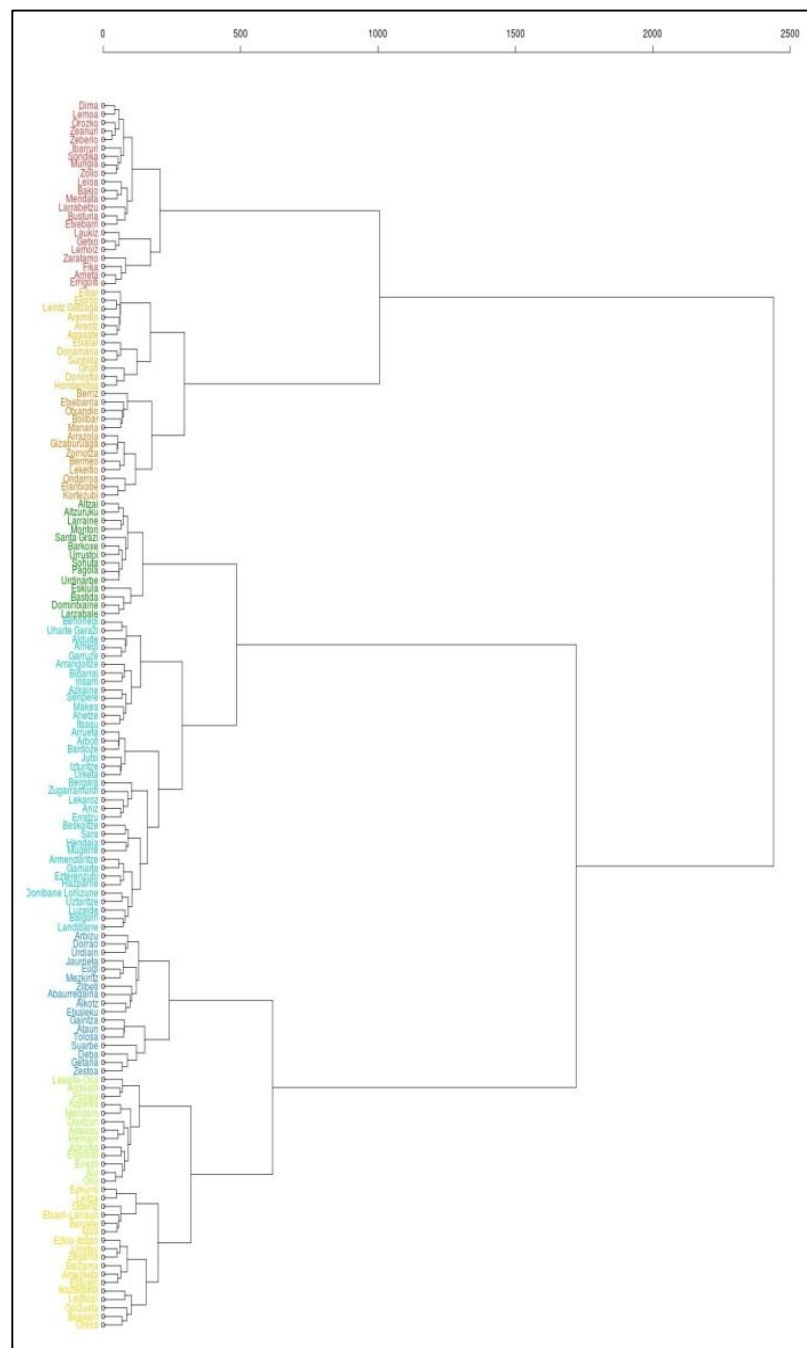
14. Analysis of the data: linguistic distance using PhRs

- $-o+ak > -oag$ distance: 1 (D level)
(one PhR needed to pass from $-oak$ to $-oag$)
- $-o+ak > -ook$ distance: 2 (B and C)
(two PhRs needed to pass from $-oak$ to $-ook$)
- $-o+ak > -ok$ distance: 3 (B, C and E)
(three PhRs needed to pass from $-oak$ to $-ok$)
- $-o+ak > -uk$ distance: 4 (A, B, C and E)
(four PhRs needed to pass from $-oak$ to $-uk$)





15. Phonetic distance

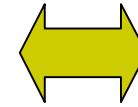


1 map: EHHA-morphology -51 questions (phonetic distance)
 Orthographic answers
 Levenshtein distance
 Cluster analysis
 Ward method-7

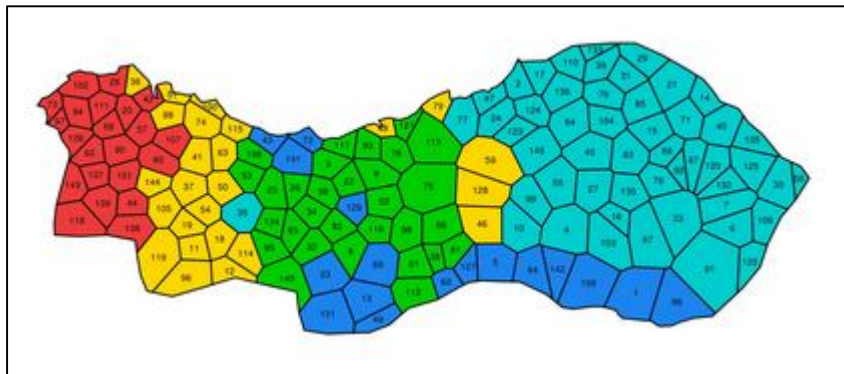
16. Phonetic distance: comparison



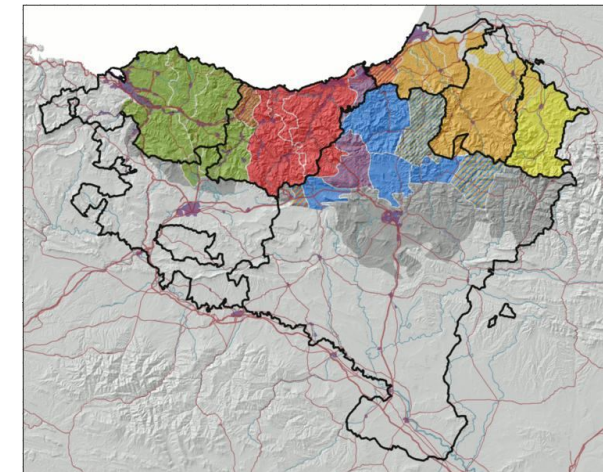
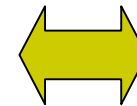
Map 2: Cluster (Levenshtein dist., ortog., Ward-7)



L.L. Bonaparte (1868)



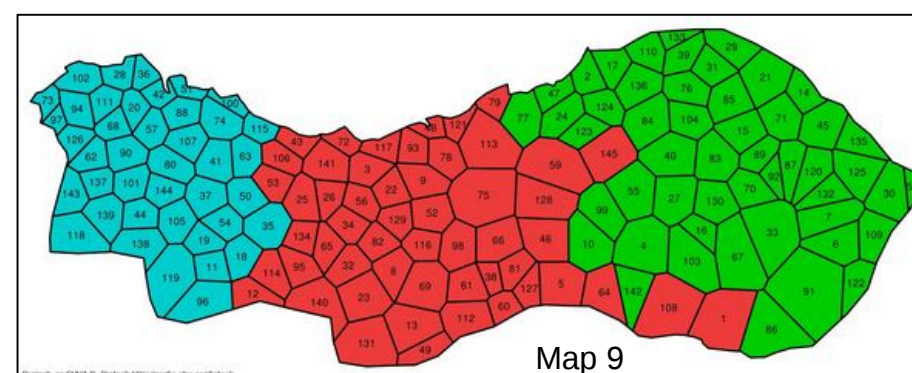
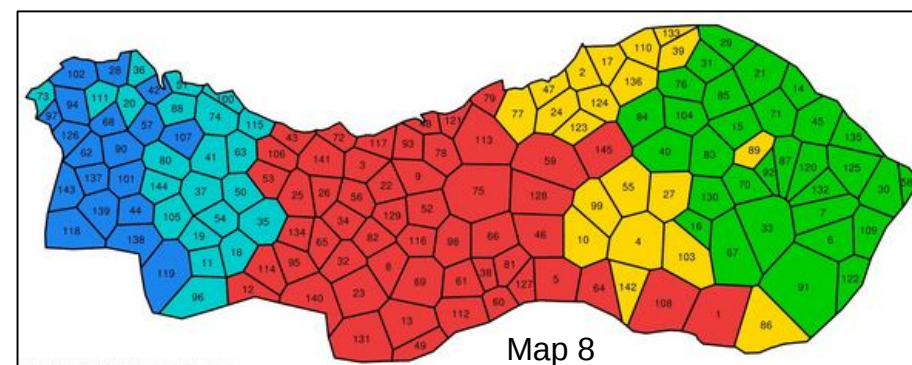
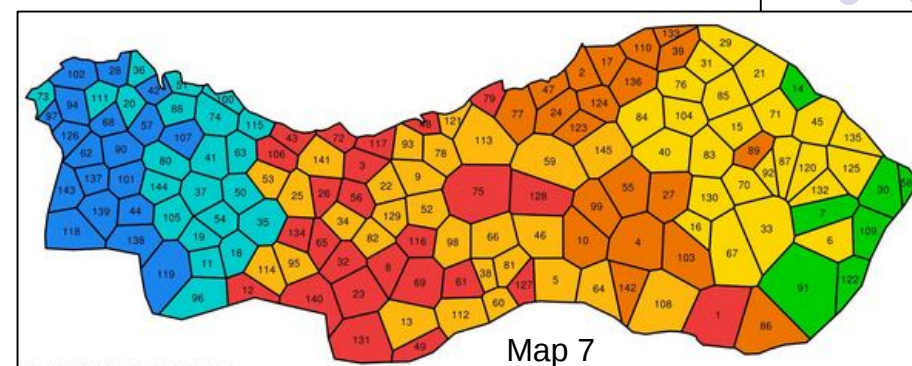
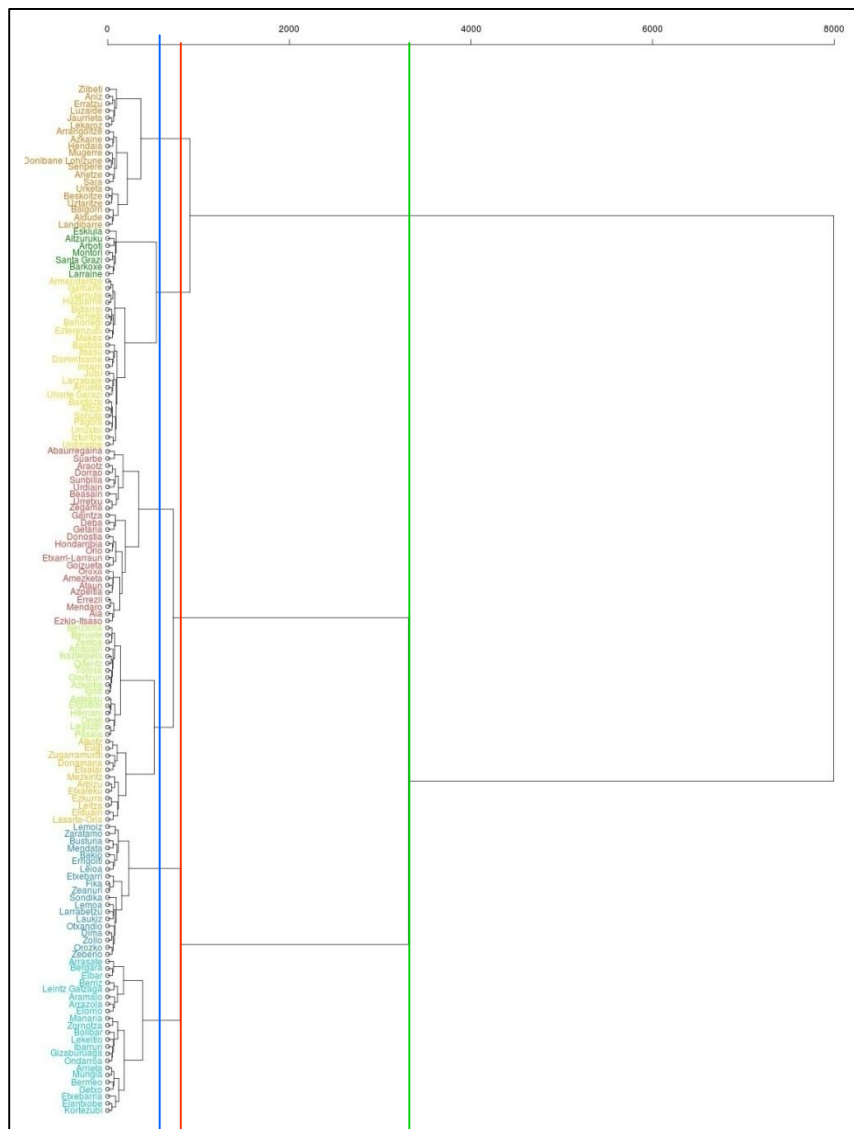
Map 3: Cluster (Levenshtein dist., ortog., Ward-5)



Zuazo (1998)



17. Phonological distance





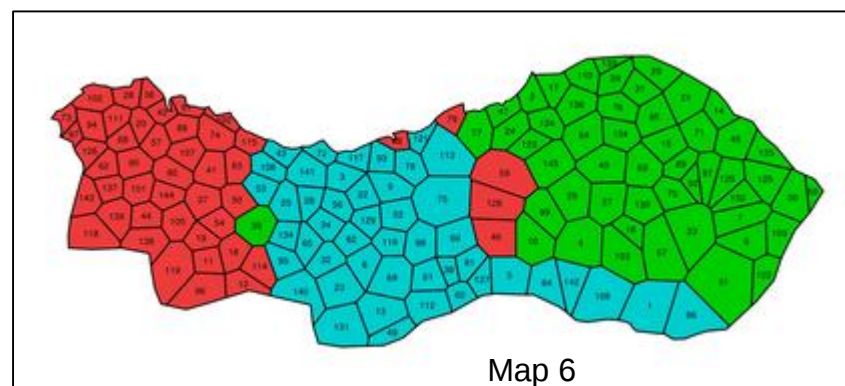
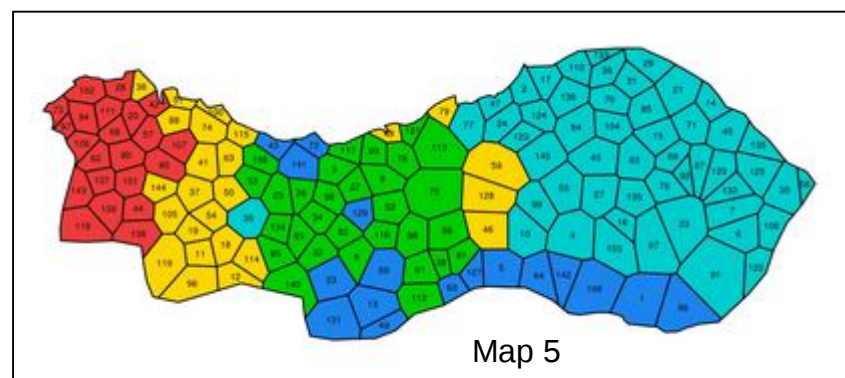
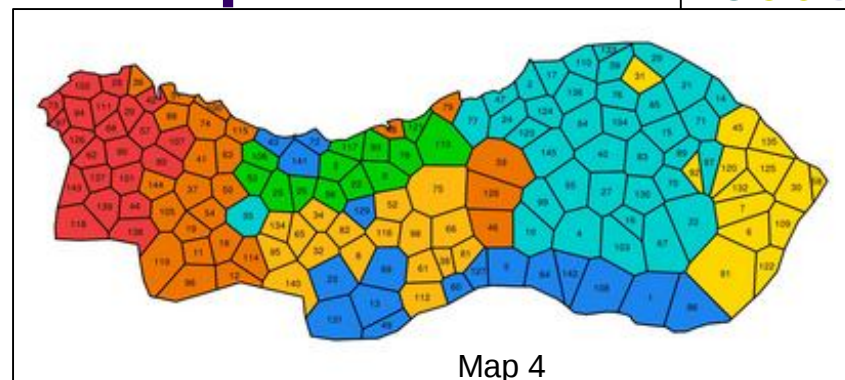
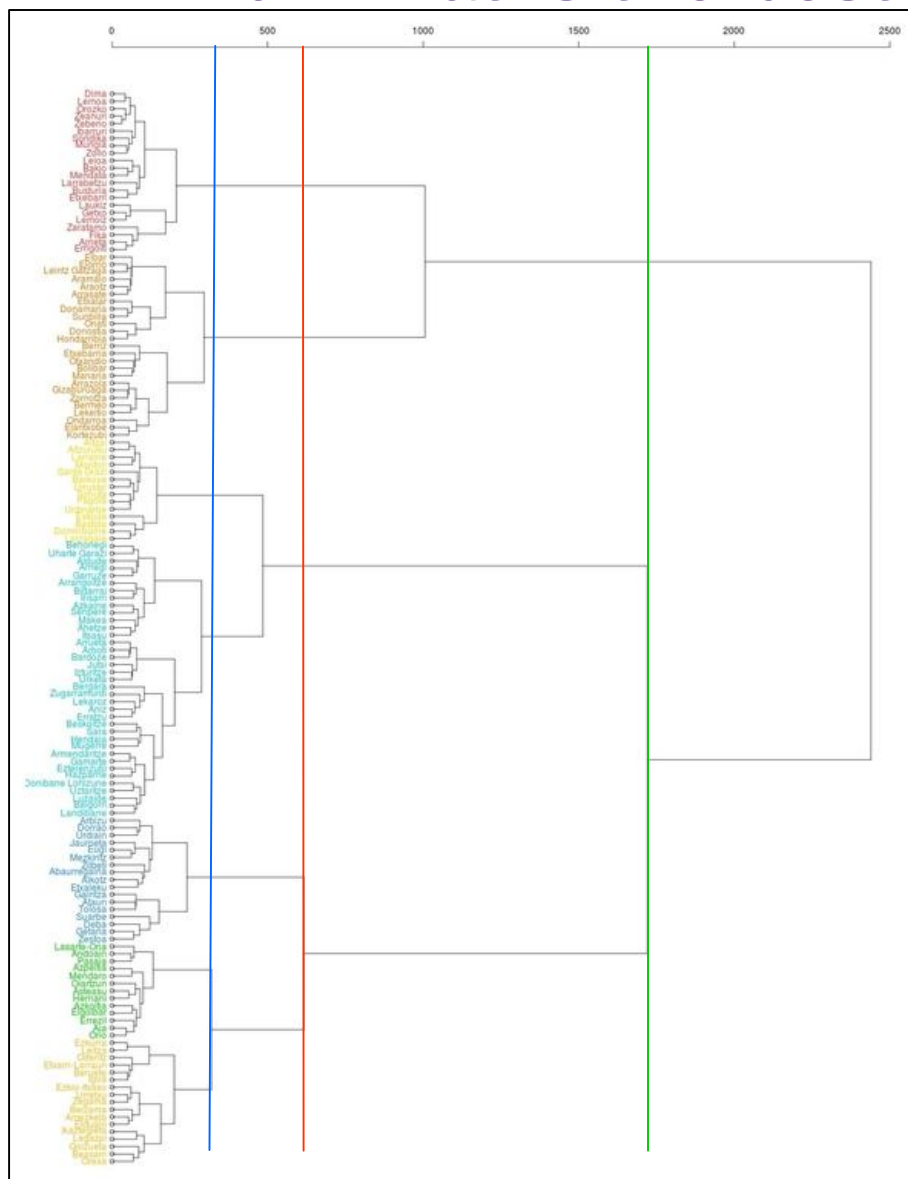


19. Discussion

- Which is the best cluster partition?
- Dialectologists have made great progress quantifying linguistic distances and drawing dialectal areas
- Have we make similar efforts in the theoretical aspects of linguistic variation?
- The comparability of the outcomes...



20. What is the best cluster partition?





21. Discussion

-
- Dialectologists have made great progress quantifying linguistic distances and drawing dialectal areas
- Have we make similar efforts in the theoretical aspects of linguistic variation?
- The comparability of the outcomes...



22. Conclusions

- The first time we use data from the Linguistic Atlas of the Basque (EHHA) project;
- I have shown the hierarchical classification of the Basque dialects using two data types (phonetic and phonological) and two linguistic distances (Levenshtein and RIV distances);
- I have shown the contrast between two distances.



References

- Alvarez Enparantza, J. L. "Txillardegí" & Aurrekoetxea, G. 1987, *Euskal dialektologiaren hastapenak*, Bilbao: UEU, [www.inguma.org].
- Aurrekoetxea, G., 1995, *Blzkaieraren egituraketa geolinguistikoa [The geolinguistic structure of Biscayan dialect]*, Bilbao: UPV/EHU.
- Aurrekoetxea, G. & Ch. Videgain, 2014, "Outils por la géolinguistique automatisée", In Fabio Tosques, Fabio (ed.), 2014, 20 *Jahre digitale Sprachgeographie - Tagungsband (Berlin 02. bis 03. November 2012)*, Berlin: Humboldt-Universität zu Berlin, Institut für Romanistik (http://www2.hu-berlin.de/vivaldi/tagung/beitraege/pdf/04_aurrekoetxea_videgan.pdf)
- Aurrekoetxea, G.; Karmele Fernandez-Aguirre; Jesus Rubio; Borja Ruiz; Jon Sanchez, 2013, "'DiaTech': A new tool for dialectology", *Literary and Linguistic Computing*; doi: 10.1093/lc/fqs049
- Euskaltzaindia, 1993, *Euskal Gramatika Laburra: Perpaus Bakuna [Brief grammar of the Basque]*, Bilbao: Euskaltzaindia.
- Euskaltzaindia, 2010-2013, *Euskararen Herri Hizkeren Atlas I-V [Linguistic Atals of the Basque Language I-V]*, Bilbao: Euskaltzaindia("http://www.euskaltzaindia.net/"www.euskaltzaindia.net])
- Clua, E., 2010, "Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal", in G. Aurrekoetxea & J. L. Ormaetxea (eds.), *Tools for linguistic variation*, Bilbao: UPV/EHU, 151-166.
- Goebel, H., 2013, "Le Baiser de la Belle au bois dormant our: des péripéties encourues par la géographie linguistique depuis Jules Gilliéron", *Corpus 12 "Dialectologie: corpus, atlas, analyses"* (numéro coordonné et présenté par Rita Caprini), 61-84.
- Hyvönen, S., Leino, M., Salmenkivi, M., 2007, "Multivariate Analysis of Finnish Dialect Data: An Overview of Lexical Variation", *Literary and Linguistic Computing* 22 (3), 271-290.
- Hualde, J.I., 1997b, "Rules vs. Constraints: Palatalization in Biscayan Basque and Related Phenomena" in F. Martínez-Gil eta A. Morales-Front (arg.), *Issues in the Phonology and Morphology of the Major Iberian Languages*. Georgetown University Press: Washington.
- Hualde, J. I. & Ortiz de Urbina, J. (eds.), 2003, *A Grammar of Basque*, Berlin: Mouton de Gruyter.
- Martínez Areta, M., 2013, "Basque dialects". In Mikel Martínez-Areta (ed.), *Basque and Proto-Basque, Mikroglottika. Minority language Studies* 5, 31-87.
- Laka, I., 1994, *A brief grammar of Euskara, the Basque language*, in http://www.ei.ehu.es/p056-12532/eu/contenidos/informacion/euskara_inst_lexiko_gramatika/eu_lex_gram/adjuntos/Laka2.pdf
- San Martin, I., 1998, "An OT Account of the Formation of Definite Forms in the Vizcayan Basque Dialect of Markina", *University of Maryland Working Papers in Linguistics* 7.
- Spruit, M.R., Nerbonne, J., Heeringa, W., 2008, "Associations among linguistic levels", *Lingua, Special issue on Syntactic databases. Selected papers presented in the special session Comparing Aggregate Syntaxes, Digital Humanities conference, Paris, July 6, 2006*, 65-99.