Historical and laboratory study of sound change typology (with a focus on the reconstruction of Basque)

Habilitation à diriger des recherches (HDR)
Université Paris Cité

Présentée par

Ander Egurtzegi
Chargé de Recherche Classe Normale CNRS
IKER — UMR 5478
2023-2024

Table of Contents

Acknowledgments *4*

1. Curriculum Vitae *6*

Education **6**
Languages 6

Previous professional appointments **7**

Awards **7**

Participation in grants and projects **7**
As a Principal Investigator 7
As a research team member 8

Teaching appointments **8**

Supervision **9**
Internship 9
BA thesis 9
MA thesis 9
PhD thesis 9
Post-doctorate 10

International research stays **10**

Service to the profession **10**
Journal manuscript reviewer 10
Book chapter reviewer 10
Project reviewer 11
Conference and workshop organization 11
Conference session chair 11
Member of scientific committees 11

Community involvement / Outreach **12**
Media consultations/interviews 12

2. Publications *13*

Monograph **13**

Refereed journal articles **13**

Book chapters **15**

Refereed conference proceedings **15**

Outreach publications **16**

Preprints and other publications **16**

Oral contributions **17**
Invited talks 17
Papers presented (selected) 20

3. PhD diploma *23*

4. Research program including selected previous works *24*
Introduction 24

4.1 Basque historical linguistics **24**

4.2 Phonological theory and sound change typology **27**

Establishing typologically rare phonological contrasts 28
Proposing alternative analyses for uncommon sound patterns 28
Documenting and analyzing allegedly impossible sound patterns 29

4.3 Phonetics 32
Phonetics of Basque sound patterns 33
Innovative methods in acoustic phonetics 37
Phonetics of sound change in progress 38

4.4 Concluding remarks 39

5. Activities as principal investigator 40

6. Research management 41

7. Research projects 43

7.1- *Modern approaches to diachronic phonology applied to Basque* 43
7.1.1 Positioning and context 43
7.1.2 Objectives 44
7.1.3 Methodology 44
7.1.4 Interest and impact 46
7.1.5 Example study 1: Using acoustic data for the reconstruction of historical sound change: Nasalance-based studies of Zuberoan Basque 47
7.1.6 Example study 2. Using corpus phonology for the reconstruction of historical sound change. Data-driven philological studies of sibilant oppositions in 18th century Gipuzkoan Basque 49

7.2 Syllables 52
i. Evidence for syllables 53
ii. Phonetic vs. phonological syllables 54
iii. Languages without syllables? 54
iv. Syllables and sonority 54
v. Syllable-internal structure 55
vi. Syllable weight 55
vii. Syllabification and related processes 55
viii. The evolution of syllable structure 56
ix. Syllable typology and universals 56

7.3 Looking for the origin of recurrent sound changes: Metathesis 56

References 60

Appendix: Selected publications 69
Blevins, J., Egurtzegi, A. & Ullrich, J. 2020. "Final obstruent voicing in Lakota: Phonetic evidence and phonological implications". *Language* 96.2, 294-337 (44 pgs.) 70
Egurtzegi, A. 2017. "Phonetically conditioned sound change: Contact induced /u/-fronting in Zuberoan Basque". *Diachronica* 34.3, 331-367 (37 pgs.) 114
Egurtzegi, A. 2018. "On the phonemic status of nasalized /ɦ̃/ in Modern Zuberoan Basque". *Linguistics* 56.6, 1353-1367 (15 pgs.) 151
Egurtzegi, A. & Carignan, C. 2020. "An acoustic description of Mixean Basque". *Journal of the Acoustical Society of America* 147.4, 2791-2802 (12 pgs.) 166
Krajewska D., Zuloaga, E. & Egurtzegi, A. 2022. "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1-31 (31 pgs.) 178

Acknowledgments

I will use this space to thank the people without whom I would not be presenting this *Habilitation à diriger des recherches*, nor have the trajectory I have had thus far.

I want to start by thanking my companion in this journey, Beatriz Umbon, who has provided me with as much care, patience, and understanding as I could have needed. She has listened to so much linguistic digression by now that a bright mind like hers could probably write a dissertation without attending a single lecture. Elkarrekin lortu dugun guztia ezin zenba dezaket, ehun mila esker.

Academically, I owe everything to my three mentors, who accompanied me in the different parts of this journey. First Joseba Lakarra, then Juliette Blevins also from early on, and finally Jonathan Harrington in the more recent years have taught me almost everything I know about Basque, (historical) phonology, typology, and phonetics. Your guidance has shaped me as the researcher I am today, and I will always be ineffably thankful to the three of you for that.

This HDR is only possible because of Ioana Chitoran, who offered to be my garante, fully aware of the workload she had to take on her shoulders to this end. I am very grateful for that, and it is an honor for me to have you as part of my journey from this point on. I hope we can walk the future way together with many engaging projects!

Likewise, this HDR would not be possible without the generous help of all the other committee members: the rapporteurs Marianne Pouplier and Jane Stuart-Smith, the examiners Jonathan Harrington, Gorka Elordieta and Annie Rialland, and the president of the jury Ricardo Etxepare. Some of you have played a crucial role in my development as a scientist for which I cannot thank you enough, while others I just had the opportunity to meet recently. I am indebted to all of you for accepting to be part of this HDR and forming an excellent committee.

I would also want to thank all my research collaborators and colleagues from former and present labs, some of whom have become friends and provided me with enormous support in addition to sharing their outstanding scientific skill and wisdom. These include (but are not limited to) Aurora Troncoso-Ruíz, Rubén Pérez de Ramón, Dorota Krajewska, Borja Ariztimuño and Eneko Zuloaga in Gasteiz; Raphael Winkelmann, Chris Carignan, Michele Gubian, Johanna Cronenberg and Andrea García-Covelo in Munich; and Irantzu Epelde, Maia Duguine, Aritz Irurtzun, Urtzi Etxeberria and Iñigo Urrestarazu-Porta in Baiona.

I do not want to forget other researchers and friends that I have encountered along the way and have been a great influence for me (and often also a big support, even if it was from far away). Here, especial mention is due to Ana Estrada, Carlota de Benito, Nuria Polo, Stefano Coretta, Timo Roettger, and Márton Sóskuthy. I met all of you a long time ago and I am very grateful to keep seeing you almost every year.

Last, but not least, I want to thank my parents Nekane and Marce, and siblings Argia and Aitor, who planted and nurtured the seed of curiosity for linguistics and the ability to find and appreciate the beauty in linguistic (and musical) sounds that has brought me all the way to this point. Mom, you once told me about a 3 year old kid that wanted to learn Latin after his first school day. That kid has now many stories to pass on to the next litter of future Basque linguists.

1. Curriculum Vitae

Updated to XII-2023

Last name: Egurtzegi Martínez	Name: Ander	Date of birth: 03/22/1986
ID: 45679754-Y	Nationality: Spanish	Gender: M

Organization:	Centre National de la Recherche Scientifique (CNRS)
Faculty, School or Institute:	IKER (Unité Mixte de Recherche, UMR 5478)
Postal address:	15 Place Paul Bert, 64100 Bayonne (France)
E-mail:	ander.egurtzegi@iker.cnrs.fr
Webpage:	https://egurtzegi.github.io/

Specialization: Historical Linguistics (5702.01), Phonology (5705.06), Phonetics (5705.05)

Professional Category: Chargé de Recherche (tenured researcher) Starting date: 01/10/2019

Education

2014 PhD., *Linguistics*, supervised by Joseba A. Lakarra & Juliette Blevins (CUNY). Dissertation: "Towards a phonetically grounded diachronic phonology of Basque" [International, maximum grade]. University of the Basque Country (UPV/EHU).

2012 M.A., *Phonetics and Phonology*, Consejo Superior de Investigaciones Científicas (CSIC).

2010 M.A., Linguistics, University of the Basque Country (UPV/EHU).

2010 Licenciatura (B.A.), Basque Philology, University of the Basque Country (UPV/EHU).

2008 Licenciatura (B.A.), German Philology, University of the Basque Country (UPV/EHU).

Languages

Spanish (nat.), Basque (nat.), English (C2), German (C2), French (B1), Latin (r.), Swedish (r.).

Previous professional appointments

2019	Post-doctoral Scholarship Holder	Ludwig-Maximilian University of Munich (LMU) (ERC-Advanced Grant to Prof. J. Harrington)
2017-2018	Post-doctoral Scholarship Holder	Ludwig-Maximilian University of Munich (LMU) (Alexander von Humboldt Foundation)
2015-2016	Post-doctoral Scholarship Holder	Complutense University of Madrid (UCM) (Spanish Government, Juan de la Cierva-F.)
2015	Post-doctoral Scholarship Holder	University of the Basque Country (UPV/EHU)
2011-2014	Pre-doctoral Scholarship Holder	UPV/EHU (Basque Government)
2009	Researcher (temporary)	University of the Basque Country (UPV/EHU)
2007-2008	Collaboration Fellowship	UPV/EHU (Basque Government)

Awards

2014: Dissertation award. Amount: 4,500€. Awarded by: Basque Government.

2019-2022: Doctoral and Research Incentive Award (PEDR). Amount: 14,000€. Awarded by: CNRS.

2023-2026: RIPEC personal award (prime individuelle RIPEC). Amount: 14,000€. Awarded by: CNRS.

Participation in grants and projects

As a Principal Investigator

2021-2024: *Modern approaches to diachronic phonology applied to Basque* (MADPAB) (ANR-AAPG2020).

Grant number: ANR-20-CE27-0007. Project members: 7. Amount granted: 272,922.18€.

***2024-2026:** Additional support for dissemination and promotion of MADPAB was granted by the ANR (SAPS-CSTI 2020) through the proposal DIRE II UPPA.

Short-term research projects financed by the Alexander von Humboldt Foundation:

2023: *Andalusian post-aspiration and the life cycle of sound change*. Amount granted: 3000€. LMU Munich.

2021: *Andalusian /s/ aspiration at the word boundary*. Amount granted: 3000€. LMU Munich.

As a research team member

2023-2027: *Corpus oral y sonoro del español rural COSER: Edición digital y análisis lingüístico* (AEI, Spain, PI: Inés Fernández-Ordóñez).

2023-2027: *Language Abilities of Bilingual children with ASD* (LABA) (ANR, France, PI: Philippe Prévost).

2022-2025: *Non-participant addressees in Basque: System, variation and attrition* (TO2NO) (ANR, France, PI: Irantzu Epelde).

2023: *Paramètres Linguistiques des Rites Surnaturels* (PLRS) (InSHS–CNRS, France, PI: Aritz Irurtzun).

2018-2022: *Human interaction and the evolution of spoken accent* (ERC Advanced, Europe, PI: Jonathan Harrington).

2016-2020: *Estudios de morfosintaxis nominal: Lenguas paleohispánicas e indoeuropeas antiguas* (MINECO, Spain, PI: Eugenio Luján).

2016-2018: *Grupo de Investigación Consolidado: Hizkuntzalaritza Teorikoko Taldea* (Basque Government, Spain, PI: Myriam Uribe-Etxebarria).

2011-2024: *Monumenta Linguae Vasconum* III, IV, V & VI (4 consecutive calls. MINECO/AEI, Spain, PIs: Joseba Lakarra, Blanca Urgell).

2009-2015: *GIC: Lingüística histórica e historia de la lengua vasca* (3 consecutive calls. Basque Government, Spain, PI: Joaquín Gorrotxategi).

Teaching appointments

University of Pau and Pays de l'Adour (UPPA) 2019-2022

Phonetics and Phonology of Basque (BA, 2020, 2021, 2022).

Historical Phonology of Basque (BA, 2021).

University of the Basque Country (UPV/EHU) 2018-2021

Introduction to Phonetics and Phonology (MA, 2018), guest lecturer.

Sound Change (MA, 2019, 2020, 2021), guest lecturer (full responsibility).

Ludwig Maximilian University of Munich (LMU) 2017-2019

Historical Phonology of Basque (MA, 2018).

Sound Change (MA & BA, 2017-18, 2018-19).

Complutense University of Madrid (UCM) 2015-2016

Language Contact in the Antiquity (MA, 2016), with Dr. Eugenio Luján.

Introduction to Historical Linguistics (BA, 2016), with Dr. Juan Antonio Álvarez-Pedrosa.

Basque I (BA, 2015-16).

Basque II (BA, 2016), with Dr. Carlos Cid.

Basque Culture (BA, 2016).

City University of New York – Graduate Center (CUNY GC) 2013, 2018

Phonology I/II (MA, 2013, 2018), guest lecturer, invited by Prof. Juliette Blevins.

Supervision

Internship

2018 Co-supervision (alongside Prof. Jonathan Harrington) of Pia Greca's internship at the IPS-LMU Munich in 2018. Successful application for pre-doctoral funding at the Class of Languages (LMU).

BA thesis

2019 Supervision of Andrea García Covelo's BA thesis at the Ludwig-Maximilian University of Munich. Thesis: *The Development of the Diminutive Suffix -iño/-inho in Galician and Portuguese: From Prehistory to Modern Times*. Defense date: August 28 2019. Grade: 1.0 (maximum).

MA thesis

2020 Co-supervision of Andrea García Covelo's MA thesis at the Ludwig-Maximilian University of Munich alongside Dr. Ryan Sandell. Thesis: *Phonetic and Phonological Developments of Obstruent-Lateral Clusters in Ibero-Romance*. Defense date: March 3 2020. Grade: 1.0 (maximum).

PhD thesis

2027 Rosalie Diefenthaler. Thesis: "/s/-aspiration and its effect on preceding vowels in the Spanish variety of Las Palmas de Gran Canaria". Ludwig-Maximilian University of Munich. Co-supervised by Prof. Jonathan Harrington.

2025 Andrea García-Covelo. Thesis: "The development of obstruent plus lateral clusters in Ibero-Romance: A historical-phonetic approach to cluster

palatalization". Co-tutelle: University of Pau & Ludwig-Maximilian University of Munich. Co-supervised by Prof. Marianne Pouplier.

2025 Iñigo Urrestarazu-Porta. Thesis: "The reconstruction of the phonetics of Common Basque: Sibilants". Co-tutelle: University of Pau & University of the Basque Country. Co-supervised by Prof. Joseba A. Lakarra.

Post-doctorate

2021-2023 Dorota Krajewska. External post-doc integrated in the MADPAB project. Now associate professor at the University of the Basque Country.

International research stays

2023 Institute of Phonetics and Speech Processing LMU Munich (Germany). 1 month, guest researcher.

2021 Institute of Phonetics and Speech Processing LMU Munich (Germany). 1 month, guest researcher.

2018 City University of New York - Graduate Center (USA). 1 month, guest researcher.

2016 Institute of Phonetics and Speech Processing LMU Munich (Germany). 30 months, post-doc.

2015 Complutense University of Madrid (Spain). 12 months, post-doc.

2014 City University of New York - Graduate Center (USA). 1 month, guest researcher.

2013 City University of New York - Graduate Center (USA). 6 months, guest researcher.

2006 Universität Leipzig (Germany). 9 months, Erasmus student.

Service to the profession

Journal manuscript reviewer

Diachronica, *Laboratory Phonology*, *Journal of Phonetics*, *Journal of the IPA*, *Phonetica*, *Journal of Linguistics*, *Glossa*, *Natural Language & Linguistic Theory*, *Linguistic Typology*, *Oceanic Linguistics*, *PlosONE*, *Languages*, *Ampersand*, *Loquens*, *ASJU*, *Fontes Linguae Vasconum*.

Book chapter reviewer

Oxford University Press, De Gruyter.

Project reviewer

National Science Center (Poland).

Conference and workshop organization

2023 Workshop *Let's talk about stats!* (CNRS – IKER-UMR5478; Bayonne, France).

2023 Workshop *Let's talk about R!* (CNRS – IKER-UMR5478; Bayonne, France).

2022 Workshop *Understanding Sound Change* (University of Deusto in Donostia/San Sebastián).

2022 *Old World Conference on Phonology* (OCP) 19 (Donostia/San Sebastián).

2015 *International workshop on Basque-Japanese and Neighbouring Languages* (Vitoria-Gasteiz).

2014 *Jornadas de la UFI 11/14 / UFI 11/14aren jardunaldiak* (UPV/EHU, Vitoria-Gasteiz, Spain).

Conference session chair

2022 25th International Conference on Historical Linguistics (University of Oxford, UK).

2022 Workshop Understanding Sound Change (University of Deusto, Donostia/San Sebastián).

2022 Old World Conference on Phonology (OCP) 19 (University of Deusto, Donostia/San Sebastián).

2017 50th Annual Meeting of the Societas Linguistica Europaea (University of Zurich, Switzerland).

2017 Abstraction, Diversity and Speech Dynamics (University of Munich, Germany).

2015 International Workshop on Basque-Japanese and Neighbouring Languages (Vitoria, Spain).

2014 Jornadas de la UFI 11/14 / UFI 11/14aren jardunaldiak (UPV/EHU, Vitoria-Gasteiz, Spain).

Member of scientific committees

2019 Fontes Linguae Vasconum 50: New methods and trends in (Basque) linguistics (Pamplona, Spain).

Community involvement / Outreach

2014 & 2015 Participation in two outreach workshops for the scientific divulgation of the Basque language, Basque Summer University (UEU).

2012 & 2013 Participation in the 12th and 13th editions of the scientific outreach sessions *Semana de la ciencia, la tecnología y la innovación*.

2009 Participation in a talk on studying linguistics at the Axular Institutua high school.

Media consultations/interviews

National Geographic America, ABC, Cadena Ser, Euskadi Irratia.

2. Publications

(check egurtzegi.github.io/publications/ for open-access pdfs)

Monograph

In preparation [J. Blevins & **A. Egurtzegi**] *Syllables* (Oxford Studies in Phonology and Phonetics). Under contract with Oxford University Press.

Refereed journal articles

Accepted [**A. Egurtzegi**, D. Krajewska, C. Carignan & I. Urrestarazu-Porta] "An acoustic exploration of sibilant contrasts and sibilant merger in Mixean Basque". *Journal of the International Phonetic Association* [pending minor revisions].

2023 [J. Blevins & **A. Egurtzegi**] "Refining explanation in Evolutionary Phonology: Macro-typologies and targeted typologies in action". *Linguistic Typology* 27.2, 289-311.

2023 [**A. Egurtzegi** & G. Elordieta] "A history of the Basque prosodic systems". *Diachronica* 40.1, 30-72.

2023 [S. Coretta, J. Casillas, (…) , **A. Egurtzegi**, (…), & T. Roettger] "Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human speech analyses". *Advances in Methods and Practices in Psychological Science* 6.3.

2023 "/ɦ̃/ hasperen sudurkarituaren inguruan" [On the nasalized aspiration /ɦ̃/]. In press in *International Journal of Basque Linguistics and Philology* (*ASJU*) 57.

2022 [D. Krajewska, E. Zuloaga & **A. Egurtzegi**] "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1-31.

2022 "Goizuetako azentueraren garapena kokatzen Euskara Batu Zaharraren prosodia zehazteko: Proposamenen ebaluazioa" [Placing the development of the Goizueta prosodic system to define the prosody of Common Basque: Evaluation of the proposals]. *Fontes Linguae Vasconum* 133, 9-35.

2020 [J. Blevins, **A. Egurtzegi** & J. Ullrich]. "Final obstruent voicing in Lakota: Phonetic evidence and phonological implications". *Language* 96.2, 294-337.

2020 [**A. Egurtzegi** & C. Carignan] "An acoustic description of Mixean Basque". *Journal of the Acoustical Society of America* 147.4, 2791-2802.

2019 "Metathesis of aspiration as the source of anticipatory voicelessness in Basque". *Journal of French Language Studies* 29.2, 265-279.

2018 "On the phonemic status of nasalized /h̃/ in Modern Zuberoan Basque". *Linguistics* 56.6, 1353-1367.

2018 [**A. Egurtzegi** & I. San Martin] "Morfema mugetako hots aldaketak Zaldibarko euskaran: Azterketa akustikoa" [Morpheme-boundary sound changes in Zaldibar Basque: Acoustic study]. *Lapurdum* 21, 3697.

2018 "Herskarien ustezko ahoskabetasun asimilazioa eta euskal herskari zaharren gauzatzea" [The apparent assimilation of voicelessness and the realization of old Basque stops]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 52, 189-206.

2017 [J. Blevins & **A. Egurtzegi**] "Unexpected obstruent loss in initial obstruent–sonorant clusters: An apparent example from Basque". *Phonology* 34.3, 507-522.

2017 "Phonetically conditioned sound change: Contact induced /u/-fronting in Zuberoan Basque". *Diachronica: International Journal for Historical Linguistics* 34.3, 331-367.

2015 "Different distributions of contrastive vowel nasalization in Basque". *Loquens: Spanish Journal of Speech Sciences* 2.1, e017.

2013 "Ekialdeko /o/-ren igoeraren testuinguru fonologikoaren bila" [Looking for the phonological context of /o/-raising in Eastern Basque]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 48.2, 257-266.

2013 [**A. Egurtzegi** & B. Ariztimuño] "Remarks on the etymology of the Basque word for 'swallow' and its potential relation to Celtic". *Ériu* 63, 79-90.

2012 "Hugo Ernst Schuchardt: Bibliografía vasca" [Hugo Ernst Schuchardt: Basque Bibliography]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 46, 161-176.

2011 "Euskal metatesiak: abiaburua haien ikerketarako" [Basque metatheses: starting point for their research]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 45, 1-79.

2011 [B. Ariztimuño & **A. Egurtzegi**] "Enarak, elaiak eta testuinguruaren garrantzia" [Swallows and the importance of context]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 45, 403-410.

2009 [M. Ayerbe Linares & **A. Egurtzegi**] "La formación del plural en alemán desde una perspectiva morfofonológica" [The development of the German plural from a morpho-phonological perspective]. *Estudios Filológicos Alemanes* 19, 237-268.

Book chapters

Forthcoming "Metathesis". In E. Aldridge, A. Breitbarth, K.É. Kiss, A. Ledgeway, J. Salmons & A. Simonenko (eds.), *The Wiley Blackwell Companion to Diachronic Linguistics*. Chapter submitted. Publication contract signed in 2022.

2022 [G. Elordieta & **A. Egurtzegi**] "Bizkaiko iparraldeko euskararen doinu-azentuaren berrikusketa" [A revision of the pitch accent of Northern Bizkaian Basque]. In Barberia Aurrekoetxea & Iribar Ibabe (eds.), *Papers on Basque linguistics in honor of Jon Franco*. Bilbao: Deusto University Press, 319-337.

2019 [**A. Egurtzegi** & B. Ariztimuño] "/f/-ren sarrera eta sorrera euskaran" [Introduction and inception of /f/ in Basque]. In Epelde & Jauregi (eds.), *Bihotz ahots. M.L. Oñederra irakaslearen omenez*. Bilbao: UPV/EHU, 197-213.

2018 "Nasalized mid back vowel raising in Gascon and Basque". In Recasens & Sánchez-Miret (eds.), *Production and perception mechanisms of sound change*. Munich: Lincom, 173-187.

2015 "Zubereraren /y/: bere sorrera eta hedapenaren inguruan" [Zuberoan /y/: on its creation and spread]. In Ezeizabarrena & Gómez (eds.), *Eridenen du zerzaz kontenta*. Bilbao: UPV/EHU, 149-162.

2013 "Phonetics and Phonology". In Martínez-Areta (ed.), *Basque and Proto-Basque. Language-Internal and typological approaches to linguistic reconstruction*. Frankfurt am Main: Peter Lang, 119-172.

2013 "Diferentes tipos de aspiración en vasco (con análisis espectrales del dialecto suletino actual)" [Different kinds of aspiration in Basque (with spectral analyses of the modern Zuberoan dialect)]. In Blasco et al. (eds.), *Iberia e Sardegna*. Firenze: Le Monnier: 151-169.

Refereed conference proceedings

2023 [**A. Egurtzegi**, A. García-Covelo & I. Urrestarazu-Porta] "A nasalance-based study of the /h/ vs. /h̃/ opposition in Zuberoan Basque". In Radek Skarnitzl & Jan

Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague: Guarant International, 3427-3431.

2015 "[õ] sudurkarituaren igoera ekialdeko euskalkietan" [Raising of nasalized [õ] in the eastern Basque dialects]. *Ikergazte: International research in Basque*, 39-45, Bilbao: UEU.

2013 [**A. Egurtzegi** & G. Elordieta] "Euskal azentueren historiaz" [On the history of Basque accentuation]. In Gómez et al. (eds.), *3rd Conference of the Luis Michelena Chair*. Vitoria-Gasteiz: UPV/EHU, 163-186.

Outreach publications

2018 "Can speech errors become part of a language?". *Berichte 2017*. Munich: Internationales Begegnungszentrum der Wisseschaft München e.V., 10-13.

Preprints and other publications

2021 [C. Carignan & **A. Egurtzegi**] "Principal components variable importance reconstruction (PC-VIR): Exploring predictive importance in multicollinear acoustic speech data". *arXiv:2102.04740 [stat.ME]*.

2021 *Nazioarteko Alfabeto Fonetikoa (2021ean moldatua)* [Basque translation of *The International Phonetic Alphabet*]. International Phonetic Association. Online: https://www.internationalphoneticassociation.org/IPAcharts/IPA_chart_trans/pdfs/IPA_Kiel_2021_full_eus.pdf.

Oral contributions

Invited talks

2023 "Does Basque have an impossible phonological opposition?". *Speech and Bilingualism Talks* (Basque Center on Cognition, Brain, and Language (BCBL), Spain).

2023 "An 'impossible' opposition: /h/ vs. /ɦ̃/ in North-Eastern Basque". *53rd Linguistic Symposium on Romance Languages* (LSRL 53), *Phonetics and phonology – Beyond Romance* (Paris, France).

2023 "A LabPhon account of Basque /ɦ̃/". *CLT Seminars* (Universitat Autònoma de Barcelona, Spain).

2023 "La variación intergeneracional como reflejo del cambio en curso: La post-aspiración de /s/ + oclusiva en andaluz". *Linguistic Variation Talk Series* (Universidad Autónoma de Madrid, Spain).

2023 [with J. Cronenberg & J. Harrington] "Andalusian aspiration as a path into sound change research". Talk in the *Linguistics MA* program (University of the Basque Country, Vitoria-Gasteiz, Spain).

2023 [with A. García-Covelo & I. Urrestarazu-Porta] "Endangered language documentation: A nasalance-based study of the /h/ vs. /ɦ̃/ opposition in Zuberoan Basque". *Linguistics MA* program (UPV/EHU).

2022 [with J. Cronenberg & J. Harrington] "Andalusian pre-to-post-aspiration and the life cycle of sound change". *Workshop on Sound Change*. Departamento de Lengua Española y Lingüística General (UNED, Madrid, Spain).

2022 "When time is the key: Analyzing time-varying data in laboratory phonology". *LPRG - Laboratory Phonology Research Group* (Université Paris Cité, Paris, France).

2022 "Phonological features conditioned by contact with Gascon in Zuberoan Basque". *Basque Summer Tutorial in Language Documentation* (U. Bordeaux-Montaigne & Uni Konstanz, Larraine, France).

2021 [with D. Krajewska] "Sibilant contrasts and sibilant merger in Mixean Basque". *Lancaster Phonetics Laboratory* (U. of Lancaster, UK).

2021 "Final obstruent voicing in Lakota: Phonetic evidence and phonological implications". *Séminaires de Recherches en Phonétique et Phonologie* (U. Sorbonne Nouvelle, Paris, France).

2021 "Reconstructing Proto-Basque prosody and the development of the modern Basque prosodic diversity". *Séminaire CLLE-ERSSàB* (Université Bordeaux Montaigne, France).

2019 "Amiküzeraren deskribapen akustikoa" [Acoustic description of Mixean Basque]. Workshop: *MLV5en 3. mintegia* (University of the Basque Country, Vitoria-Gasteiz, Spain).

2019 "Final obstruent voicing in Lakhota: Phonetic evidence". *Workshop of the IPS* (Venice International University, Venice, Italy).

2019 "Reconstruyendo el protovasco: Hacia una nueva metodología para la reconstrucción fonológica de lenguas aisladas". *Departamento de Lengua Española y Lingüística General* (UNED, Madrid, Spain).

2019 "How phonetics can complement the comparative method in sound change research: Reconstructing Basque laryngeals". Workshop: *Recent Advances in Comparative Linguistic Reconstruction* (SOAS University of London).

2019 "A case of perceptual reanalysis: The ongoing merger of /h/ and /ɦ̃/ in Mixean Basque". Workshop: *Whither reanalysis?* (Humboldt University Berlin).

2018 "Re-reconstructing Proto-Basque: Towards a new methodology for the reconstruction of a linguistic isolate". Talk at *Laboratoire Parole et Langage* (CNRS-Aix-en-Provence, France).

2018 "Contact-induced shift in the laryngeal features of Proto-Basque stops". *Atelier de Phonologie* (SFL CNRS-Paris, France).

2018 "Reconstructing the phonetic realization of Proto-Basque stops". *Endangered Language Initiative Talk Series* (City University of New York, USA).

2018 "Looking for the cognitive basis of a little-studied sound change: reciprocal metathesis". *Nijmegen Linguistics* (Radboud University Nijmegen, Netherlands).

2018 "Origen y evolución de la metátesis recíproca en castellano" [The origin and evolution of reciprocal metathesis in Spanish]. (University of Liege, Belgium).

2017 "Euskal herskari zaharren gauzatze fonetikoa" [The phonetic realization of old Basque stops]. *MLV5en 1. mintegia* (UPV/EHU, Vitoria-Gasteiz, Spain).

2017 "Can speech errors become part of a language?". *The Humboldt Salon* (Munich, Germany).

2017 "A case of necessarily sporadic sound change: Reciprocal metathesis". Workshop: *The regularity principle in sound change* (University of Cologne, Germany).

2017 "Metátesis recíproca: La transposición de dos segmentos no consecutivos" [Reciprocal metathesis: The transposition of two non-consecutive sounds]. *Oberseminar Romanistische Linguistik* (University of Munich, Germany).

2015 [with G. Elordieta] "Basque accentuation: From word-level to phrase-level and back". *1st International Workshop on Basque-Japanese and Neighbouring Languages* (UPV/EHU, Vitoria-Gasteiz, Spain).

2015 "Euskararen hotsak: orain eta lehen, han eta hemen" [The sounds of Basque: Now and then, here and there]. *Euskararen historian barna: uste dugunetik dakigunera* (UEU, Vitoria-Gasteiz, Spain).

2014 "Towards a phonetically based historical phonology of Basque". *Jornadas de la UFI 11/14 / UFI 11/14aren jardunaldiak* (UPV/EHU, Vitoria-Gasteiz, Spain).

2014 "Phonetically conditioned (and contact-induced) /u/-fronting in Zuberoan Basque". *NOSH: New Observations in Speech and Hearing* (University of Munich, Germany).

2014 "Euskararen laringalen historiaz (edo zergatik bi hatxe behar ditugun zuberera modernoan)" [On the history of Basque laryngeals (or why we need two aitches in Modern Zuberoan)]. *IKER UMR-5478 Talks* (Bayonne, France).

2014 "Euskararen hots 'bitxiak'" [The 'strange' sounds of Basque]. *Euskararen lekua munduan eta historian* (Udako Euskal Unibertsitatea, Eibar, Spain).

2013 [with J. Blevins] "Advances in Proto-Basque Reconstruction: Extensions of the CVC Root Theory". Plenary talk. *International Conference on Historical Linguistics 21* (University of Oslo, Norway).

2013 "Accentogenesis in Basque: From phrasal pitch accent to word-level stress". *Endangered Language Initiative Talk Series* (City University of New York, USA).

2012 [with G. Elordieta] "Euskal azentueren historiaz" [On the history of Basque accentuation]. *3. Conference of the Koldo Mitxelena Chair* (UPV/EHU, Vitoria-Gasteiz, Spain).

2012 "Avances en fonología (pre)histórica vasca: Diferentes tipos de aspiración en vasco" [Advances in Basque (pre)historical Phonology: Different kinds of aspiration in Basque]. *Gorosti U5B3* (Università degli Studi di Cagliari, Italy).

2011 "Origini e caratteristiche della lingua basca. Una politica linguistica di successo" [Origin and characteristics of the Basque language. A successful linguistic policy]. *Gorosti 2* (Università degli Studi di Cagliari, Italy).

Papers presented (selected)

2023 [with A. García-Covelo & I. Urrestarazu-Porta] A nasalance-based study of the /h/ vs. /h̃/ opposition in Zuberoan Basque. *International Conference on Phonetics Sciences* (*ICPhS*; Prague, Czech Rep.).

2023 [with A. García-Covelo & I. Urrestarazu-Porta] Documenting nasalized aspirates in Zuberoan Basque. *9th International Conference on Experimental Phonetics* (*CIFE*; University of Vigo, Spain).

2023 [with D. Krajewska, C. Carignan & I. Urrestarazu-Porta] "An acoustic description of sibilant contrasts and sibilant merger in Mixean Basque". *9th International Conference on Experimental Phonetics* (*CIFE*; University of Vigo, Spain).

2023 [with E. Zuloaga & D. Krajewska] "Textual genres and linguistic variation: Socio-historical explanations for Basque sibilant mergers". *Historical Sociolinguistics Network Conference* (*HiSoN*-12) (Vrije Universiteit Brussel, Belgium).

2022 [with D. Krajewska & E. Zuloaga] "A quantitative analysis of sibilant mergers in 18th-century Basque". *25th International Conference on Historical Linguistics* (*ICHL*; University of Oxford, UK).

2022 [with G. Elordieta] "A history of the Basque prosodic systems". *25th International Conference on Historical Linguistics* (*ICHL*; University of Oxford).

2022 [with J. Cronenberg, M. Gubian, J. Harrington & H. Ruch] "La aspiración de /s/ en andaluz: el largo camino de la post-lexicalidad a la lexicalización". *CILPR* 30 (University of La Laguna, Spain).

2022 [with J. Cronenberg, M. Gubian, J. Harrington & H. Ruch] "From post-lexicality to phonologization: Andalusian /s/-aspiration at the word boundary". *LabPhon18* [held online].

2022 [with G. Elordieta] "A new proposal for the tonal representation of phonological phrases in Northern Bizkaian Basque". 31st *Colloquium of Generative Grammar* (University of the Balearic Islands, Spain).

2022 [with G. Elordieta] "Lexical and non-lexical tones in Basque revisited". *Old World Conference on Phonology* 19 (University of Deusto at Donostia/San Sebastián).

2021 [with G. Elordieta] "Lexical and non-lexical tones in Northern Bizkaian Basque". *UD Workshop on Basque Linguistics* (University of Deusto, Spain).

2020 [with C. Carignan] "Rhinoglottophilia is alive and well in Mixean Basque". *LabPhon17* (University of British Columbia, Vancouver, Canada [held online]).

2017 "The phonetic realization of Proto-Basque stops and the reconstruction of laryngeal contrasts". *3rd Symposium on Historical Phonology* (University of Edinburgh, UK).

2017 "The loss of a typologically rare opposition in two endangered dialects: /h/ vs. /h̃/ in Mixean and Zuberoan Basque". *23th International Conference on Historical Linguistics* (*ICHL*; University of Texas San Antonio, USA).

2017 "Reciprocal metathesis: Typology and phonetic basis of a little-studied process". *50th Annual Meeting of the Societas Linguistica Europaea* (University of Zurich, Switzerland).

2017 "Reciprocal metathesis: The case of Basque". *Abstraction, Diversity and Speech Dynamics* (Herrsching am Ammersee, Germany).

2017 "Sporadic nasalization of /h/: the loss of a rare contrast in Mixean Basque". *4th International Workshop on Sound Change* (University of Edinburgh, UK).

2016 "Nasalized [õ]-raising in Basque and Gascon". *3rd Workshop on Sound Change* (University of Salamanca, Spain).

2016 [with J. Blevins] "The nature of obstruent loss in Basque borrowings with initial OR clusters". *CUNY Phonology Forum* (City University of New York, USA).

2015 "The evolution of the different distributions of contrastive vowel nasalization in Basque". *2nd Symposium on Historical Phonology* (University of Edinburgh, UK).

2015 "Anticipatory distant 'voicelessness' assimilation in Basque". *22th International Conference on Historical Linguistics* (*ICHL*; Naples, Italy).

2015 "The role of prosodic prominence in Basque aspirate distribution". Workshop: *Strong versus weak prosodic positions* [*37th Meeting of the DGfS*] (Universität Leipzig, Leipzig, Germany).

2014 "On the historical origin of Basque glottal fricative(s): phonetic evidence". *Societas Linguistica Europaea* 47 (*SLE*; Adam Mickiewicz University, Poznan, Poland).

2014 “Phonetically conditioned sound change: /u/-fronting in Zuberoan Basque”. *Sound Change in Interacting Human Systems* (University of California, Berkeley, USA).

2012 [with G. Elordieta] “Historical development of modern Basque accentuation systems”. *Tone: Theory and Practice* (Max Planck Institute, Leipzig, Germany).

2011 “Metathesis processes in Basque”. *Phonetics and Phonology in Iberia* (*PaPI*; Universitat Rovira i Virgili, Tarragona, Spain).

2010 “Sonorant and vowel metatheses in Basque”. *LESCOL II* (University of Leipzig, Germany).

3. PhD diploma

eman ta zabal zazu

UPV EHU

Felipe VI, Rey de España

jaunak eta haren izenean

y en su nombre el

nik, Euskal Herriko Unibertsitateko errektoreak,

Rector de la Universidad del País Vasco

indarrean dagoen legeriak jasotako xedapen eta baldintzen arabera, beherago adierazitako titulua ematen diot honako honi:

considerando que, conforme a las disposiciones y circunstancias previstas por la legislación vigente,

ANDER EGURTZEGI MARTINEZ

Jaiotze data: 1986ko martxoaren 22a
Jaiotze lekua: Barakaldo (Bizkaia)
Nazionalitatea: espainiarra

nacido el día 22 de marzo de 1986 en Barakaldo (Bizkaia), de nacionalidad española,

Izan ere, 2014ko abenduan ikasketak gainditu ditu unibertsitate honetan titulu hau eskuratzeko:

ha superado en diciembre de 2014 los estudios conducentes al título universitario oficial de

Euskal Herriko Unibertsitateko doktorea

Doctor por la Universidad del País Vasco

Aipamena: Cum Laude eta Nazioarteko doktorea

Mención: Cum Laude y Doctor Internacional

Doktoregoko Programa Ofiziala:
Hizkuntzalaritza

dentro del Programa Oficial de Doctorado:
Lingüística

eta Ministro Kontseiluaren erabakiaz onartua (2011/02/11).

establecido por Acuerdo de Consejo de Ministros de 11 de febrero de 2011,

Titulu ofizial honek estatu osoan du balioa, eta indarreko legeriaren arabera dagozkion eskubide guztiak ematen dizkio tituludunari.

expide el presente título oficial con validez en todo el territorio nacional, que faculta al interesado para disfrutar los derechos que a este título otorgan las disposiciones vigentes.

Leioan, 2014ko abenduaren 23an emana.

Dado en Leioa, a 23 de diciembre de 2014.

Interesduna,
El Interesado,

Errektorea,
El Rector,

Josebe Iñaki Goirizelaia Ordorika

Tituluen Bulegoko burua,
La Jefa del Negociado de Títulos,

Yolanda Fernández Santamaría

020A-033070

Registro Nacional de Títulos	Código de CENTRO	Registro Universitario de Títulos
2015091826	48012735	201500104

SIGNE, S.A.

4. Research program including selected previous works

(check egurtzegi.github.io/publications/ for open access pdfs)

Introduction

My research concerns the study of sound change from all the potentially relevant theoretical angles and subdisciplines of linguistics, and it has already brought new insights and important results to the fields of historical linguistics, theoretical and laboratory phonology, linguistic typology, and to the reconstruction of the Basque language. It involves a holistic approach to historical explanation in which phonological analysis is combined with philological study, typological assessment, phonetic experimentation and analysis, sociolinguistics, linguistic contact and acquisition to reach a complete account of sound change and reconstructed sound patterns.

Given that my research has a clear focus on sound change, I consider myself a historical laboratory phonologist, as well as a bascologist. Two major topics emerge from my work: how sound change is conditioned by phonetic detail as well as external factors such as language contact, and the reconstruction of Proto-Basque phonology. These are also the topics that will be the guidelines for my future research projects. I have worked on improving our understanding of these issues by investigating instances of sound change that have not been appropriately addressed thus far, especially within Basque but sometimes also working on other languages, including the Romance languages Gascon and Andalusian Spanish or the Siouan language Lakota. With these interests in mind, I have advanced the knowledge of the linguistic community on four related but different disciplines, namely historical linguistics, sound change typology, phonological theory, and acoustic phonetics. In this section, I'll discuss my early, recent and ongoing research in the context of the disciplines it adheres to and show that my previous experience has prepared me to carry out the research outlined in the section on future research perspectives. Not all my papers are equally discussed here. Instead, I focus on a number of recent papers that have been more influential and I use them to showcase the different approaches and disciplines present in my previous and current work, which I also plan to incorporate to my future research projects.

4.1 Basque historical linguistics

(Egurtzegi 2013a, 2014, 2015, 2018b, 2018c, 2022, 2023; Egurtzegi & Ariztimuño 2013, 2019; Egurtzegi & Elordieta 2013, 2023; Elordieta & Egurtzegi 2022, in prep., Egurtzegi et al. 2023, in prep., etc.)

My research program has resulted in a great leap forward in the reconstruction of the phonology of Proto-Basque and its further evolution to these of the modern Basque varieties. Major advancements have been reached in our understanding of the evolution of its vowel inventory (Egurtzegi 2015, 2017), stop oppositions (Egurtzegi 2019), laryngeal contrasts (Egurtzegi 2018a), and prosodic system (Egurtzegi & Elordieta 2023). These advancements are especially important, given that the field of Basque historical phonology was at a point considered stuck after the major works by L. Michelena in the 1970's, and my work alongside that of J.A Lakarra's and J.I. Hualde's have revitalized it in more recent times.

Basque has been the main language of study in my work, and I've mostly studied it from a historical perspective. This has been so since the beginning of my career as a researcher, when I mainly put forward studies on Basque historical phonology (Egurtzegi 2011, 2013a, 2013c; Egurtzegi & Elordieta 2013) and etymology (Ariztimuño & Egurtzegi 2011; Egurtzegi & Ariztimuño 2013). One of my first papers (Egurtzegi 2013a) involved an updated general overview of the historical phonology of the language, and it is still the most recent reference chapter on the historical phonology of Basque published in English. The research I made to write this book chapter enabled me to work on the historical phonology of Basque on a thoroughly informed manner from my early days as a doctoral student, and paved the way for my doctoral thesis.

I defended my dissertation *Towards a phonetically grounded historical phonology of Basque*, supervised by Prof. Joseba Lakarra (University of the Basque Country – UPV/EHU) and Prof. Juliette Blevins (City University of New York - CUNY), in December 2014. It already incorporates my main lines of investigation with the Basque language as a subject of study. The dissertation analyzes Basque diachronic phonology within a phonetically informed approach to sound patterns and sound change in line with *Evolutionary Phonology* (Blevins 2004, 2015). It assesses Basque phonological developments in terms of cross-linguistic typological tendencies, and incorporates recent experimental results from laboratory phonology, recognizing the importance of phonetics in phonology and its role in explaining instances of sound change.

My dissertation constituted the first extensive general work on the historical phonology of Basque since *Fonética Histórica Vasca* was written in 1961 (extended second edition: Michelena 1977). Sound patterns discussed in this dissertation include: the development of the Basque accentuation systems, including systems based on word-level stress and phrase-level pitch accent (see also Egurtzegi 2022; Egurtzegi & Elordieta

2013, 2023; Elordieta & Egurtzegi 2022, in prep.); the evolution of Basque laryngeals /h/ and /ɦ̃/ (see also Egurtzegi 2018a, 2023; Egurtzegi et al. 2023); the evolution of dialectal /y/ and phonetic conditions inhibiting u-fronting (see Egurtzegi 2017); the development of contrastively nasalized vowels in Basque dialects (Egurtzegi 2015), explanations for their distribution, and a phonetic account of [õ]-raising (see Egurtzegi 2013c and 2018b, which places this sound pattern in the context of the linguistic contact between Gascon and Basque); and a study of distinct metathesis types in the history of Basque, many of which have clear phonetic bases (Egurtzegi 2011; Egurtzegi forth.). In addition to detailed analyses of these processes, I provide chronologies of these sound changes, and highlight cases where contact may play a role.

While some of the topics discussed in the dissertation were not very present in the previous literature on Basque historical phonology, other topics (such as u-fronting and laryngeals) had been widely discussed by previous authors, yet a new perspective has been presented by my research, yielding new results and more accurate interpretations of the previously available historical data. At the same time, unique sound changes in the history of Basque are studied closely, in order to understand how they can inform theories of phonetically based sound change. Work that derived from some of the chapters of the dissertation has resulted in journal articles and book chapters, including publications in prestigious linguistic journals such as *Diachronica* and *Linguistics*.

As can be inferred from many of the works referred to in the previous paragraphs, I did not stop working on the topics present in my dissertation after the defense of the thesis, but I keep gradually improving our understanding of many of these topics even today, given that I consider topics such as the opposition between laryngeals (Egurtzegi et al. 2023, in prep.; Egurtzegi & Carignan in prep.) or the development of the different accentuation systems in Basque (Egurtzegi 2022; Egurtzegi & Elordieta 2023) of great importance beyond the field of Bascology. Among others, these studies have been proven important for phonological theory (Egurtzegi 2018a; Blevins & Egurtzegi 2017, etc.), to improve our understanding of the evolution of prosodic systems (Egurtzegi 2022; Egurtzegi & Elordieta 2023, etc.), sound change in situations of linguistic contact (Egurtzegi 2015, 2017, etc.), as well as for general typologies of sound change, to which most of my works have contributed.

My study on Zuberoan Basque /u/-fronting is a good example of the approach to historical phonology that I developed in my dissertation. This sound change has been widely described in the literature since the earliest research on Basque (cf. Uhlenbeck

1903; Gavel 1920; Lafon 1962a, 1962b; Michelena 1977, etc.). However, the split of Common Basque *u into Zuberoan /u/ and /y/ has puzzled bascologists for decades, with no complete account being proposed (Zarabozo 1972; Oñederra 2009a, 2009b, etc.). While u-fronting occurs in neighboring languages such as Gascon or French, neither of them shows the phonological context found in Zuberoan Basque, making an explanation solely based on contact impossible. In Basque, u-fronting is blocked before the coronal segments /ɾ/, /s̪/ and the clusters /rtʰ/, /rt/, /rd/, but not before any other segment, including the coronals /t/, /d/, /l/, /ʎ/, /s̪/, /ts̪/, /ʃ/, /ʧ/ and /r/ (cf., for instance, *hur* /huɾ/ 'water' vs. *hür* /hyr/ 'hazelnut'). The set of segments that inhibit fronting is particularly difficult to define by means of phonological natural classes, while excluding the segments that do not block the process. I argue that the inhibition of the fronting process occurred due to consonant coarticulation effects on the preceding high back vowel. More specifically, the inhibitory segments and clusters require active tongue dorsum lowering and backing to perform the fine movements of the tongue tip involved in their production. This research (Egurtzegi 2017) includes a philological study of the oldest Zuberoan texts and an account of the fronting process based on linguistic contact and phonetic detail. It was published in the journal *Diachronica* in 2017, and it has been discussed in detail in a recent monograph on sound change (Salmons 2021: 85-86) and in the chapter on contact of a recent reference handbook (Gardani 2022).

4.2 Phonological theory and sound change typology

Much of my work on Basque (as well as on other languages) has important implications for general theoretical phonology and, especially, for theories of phonetically based sound change. Among others, the proposed approach to the study of sound change has resulted in a better understanding of metathesis (Egurtzegi forth.) as well as of the interaction between historically sustained bilingualism and language-specific coarticulatory constraints in conditioning the development of contact-based historical change, as in the research on Zuberoan /u/-fronting discussed above (Egurtzegi 2017). Nonetheless, other research that I have put forward has also had an impact on phonological theory and phonological typology in topics such as unlikely phonological oppositions, lesser described yet recurrent sound changes, alternative analyses of uncommon sound patterns, and allegedly impossible sound changes.

Establishing typologically rare phonological contrasts
(Egurtzegi 2018a; Egurtzegi et al. 2023, in prep.-2; Blevins & Egurtzegi 2023)

Basque laryngeals have been a usual topic in the literature on the history of Basque (cf. Michelena 1977; Trask 1997; Lakarra 2018; Igartua 2015; Hualde 2018; Blevins 2018, etc.). Nevertheless, one of the results of my dissertation was establishing for the first time that a typologically rare phonological opposition between /h/ and /h̃/ exists in modern Zuberoan and Mixean varieties of Basque, and can indeed be traced back to Old Common Basque. In these dialects, in sequences such as [ãh̃ã]-, nasalization extends throughout the vowel-aspirate-vowel sequence, which has made many authors conclude that both vowels surrounding the aspirate are contrastively nasalized (cf. Michelena 1977). Hualde (1993) proposed that the phonologically nasal segment was /h̃/ instead, but little evidence was presented for this claim. I argued for Hualde's position with evidence that strongly suggests an opposition between /h/ and /h̃/. Evidence for the contrast comes mainly from two sources. First, a newly described process of assimilation of /h/ to /h̃/ in nasal environments which then serves as a basis for the analogical extension of the nasalized aspirate in a context where it cannot be phonologically derived. And second, from the neighboring Mixean Low Navarrese dialect (Camino 2016), a variety that shows the nasalization contrast in aspirates but lacks contrastively nasalized vowels, leaving no potential coarticulatory source for /h̃/. The phonological part of the research was published in the journal *Linguistics* (Egurtzegi 2018a) in 2018, and a first nasalance-based study is already available as an *ICPhS* proceedings paper (Egurtzegi et al. 2023). I'm currently working on a future, more detailed paper on the phonetics behind this rare phonological opposition (Egurtzegi et al. in prep.-2).

Proposing alternative analyses for uncommon sound patterns
(Egurtzegi 2019; Blevins & Egurtzegi 2017)

A number of phonological processes have been accounted for in terms that make them extraordinarily typologically unusual. At least two examples of this can be found in Basque, and I have proposed alternative, more explanatory accounts for both of them.

First, in Blevins & Egurtzegi (2017), we analyze an unexpected pattern of obstruent deletion in the adaptation of obstruent-sonorant onset clusters in loanwords from Romance into Basque (cf. Lat. *flore* >> Basque *lore* 'flower') by linking the obstruent deletion to the process of cluster palatalization in Romance (cf. Old Spanish *llorado* 'bloomed' in contrast to the learned word present in modern Spanish, *florecido* 'bloomed'). Instead of deletion of the cluster-initial obstruent during an early loanword

adaptation from Latin (i.e. Latin /fl-/ > Basque /l-/), we propose that Basque borrowed these loanwords after the beginning of Romance cluster palatalization, and adapted /ʎ/ to /l/ due to the lack of the former segment in older stages of Basque. This study was published in *Phonology* in 2017.

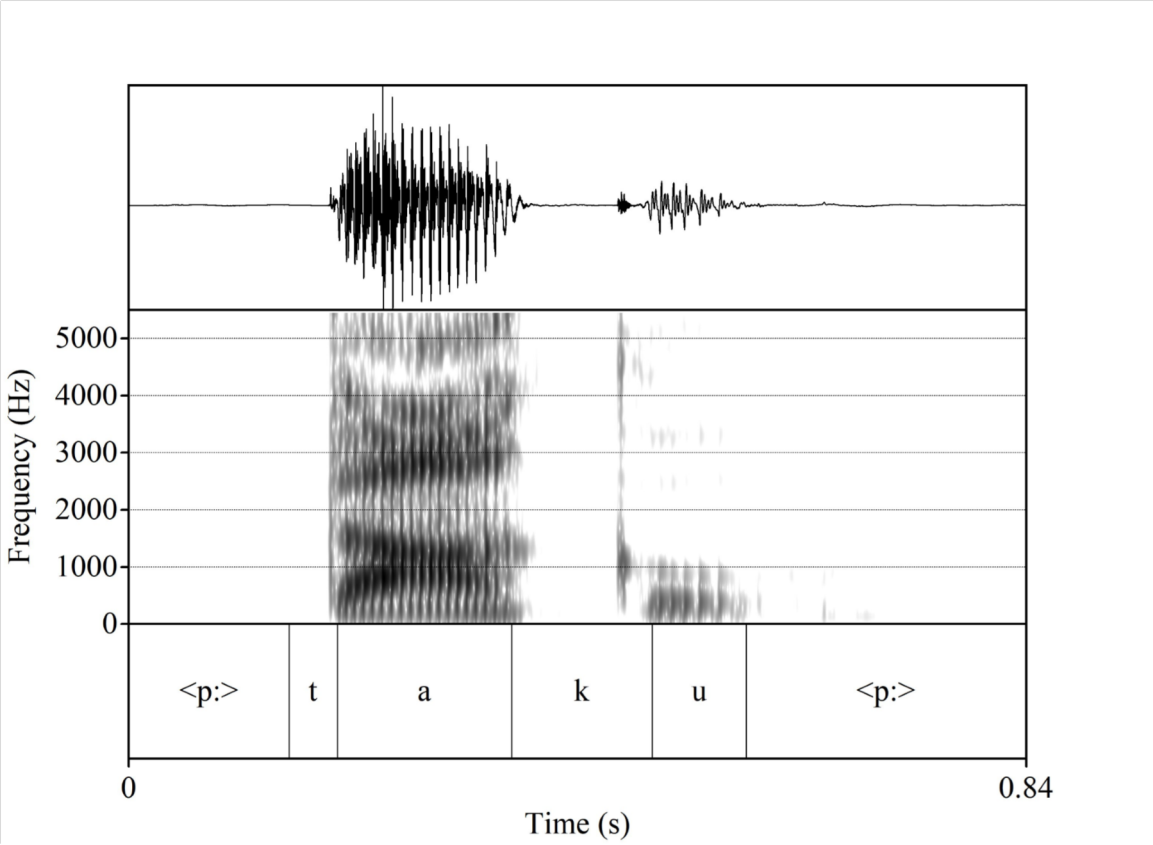
Second, Egurtzegi (2019) presents a new interpretation of a cross-linguistically infrequent process of distant anticipatory assimilation of voicelessness in the French dialects of Basque. This process can be observed when comparing dialectal variants such as *gathea* vs. *khatea* 'chain'. While assimilation of voicing or voicelessness is common when the affected segments are in contact, it is almost unheard of at a distance (cf. Hansson 2004). Instead, I analyze this process as a metathesis of stop aspiration, an explanation that also accounts for the distribution of aspirated voiceless stops in the modern dialectal variants. This paper was published in *Journal of French Language Studies* in 2019.

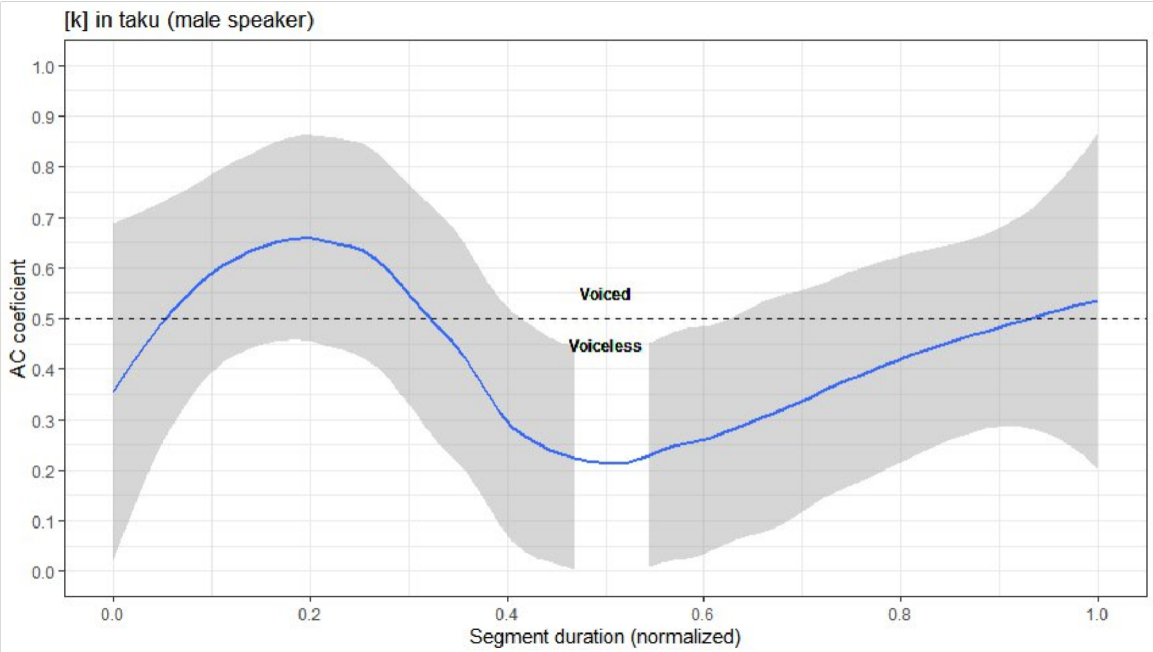
Documenting and analyzing allegedly impossible sound patterns
(Blevins, Egurtzegi & Ullrich 2020; Blevins & Egurtzegi 2023)

Some sound patterns are very frequent, while others are very uncommon or even non-occurring. Blevins (2006a) discussed the high frequency of sound patterns such as word-final obstruent devoicing in comparison to the virtually unattested opposite process of word-final obstruent voicing. She proposed that the recurrence of this sound pattern in the world's languages is a consequence of obstruent devoicing being the outcome of a number of phonetically natural pathways, i.e., a number of known phonetic biases are likely to result in the devoicing of word-final obstruents. Nevertheless, Evolutionary Phonology predicts that word-final voicing is possible as the result of the combination of multiple sound changes, but, not being due to a single, phonetically natural change, it is very uncommon (cf. Blevins 2006a). This prediction contrasts with approaches based on markedness that deem a synchronic rule of word or syllable-final obstruent voicing impossible (cf. Kiparsky 2006).

In this research, Prof. Juliette Blevins and I analyze a sound pattern of syllable and word-final stop voicing in the Siouan language Lakota with the assistance of Dr. Jan Ullrich, an expert on the Lakota language and responsible of the main dictionary of the language, the *New Lakota Dictionary* (Ullrich 2011). In this language, words ending in a TV sequence (T being a voiceless stop and V a vowel) are produced with final vowel loss and final obstruent voicing under two different conditions: 1- in derivative and

reduplicated forms, as in *sápA* ‘black’, redup. form *sabsápA* ‘black (pl.)’, adv. *sabyá* ‘blackly’; and 2- in fast speech forms, as in *sáb* ‘black (fast speech)’. These alternations are illustrated by the waveforms, spectrograms and Auto-Correlation coefficients (values from 0 to 1, values closer to 0 being more voiceless and these closer to 1 more voiced) as a function of normalized time of the morpheme-alternant pair of words in Figure 1.





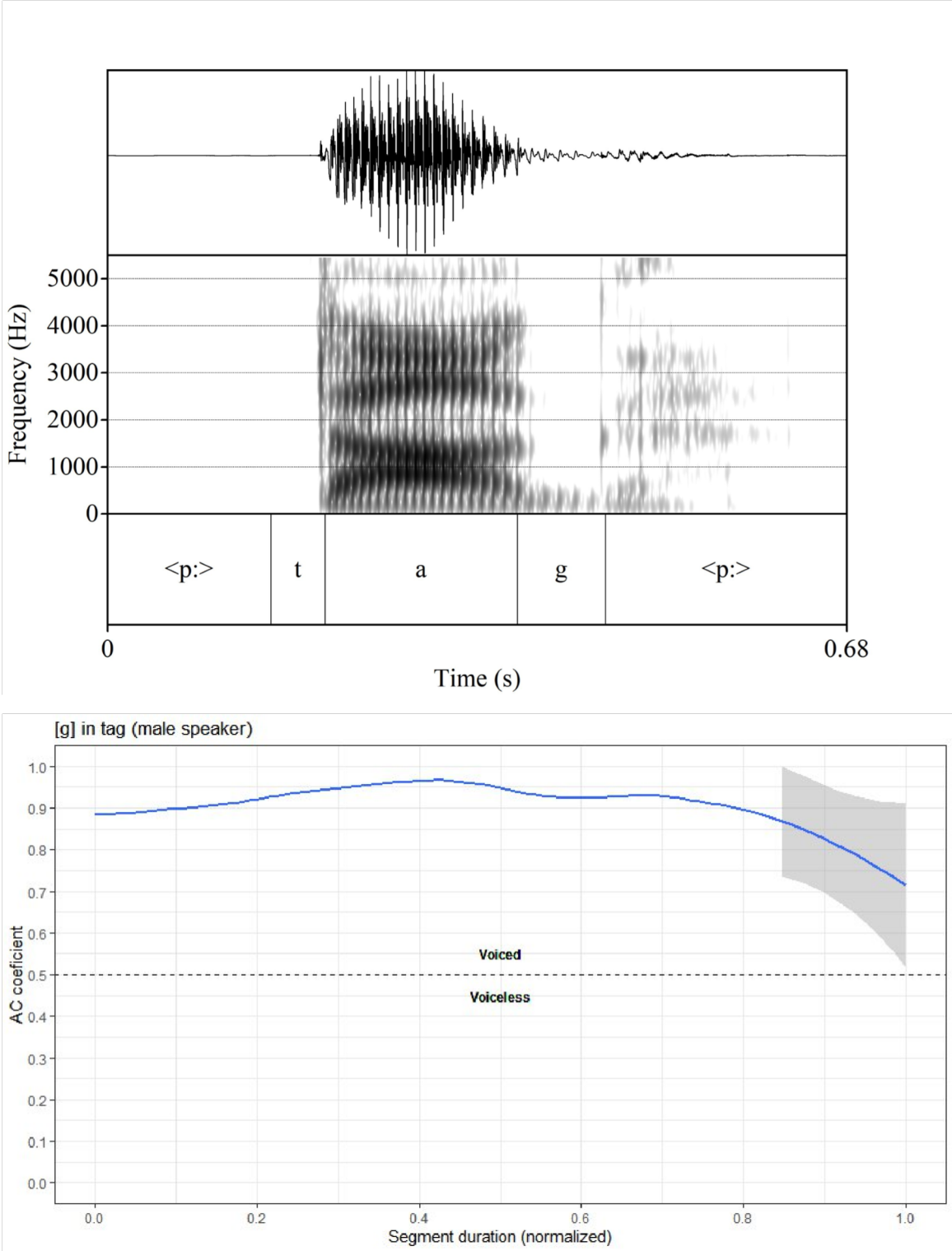


Figure 1 – Comparison of consonants in a morpheme-alternant pair: taku 'what' vs. tag 'what (CONT)'

To reach a thorough understanding of Lakota final voicing, we study the historical pathway that could have resulted in final voicing of voiceless stops as well as the synchronic realization of these stops, in order to show that they are realized as actual stops and that they are phonetically voiced. Thus, this paper presents a synchronic account, an

acoustic analysis and the historical evolution of an allegedly impossible synchronic pattern of stop voicing in syllable codas as found in the Native American language Lakota. The manuscript that resulted from this research was published in the journal *Language* in 2020. See also Blevins & Egurtzegi (2023) on the importance of lesser-studied languages and varieties in the documentation of allegedly impossible sound patterns and my work on the nasalization opposition in aspirates in Zuberoan and Mixean Basque (above, and again below).

4.3 Phonetics

With regard to the field of phonetics and laboratory phonology, my research has brought rare sound patterns and endangered languages and varieties to the spotlight. This has been done by studying and documenting allegedly impossible sound changes (Blevins et al. 2020) and phonological oppositions (Egurtzegi 2018a; Egurtzegi et al. 2023) that would otherwise remain unattested in largely under-researched languages. My work has been the first to present acoustic evidence of a process of final obstruent voicing (in Lakota, see Blevins et al. 2020 and above), and acoustic and nasalance evidence of an opposition between a nasalized /h̃/ and an oral /h/ (in Mixean and Zuberoan Basque, respectively; see Egurtzegi and Carignan in prep.; Egurtzegi et al. 2023 and below for summaries of these works). In addition, my phonetic research has focused in documenting the most endangered varieties of Basque (such as Zuberoan and Mixean; Egurtzegi & Carignan 2020), of which we had very limited acoustic data and will be drastically altered in a generation or two (if not lost altogether). The last important results from my work on phonetics come from the testing of phonological theories with sound changes currently underway (Egurtzegi et al. in prep.-1).

As mentioned above, my dissertation presented evidence of the necessity of incorporating phonetic detail into the study of historical sound change while discussing a number of typologically rare sound patterns. While sound change and phonological typology have been my main research interests, my recent research has incorporated a bigger phonetic component, as I have gained knowledge and experience from my 3-year working experience at the IPS phonetics laboratory of the University of Munich (LMU), one of the most prominent phonetics laboratories in Europe, as a post-doc under the supervision of Professor Jonathan Harrington, director of the IPS-LMU. An indicator of the growing importance of phonetics in my research is the number studies that I am

currently conducting within this field. These include research on palatal stops in Bizkaian Basque (Egurtzegi & Egurtzegi in prep.) and nasalized aspirates in Mixean Basque (Egurtzegi & Carignan in prep.) and Zuberoan Basque (Egurtzegi et al. 2023; in prep.-2), documentation of lesser studied languages/varieties (Egurtzegi & Carignan 2020; Egurtzegi et al. 2023, etc.), and studies of sound change in progress (Egurtzegi et al. in prep.-1), as well as the development of new methods for acoustical analysis (Carignan & Egurtzegi 2021).

Phonetics of Basque sound patterns

(Egurtzegi 2013b; Egurtzegi & Carignan 2020, in prep.; Egurtzegi & Egurtzegi in prep.)

While I started working on acoustic phonetics early in my career (specifically during my master's, cf. Egurtzegi 2013b), it was not until I spent some time working at the IPS-LMU phonetics laboratory in Munich that I returned to work on a purely phonetic research project, as opposed to my usual research on phonetically based historical phonology, as discussed above. Now, after spending more than two years on a phonetics laboratory, I am involved in a considerable number of projects in the field of phonetics; many of them, although not all, involving the Basque language. In a recently started study, for instance, Aitor Egurtzegi (University of Zurich - UZH) and I study the phonetic realization of voiceless palatal stops in the variety of Bizkaian Basque spoken in Aulesti. We find that /c/ is not always produced as voiceless, being fully voiced in some instances, and neither it is always produced as a stop, as it is sometimes pronounced as an affricate (Egurtzegi & Egurtzegi in prep.).

A more advanced study that I am currently working on involves the phonetic realization of the /h/ vs. /ɦ̃/ opposition in the endangered Mixean variety of Basque. This research has been done alongside Dr. Christopher Carignan from the Phonetics Laboratory at the University College London (Egurtzegi & Carignan in prep.). While I previously phonologically argued for the phonemic status of /ɦ̃/ in Zuberoan and Mixean Basque (Egurtzegi 2018), no study had presented phonetic evidence for the opposition between /h/ and /ɦ̃/ in these dialects, nor in any other language of the world, this opposition being as cross-linguistically uncommon as it is. In addition, previous dialectological research (Camino 2016) has suggested an ongoing merger of /h/ and /ɦ̃/ in Mixean Basque. This merger would be due to a generalization of nasality in aspirates that could potentially result in all instances of /h/ becoming /ɦ̃/. The nasalization in /ɦ̃/ would now be seen as a means of perceptually enhancing the aspirate, instead of a

contrastive feature that opposes /h/ to /h̃/. According to Ohala (1975), the coarticulatory cues of aspiration and nasalization present in vowels in contact to these segments are perceptually similar, a relationship usually referred to as rhinoglottophilia (Matisoff 1975). This ambiguity has resulted in a number of sound changes in many unrelated languages (Blevins 2004, see Igartua 2015 for Basque).

We present an acoustic study to evaluate the current state in the evolution of the claimed ongoing merger and evaluate the gradual nature of the extension of perceptual reanalysis to production. The (acoustic) degree of nasalization in the speech of six speakers of Mixean Basque was measured in the following four contexts: in /h/ adjacent to nasals (which are expected to be phonetically nasalized by coarticulatory assimilation), in etymologically nasalized items (i.e., phonemic /h̃/), in non-etymologically nasalized items (i.e., phonemic /h/ with no nasalization expected), and in vowels adjacent to oral stops (/p/, /t/, /k/, etc.) in which no nasalization is expected. These four contexts were chosen in order to observe whether there are any differences in the degree of nasalization between /h̃/ and /h/, and whether any of these segments is produced with no acoustic nasalization (as vowels adjacent to oral stops are expected to be). The results of the acoustic analysis reveal a two-way distinction in the data: nasalized contexts and non-nasalized contexts. The nasalized contexts include /h/ adjacent to nasals, as expected by the process of assimilation of nasality in aspirates described in Egurtzegi (2018), but also etymologically nasalized /h̃/. The non-nasalized contexts include vowels adjacent to oral consonants, as expected, but also etymologically oral /h/. This is shown by Figure 2, where A represents the degree of nasalization of assimilated [h̃] (e.g. /h/ in the vicinity of a nasal obstruent), E represents etymological /h̃/, N represents non-nasal [h], and V represents the group of oral vowels.

These results suggest that the phonological opposition between nasalized /h̃/ and oral /h/ still exists in Mixean Basque. This is an important result, given that such an unusual opposition has never been phonetically studied neither in Basque nor in any other language. Furthermore, we do not observe evidence that etymologically oral /h/ is ever produced with nasalization; otherwise, we would expect to find a three-way distinction in the acoustic analysis, i.e., etymologically nasalized /h̃/ is produced with more nasalization than etymologically oral /h/, which in turn is produced with more nasalization than oral vowels. Rather, no difference is observed in the degree of nasalization between etymologically oral /h/ and oral vowels, suggesting that etymologically oral /h/ is not nasalized in the language.

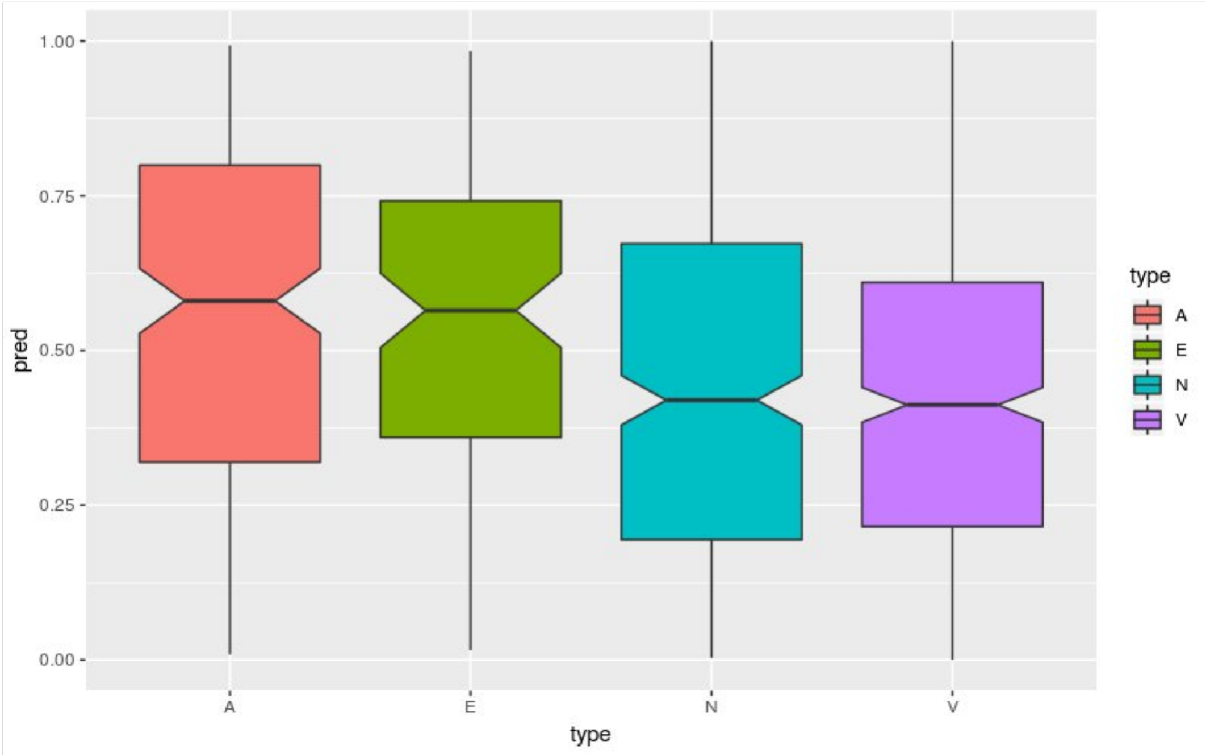


Figure 2 Degree of nasalization of assimilated [h̃]*, etymological* /h̃/*, non-nasal* [h]*, and oral vowels in Mixean Basque*

Contrary to these expectations, the acoustic analysis shows evidence of a stable opposition between /h/ and /h̃/ in Mixean Basque. We do not discard the possibility of an ongoing merger of these sounds, but we suggest that, if there is one, it is still on its incipient stage and does not reflect in the results of our analysis. We plan on submitting this research to the journal *Laboratory Phonology* in the near future.

While working in the aforementioned study on Mixean Low Navarrese, we also used the opportunity to perform a general phonetic description of this endangered dialect of Basque, which will be published as Egurtzegi & Carignan (2020). The manuscript presents an acoustic analysis of 10 speakers of the Mixean variety coming from 10 different towns in the region. In addition to our acoustic analysis, the paper includes an overview of previous acoustic studies performed on different Basque varieties, in order to synthesize the sparse acoustic descriptions of the language that are available. This synthesis serves as a basis for the acoustic analysis performed on the Mixean data, in which we replicate the various acoustic analyses given in previous studies in a single, cohesive general acoustic description of Mixean Basque. The analyses include formant and duration measurements for the six-vowel system, voice onset time measurements for the three-way stop system, spectral center of gravity for the sibilants, and number of

lingual contacts in the alveolar rhotic tap and trill. Important findings include: a centralized realization ([ʉ]) of the high-front rounded vowel usually described as /y/ (see Figure 3); a data-driven confirmation of the three-way laryngeal opposition in the stop system (see Figure 4); evidence in support of an alveolo-palatal to apical sibilant merger; and the discovery of a possible incipient merger of rhotics.

These results show how using experimental acoustic methods to study under-represented linguistic varieties can result in revelations of sound patterns otherwise undescribed in more commonly studied varieties of the same language. This study was published in the *Journal of the Acoustical Society of America* in 2020. An in-depth study of the sibilant system of this variety will appear in the near future in the *Journal of the International Phonetic Association*.[1]

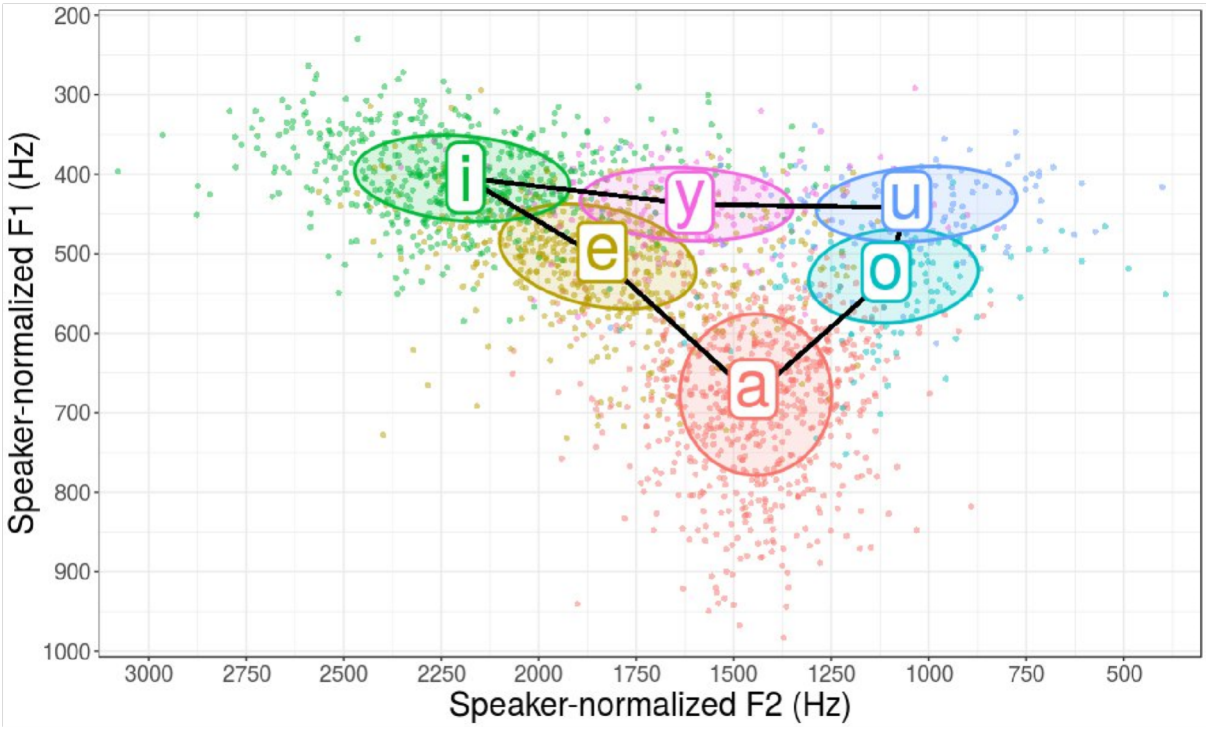


Figure 3 Distribution of the vowel space in Mixean Basque

1 The acoustic data we used for the studies of this endangered variety of Basque comes from Camino's (2016) continued work on the Mixe region (Amiküze) that spanned more than two decades of recordings.

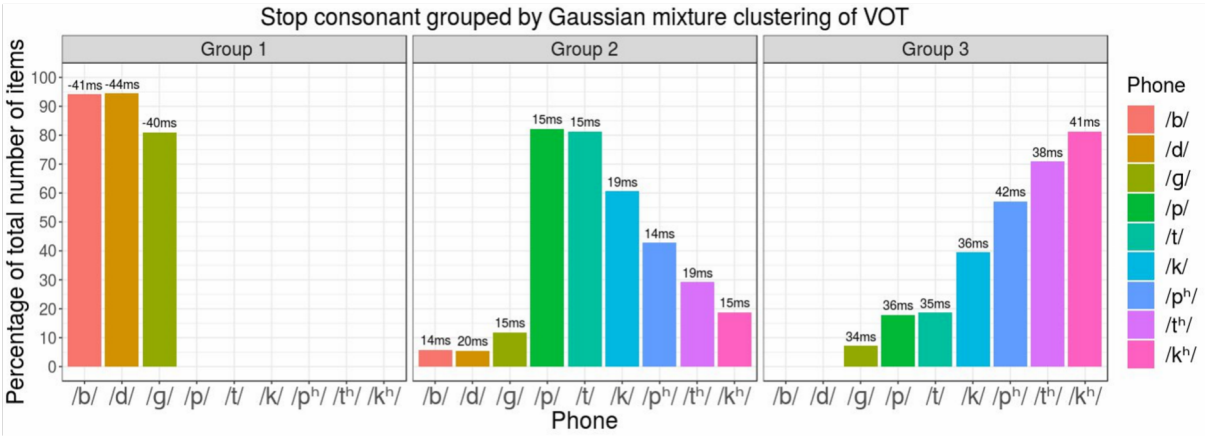


Figure 4 Three-way stop opposition in Mixean Basque

Innovative methods in acoustic phonetics

(Carignan & Egurtzegi 2021)

While working on the phonetics of the /h/ vs. /h̃/ opposition in Mixean Basque, Dr. Carignan and I faced the difficulty of measuring acoustic nasalization in a reliable way. A number of acoustic studies rely on individual measures, such as A1-P0 (cf. Scarborough et al. 2015; Zellou & Tamminga 2014, etc.). Nevertheless, such values are not easy to measure, and their interpretation is not always conclusive. Work by Styler (i.a. 2017) has shown that none of the individual acoustical features of nasality is a reliable indicator in all cases he analyzed. In light of this shortcoming, Carignan & Egurtzegi (2021) present a method based on applying a Principal Component Analysis (PCA) to 20 acoustical features that have been used as indicators of phonetic nasality in the literature (see Styler 2017). The PCAs are done separately for each speaker, retaining the principal components that cumulatively explain 80% of the total variance. We call this method *Principal Components Variable Importance Reconstruction* (PC-VIR). Figure 5 shows a plot with the importance of the different acoustic features after applying the PC-VIR method. It shows that not all indicators of nasality are equally good as predictors, some of them contributing to the analysis more than the rest. Variables in the left side (green field) can be considered strongly significant predictors and variables in the middle (yellow field) are considered moderately significant predictors, while variables in the right side (red field) are considered non-significant predictors. The inter-speaker variation is plotted as a boxplot, the horizontal black bar and the grey circle in the boxes showing the median and mean values, respectively. Note that the two significance thresholds are marked by the discontinuous horizontal lines (moderate importance with a dotted line and strong importance with a dashed line), with values

above zero representing positive correlation while these below zero represent negative correlation.

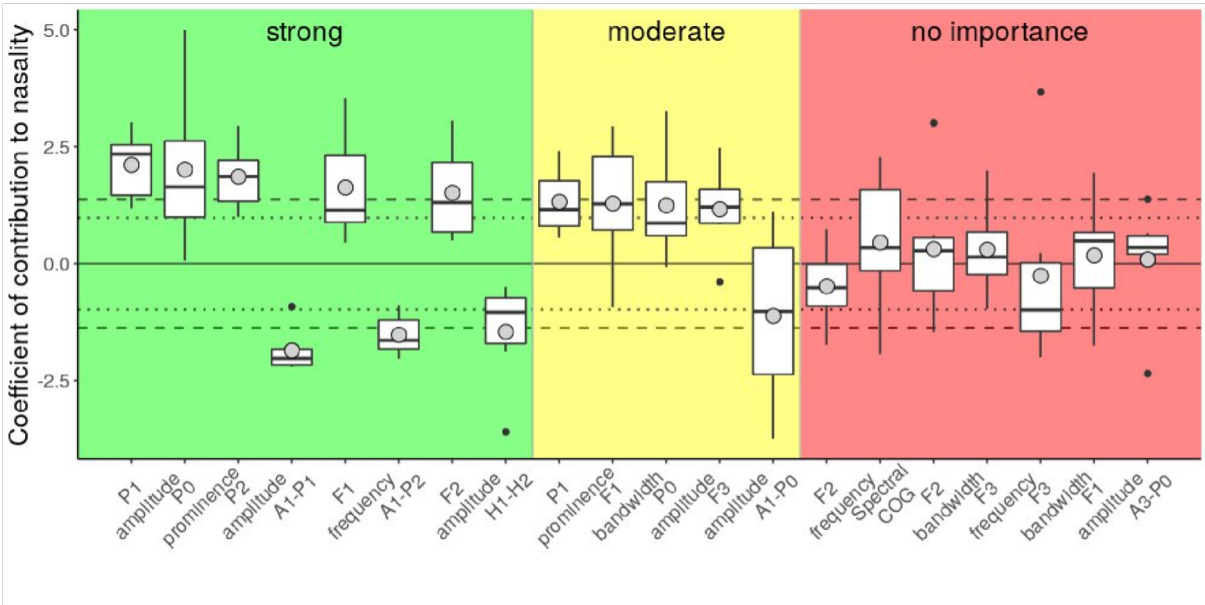


Figure 5 PC-VIR coefficients that contribute to discrimination of vocalic nasality. Coefficients are displayed from left to right in descending order of the mean value.

This new methodology has enabled us to reliably distinguish phonological contrasts solely based on nasality (such as distinguishing /h̃/ from /h/) in cases in which audio recordings are accessible, but obtaining physiological data proves difficult. The development of the PC-VIR method was a timely solution, given the difficulties of obtaining articulatory information of nasalized sounds in endangered languages or varieties with an aged speaker population, as in the case of Mixean Basque discussed above.

Phonetics of sound change in progress

(Egurtzegi et al. in prep.-1)

In addition to my work on Basque phonetics, I am also currently working on an instance of sound change in progress in Andalusian Spanish alongside Dr. Johanna Cronenberg from the Université Paris Cité and Prof. Jonathan Harrington from the IPS-LMU Munich. We are studying the shift from pre-aspiration to post-aspiration, an instance of gradual local metathesis in progress in this variety (Egurtzegi forth.). This shift can be observed in sequences that etymologically involved sibilant-stop heterosyllabic clusters, as in words such as Sp. *pasta* ‘pasta, paste’ ([pa^{h}ta] ~ [pat^{h}a]), *esto*

'this' ([e^{h}to] ~ [et^{h}o]), *kiosko* 'kiosk' ([kjo^{h}ko] ~ [kjok^{h}o]) or *espeso* 'dense' ([e^{h}peso] ~ [ep^{h}eso]). We are currently analyzing speech data from 80 participants, and comparing the realizations of younger and older speakers from the regions of Sevilla and Granada. In previous studies on this sound change, both age and geographical origin have been considered significant predictors in the distribution of pre-aspirated vs. post-aspirated realizations of this sequence (cf. Cronenberg et al. 2020; Ruch & Harrington 2014; Ruch & Peters 2016).

For almost all experimental investigations of this sound change, the focus has been on isolated words (Parrell 2007; Ruch & Harrington 2014; Cronenberg et al. 2020). The new angle in the present research project is to analyze whether there is evidence of post-aspiration across word-boundaries (*las tapas* 'the tapas' → [lathapah]). The purpose of doing so is to assess the extent to which this sound change in progress has been phonemicized/lexicalized. According to proponents of a modular feedforward architecture (Bermúdez-Otero 2015; Bermúdez-Otero & Trousdale 2012; Fruehwald 2017; Ramsammy 2015) and the closely related lexical phonology framework (Kiparsky 1982, 2015), sound change proceeds strictly bottom-up from least to more abstract domains. In these models, sound change in its initial stages is post-lexical and during its life-cycle subsequently applies in increasingly narrower domains as it is lexicalized. We sought to determine whether Andalusian Spanish stop post-aspiration occurred word-initially i.e. following a preceding underlying /s/: if so, this would point to Andalusian Spanish post-aspiration being post-lexical, and thus not yet phonemicized. This research was presented in the last instance of *LabPhon*, and we are currently working on its manuscript.

4.4 Concluding remarks

In short, my research has updated the historical phonology of the Basque language by applying a modern, interdisciplinary approach to historical linguistics that involves the integration of phonetic and typological research as well as contact linguistics in historical phonology. In addition, my research agenda contributes to the fields of theoretical phonology and linguistic typology in a meaningful way, as well as contributing to theories of phonetically based sound change.

5. Activities as principal investigator

I received the funding for my PhD thesis through a competitive call of the Basque Government for which I had to design my own thesis project and defended it in front of a committee, which gave me my first experience in planning and conducting my own research project.

After that first experience, I had two post-doctoral fellowships that also involved competitive individual funding in which I was the PI of my (individual) project: a Juan de la Cierva Fellowship (the main government-funded post-doctoral fellowship in Spain) to work at the Complutense University of Madrid and a Humboldt Fellowship (awarded by the prestigious German Alexander von Humboldt Foundation) to be part of the Institute of Phonetics and Speech Processing of the University of Munich (LMU). It was during this second fellowship that my career as a researcher really took a leap forward and I was able to produce my first important publications, that furthered on topics that I started working on for my dissertation (e.g. Egurtzegi 2017, 2018; Blevins & Egurtzegi 2017).

I was also the recipient of 2 individual grants by the Humboldt Foundation for short research projects that I conducted at the LMU Munich (2021 & 2023; 3000€ each) with Prof. J. Harrington (*Andalusian /s/ aspiration at the word boundary* & *Andalusian post-aspiration and the life cycle of sound change*, respectively). In both of these short projects, we advanced work on Andalusian pre-to-post aspiration, which was presented in the last *LabPhon* conference in 2022. We are currently working on a paper based on that research, as well as designing on a future project on the topic.

I am currently the PI of the ANR project *Modern approaches to diachronic phonology applied to Basque* (ANR-20-CE27-0007; 2021-24, national, France; 272 922€), which is still underway until the end of the year 2024. In addition, my project was granted additional grant support for its dissemination and promotion (ANR SAPS-CSTI 2020; 2024-26), so that its results will be disseminated through different initiatives by the services to be contracted to that end and the responsible dissemination team at the local university (UPPA Pau). Results of this project include most of my publications of the last 3 years (Egurtzegi 2022, 2023, forth.; Egurtzegi & Elordieta 2023; Egurtzegi et al. 2023, accep.; Blevins & Egurtzegi 2023; Elordieta & Egurtzegi 2022; Krajewska et al. 2022). For details on the project, see egurtzegi.github.io/projects/madpab/.

6. Research management

I consider myself a historical laboratory phonologist who specializes in sound change, typologically rare sound patterns, and the reconstruction of Proto-Basque. I pursue a very interdisciplinary approach to the study of sound change, which I place at the intersection between historical linguistics, theoretical phonology, laboratory phonology / experimental phonetics, phonological typology, philology, sociolinguistics, and contact linguistics. In addition, my work has become more technically oriented in the last years.

I am a chargé de recherche (permanent researcher, French State civil servant) at the French National Centre for Scientific Research (CNRS) since 2019. Before that, I was a post-doc at the University of Munich with a competitive scholarship (Alexander von Humboldt Foundation) and then as part of an European Research Council (ERC) project to Prof. Jonathan Harrington. I also held a Juan de la Cierva post-doctoral grant at the Complutense University of Madrid and a short post-doc at the University of the Basque Country (UPV/EHU). I finished my PhD thesis in Linguistics, *Towards a phonetically grounded diachronic phonology of Basque*, in 2014, under the supervision of Prof. Joseba Lakarra (UPV/EHU) and Prof. Juliette Blevins (City University of New York). I hold BAs in *German Philology* and *Basque Philology*, and MAs in *Linguistics* and *Phonetics & Phonology*.

The background that I started developing under the supervision of the Basque historical linguist Joseba Lakarra (UPV/EHU) and the theoretical phonologist and typologist Juliette Blevins (CUNY) during my dissertation in Vitoria-Gasteiz (University of the Basque Country) was later extended to laboratory phonology and more experimental approaches under the supervision of Jonathan Harrington (IPS-LMU) in the three years I spent as a post-doc in Munich. This unique skillset has helped me achieve an international presence in different disciplines (mostly in historical linguistics but, more recently, also in laboratory phonology).

I have published research articles in some of the most prestigious journals in the fields that are relevant to my work. These include *Diachronica* for historical linguistics (two papers), *Phonology* and *Phonological Data & Analysis* for phonology, *Journal of the Acoustical Society of America* and *Journal of the International Phonetic Association* for phonetics, *Linguistic Typology* for typology and *ASJU* and *FLV* for Basque linguistics,

as well as major general linguistics journals such as *Language* or *Linguistics*. I have also coauthored a big methodology-focused paper involving many laboratories in *Advances in Methods and Practices in Psychological Science*. I am currently working on a reference book *Syllables* coauthored with Prof. Juliette Blevins (under contract with Oxford University Press), and a just-submitted reference chapter *Metathesis* to be published in *The Wiley Blackwell Companion to Diachronic Linguistics*.

I have offered 35 invited talks at renowned universities and international conferences and workshops (including a co-authored plenary at the *International Conference on Historical Linguistics* 21 and an invited talk at "Phonetics and phonology – Beyond Romance", part of the 53rd *Linguistic Symposium on Romance Languages*) and 25+ peer-reviewed talks, including some in the main conferences of the relevant fields such as *Laboratory Phonology* (2 times), and *International Conference on Historical Linguistics* (5 talks). I have also organized 3 workshops (*Understanding Sound Change*, *Let's talk about stats!*), co-organized 2 more, and co-organized an international conference (*Old World Conference on Phonology* 19). I regularly collaborate with multiple internationally recognized linguists.

I currently supervise or have supervised the work of Bachelor students, Master students, research assistants, PhD students (3), visiting PhD students (2), and post-doctoral researchers (2). These had diverse backgrounds (Basque Studies, Indo-European Studies, Romance Studies, etc.) and came from different fields within (historical) linguistics (historical linguistics, phonology, morphology, syntax…). One of my students, Andrea García-Covelo, has been under my supervision during her BA, MA, and now PhD, which is co-supervised by Prof. Marianne Pouplier at the LMU Munich. Important research conducted with PhD students or post-docs under my supervision includes Krajewska et al. (2022); Egurtzegi et al. (2023), and Egurtzegi et al. (accep.), as well as many papers currently under way.

7. Research projects

7.1- *Modern approaches to diachronic phonology applied to Basque*

ANR Project MADPAB. ANR-AAPG2020; grant number: ANR-20-CE27-0007

7.1.1 Positioning and context

Historical linguistics is arguably the oldest and most successful branch of the science of language, showing considerable success in reconstructing proto-languages spoken over 5,000 years ago. During the last 40 years, a nuanced theory of sound change, based on articulatory, aerodynamic, and acoustic properties of speech, has evolved from interdisciplinary research (Ohala 2003; Blevins 2004). My short-term research plans aim to build on my previous work, which focused on improving our current understanding of Basque historical phonology by incorporating recent advances in theories of phonetically based sound change, contact linguistics and phonological typology and by using state-of-the-art quantitative and experimental techniques. More generally, I aim to contribute to the phonetically-informed theories of sound change, to typologies of lesser-attested sound changes, and to our understanding of sound change under linguistic contact. My current research project also aims to make an important methodological contribution to the historical study of the phonologies of language isolates and lesser attested languages. The new approach is to combine the historical-comparative method with phonetic detail and quantitative typological data for studying the historical sound patterns of a language without a family (i.e. isolate) as Basque, in which the comparative method finds most difficulties (Campbell 2017).

The classic approach to historical phonology has brought important results to the history of the Basque language in the past, but the advancements in the reconstruction of the phonological systems of older stages of the language have gradually slowed down since the 1970s until very recent times. Most work on Basque historical phonology (Michelena 1977) predates our current understanding of sound change and the use of phonetic methods in historical phonology. My dissertation (Egurtzegi 2014) was the first general study to apply these methods to a wide range of historical sound patterns of Basque, and it paved the way for new research questions (Egurtzegi 2017, 2018, 2019). The new angle was to ground Basque historical phonology far more in phonological typology (Kümmel 2015), synchronic phonetic variation (Labov 1994) and experimental analysis (Yu 2015) than was hitherto possible. Ohala (2003, i.a.) extensively argued for the integration of modern scientific phonetics in the comparative method. The integration

of phonetics (alongside language contact, acquisition or typology) gave rise to a new approach to historical phonology that is reflected, among others, in the theoretical framework Evolutionary Phonology (Blevins 2004, 2015).

In addition, recent research has shown the viability of the "experimental historical phonology" research program. For example, Carignan (2018) exposed naïve listeners of Australian English to stimuli created by speakers of Southern French, effectively finding evidence for sound changes involving nasalized vowels. A phonetically informed approach can also help us solve some of the longest-standing problems in Basque historical linguistics.

7.1.2 Objectives

The objectives of my project include achieving a better understanding of sound patterns that have been subject of discussion for years in Basque historical phonology, and advancing our knowledge of general typologies of sound change with a thorough analysis of the uncommon changes that will be studied. The areas of interest will at least cover the following topics: 1-the evolution of the opposition between /h/ and /h̃/ (only documented in two languages), 2-changes in the phonetic cues underlying the opposition between the stop series (from [spread glottis] to voicing), 3-changes in the place of articulation of sibilants, 4- vowel shifts, and 5-accentual systems (from phrasal pitch accent to word-level systems, based on pitch or stress depending on the dialect). The topics are very different in nature, as the aim is to reach a new global picture of the historical phonology of Basque as a main result of the project. In addition, most of the sound-patterns to be studied are typologically uncommon and theoretically challenging, and all have been subject to much discussion (but no real consensus) within Basque historical phonology. The expected success of the study of each proposed research topic is due to my previous research on connected topics as well as my overall expertise on the phonology of Basque (cf. Egurtzegi 2013, 2015, 2017, 2018, 2019; Egurtzegi & Elordieta 2013; Egurtzegi & Carignan 2020).

7.1.3 Methodology

For each topic under study, the employed methodology will combine philological work comparing the historical dialectal variants attested in older stages of the language and analysis of acoustic and articulatory data of the modern varieties. First, given that in

most cases the precise acoustic and/or articulatory realizations of a given segment have not been described, varieties of interest will be found in the literature in order to obtain acoustic and physiological data in the field encompassing all relevant phonological contexts. After that, experiments will be designed to elicit the data including words containing the kind of phonetic information required in each particular case (as well as distractors, filler words, etc.). Once the experiments are designed, recordings will be performed by eliciting the responses of interest with help of the computer program *Speech Recorder* (Draxler & Jänsch 2019). Then, the audios will be automatically segmented with *MAUS* (Kisler et al. 2017) and hand-corrected as required. After this, precise phonetic information will be extracted from the recordings by using *Praat* (Boersma & Weenink 2019) and the *R* packages *emuR* (Winkelmann 2017) and *wrassp* (Winkelmann et al. 2019). In each case, the phonetic metrics of interest will be measured and analyzed, and appropriate statistic models will be applied with regard to the specific research questions of each study, again by using the software *R* (R Core Team 2019), thanks to different packages widely used in comparable research programs. These will likely include *brms* (Bürkner 2017) and *lme4* (Bates et al. 2015), the packages most frequently used for Bayesian and frequentist linear mixed models respectively. Post-hoc tests, when appropriate, will likely be performed with packages such as *emmeans* (Lenth 2023).

All databases generated from both recordings as well as physiological measurements will be gender-balanced (i.e. a similar number of male and female subjects will be maintained through all tests). Differences regarding age-group will be explored whenever there is evidence for a sound change in progress (following the methodology of apparent-time studies, see Bailey et al. 1991). Whenever it is deemed of interest, laboratory experiments will be devised to find coarticulatory patterns that have been proposed to yield particular sound changes in the past, including analyzing reconstructed sequences that are not found in the modern language (e.g. cluster coarticulation for the sibilant study) as pronounced by speakers of modern Basque, in order to ascertain the likelihood of reconstructed changes given our current understanding of the mechanisms behind them. While in some cases the recordings for these experiments can also be made beforehand, it is likely that in other cases these recordings will follow the initial, exploratory study in a second phase within a given research topic. The project methodology will also include other data-based approaches, such as computer-based segment co-occurrence searches (by using regular expressions) of big historical corpora (such as the list of Aquitanian names from the 1st-3rd centuries, cf. Gorrochategui 1984 or

project Hesperia, de Hoz et al. 2019) to find potential biases in the phonotactic distribution of the segments under study and searches for typological parallels of each proposed phonological process (Kümmel 2007).

These will be especially important in cases in which the modern phonetic result of the hypothesized sound pattern is no longer useful for reconstructive inference, as in the case of complete loss of the segment of interest in a given phonological context (as in the case of wordinitial stops, cf. Michelena 1977). The role of contact will be assessed by looking for convergent evolutions in neighboring languages. Audio data will be recorded both in the field (in France and Spain) and in phonetic laboratories (as the UPV/EHU phonetics lab in Vitoria-Gasteiz, Spain). Participant data from both recordings and experiments will be anonymized and stored on password-protected mass storage devices. The materials collected for this project will be archived in ANPERSANA, a server hosted by IKER. Data will be made freely available for scientific purposes upon completion of the project, resulting publications will be free to access, and resulting papers will be presented at international conferences.

7.1.4 Interest and impact

The approach of this research project is both innovative in theory and method. Phonetically based accounts of sound change have not been applied to the history of Basque or, more generally, to the reconstruction of the phonology of any other linguistic isolate. This project will set a precedent in the methodology to be applied to the reconstruction of the phonologies of languages in which the comparative method finds most difficulties. It will also generate knowledge for theories of phonetically-based sound change, phonological typology, contact linguistics and historical linguistics. The general theoretical interest of the project is a consequence of the unique history of the Basque language, its long-standing contact with different Romance varieties, the interest of the typologically uncommon sound patterns under study and the applicability of the theoretical framework to modern approaches to historical phonology.

The long-term significance of the project lies in contributing to the reconstruction of the phonology of Proto-Basque as well as to both typologies of phonetically based sound change and sound change in situations of linguistic contact. An indirect and major benefit of the project is that it will provide a new methodology for the study of the historical phonologies of language isolates.

7.1.5 Example study 1: Using acoustic data for the reconstruction of historical sound change: Nasalance-based studies of Zuberoan Basque

Eastern varieties of Basque have been described with an extremely rare opposition between an oral /h/ and a nasalized aspirate /h̃/ (Hualde 1993; Egurtzegi 2018), which some researchers considered impossible (see Walker & Pullum 1999). We have already collected some data and we are currently working on the first nasalance-based study of the /h/ vs. /h/ contrast in Basque (and likely any other language), with data from the severely endangered Zuberoan variety. We have conducted a first study (with only 5 speakers analyzed) that has been reported in Egurtzegi et al. (2023).

For this study, we have recorded speakers of the village of Larraine, in Zuberoa with a nasalance device (Glottal Ent.). Elicitation was performed by prompting isolated Basque words (alongside their French translation, for potential disambiguation) in a computer screen though *SpeechRecorder*. The speakers only read the Basque words. Acoustic processing was performed in *Praat* — band-pass filtering the signal (80 Hz-10000 Hz) and computing two metrics that described the degree of nasalization—, and data wrangling and statistical analysis were performed in *R*. The median of the z-transformed measurements of each relevant token was calculated and submitted to two Bayesian Gaussian models, one for each metric we used (nasalance and mean-corrected nasal amplitude).[2]

Our preliminary results suggest that the /h/ vs. /h̃/ opposition is still present in Larraine Basque. Figure 6 plots the posterior distributions of the three categories under study: etymologically nasalized —/h̃/s that historically evolved from intervocalic *n—, assimilated —[h̃] that are nasal because they have a nasal obstruent in the vicinity, but etymologically were oral /h/s, see Egurtzegi (2018) on this assimilation—, and oral /h/s with no source of nasalization anywhere in the word in which they were produced. While the posterior distributions of the two nasal categories are comparable, they show little overlap with the distribution of the oral category.

[2] See above and Egurtzegi et al. (2023) for the packages used and further details. Future experiments within this line of research will likely follow a similar methodology.

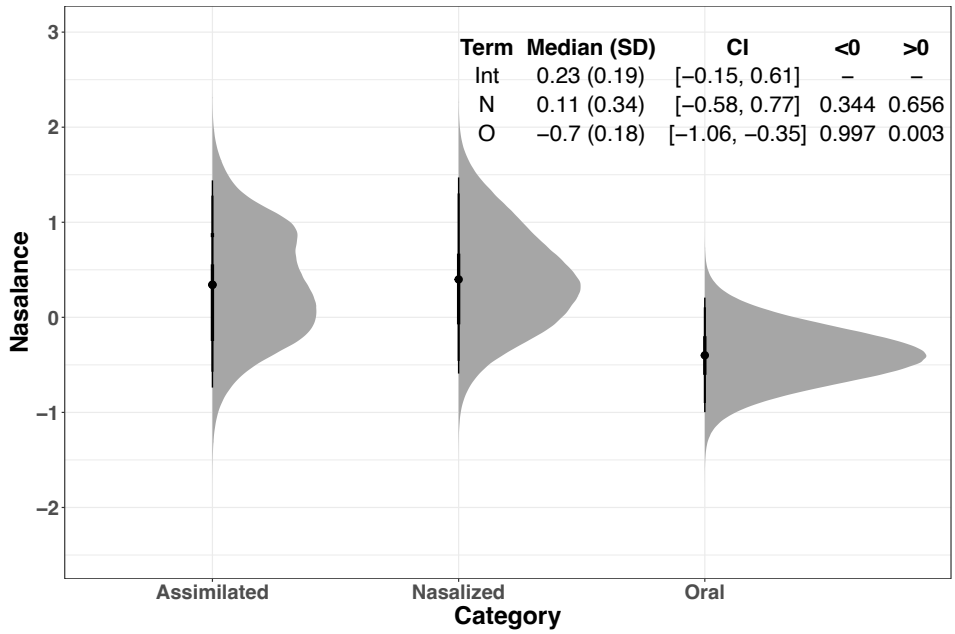


Figure 6 – Posterior distribution of the 3 aspirate categories in Egurtzegi et al. (2023).

However, we also observed that some lexical items sporadically lost nasality (see *ohol* 'wooden board', highlighted in Figure 7, which patterns with the etymologically oral words from the top instead of patterning with its nasal category), and some speakers have completely merged both segments and consistently produce /h/ in all context (as speaker 1 in Figure 8, which has oral values or wide distributions in all categories).

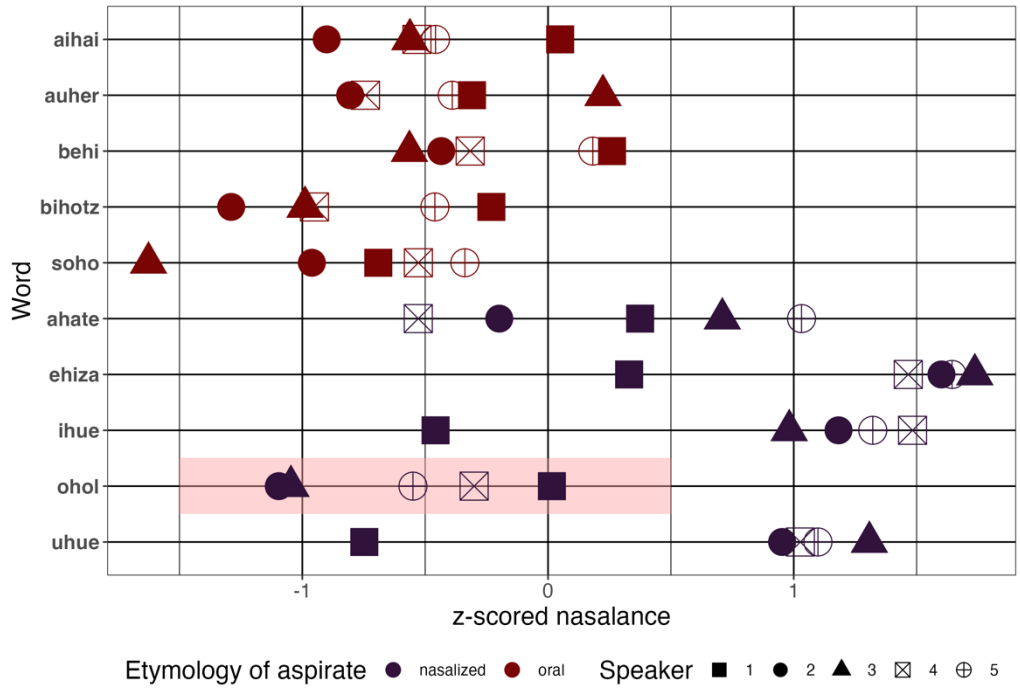


Figure 7 – Mean speaker nasalance in example words including an aspirate of oral vs. nasal origin.

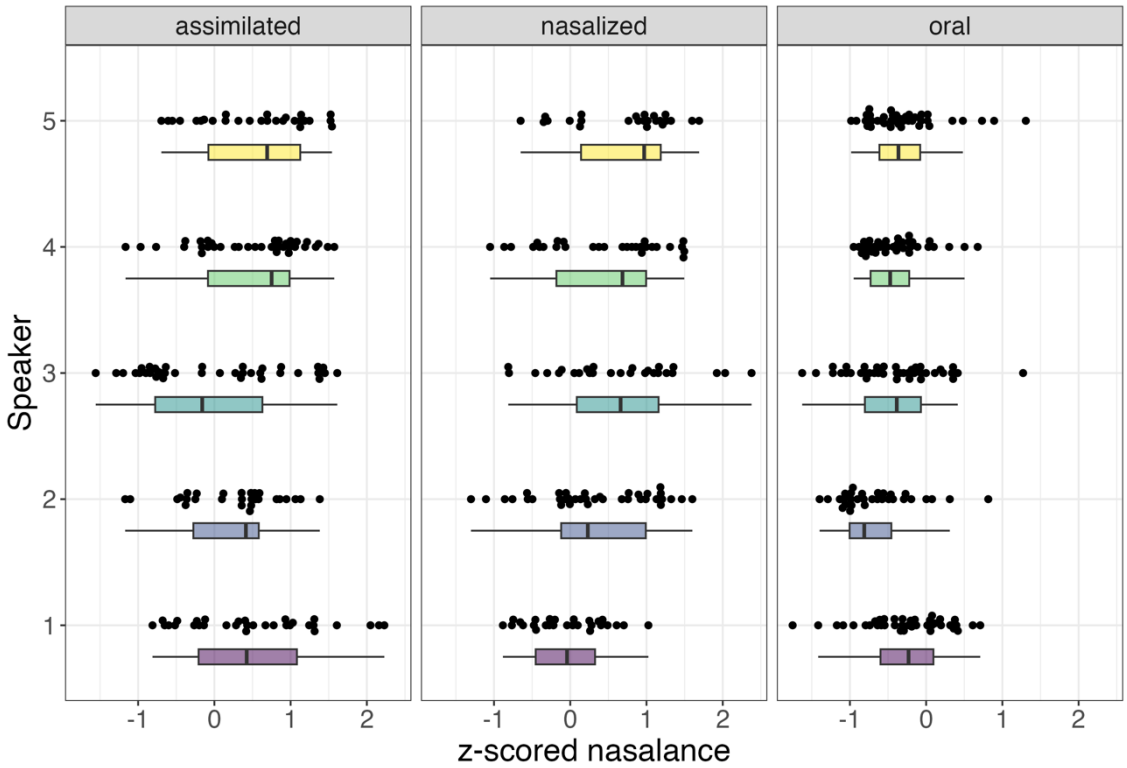


Figure 8 – Z-scored nasalance by speaker, faceted by aspirate category.

For our planned longer study, we will analyze more speakers, and we will also provide a dynamic analysis to test whether the vowels surrounding a nasal /h̃/ need to be consistently nasalized to maintain this opposition (as suggested in Blevins & Egurtzegi 2023). For the dynamic analysis we plan to use Functional Principal Component Analysis (FPCA, see Gubian et al. 2015; Cronenberg et al. 2020), a useful method for the analysis of time-varying data that I am currently using to analyze Andalusian post-aspiration data (Egurtzegi et al. in prep.-1).

7.1.6 Example study 2. Using corpus phonology for the reconstruction of historical sound change. Data-driven philological studies of sibilant oppositions in 18th century Gipuzkoan Basque

The Basque sibilant system is relatively complex: conservative varieties distinguish voiceless fricative-affricate pairs with three places of articulation —lamino-alveolar, apico-alveolar and post-alveolar (Michelena 1977; Hualde 2003). This system has undergone several changes at different stages of the language. The oldest attested development is a change initiated in the Western varieties of Basque in the 17th century

(i.e., the "Western merger"). It involves the neutralization of the laminal and apical alveolar fricative sibilants in favor of the latter, and the neutralization of the laminal and apical alveolar affricates in favor of the former (Figure 9, left). This merger then spread to some central varieties as well. In those central varieties, yet another neutralization pattern (the "Central merger") emerged between the 17th and 19th centuries: in this case, both fricative and affricate alveolar sibilants are realized as laminals (Figure 9, right). Previous philological work observed that, in the 18th and 19th centuries, the Central merger is always documented alongside the Western neutralization, i.e. in the fricative series, both patterns co-occur in the same texts (the Western pattern being more frequent). From the 19th century onwards, the Central merger became the prevailing pattern in the Guipuzkoan coastal areas and in the Urola valley (see e.g. Camino 2000; Beristain 2022).

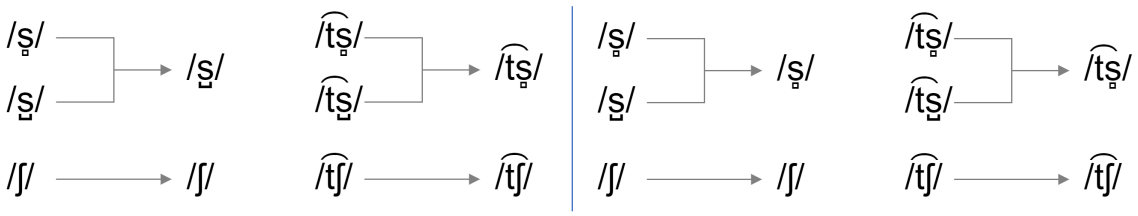


Figure 9 – The Western (left) and Central (right) merger.

Research on the history of sibilant mergers in Basque has generally been qualitative, mainly because Basque attestations preceding the 19th-century do not always allow quantitative analyses due its limited size. In this research, we study a relatively extensive manuscript, "Lubieta's dictionary", dated 1728. This Gipuzkoan text shows evidence of both patterns of merger described above.

The objective of this research is to ascertain the degree of development of the Western and Central mergers in the variety of San Sebastián in the middle of the 18th century, whether a potential relation between the two mergers may have existed, and the developments that led to the generalization of the Central merger in that area. To this end, we created a database with all the words including a sibilant in Lubieta's dictionary after the careful philological work of transcription of pictures of the original hand-written document. With the database ready, we fitted a generalized linear mixed-effects model to the data extracted from the text, and analyzed the following variables: the phonetic context in which the sibilant appears, the frequency of the lexeme and its type (lexical vs. grammatical).

The results of this first study suggested that both sibilant mergers were still in progress in the relevant variety and period, and that they were interrelated processes. We suggest that the Western merger started as a phonetically-conditioned sound change due to coarticulation to a following (alveolar) stop. Figure 10 shows the higher degree of confusion towards the apical before stops, the highest being preceding /t/. As this neutralization extended to other contexts, a hypercorrective change was initiated in some Central varieties, which eventually resulted in a mirror-image process, namely a change from apical to laminal alveolar sibilant fricatives that gave rise to a second pattern of merger. This mirror-image change is less frequent in the contexts where we expect most CC coarticulation, and strongest elsewhere (e.g. intervocalically or word-initially), as depicted in Figure 11.

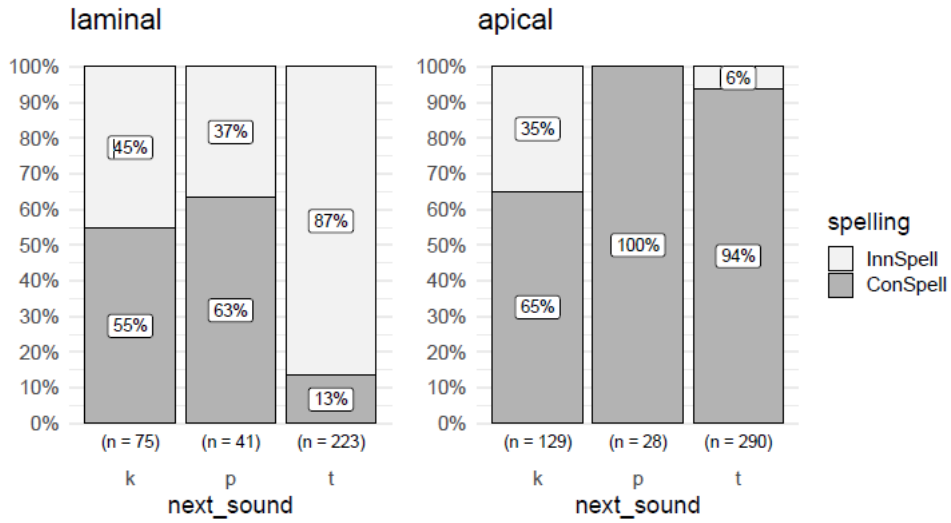


Figure 10 – Proportion of innovative vs. conservative spelling suggests apicalization of laminals before /t/.

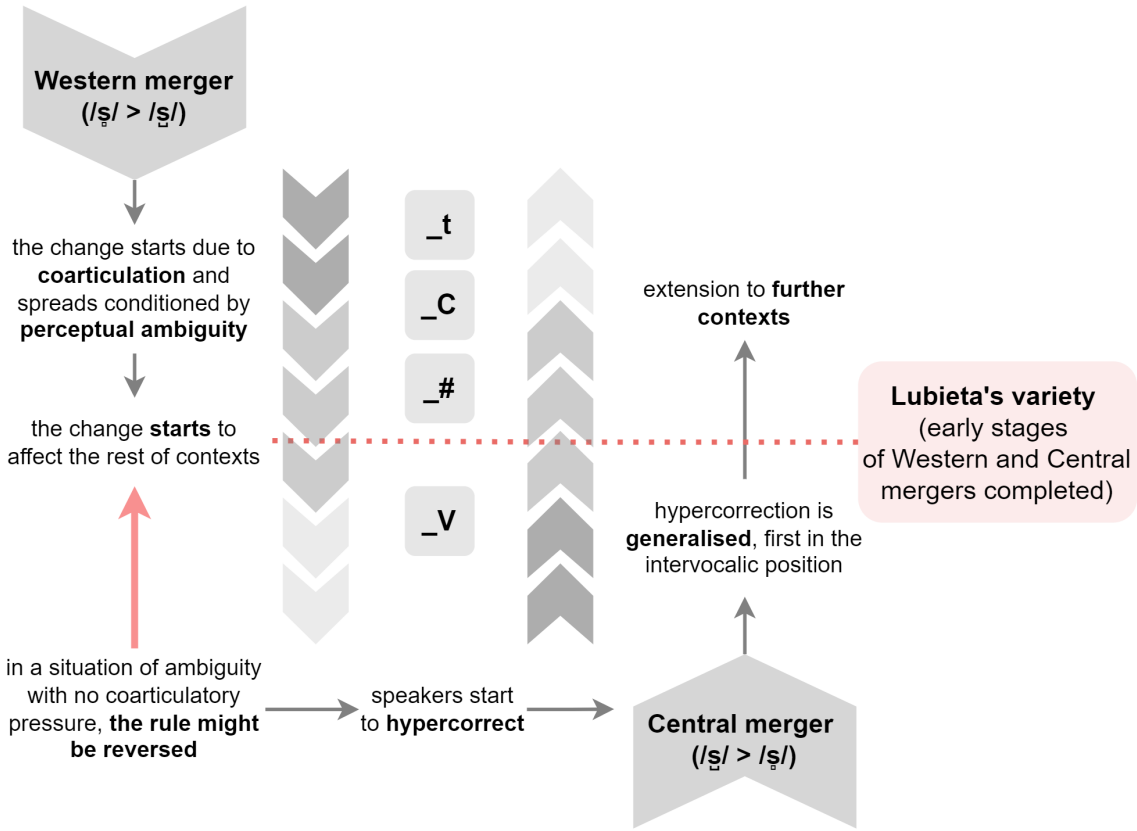


Figure 11 – The proposed evolution of the Western (left) and Central (right) merger in Gipuzkoan Basque.

This first study, which was published in *Phonological Data & Analysis* (Krajewska et al. 2022), showed that we can learn a lot about the direction and gradual extension of a (neutralizing) sound change by looking at the deviations from previous stages of the language in a geographically and historically precisely-located hand-written manuscript —i.e., where no changes have been implemented to the text by any third person with a different variety of the language. Following the same methodology, we plan to analyze other 3-4 texts from the same region but slightly later chronology, in order to better understand the evolution of this sibilant merger and corroborate or correct our initial predictions based on a single written document. Being able to perform a comparable analysis in texts that cover a wide timespan would give us a clear historical picture of this sound change, although we are inevitably constrained by the current historical textual documentation, which, although it has increased in the last decade with the discovery of "new" old texts, is not nearly as extensive as for other European languages.

7.2 Syllables

(Juliette Blevins & Ander Egurtzegi; book project for Oxford University Press)

Though an enormous amount has been written on the syllable and syllabification, books on the syllable are framed within a particular theory (e.g. Féry, C. & R. van de Vijver. 2003. *The syllable in Optimality Theory*. Cambridge University Press.), or focus primarily on the issue of structure in one or two languages (e.g. Duanmu, S. 2008. *Syllable Structure: The limits of Variation*. Oxford University Press, which treats mainly English and Chinese dialects). The range of exemplification in the planned book will be broader than other treatments, and not embedded within a single theoretical framework. To our knowledge, there is no book-length treatment of the syllable and syllabification which attempts to provide a balanced presentation of approaches, while, at the same time, integrating data and analysis from typological studies, language descriptions, theoretical treatments, laboratory phonology, phonetics, psycholinguistic experiments, writing systems, speech errors, language acquisition, language games, computational modeling, sound change, loan word phonology, and sign language. This project aims to fill that gap.

This book will examine current and past approaches to the syllable as a constituent, syllable-internal structure, sonority, syllable weight, syllabification, and a range of syllable-related alternation types. It will offer a reasoned account of the theories and methods that lead to the most accurate account of the facts in these areas, present a balanced view of alternative accounts, and suggest degrees of contention and consent on the same topics. The discussion of the syllable will be set within the context of wider developments in the field, especially where aspects of general cognition and innate knowledge have been invoked. Data and analysis will be drawn from cross-linguistic typological studies, detailed language descriptions, theoretical phonology, acoustic, articulatory and aerodynamic phonetics, psycholinguistic experiments, writing systems, speech errors, language acquisition, language games, computational modeling, sound change, and comparisons between spoken and signed language syllables. Brief outlines of some of the likely topics follow.

i. Evidence for syllables

Syllables are widely recognized units of sound structure. In this chapter, evidence from a range of languages and phenomena are gathered to demonstrate that there are units larger than single segments, and smaller than metrical feet. Arguments for the syllable include: processes in which the syllable is the domain of a phonetic or phonological feature (stress, tone, pharyngealization, laryngealization, rhoticization, etc.); distributional aspects of sound patterns which refer to syllable edges (e.g. syllable-initial

aspiration; syllable-final glottalization); secret languages and language games where syllables are moved, replaced, targets of affixation, etc.; and native speaker intuitions regarding the existence of syllables, and the ability to count, for example, the number of syllables in a word.

ii. Phonetic vs. phonological syllables

The majority of literature on syllables is framed within phonological models where reference is to abstract features, constituency, and hierarchical structure. In some languages there are clear mismatches between phonological syllables of this kind, and surface pronunciations where phonological vowels are missing, or where consonants are vocalized (Pouplier & Beňuš 2011). This chapter explores the range of mismatches known to exist, and experimental evidence bearing on the saliency of each in the mind of the listener. Work on loanword phonology is introduced to illustrate how surface phonetic syllables in one language are rendered as multisyllabic sequences in another. We also bring the question of 'minimal utterance': while, in some languages, syllables constitute minimal utterances, in others the minimal utterance is a stressed syllable, or a disyllabic foot.

iii. Languages without syllables?

Case studies of a range of unrelated languages suggest no evidence of organization at the level of the syllable. Perhaps the best known of these is Gokana, from the work of Hyman (1983, 1985). This chapter explores the range of evidence (or lack of evidence) for syllables in Gokana and other languages claimed to lack syllables. Moraic approaches, like that proposed by Hyman for Gokana, are compared with syllable-based approaches, and assessed in terms of their predictions for distinct experimental tasks.

iv. Syllables and sonority

The melodic organization of segments into syllables is based, in many languages, on the inherent sonority of segments. While sonority is roughly defined as the loudness of a sound relative to other sounds produced with the same input energy, sonority scales are not widely agreed upon in their entirety, and sometimes even put into question altogether. In addition to summarizing the range of proposals in this area, evidence from acoustic studies (e.g. Jany et al. 2007; Chitoran 2016; Crouch et al. 2023) is shown to

support language-specific areas of variation. Evidence from experimental paradigms in speech perception are also evaluated (e.g. Berent et al. 2007, Berent et al. 2008).

v. Syllable-internal structure

This chapter presents a range of proposals concerning the internal hierarchical structure of syllables, including flat structures, moraic approaches, nucleus/coda approaches, empty nucleus approaches, demi-syllables, and more. Evidence for sub-syllabic constituency is evaluated, and the strongest cases for a range of constituents (onset, nucleus, rhyme, coda) is presented. Segmental constraints associated with these constituents are also explored, e.g. the preference for high sonority codas. Finally, languages claimed to have 'sesqui-syllables' and 'degenerate' syllables are reconsidered in light of the minimal defining properties that phonetic and phonological syllables must exhibit.

vi. Syllable weight

In many languages, a distinction must made between 'light' syllables and 'heavy' syllables. Light syllables are those which do not attract stress, or which have only a single tone-bearing unit, while heavy syllables attract stress, or can bear multiple tones. Work in this area has progressed steadily since the publication of Gordon (2006), where strong evidence is presented for a relationship between 'weight' and the phonetic feature born by the syllable in question. This chapter lays out the range of proposals and distinctions required in weight-sensitive systems, and highlights the relationships between phonetic feature (stress, pitch-accent, tone) and the kind of segmental string bearing the phonetic feature. Finally, the role of onsets in syllable weight is also considered, with special attention to Gordon (2005) on onset-sensitive stress patterns.

vii. Syllabification and related processes

How are strings of vowels and consonants parsed into syllables? Are there cross-linguistic generalizations regarding the segmentation of speech into smaller syllable-sized units? Are these parsings consistent across speakers of a language, or is there variation? These questions will be tackled in this chapter through experimental data.

viii. The evolution of syllable structure

Most research on syllable structure and syllabification focuses on the syllable as a static synchronic unit of grammar. However, syllable structure, like all aspects of speech, changes over time (Blevins 2004). In this chapter, a range of evolutionary pathways are illustrated, showing the evolution of: closed syllables; open syllables; complex onsets; complex codas; complex nuclei; syllabic consonants; well-formed sonority sequences; sonority sequencing violations; consonant and vowel epentheses; and the origins of aberrant syllabifications. Contact-induced changes in syllable structure are also examined, as are syllable types that appear to be highly stable across time. By charting a range of sound changes and their implications for syllable structure, one is better able to assess whether certain recurrent aspects of syllable structure are secondary effects of independent phonetic processes, or primary features of language structure.

ix. Syllable typology and universals

Finally, findings of the previous chapters will be summarized in terms of their relationship to general syllable typology and universals (Blevins 2006b). An introductory section on phonological typology discusses several databases (e.g. *Phoible*, *WALS*) and what they may, or may not, tell us about syllable structure. Another introductory section on phonological universals reviews the differences between absolute universals and statistical universals (cf. Hyman 2008), and also the implications of formalizing a universal as a defeasible constraint within a grammar.

7.3 Looking for the origin of recurrent sound changes: Metathesis

(Egurtzegi forth., in prep.)

Metathesis has historically been one of the lesser understood processes in phonology. My last main projects for the near future focus on metathesis: a first leg, a general review chapter of metathesis, was recently submitted for publication in *The Wiley Blackwell Companion to Diachronic Linguistics* (Egurtzegi forth.). This research will soon be followed by a more specific account of reciprocal metathesis.

Reciprocal metathesis refers to the cases of metathesis in which two different non-adjacent segments exchange their positions with one another without directly affecting the rest of the sequence (at least in the resulting form, cf. Lat. *leriquiae* < Lat. *rēlĭquĭae* ‘relic’ or Fr. *moustique* vs. Sp. *mosquito* ‘mosquito’). For this project, I am compiling a

great number of cases of reciprocal metathesis from paper and grammars on languages from all over the world and I hypothesize a potential origin to account for its recurrent nature.

In order for reciprocal metathesis to occur, the two affected segments typically need to be in the same syllabic position —i.e., onset, nucleus or coda— and they need to share some crucial phonological features. In languages such as Basque, reciprocal metathesis affects not only consonants, as in ***bage*** > ***gabe*** 'without' or *era**kut**si* > *era**tsu**ki* 'show', but vowels as well, as shown by ***a**tera* > ***eta**ra* 'come out' or *alk**a**ndora* > *alk**o**nd**a**ra* 'shirt' (Egurtzegi 2014; Egurtzegi forth.). After compiling a typological survey of this process in already more than 50 languages (and counting), I hypothesize that reciprocal metathesis originates in motor planning errors, being comparable to the speech errors usually regarded as *spoonerism*. These speech errors have been proposed to be the consequence of the influence of some planning elements —such as gestures— in others through priming, coactivation, inhibition, etc. (Garrett & Johnson 2013).

The comparison between spoonerisms and reciprocal metathesis reveals a lexical bias: while both processes tend to target low-frequency words (i.e., the less familiar the word, the more likely for errors to be made), the resulting word tends to be a familiar word in the case of spoonerisms but a new word in the case of reciprocal metathesis. In addition, after calculating the relative frequency of the words affected by reciprocal metathesis in different corpora of Spanish, I have observed a second bias related to usage: the directionality of a given change (e.g. /gVlV/ becomes /lVgV/ vs. /lVgV/ becomes /gVlV/) in Spanish seems to be correlated with a higher relative frequency of the sequence of sounds (/lVgV/ vs. /gVlV/) in the result of the metathesis within the language. This is exemplified by Figure 12, which shows the relative frequencies in the CREA corpus of the sound pairs in both the input and output of instances of reciprocal metathesis attested in Spanish. These relative frequencies tend to be higher for the sequences of sounds found in metathesis outcomes than they are for the etymological order of the segments.

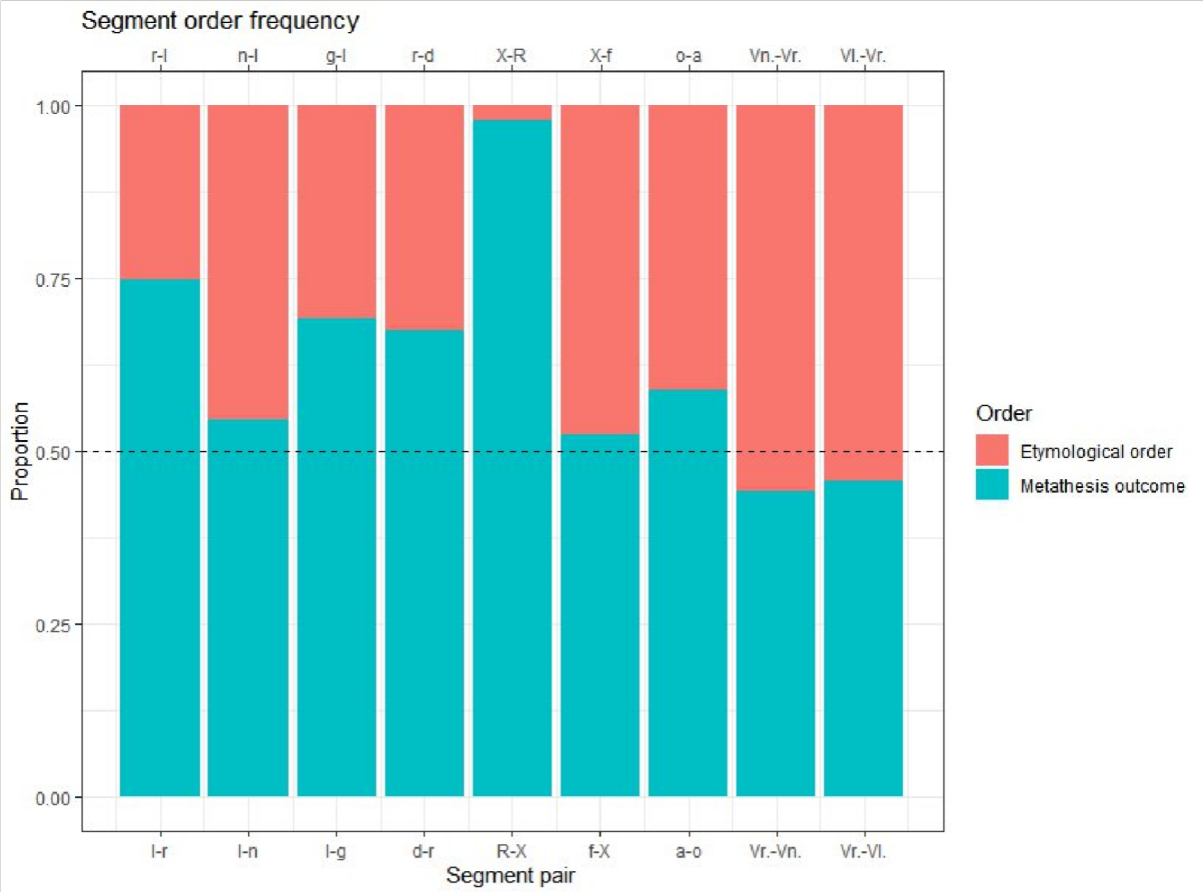


Figure 12 - Proportion of words including pairs of segments with the sequential order and phonological context found in words that were affected by metathesis (above) vs. the proportion of the inverse sequential order that resulted from metathesis (below)

In a second, related study on the same topic, I am currently compiling examples of spoonerism-like speech errors made by children acquiring Spanish as their first language, i.e., slips of the tongue that are superficially similar to reciprocal metatheses as produced by Spanish-learning children. Interestingly, the lexical bias is not as apparent in the survey of speech errors compiled from Spanish-learning children: while apparent cases of reciprocal metathesis are phonologically similar in children, the lexical bias favoring lesser attested words observed for phonologized reciprocal metatheses does not seem to hold in infant speech. I suggest that the similar tendencies found in spoonerisms and reciprocal metathesis reflect a common psychological origin of the two processes, while the few differences between them are related to the human capacity to retrieve lexical items from our memory: if an erroneously produced/heard word deviates too much from the target word, or if the erroneously produced/heard word already exists in the language, the error will be quickly corrected by the speaker and/or the hearer and will not be retrieved for future use. If the affected word is a low frequency item, and the resulting

erroneously produced word is not too different from the target and it is not part of the language, it might, in rare occasions, go uncorrected, be further reproduced and be ultimately integrated into the language variety as a new variant of that word.

References

Ariztimuño, B. & Egurtzegi, A. 2011. Enarak, elaiak eta testuinguruaren garrantzia [Enarak, elaiak [swallows] and the importance of the context]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 45, 403-410.

Bailey, G., Wikle, T., Tillery, J. & Sand, L. 1991. The apparent time construct. *Language Variation and Change* 3, 241-264.

Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67, 1-48.

Berent, I., D. Steriade, T. Lennertz, & V. Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104, 591-630.

Berent, I., T. Lennertz, J. Jun, M. Moreno & P. Smolensky. 2008. Language universals in human brains. *Proceedings of the National Academy of Sciences* 105, 5321-5325.

Beristain, A. 2022. Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque. *Journal of the International Phonetic Association* 52.3, 421-452.

Bermúdez-Otero, R. 2015. Amphichronic explanation and the life cycle of phonological processes. In P. Honeybone, & J. Salmons (eds.), *The Oxford Handbook of Historical Phonology*, Oxford: OUP.

Bermúdez-Otero, R. & G. Trousdale. 2012. Cycles and continua: on unidirectionality and gradualness in language change. In Nevalainen & Traugott (eds.), *The Oxford Handbook of the History of English*, Oxford: OUP.

Blevins, J. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: CUP.

Blevins, J. 2006a. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32, 117-165.

Blevins, J. 2006b. Syllable typology. Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edition, Volume 12. Oxford: Elsevier. 333-337.

Blevins, J. 2015. Evolutionary Phonology: A holistic approach to sound change typology. In P. Honeybone and J. Salmons (eds.), *Handbook of Historical Phonology*. Oxford: OUP, 485-500.

Blevins, J. 2018. *Advances in Proto-Basque Reconstruction with Evidence for the Proto-IndoEuropean-Euskarian Hypothesis*. New York: Routledge.

Blevins, J. & Egurtzegi, A. 2017. Unexpected obstruent loss in initial obstruent–sonorant clusters: an apparent example from Basque. *Phonology* 34, 507-522.

Blevins, J., Egurtzegi, A. & Ullrich, J. 2020. Final obstruent voicing in Lakota: Phonetic evidence and phonological implications. *Language* 96.2, 294-337.

Blevins, J. & Egurtzegi, A. 2023. Refining explanation in Evolutionary Phonology: Macro-typologies and targeted typologies in action. *Linguistic Typology* 27.2, 289-311.

Boersma, P. & Weenink, D. 2019. *Praat: doing phonetics by computer*. Computer software available from http://www.praat.org/.

Bürkner, P. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80, 1-28.

Camino, I. 2000. Abiaburu bat Donostiako euskera aztertzeko. In Koldo Zuazo (ed.), *Dialektologia gaiak*, 53-84. Vitoria-Gasteiz: Arabako Foru Aldundia.

Camino, I. 2016. *Amiküze eskualdeko heskuara* [*The Basque of the region of Amiküze* (*Mixe*)] (*Mendaur* 11). Bilbao: Euskaltzaindia.

Campbell, L. 2017. *Language isolates*. New York: Routledge.

Carignan, C. 2018. Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Laboratory Phonology* 9, 18.

Carignan, C. & Egurtzegi, A. 2021. *Principal components variable importance reconstruction (PC-VIR): Exploring predictive importance in multicollinear acoustic speech data*. *arXiv:2102.04740 [stat.ME]*. https://arxiv.org/pdf/2102.04740.pdf.
R code and data: https://github.com/ChristopherCarignan/PC-VIR.

Chitoran, I. 2016. Relating the sonority hierarchy to articulatory timing patterns: A cross-linguistic perspective. In M.J. Ball and N. Müller (eds.), *Challenging sonority: Cross-linguistic evidence* [*Studies in phonetics and phonology*]. Equinox Publishing, 45-62.

Cronenberg, J., M. Gubian, J. Harrington & H. Ruch. 2020. A dynamic model of the change from pre- to post-aspiration in Andalusian Spanish. *Journal of Phonetics* 83, 101016.

Crouch, C., Katsika, A. & Chitoran, I. 2023. Sonority sequencing and its relationship to articulatory timing in Georgian. *Journal of the International Phonetic Association*.

De Hoz, J., Gorrochategui, J., Luján, E. & Beltrán, F. 2019. *Hesperia. Banco de datos de lenguas paleohispánicas*. Online database: http://hesperia.ucm.es/.

Draxler, C. & Jänsch, K. 2019. *SpeechRecorder*. Computer software available from http://www.bas.uni-muenchen.de/Bas/software/.

Egurtzegi, A. 2011. Euskal metatesiak: abiaburua haien ikerketarako. *International Journal of Basque Linguistics and Philology* (*ASJU*) 45, 1-79.

Egurtzegi, A. 2013a. Phonetics and Phonology. In M. Martínez-Areta (ed.), *Basque and ProtoBasque* [*Mikroglottika* 5]. Frankfurt am Main: Peter Lang, 119-172.

Egurtzegi, A. 2013b. Diferentes tipos de aspiración en vasco. In E. Blasco, P. Francalacci, A. Nocentini & G. Tanda (eds.), *Iberia e Sardegna*. Firenze: Le Monnier, 151-169.

Egurtzegi, A. 2013c. Ekialdeko /o/-ren igoeraren testuinguru fonologikoaren bila [Looking for the phonological context of /o/-raising in Eastern Basque], *International Journal of Basque Linguistics and Philology* (*ASJU*) 48, 257-266.

Egurtzegi, A. 2014. *Towards a Phonetically Grounded Diachronic Phonology of Basque*. PhD dissertation, University of the Basque Country.

Egurtzegi, A. 2015. Different distributions of contrastive vowel nasalization in Basque. *Loquens* 2, e017.

Egurtzegi, A. 2017. Phonetically conditioned sound change: Contact induced /u/-fronting in Zuberoan Basque. *Diachronica* 34, 331-367.

Egurtzegi, A. 2018a. On the phonemic status of nasalized /h̃/ in Modern Zuberoan Basque. *Linguistics* 56, 1353-1367.

Egurtzegi, A. 2018b. Nasalized mid back vowel raising in Gascon and Basque. In D. Recasens & F. Sánchez-Miret (eds.), *Production and perception mechanisms of sound change*. Munich: Lincom, 173-187.

Egurtzegi, A. 2018c. Herskarien ustezko ahoskabetasun asimilazioa eta euskal herskari zaharren gauzatzea. *International Journal of Basque Linguistics and Philology* [*ASJU*] 53, 171-188.

Egurtzegi, A. 2019. Metathesis of aspiration as the source of anticipatory voicelessness in Basque. *Journal of French Language Studies* 29.2, 265-279.

Egurtzegi, A. 2022. Goizuetako azentueraren garapena kokatzen Euskara Batu Zaharraren prosodia zehazteko: Proposamenen ebaluazioa [Placing the development of the Goizueta prosodic system to define the prosody of Common Basque: Evaluation of the proposals]. *Fontes Linguae Vasconum* 133, 9-35.

Egurtzegi, A. 2023. /ɦ̃/ *hasperen sudurkariaren inguruan* [On the nasalized aspiration /ɦ̃/]. In press in the *International Journal of Basque Linguistics and Philology* (*ASJU*) 57.

Egurtzegi, A. Forthcoming. "Metathesis". In E. Aldridge, A. Breitbarth, K.É. Kiss, A. Ledgeway, J. Salmons & A. Simonenko (eds.), *The Wiley Blackwell Companion to Diachronic Linguistics*.

Egurtzegi, A. In preparation. *Reciprocal Metathesis*. Manuscript.

Egurtzegi, A. & Ariztimuño, B. 2013. Remarks on the etymology of the Basque word for 'swallow' and its potential relation to Celtic. *Ériu* 63, 79-90.

Egurtzegi, A. & Carignan, C. 2020. An acoustic description of Mixean Basque. *Journal of the Acoustical Society of America* 147.4, 2791-2802.

Egurtzegi, A. & Carignan, C. In preparation. *A typological rarity: The /h/-/ɦ̃/ contrast of Mixean Basque*. Manuscript. Paper presented at *LabPhon* 17.

Egurtzegi, A., Cronenberg, J. & Harrington, J. In preparation-1. *From post-lexicality to phonologization: Andalusian /s/-aspiration at the word boundary*. Manuscript. Paper presented at *LabPhon* 18.

Egurtzegi, A. & Egurtzegi, A. In preparation. *A study of palatal stops in Aulestia Basque*. Manuscript.

Egurtzegi, A. & Elordieta, G. 2013. Euskal azentueren historiaz. In R. Gómez, J. Gorrochategui, J.A. Lakarra & C. Mounole (eds.), *3rd Conference of the Luis Michelena Chair*. Bilbao: University of the Basque Country, 163-186.

Egurtzegi, A. & Elordieta, G. 2023. A history of the Basque prosodic systems. *Diachronica* 40.1, 30-72.

Egurtzegi, A., A. García-Covelo & I. Urrestarazu-Porta. 2023. A nasalance-based study of the /h/ vs. /ɦ̃/ opposition in Zuberoan Basque. In Radek Skarnitzl & Jan Volín

(eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague: Guarant International, 3427-3431.

Egurtzegi, A., A. García-Covelo & I. Urrestarazu-Porta. In preparation-2. */h/ vs. /h̃/ in Zuberoan Basque: A phonetic-typological account of an 'impossible' opposition*. Manuscript.

Egurtzegi, A., D. Krajewska, C. Carignan & I. Urrestarazu-Porta. Accepted. An acoustic exploration of sibilant contrasts and sibilant merger in Mixean Basque. *Journal of the International Phonetic Association*.

Elordieta, G. & Egurtzegi, A. 2022. Bizkaiko iparraldeko euskararen doinu-azentuaren berrikusketa [A revision of the pitch accent of Northern Bizkaian Basque]. In press in *Papers on Basque Linguistics in Honor of Jon Andoni Franco Elorza*.

Elordieta, G. & Egurtzegi, A. In preparation. *A revision of the tonal configuration of the lexical accent and phonological phrases in Northern Bizkaian Basque*. Manuscript.

Fruehwald, J. 2017. The role of phonology in phonetic change. *Annual Review of Linguistics* 3, 25-42.

Gardani, F. 2022. Contact and borrowing. In Adam Ledgeway & Martin Maiden (eds.), *The Cambridge handbook of Romance linguistics*. Cambridge: Cambridge University Press.

Garrett, A. & Johnson, K. 2013. Phonetic bias in sound change. In A.C.L. Yu (ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press, 51-97.

Gavel, H. 1920. *Éléments de phonétique basque*. Paris: Édouard Champion.

Gordon, M. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23, 595-653.

Gordon, M. 2006. *Syllable weight: Phonetics, phonology, typology*. Oxford: Routledge.

Gorrochategui, J. 1984. *Estudio sobre la onomástica indígena aquitana*. Bilbao: University of the Basque Country.

Gubian, M., Torreira, F. & Boves, L. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49, 16-40.

Hansson, G.Ó. 2004. Long-distance voicing agreement: an evolutionary perspective. *Proceedings of the Berkeley Linguistic Society* 30, 130-141.

Hualde, J.I. 1993. Topics in Souletin phonology. In J.I. Hualde & J. Ortiz de Urbina (eds.), *Generative studies in Basque linguistics*. Amsterdam: John Benjamins, 289-327.

Hualde, J.I. 2018. Aspiration in Basque. *Papers in Historical Phonology* 3, 1-27.

Hualde, J.I. 2003. Segmental phonology. In J.I. Hualde & J. Ortiz de Urbina (eds.), *A grammar of Basque*, 15-64. Berlin: De Gruyter Mouton.

Hyman, L.M. 1983. Are there syllables in Gokana? In Kaye, J. et al (eds.), *Current issues in African linguistics* 2, 171-179. Dordrecht: Foris.

Hyman, L.M. 1985. *A Theory of Syllable Weight*. Stanford: CSLI.

Hyman, L.M. 2008. Universals in phonology. *The Linguistic Review* 25, 83-137.

Igartua, I. 2015. Diachronic effects of rhinoglottophilia, symmetries in sound change, and the curious case of Basque. *Studies in Language* 39, 635-663.

Jany, C., M. Gordon, C.M. Nash, and N. Takara. 2007. How Universal is the Sonority Hierarchy?: A Cross-Linguistic Acoustic Study. *Proceedings of the 16th International Congress of Phonetic Sciences*. ICPhS Saarbrücken, Germany, 6-10 August 2007.

Kiparsky, P. 2006. The Amphichronic Program vs. Evolutionary Phonology. *Theoretical Linguistics* 32, 217-236.

Kiparsky, P. 1982. From Cyclic to Lexical Phonology. In H. van der Hulst and N. Smith (eds.), *The structure of phonological representations*. Vol. 1. Dordrecht: Foris, 131-175.

Kiparsky, P. 2015. Phonologization. In P. Honeybone and J. Salmons (eds.), The Oxford *Handbook of Historical Phonology*. Oxford: OUP.

Kisler, T., Reichel, U.D. & Schiel, F. 2017. Multilingual processing of speech via web services. *Computer, Speech & Language* 45, 326-347.

Krajewska, D., E. Zuloaga & A. Egurtzegi. 2022. Sibilant mergers in 18th-century Basque: A quantitative study. *Phonological Data & Analysis* 4.5, 1-31.

Krajewska, D., E. Zuloaga & A. Egurtzegi. In preparation. *The evolution of the sibilant merger(s) in San Sebastian Gipuzkoan Basque*. Manuscript.

Kümmel, M. 2007. *Konsonantenwandel*. Wiesbaden: Reichert.

Kümmel, M. 2015. The role of typology in historical phonology. In P. Honeybone & J.C. Salmons (eds.), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press, 121-132.

Lafon, R. 1962a. Contact de langues et apparition d'une nouvelle voyelle: *u* et *ü* en basque. *Xe Congrès International de Linguistique et Philologie Romanes III. Strasbourg*, 901-909.

Lafon, R. 1962b. Sur la voyelle *ü* en basque. *Bulletin de la Société de Linguistique de Paris* 57, 83-102.

Labov, W. 1994. *Principles of linguistic change. Vol. 1: Internal factors*. Oxford: Blackwell.

Lakarra, J. 2018. La prehistoria de la lengua vasca. In J. Gorrochategui, I. Igartua & J.A. Lakarra (eds.), *Historia de la lengua vasca*. Vitoria-Gasteiz: Gobierno Vasco, 23-244.

Lenth, R. 2023. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.7, https://CRAN.R-project.org/package=emmeans.

Matisoff, J.A. 1975. Rhinoglottophilia: The mysterious connection between nasality and glottality. In C.A. Ferguson, L.M. Hyman & J.J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford: Stanford University Press, 265-287.

Michelena, L. 1977. *Fonética histórica vasca*. San Sebastián: Diputación Foral de Guipúzcoa.

Oñederra, M.L. 2009a. Diacronía y motivación fonética. In F. Sánchez-Miret (ed.), *Romanística sin complejos. Homenaje a Carmen Pensado*. Bern: Peter Lang, 47-67.

Oñederra, M.L. 2009b. Ohar batzuk bokal batez: Zubereraren [y]. *International Journal of Basque Linguistics and Philology* (*ASJU*) 43, 663-678.

Ohala, J.J. 1975. Phonetic explanations for nasal sound patterns. In C.A. Ferguson, L.M. Hyman & J.J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford: Stanford University Press, 289-316.

Ohala, J.J. 2003. Phonetics and historical phonology. In B. Joseph & R. Janda (eds.), *The handbook of historical linguistics*. Malden: Wiley-Blackwell, 669-686.

Parrell, B. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics* 40, 37–45.

Pouplier, M., Beňuš, Š. 2011. On the Phonetic Status of Syllabic Consonants: Evidence from Slovak. *Journal of Laboratory Phonology* 2, 243-273.

Ramsammy, M. 2015. The life cycle of phonological processes: accounting for dialectal microtypologies. *Language and Linguistics Compass* 9, 33-54.

R Core Team. 2023. *R: A language and environment for statistical computing*. Computer software available from http://www.R-project.org.

Ruch, H. & Harrington, J. 2014. Synchronic and diachronic factors in the change from preaspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics* 45, 12-25.

Ruch, H. & Peters, S. 2016. On the origin of post-aspirated stops: production and perception of /s/+voiceless stop sequences in Andalusian Spanish. *Laboratory Phonology* 7, 1-36.

Salmons, J. 2021. *Sound Change*. Edinburgh: Edinburgh University Press.

Scarborough, R., Zellou, G., Mirzayan, A. & Rood, D. 2015. Phonetic and phonological patterns of nasality in Lakota vowels. *Journal of the International Phonetic Association* 45, 289-309.

Styler, W. 2017. On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America* 142, 2469-2482.

Trask, R.L. 1997. *The history of Basque*. London: Routledge.

Uhlenbeck, C.C. 1903. *Beiträge zu einer vergleichenden Lautlehre der baskischen Dialecte*. Amsterdam: Johannes Müller.

Ullrich, Jan F. (ed.). 2011. *New Lakota Dictionary*. Second edition. Bloomington, Indiana: Lakota Language Consortium.

Walker, R. & G.K. Pullum. 1999. Possible and impossible segments. *Language* 75, 764-780.

Winkelmann, R. 2017. *The EMU-SDMS*. PhD thesis, University of Munich (LMU). https://cran.r-project.org/package=emuR.

Winkelmann, R., Bombien L., & Scheffers, M. 2019. *wrassp: Interface to the 'ASSP' library*. https://cran.r-project.org/package=wrassp.

Yu, A.C.L. 2015. The role of experimental investigation in understanding sound change. In P. Honeybone & J. Salmons (eds.), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press, 410-428.

Zarabozo, C. 1972. Two borrowed sound changes in Basque. *International Journal of Basque Linguistics and Philology* (*ASJU*) 6, 174-182.

Zellou, G. & M. Tamminga. 2014. Nasal coarticulation changes over time in Philadelphia. *Journal of Phonetics* 47, 18-35.

Appendix: Selected publications

This appendix includes 5 of my most important publications in alphabetical order. Most of my publications can be downloaded from https://egurtzegi.github.io/publications/.

List of included publications:

Blevins, J., Egurtzegi, A. & Ullrich, J. 2020. "Final obstruent voicing in Lakota: Phonetic evidence and phonological implications". *Language* 96.2, 294-337 (44 pgs.).

Egurtzegi, A. 2017. "Phonetically conditioned sound change: Contact induced /u/-fronting in Zuberoan Basque". *Diachronica* 34.3, 331-367 (37 pgs.).

Egurtzegi, A. 2018. "On the phonemic status of nasalized /ɦ̃/ in Modern Zuberoan Basque". *Linguistics* 56.6, 1353-1367 (15 pgs.).

Egurtzegi, A. & Carignan, C. 2020. "An acoustic description of Mixean Basque". *Journal of the Acoustical Society of America* 147.4, 2791-2802 (12 pgs.).

Krajewska D., Zuloaga, E. & Egurtzegi, A. 2022. "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1-31 (31 pgs.).

FINAL OBSTRUENT VOICING IN LAKOTA: PHONETIC EVIDENCE AND PHONOLOGICAL IMPLICATIONS

Juliette Blevins
The Graduate Center, City University of New York

Ander Egurtzegi
Centre National de la Recherche Scientifique / IKER (UMR5478)

Jan Ullrich
The Language Conservancy

Final obstruent devoicing is common in the world's languages and constitutes a clear case of parallel phonological evolution. Final obstruent voicing, in contrast, is claimed to be rare or nonexistent. Two distinct theoretical approaches crystalize around obstruent voicing patterns. Traditional markedness accounts view these sound patterns as consequences of universal markedness constraints prohibiting voicing, or favoring voicelessness, in final position, and predict that final obstruent voicing does not exist. In contrast, phonetic-historical accounts explain skewed patterns of voicing in terms of common phonetically based devoicing tendencies, allowing for rare cases of final obstruent voicing under special conditions. In this article, phonetic and phonological evidence is offered for final obstruent voicing in Lakota, an indigenous Siouan language of the Great Plains of North America. In Lakota, oral stops /p/, /t/, and /k/ are regularly pronounced as [b], [l], and [g] in word- and syllable-final position when phrase-final devoicing and preobstruent devoicing do not occur.*

Keywords: final voicing, final devoicing, markedness, Lakota, rare sound patterns, laboratory phonology

1. Final obstruent devoicing and final obstruent voicing in phonological theory. There is wide agreement among phonologists and phoneticians that many of the world's languages show evidence of final obstruent devoicing (Iverson & Salmons 2011). Like many common sound patterns, final obstruent devoicing has two basic instantiations: an active form, involving alternations, and a passive form, involving static distributional constraints. In languages with active final obstruent devoicing, voiced obstruents like /b/, /d/, and /g/ are pronounced as voiceless [p], [t], and [k] in word- or syllable-final position. One language with this pattern is Czech (Šimáčková, Podlipský, & Chládková 2012), as illustrated in Table 1, where words are given in orthographic form, followed by International Phonetic Alphabet (IPA) transcriptions in square brackets.[1]

In languages with static final obstruent voicelessness, the contrast between voiced and voiceless obstruents is neutralized in favor of the voiceless series in word- or syllable-final position, though there is no synchronic evidence of productive alternations. This pattern is illustrated by the representative Basque data in Table 2. Though Basque has a contrast between voiced /b/, /d/, /g/ and voiceless /p/, /t/, /k/, the voicing contrast

* We are grateful to the Lakota Language Consortium for allowing us to access and analyze sound files from the *New Lakota dictionary* for the purposes of this study, and to Ben Black Bear, Jr. and Iris Eagle Chasing, whose contributions to the *New Lakota dictionary* made this research possible. We also want to thank *Language* co-editor Megan Crowhurst, associate editor Khalil Iskarous, and three anonymous referees for their insightful comments, which have greatly improved this article. The first author pursued this work under the auspices of the Endangered Language Initiative, generously supported by the Provost's Office of The Graduate Center, CUNY. The second author would like to thank Christopher Carignan, Aitor Egurtzegi, and Raphael Winkelmann for their helpful suggestions. The work of the second author was supported, in part, by an ERC Advanced Grant to Jonathan Harrington.

[1] Throughout this article, we use the following abbreviatory conventions: ACC: accusative case, ADV: adverb, adverbial, C: consonant (nonsyllabic segment), CAUS: causative, CONT: contracted (or truncated) form, diss.: dissimilation, Eng.: English, e.o.: 'everyone', Fr.: French, GEN: genitive case, INAN: inanimate, IPA: International Phonetic Alphabet, n.a.: not applicable, MVS: Mississippi Valley Siouan, *NLD*: *New Lakota dictionary*, PL: plural, PRS: present, RED: reduplication, resyllab.: resyllabification, sb.: 'somebody', SG: singular, sth.: 'something', SUFF: suffix, V: vowel (syllabic segment), VBZ: verbalizer, VOT: voice onset time, wd.: word, <p:>: pause (in phonetic transcriptions), '.' (period): used to mark syllable boundary.

Printed with the permission of Juliette Blevins, Ander Egurtzegi, & Jan Ullrich. © 2020.

	INITIAL			MEDIAL			FINAL		
/b/	bát se	[**b**a:tsɛ]	'to fear'	chleba	[xle**b**a]	'bread.ACC'	chléb	[xlɛ:**p**]	'bread'
/p/	pád	[**p**a:t]	'a fall'	čápi	[tʃa:**p**i]	'stork.PL'	čáp	[tʃa:**p**]	'stork'
/d/	dát	[**d**a:t]	'give'	ledu	[lɛ**d**u]	'ice.GEN'	led	[lɛ**t**]	'ice'
/t/	tát	[**t**a:t]	'melt'	letu	[lɛ**t**u]	'flight.GEN'	let	[lɛ**t**]	'flight'
/g/	gáza	[**g**a:za]	'gauze'	kolega	[kole**g**a]	'colleague'	gong	[go:ŋ**k**]	'gong'
/k/	kát se	[**k**a:tsɛ]	'repent'	žáky	[ʒa:**k**i]	'pupil.PL'	žák	[ʒa:**k**]	'pupil'

Table 1. Final obstruent devoicing in Czech.

is possible only in word-initial and word-medial position (Egurtzegi 2013).[2] In word-final position, of the oral stops, only /p/, /t/, and /k/ are attested, though [p] is restricted to sound-symbolic words, [t] is rare, and [k] is primarily found in a handful of highly productive suffixes (e.g. *-ak* PLURAL, *-k* ERGATIVE, *-tik* ABLATIVE). Forms in Table 2 are written in the standard Basque orthography, which is phonemic: word-final /p/, /t/, and /k/ represent [p], [t], and [k], respectively.

	INITIAL		MEDIAL		FINAL	
/b/	**b**eldar	'caterpillar'	za**b**al	'wide'	—	
/p/	**p**eldo	'wild mint'	za**p**al	'crushed'	za**p**	[sound of hit, knock]
/d/	**d**u	'have.3SG.PRS'	a**d**ar	'horn, branch'	—	
/t/	**t**u	'saliva, spit'	a**t**al	'section, piece'	ba**t**	'one'
/g/	**g**ehi	'more, a lot'	e**g**arri	'thirst'	—	
/k/	**k**e	'smoke, steam'	e**k**arri	'to bring'	zuti**k**	'standing, upright'

Table 2. Final obstruent voicelessness in Basque.

Sound patterns similar to Czech and Basque active and static final obstruent devoicing have been described for many other languages. Blevins (2006a) lists over a dozen modern Indo-European languages with final obstruent devoicing, including Bulgarian, Catalan, Dutch, Lithuanian, Polish, Russian, and Zaza. At the same time, she illustrates that the sound pattern has clearly evolved independently in unrelated languages around the world, from Afar, a Cushitic language spoken in the Horn of Africa, to Awara, a Finisterre-Huon language of Morobe Province, Papua New Guinea.[3] Further, as new languages are described, new cases of final obstruent devoicing continue to be discovered: acoustic analysis confirms the sound pattern in Camuno, a Gallo-Romance language of Valcamonica (Cresci 2014), and in Ganza, an Omotic language of Ethiopia and the Sudan (Smolders 2016).[4]

Final obstruent devoicing has received a great deal of attention in the phonological literature, since the understanding of this common sound pattern has important implica-

[2] Basque dialects differ as to whether /p, t, k/ are voiceless unaspirated or voiceless aspirated in onset position. In aspirating dialects, aspiration is limited to prevocalic position.

[3] The description of Ingush, a Nakh-Daghestanian language of the Caucasus, reported to have final obstruent devoicing in Blevins 2006a, should be slightly modified. Following Nichols (2011:80), Ingush has only partial devoicing of voiced obstruents in word-final position. Acoustically, neutralization with the voiceless series is incomplete, though some speakers cannot perceive a contrast (Nichols 2011:8).

[4] Final obstruent devoicing can be viewed as one of five common domain-final laryngeal sound patterns. The other four are: final deaspiration, as in Marathi (Houlihan & Iverson 1979); final aspiration, as in Sierra Popoluca (Elson 1947, de Jong Boudreault 2009); final deglottalization, as in Ganza (Smolders 2016); and final glottalization, as in Standard Thai (Henderson 1964, Harris 2001). In this study, we restrict discussion to final obstruent devoicing in languages with true obstruent voicing contrasts. In true voicing languages, the voiced series show significant voiced closure durations, and the voiceless series have insignificant VOTs. See Jansen 2004 and Beckman, Jessen, & Ringen 2013 for further discussion of true voicing languages, and Kakadelis 2018 for analysis of domain-final laryngeal sound patterns in 'no-voicing' languages.

tions for phonological theory, as summarized in Iverson & Salmons 2011. One important area of research focuses on explanations for the sound pattern itself and asks why final obstruent devoicing is a common sound pattern crosslinguistically. Two distinct theoretical approaches offer two very different answers to this question.

Under traditional markedness accounts inspired by Trubetzkoy (1939) (e.g. Wetzels & Mascaró 2001) and modern OPTIMALITY treatments (e.g. Kager 1999, Lombardi 1999), final obstruent devoicing is viewed as a direct consequence of universal phonological markedness constraints. Traditionally, voiced obstruents are marked, voiceless obstruents are unmarked, and final devoicing, as neutralization, constitutes a shift to the unmarked. In modern optimality terms, a markedness constraint prohibiting voicing in obstruents combines with positional markedness or faithfulness constraints. As components of universal grammar, these markedness constraints determine that obstruent voicing will be generally disfavored, and particularly disfavored in final (or noninitial) position.[5] The same kinds of markedness accounts make explicit predictions that final obstruent voicing should not exist (Kiparsky 2006, 2008): a language with /p/, /t/, /k/ regularly pronounced as [b], [d], and [g] in word- or syllable-final position is ruled out, since, under any analysis, the voiced obstruents are marked in contrast to their voiceless counterparts.

In contrast to markedness theories, phonetic-historical approaches to final obstruent devoicing, like EVOLUTIONARY PHONOLOGY (Blevins 2004, 2006a,b, 2008, 2015, 2017), attempt to explain the recurrent sound pattern as phonologized instances of natural phonetic processes. Under these accounts, inspired by the Neogrammarian tradition and the early work of John J. Ohala (e.g. 1981, 1983), final obstruent devoicing is common crosslinguistically because of the way we speak and the way we perceive speech. More specifically, phrase-final laryngeal gestures, phrase-final lengthening, and final consonant nonrelease, along with perception and phonologization of these articulatory routines, can all give rise to voicelessness, or perception of voicelessness, in final obstruents, yielding sound patterns like those illustrated in Tables 1 and 2 above. Within evolutionary phonology, nothing prohibits sound patterns of final obstruent voicing, though they are expected to be rare, due to the articulatory and perceptual factors just mentioned that yield devoicing (Blevins 2006a,b).

The debate between markedness and phonetic-historical approaches has led to a special interest in languages that may show evidence of word- or syllable-final voicing. Lezgian, a Nakh-Daghestanian language, is one of these. In Lezgian, there is a contrast between plain voiceless, voiceless aspirated, voiced, and glottalized stops, with plain voiceless stops alternating with voiced stops word-finally. Yu (2004) provides acoustic and phonological evidence for a synchronic process of final obstruent voicing and lengthening. In an attempt to maintain predictions of markedness theory, Kiparsky (2006, 2008) offers an alternative analysis of the Lezgian sound pattern: final voiced stops are taken as basic, analyzed as phonologically voiced geminate stops, and degeminated and devoiced in syllable onsets. Another language with possible word-final voicing of obstruents is Somali (Blevins 2006a,b). However, there is a great deal of variation in how final obstruents are pronounced, and Kiparsky (2006, 2008) chooses to analyze Somali final stops as lenis unaspirated, in contrast to aspirated stops that occur syllable-initially. Though Iverson and Salmons (2011:1638) conclude that Kiparsky's proposed markedness universal 'does not ultimately hold up empirically', the absence of any constraint

[5] More abstract universalist approaches, like GOVERNMENT PHONOLOGY, propose empty nuclei where other frameworks have coda consonants. See, for example, Brockhaus 1995 on German final devoicing.

against final obstruent voicing within evolutionary phonology has led us to search for more convincing examples of this sound pattern. In this context, we offer the present study of Lakota, an indigenous Siouan language of the Great Plains.

Our central argument is that Lakota has a true synchronic process of syllable-final obstruent voicing. This argument is supported by phonological and phonetic evidence that Lakota voiceless oral stops /p/, /t/, and /k/ are regularly pronounced as [b], [l], and [g], respectively, in syllable-final position. Section 2 provides an introduction to the Lakota language, its speakers, and the phonology of the language, with a focus on the distribution of obstruent voicing, and a brief summary of earlier analyses. Acoustic analyses of Lakota segments in different positions of the word are presented in §3, demonstrating a voicing contrast in prevocalic position and a neutralization of oral stops to the voiced series in syllable-final position. Other patterns of note are optional phrase-initial devoicing, gradient phrase-final devoicing, regressive devoicing of oral stops followed by voiceless segments, categorical syllable-final fricative devoicing, and presonorant oral stop voicing, though our acoustic analysis is focused on showing that /p/ and /k/ undergo final voicing. In §4 we suggest that final obstruent voicing in Lakota is a continuation of an earlier coarticulatory sound change that voiced *p, *t, *k to [b], [d], [g] intervocalically before final unstressed vowels concomitant with devoicing and loss of those vowels. This sound change was followed by a shift of *d > l in Lakota. Under this account, the historical origins of final stop voicing are tied to retiming of the final vowel gesture. Section 5 summarizes the implications of this study for phonological theory.

2. Lakota obstruent voicing patterns.

2.1. A brief introduction to the Lakota language. Lakota (a.k.a. Lakhota) is an endangered indigenous language of North America. Today, it is mainly spoken on five reservations in North and South Dakota. The number of fluent speakers of Lakota has been declining steadily since the 1950s, and intergenerational transmission of the language ended during the 1960s, with a very small and decreasing number of isolated families continuing to speak Lakota to their children up to the 1990s. According to the Lakota Language Consortium, since that time, the number of first language speakers has decreased from approximately 6,000 to about 2,000 speakers today (Ullrich 2018:33).

Lakota is a member of the Siouan language family,[6] and within Siouan, it is usually classified as a member of the Mississippi Valley subgroup. Siouan languages were spoken primarily in the Great Plains, and in the Ohio and Mississippi valleys. Lakota is a member of a dialect continuum that includes five distinguishable languages, namely: Lakȟóta, the subject of this study, Western Dakhóta (a.k.a. Yankton-Yanktonai), Eastern Dakhóta (a.k.a. Santee-Sisseton), Assiniboine Nakhóta, and Stoney Nakhóta. Some phonological differences between these languages are illustrated in Table 3.

Where Lakota has /l/, it regularly corresponds to /d/ in all languages but Assiniboine, where it corresponds instead to /n/. (In final position /-n/ has diffused to former /-d/ dialects.) For some morphemes, however, like the Lakota suffix /-la/ illustrated in the word for 'lizard', unexpected *n*-forms occur in Yanktonai, Yankton, and Sisseton, which are *d*-dialects. (In most instances, these can be attributed to lexical diffusion.) Note also that Lakota /bl-/ corresponds to Assiniboine /mn-/, while Lakota /gl-/ corresponds to Yanktonai /gd-/, Yankton /kd-/, Sisseton-Santee /hd-/, and Assiniboine /kn-/. More will

[6] The Siouan language family is also sometimes referred to as Siouan-Catawban to include the more distantly related Catawban languages. See Ullrich 2018:33–35 for a brief summary of language relationships within this family, and Parks & DeMallie 1992 on Lakota-Dakota dialects.

LAKȞÓTA	YANKTONAI	YANKTON	SISSETON	SANTEE	ASSINIBOINE	GLOSS
loté	**d**oté	**d**oté	**d**oté	**d**oté	**n**oté	'throat'
-ke**l**	-ke**d**	-ke**d**	-ke**d**	-ke**d**	-ke**n**	'kind of' (ADV SUFF)
blaská	**bd**aská	**bd**aská	**bd**aská	**bd**aská	**mn**aská	'to be flat & solid'
a**gl**éška-**l**a	a**gd**éška-**n**a	a**kd**éška-**n**a	a**hd**éška-**n**a	a**hd**éška-**d**a	a**kn**éška-**n**a	'lizard'

TABLE 3. Some phonological differences across the Lakota-Dakota dialect continuum.

be said about these correspondences in our discussion of the evolution of final obstruent voicing in §4. In the subsections that follow we focus solely on Lakota.

The Lakota language can be divided into two dialects: Northern Lakota, represented by speakers of the Standing Rock Reservation and parts of the Cheyenne River Reservation, and Southern Lakota, spoken by the Oglála and Sičháŋğu tribes, who reside on the Rosebud and Pine Ridge Reservations, respectively, and by some speakers from Cheyenne River (Ullrich 2018:38). Since these two dialects show virtually no phonological variation and are characterized by only a small number of lexical variants, they are treated as one for the purposes of this study.

Descriptive and teaching grammars of Lakota include Buechel 1939, Boas & Deloria 1941, Rood & Taylor 1976, 1996, Ingham 2003, the grammar section of Ullrich 2008, and Ullrich & Black Bear 2016. This study makes extensive use of the *New Lakota dictionary* (Ullrich 2008, 2011, 2019), a publication of the Lakota Language Consortium. The app version of the dictionary contains over 40,000 entries and includes not only full forms of words and thousands of audio files, but also truncated word forms, which are important to this study.

2.2. SEGMENT INVENTORY AND ORTHOGRAPHY. The Lakota vowel system is shown in Table 4, and the consonant inventory is shown in Table 5, following Rood and Taylor (1996) and Ullrich (2011). Here and throughout, we use the orthography of the *New Lakota dictionary* (*NLD*; Ullrich 2011, 2019), sometimes called the 'Standard Lakota orthography'. Where orthographic symbols in these tables differ from symbols of the IPA, the IPA symbol is given in square brackets.

		FRONT	CENTRAL	BACK
HIGH	oral	i		u
	nasalized	<iŋ> [ĩ]		<uŋ> [ʊ̃]
MID	oral	e		o
LOW	oral		a	
	nasalized		<aŋ> [ã]	

TABLE 4. Lakota vowel contrasts.

Lakota has a basic five-vowel system /i, u, e, o, a/, along with three nasalized vowels, including high /iŋ/, /uŋ/ and nonhigh /aŋ/. Another vowel symbol in use in the *NLD* is <A>; this symbol represents a root/stem-final vowel that alternates between /a/, /e/, and /iŋ/ and is often subject to deletion in morphologically complex forms. In addition, the acute accent is used to mark a stressed vowel in this orthography.[7] It should also be noted that in addition to the three contrastively nasalized vowels shown in Table 4, Lakota also has allophonically nasalized vowels that are due to coarticulation with a preceding or following nasal consonant (Scarborough et al. 2015). In our study of properties of oral stop consonant voicing, we do not measure stops in the context of nasalized vowels, since these stops are typically nasal and voiced. Given that our study is

[7] For one of the earliest analyses of stress patterns in Dakota-Lakota dialects, see Shaw 1980.

limited to a discussion of (oral) stops adjacent to oral vowels, it provides only a partial picture of voicing alternations in Lakota.

Table 5 shows the consonant inventory of Lakota, with obstruents at the top of the table and sonorants at the bottom.

	BILAB	DENT	ALV	POST-ALV	VELAR	GLOT
OBSTRUENTS						
Stops & affricates						
voiceless unaspirated	p	t		č = [tʃ]	k	
voiceless aspirated (or Th cluster)	ph	th		čh	kh	
voiceless with velar aspiration (or Tȟ cluster)	pȟ	tȟ			kȟ	
voiceless ejective (or Tʔ cluster)	p’	t’		č’	k’	’ = [ʔ]
voiced	b				(g)	
Fricatives					(POST)VELAR	
voiceless			s	š = [ʃ]	ȟ = [x], [χ]	h
voiceless glottalized (or Sʔ cluster)			s’	š’	ȟ’ = [x’], [χ’]	
voiced			z	ž = [ʒ]	ğ = [ɣ], [ʁ]	
SONORANTS				PALATAL		
Nasals	m		n			
Lateral		l				
Approximants	w			y = [j]		

TABLE 5. Lakota consonantal contrasts.

Several comments are in order regarding the consonant contrasts in Table 5. Note that all Lakota sonorants—/m, n, l, w, y/—are voiced. In contrast, the nonlaryngeal fricatives all have voiced and voiceless counterparts. For the oral stops and affricates, there appear to be at least five contrastive laryngeal series: voiceless unaspirated, voiceless aspirated, voiceless with velar fricative release, voiceless ejective, and voiced. However, the voiceless aspirated, voiceless with velar fricative release, and voiceless ejective series have alternative analyses as clusters of plain voiceless obstruent plus /h/, /ȟ/, and /’/, respectively (Boas & Deloria 1941:5). Since the focus of this study is the plain voiceless and voiced series of obstruents, we leave the analysis of these other laryngeal series open and focus on contrastive (and noncontrastive) obstruent voicing in Lakota.

2.3. THE DISTRIBUTION OF OBSTRUENT VOICING. Let us begin by observing a voicing contrast in fricatives. As illustrated in Table 6, the pairs /s/ vs. /z/, /š/ vs. /ž/, and /ȟ/ vs. /ğ/ contrast in prevocalic (onset) position, but not word- or syllable-finally, where, as in Czech, Basque, and many other languages, there is neutralization to the voiceless series, illustrated by the bold characters in the final column. The fricative devoicing pattern exemplified in Table 6 is most obvious in full and contracted (CONT) pairs like *čhážga* ‘ice’ vs. *čhá**ȟ*** (CONT). Contracted forms are the primary source of obstruent codas in Lakota and are discussed further in §2.4. Of importance now is to observe that the voicing contrast in Lakota fricatives is unremarkable from a typological perspective. There is a voicing contrast that occurs in single-member onsets at three distinct points of articulation, and that contrast is neutralized in final position to the voiceless series.

Since it is rare to have a voicing contrast in fricatives without having a voicing contrast in oral stops, we expect Lakota to show a series of voiced stops. However, the voiced series of oral stops appears to have only a single member, /b/. Though there are very few contrasts between /b/ and /p/, data like that in example 1 argue that voicing is contrastive for prevocalic bilabial stops in Lakota.[8] In addition to potentially native

[8] Though /bu/ and /pu/ contrast in Lakota, there is no /wu/, leading some to suggest that *bu* < *wu historically. However, since *bá* ‘to blame sb.’ and *wa-* contrast (cf. *wa-*[1] indefinite object marker, *wa-*[2] ‘with a

	INITIAL (ONSET)		INTERVOCALIC (ONSET)		SYLLABLE-FINAL (CODA)	
/s/	só	'to cut sth. into strings'	pasí	'research'	khúskhus[a]	'couscous'
/z/	zomí	'to be a schemer'	kawázA	'to throw sth. up'	kawá**s** (CONT)	'to throw sth. up'
			kózA	'to wave sth.'	kó**s**kozA	'to wave sth.'
/š/	šóta	'smoke'	itȟášoša	'to spit sth. out'	itȟášoš (CONT)	'to spit sth. out'
					tóš, toštóš	'yes, surely, of course'
/ž/	žó	'to whistle'	léžA	'to urinate'	léš (CONT)	'to urinate'
					lešlížA	'to urinate often'
/ȟ/	ȟolyá	'being gray'	iyáyuȟa	'accompany sb.'	iyáyuȟ (CONT)	'accompany sb.'
					iyáyuȟya	'constantly following sb.'
/ǧ/	ǧópa	'to snore'	čháǧa	'ice'	čhá**ǧ** (CONT)	'ice'
					čha**ǧ**nážužu	'ice breaks up'

TABLE 6. Lakota voiced vs. voiceless fricatives, with final devoicing.

[a] A loan is used for illustration since all known syllable-final instances of [s] in the native vocabulary derive from /z/, as in the adverb *inák'es* 'to be flush with' from /i-na-k'ezA/, /-k'ezA/ 'to be even with, flush with'.

roots, like *bá* 'to blame someone' and *bú* 'make a deep noise', there is at least one likely loan, *bébela* 'baby' (<< Fr. *bébé*), which also suggests that Lakota /b/ was contrastive at the time of borrowing. If there was no voicing contrast between /b/ and /p/ in the language, we would expect the word to be borrowed as /pepe…/.

(1) The /b/ vs. /p/ contrast in Lakota
  a. ***bá*** 'to blame sb.' (not widely known) vs. ***pa***- 'by pushing'
  b. ***bébela*** 'baby' (<< Fr. *bébé*) vs. -***pi*** PLURAL
  c. ***bú*** 'make a deep noise' vs. ***pu***- 'by pressure'
  d. *á**b**ela* 'scattered', *á**b**eya* 'scattering' vs. *a**p**é* 'leaf'
  e. *ka**b**ú* 'to play the drum' (*ka*- 'by hitting', *bu* 'make a deep noise') vs. *ka**p**úza* 'to become dry in the wind' (*ka*- VBZ, *púzA* 'to be dry')
  f. *hi**b**ú* 'I am coming' (archaic form of 1SG of *hiyú* 'to start coming') vs. *i**p**áblaye* 'rolling pin'

In other positions within the word, the pronunciation of /p/ as [b] is predictable, as discussed further below. Though the voicing contrast for /p/ vs. /b/ has a low functional load, it is supported by the data in 1.

For the velar stop /k/, the situation is different, and this is why /g/ is in parentheses in Table 5. Though [g] is a common and predictable allophone of /k/, there are no contexts where /k/ and /g/ contrast. In example 2, we illustrate two of the three positions where /k/ has predictable allophones: prevocalically (2a), where it is voiceless unaspirated, and before sonorant consonants /l, m, n, w/, where it is voiced (2b). Notice that unlike /b/ in *bébela*, in Lakota *spakéli*, the English prevocalic [g] of [spəˈgɛri] 'spaghetti' is borrowed as /k/ (in 2a); only when [g] is in presonorant position in source words like *magnet* or *anglais* is it borrowed as [g] (in 2b).

(2) /k/ with predictable [k] and [g] allophones
  a. prevocalic [k]: *a**k**ábu* 'to drum on sth.', ***k**ibá* 'to regret', -*la**k**e* 'very, really', *spa**k**éli* 'spaghetti' (<< Eng.)
  b. presonorant syllable-initial [g]: ***g**lalú* 'to fan one's own', ***g**má* 'walnut', ***g**núni* 'to lose one's own' (< *ki-núni*), ***g**wéza* 'rippled, ridged', *ma**g**néta* 'magnet' (<< Eng.), *ša**g**láša* 'English' (<< Fr. *les Anglais*)

knife', *wa*-[3] 1SG for class I verbs), the contrastive status of /b/ seems secure in some vocalic contexts. Note that Lakota *wa*-[1] corresponds to *ba*- in Dakota dialects and to *ma*- in the two Nakota languages.

Like /k/, Lakota /t/ also lacks a voiced counterpart. Recall from Table 3 that where other closely related languages show /d/, the corresponding sound in Lakota is /l/. Internal to Lakota, there is also evidence that /l/ is, in some sense, the 'voiced' counterpart of /t/. For example, consider the data in Table 7, where contracted forms of words with medial /p/, /t/, and /k/ show final [b], [l], and [g], respectively. When the preceding vowel is nasalized, the voiced consonants [b] and [g] may be realized as [m] and [ŋ], respectively, or as partially nasalized [m^b]/[m^p] and [$ŋ^g$]/[$ŋ^k$], respectively, while [l] is often nasalized and lenited in the same context. For example, compare *núŋpa* [ˈnõ.pa] 'two, twice' with shortened forms *núŋm* [ˈnõm], *núŋp* [ˈnõm^p] and reduplicated *núŋmnuŋpa* [ˈnõ$m^ə$.ˈnõ.pa] 'by twos, two each'. We analyze these variants as nasalized instances of [b] and [g], and we focus on oral contexts in most of the discussion that follows.

	medial onset		word-final (coda)		medial coda	
/p/	tó**p**a	'four'	tó**b** (cont)	'four'	tó**b**topa	'by fours'
/t/	napó**t**A	'to wear sth. out with the feet' (of footwear)	napó**l** (cont)	'to wear sth. out with the feet' (of footwear)	napó**l**potA	'wearing sth. out with the feet'
/k/	šó**k**A	'to be thick'	šó**g** (cont)	'to be thick'	šo**g**šókA	'to be thick'

Table 7. A first glance at Lakota final stop voicing.

If the alternations in Table 7 represent a unified process, the expected pronunciation of /t/ is [d], not [l]. A further piece of evidence that Lakota [l] is, in some sense, the voiced counterpart of /t/ comes from place of articulation. As illustrated in Table 5, /t/ (along with /th/, /tȟ/, and /t'/) has a dental place of articulation, while /s/, /z/, and /n/ are alveolar. Since /l/ is a sonorant like /n/, and sometimes considered a continuant like /s/ and /z/, it might be expected to have an alveolar place of articulation. However, like /t/, it is dental. In §4 we suggest that the voicing alternations in Table 7 and the /d/ correspondent of /l/ in other Siouan languages support a historical sound change of *d > l in Lakota. In other words, the seemingly unnatural synchronic change of /p, t, k/ to [b, l, g] as opposed to [b, d, g] in Lakota is due to telescoping of two sound changes, obstruent voicing followed by *d > l.[9]

A final context where obstruent voicing is predictable is in word-initial consonant clusters. Given that word-initial position defines the beginning of a syllable at the beginning of an utterance, we take word-initial phonotactics to define (at least partially) syllable-initial phonotactics. Word-initially, any consonant in Table 5 (except [g], which, recall, is not contrastive) can constitute a single-member prevocalic syllable onset. Attested word-initial clusters, shown in Table 8, are highly restricted:[10] (i) they are limited to two consonants, #C_1C_2; (ii) C_2 can be either a plain voiceless obstruent or a sonorant; (iii) if C_2 is a plain voiceless obstruent, C_1 is also a plain voiceless obstruent; (iv) if C_2 is a sonorant, C_1 is either a voiceless fricative, a voiced oral stop [b] or [g], or [m], which we interpret as a nasalized [b] / __ n. In other words, singleton onsets show voicing contrasts in Lakota, but within consonant clusters, obstruents may not contrast in voicing. In-

[9] Within evolutionary phonology, synchronic patterns of this kind are expected. In other frameworks, the synchronic voicing of /p, t, k/ to [b, l, g] might be viewed as odd, since [b, l, g] do not appear to form a natural class. However, with /l/ specified as [−continuant], one can view the spell-out of a voiced, coronal, non-continuant in Lakota as [l]. For other examples where /l/ shows [−continuant] behavior, or patterns essentially as /d/, see Mielke 2008.

[10] For the purposes of this discussion, we follow Table 5 in treating aspirated, velar aspirated, and ejective oral stops as either single segments or clusters, indicated here by parentheses in Table 8. Whether or not these are treated as clusters, the same restrictions can be seen to be in effect.

dependent of obstruent voicing, notice that there are no sequences of identical consonants (**), there are no fricative clusters (***), and, though C_2 may be an affricate, there are no clusters with an affricate in C_1 position and a consonant other than the laryngeals /h/, /'/ in C_2 position (*$). Lakota words exemplifying the initial consonant clusters in Table 8 are shown in 3.[11]

C_1 \ C_2	p	t	k	č	s	š	ȟ/h	m	n	l	w	'
p	**	pt	—	pč	ps	pš	(pȟ)	**	mn	bl	—	(p')
t	—	**	tk	—	—	—	(tȟ)	—	**	**	—	(t')
k	kp	kt	**	kč	ks	kš	(kȟ)	gm	gn	gl	gw	(k')
č	*$	*$	*$	**	*$	*$	(čh)	*$	*$	*$	*$	(č')
s	sp	st	sk	sč	**	***	***	sm	sn	sl	sw	(s')
š	šp	št	šk	šč	***	**	***	šm	šn	šl	šw	(š')
ȟ	ȟp	ȟt	—	ȟč	***	***	**	ȟm	ȟn	ȟl	ȟw	(ȟ')

Table 8. Initial consonant clusters in Lakota.[12] (Notes: Symbols in parentheses may be treated as single segments, or as tautosyllabic onset clusters. '**' Absence of cluster may be due to constraint against sequence of identical consonants. '***' Absence of cluster may be due to constraint against sequence of fricative consonants. '*$' Absence of cluster may be due to constraint against /č/ as first member of cluster.)

(3) Illustration of word-initial consonant clusters in Table 8

a. /p/-initial: ***pté*** 'buffalo cow', ***pčéčela*** 'to be short', ***psá*** 'reed, straw', ***pšíŋ*** 'onion', (***pȟá*** 'to be bitter'), ***mní*** 'water', ***blé*** 'lake', (***p'é*** 'American elm')

b. /t/-initial: ***tké*** 'to be heavy', (***tȟápa*** 'ball'), (***t'éča*** 'to be lukewarm')

c. /k/-initial: ***kpá*** 'to be gouged out', ***kté*** 'to kill sb./sth.', ***kčeyá*** 'to broil sth. over coals', ***ksúyeya*** 'to hurt or injure sb.', ***kšú*** 'to bead sth.', (***kȟákA*** 'to clack, clatter'), ***gmá*** 'walnut', ***gnákA*** 'to lay sth. by', ***glá*** 'to loathe sth./sb.', (***k'á*** 'to dig sth.')

d. /s/-initial: ***spáyA*** 'to be wet', ***stákA*** 'tired (of bodypart)', ***sčú*** 'be shy', ***ská*** 'to be clear white', ***sm**iyáŋyaŋ* 'bare of any outside layer', ***sní*** 'it is cold', ***slá*** 'it is greasy', ***sw**aká* 'to be frayed at the edge', (***s'a*** 'as a habit')

e. /š/-initial: ***špáŋ*** 'to be burned by heat or cold', ***štákA*** 'to be melting', ***ščépȟaŋ*** (v. *sčépȟaŋ*) 'sister-in-law', ***škátA*** 'to play', ***šm**eyá* 'deeply', ***šn**ížA* 'to shrink, shrivel', ***šl**ayá* 'being bare', ***šw**okÁ* 'it is overflowing', (***š'ákA*** 'it is strong/powerful')

f. /ȟ/-initial: ***ȟp**áyA* 'to lie', ***ȟt**álehaŋ* 'yesterday', ***ȟč**á* 'to blossom', ***ȟm**iyáŋ* 'crookedly', ***ȟn**á* 'to groan, snort', ***ȟw**ayÁ* 'to cause sb. to be sleepy/bored', ***ȟl**í* 'to be muddy or slimy', (***ȟ'é*** 'it is rough')

[11] There are a small number of potential initial CCC clusters limited to fast-speech forms: *kčhí*, *kičhí* 'with sb.'; *kčhíčho*, *kičhíčho* 'to invite e.o.'; *kčhíčhopi*, *kičhíčhopi* 'a feast, party'; *kčhíšnala*, *kičhíšnala* 'with him/her/it alone'; *kčhízA*, *kičhízA* 'to fight e.o.'; *kčhó*, *kičhó* 'to invite sb., call to'. However, if /*čh*/ is an aspirated affricate, then these can all be analyzed as CC clusters.

[12] Note: though /h/ and /ȟ/ contrast word-initially (e.g. *há* 'the skin or hide of sb., sth.' vs. *ȟÁ* 'to bury sb., sth.'; *hé* 'animal horns or antlers' vs. *ȟé* 'mountain, mountain ridge'), they do not contrast after word-initial/syllable-initial stops or affricates before oral vowels. In this position, [h] is found after [tʃ] and before high vowels [i] and [u]; [χ] is found before nonhigh back/central vowels [a] and [o], and before nasalized /aŋ/ and /uŋ/; [h] and [χ] are in free variation before [e]; and [h] is found before nasalized /iŋ/, unless it is the ablaut vowel, in which case [χ] is found, resulting in a limited case of contrast. Alternations support these phonotactics. For example, we find *phíŋkpa* 'the top of anything' from *pȟá* 'the principal part of sth.' + *íŋkpa* 'tip', where *pȟ* → [ph] /__ *i*.

While words like those in 3 support the analysis of clusters in Table 8 as syllable-initial clusters, further support for these as true complex onsets comes from distinct phonetic realizations of the same phonological clusters in word-medial position, across a syllable boundary. In Table 9, word-initial and word-medial complex onsets are compared with word-medial heterosyllabic clusters.

	INITIAL TAUTOSYLLABIC	MEDIAL TAUTOSYLLABIC	MEDIAL HETEROSYLLABIC
/pt/	***pt**á.ya* 'together, collectively'	*ptá.**pt**a.ya* 'in bunches'	*ó**b.t**u* 'to be among, with' *to**b.t**ó.pa* 'by fours, four each'
/kp/	***kp**é* 'to make a sharp noise'	*kpé.**kp**e* 'sharp staccato reports'	*i.wóg.la**g.p**hi.ča* 'it is worth talking about' *o.yá**g.p**hi.ča šni* 'it cannot be told'
/tk/	***tk**é* 'to be heavy'	*tke.**tk**é* 'they are heavy'	*ȟo**l.k**í.yA* 'to make one's own gray' *kȟa**l.k**ȟá.tA* 'being hot'
/ps/	***ps**í.čA* 'to jump, leap, hop'	*psí.**ps**i.čA* 'to jump, leap, hop'	*ča**b.s**íŋ.te* 'beaver tail' *sa**b.s**á.pA* 'black' (INAN.PL)
/ks/	***ks**e.yá* 'abruptly'	*a.pá.**ks**a* 'to break sth. on sth.' *čhaŋ.**ks**á* 'a club, policeman's club'	*šuŋ**g.s**á.pA* 'black horse' *sa**g.s**á.kA* 'to be dried until hard' (INAN.PL)
/pȟ/	***pȟ**á* 'to be bitter'	*pȟa.**pȟ**á* 'they are bitter' (INAN.PL)	*ȟa**b.ȟ**á.pA* 'rustling sound'
/tȟ/	***tȟ**á.ȟča* 'deer'	*tȟa.**tȟ**áŋ.ka* 'buffalo bull'	*ȟo**l.ȟ**ó.ta* 'gray'
/kȟ/	***kȟ**ákA* 'to clack, clatter'	*kȟa.**kȟ**á.kA* 'to have the quality of clacking'	*ki.čhí.ya.ȟta**g.ȟ**tag* 'biting each other repeatedly'
/pn/	***mn**í.žA* 'to be curled, contracted'	*mni.**mn**í.žA* 'to be contracted'	*itȟóka**b.n**i* 'beforehand' *tȟa**b.n**ákȟapapi* 'football, soccer'
/p'/	***p'**ó* 'fog, mist, steam'	*ya.**p'**ó* 'to exhale steam'	*čhe**b.'**í.č'i.yA* 'to gain weight' *tȟa**b.'**á.pȟA* 'to play baseball'
/t'/	***t'**é.ča* 'to be lukewarm'	*t'a.**t'**á* 'to be mentally disabled'	*kȟa**l.'**í.slol.ye* 'thermometer' *á**l.'**a.ta.ya* 'whole in each case'
/k'/	***k'**á* 'to dig sth.'	*k'o.**k'**ó.ye.la* 'gulping hurriedly & noisily'	*tȟo**g.'**í.ya* 'to speak a foreign language' *ka.ȟlo**g.'**o.štaŋ.pi* 'vest'

TABLE 9. Tautosyllabic vs. heterosyllabic CC clusters in Lakota.

In C_1C_2 onsets where C_2 is nonnasal, C_1 is consistently voiceless, but when heterosyllabic, the same clusters may show voiced codas [b], [l], [g], as in the final column of Table 9. Also, note that the possibility of a medial coda consonant, followed by an onset consonant cluster, predicts $VC_1.C_2C_3V$ sequences in the language, provided that C_1 is [l], [s], [ʃ], [χ], [b], or [g] and C_2C_3 is one of the clusters shown in Table 8. Some ex-

amples of triconsonantal clusters are: *pȟe**lmn**á* ‘to smell of fire’, *le**šmn**á* ‘to smell of urine’ (cf. *pȟéta* ‘fire’, *léžA* ‘to urinate’, *mná* ‘to have a particular smell’), *ška**lšk**átA* ‘to play frolicking’ (cf. *škátA* ‘to play’), and *wetȟá**bsk**ala* ‘white blood cells’ (cf. *tȟápa* ‘ball’, *ská* ‘white’).

Note that the syllable structure described for Lakota is not typologically unusual. The majority of Lakota syllables are open, ending in a vowel. Assuming two major classes of consonants, obstruents with low sonority and sonorants with high sonority, one can view Lakota as a language that weakly adheres to a general sonority sequencing principle: within the syllable, there is a sonority plateau or rise to the nucleus, and an optional sonority fall after the nucleus (i.e. an optional coda). Under this analysis, Lakota single-member onsets are unrestricted, onset clusters are those shown in Table 8, and coda consonants are restricted to [l], [s], [ʃ], [χ], [b], and [g]. However, what has yet to be detailed is the origin of these coda consonants. With only a few exceptions, all Lakota roots, stems, and words end in vowels. As a consequence, the majority of coda consonants are found only in derived forms, where truncation or final vowel loss may occur under compounding, derivation, or inflection. A discussion of truncation is offered in §2.4.

Before examining the status of derived coda consonants, let us summarize the distribution of voicing for obstruents that we have seen thus far. Voicing is contrastive for all fricatives in Lakota, with evidence of syllable-final fricative devoicing in Table 6. For oral stops, the situation is different. Of the oral stops, /p/, /t/, and /k/, only /p/ has a voiced obstruent counterpart, /b/. In contrast to fricative devoicing, oral stops appear to be voiced in the same contexts, independent of whether voicing is contrastive. In syllable-final position, /p/, /t/, and /k/ are realized as [b], [l], and [g], respectively (Table 7). Syllabification of tautosyllabic vs. heterosyllabic consonant clusters may be signaled by differences in obstruent voicing (and manner), as shown in Table 8 and Table 9, where, for example, tautosyllabic /pt/, /ps/, and /pn/ are realized as [pt], [ps], and [mn], but the same heterosyllabic sequences can be realized as [b.t], [b.s], and [b.n]. Section 3 of this article supports these observations about obstruent voicing with acoustic measurements. In §2.4 we describe the truncation process that gives rise to the majority of syllable codas. Since our argument is that oral stops undergo voicing in syllable-final position, it is important to understand how truncation gives rise to syllable codas in the language. In §2.5, we briefly review other analyses of the voicing pattern in Lakota obstruents and highlight differences between our proposal and those of earlier scholars.

2.4. Truncation and syllable codas. Lakota syllables may be of the form CV, CCV, CVC, or CCVC, with representative examples in Table 10. Note that onsets appear to be obligatory, at least in careful speech: syllables that are vowel-initial phonologically are pronounced with an initial glottal stop in careful speech: *á* [ˈʔa] ‘armpit’, *alí* [ʔaˈli] ‘to climb up’, *aá* [ʔaˈʔa] ‘to be moldy’. Notice also that three of the four examples of word-final consonants in Table 10 are marked (cont), indicating that they are ‘contracted’ forms. Indeed, the majority of syllable codas in Lakota are the result of contraction or truncation, as described below.[13]

Recall from Table 6 and Table 7 data suggesting that the set of syllable codas is limited to voiceless fricatives [s], [ʃ], [χ] and to the voiced consonants [b], [l], and [g]. However, of all the morphemes described in the *NLD*, only a few appear to be truly consonant-final. Two productive suffixes that appear to be consonant-final are -*kel*, a derivational affix meaning ‘somewhat, rather, fairly, kind of, sort of’, and -*š*, a suffix used

[13] Another source of coda consonants is glottal stop insertion at the end of statements after vowel-final words: *alí* [ʔaˈliʔ], [ʔaˈli] ‘he climbed up on it’.

	monosyllable	initial	medial	final
open				
CV	*šá* ‘red’	*sá.pA* ‘black’	*tȟó.sa.pA* ‘dark blue’	*ša.šá* ‘red’ (inan.pl)
CCV	*tké* ‘heavy’	*tke.yá* ‘heavily’	*wó.tke.ya* ‘to hang things’	*tke.tké* ‘heavy’ (inan.pl)
closed				
CVC	*sáb* ‘black’ (cont)	*sab.sá.pA* ‘black’ (inan.pl)	*ǧí.sab.ye.la* ‘very dark brown’	*yu.šáb* ‘making sth./sb. dirty’ (cf. *šápA*, *šáb* ‘dirty’)
CCVC	*gléb* ‘vomiting’ (cont)	*gleb.khí.yA* ‘to make sb. vomit’	*í.ksab.ya* [ʔiksabja] ‘to be a burden for sb.’	*a.gléb* ‘vomiting on sth.’ (cont)

Table 10. Lakota syllable types.

with a number of word categories to express adversative (opposition or contrast) or emphasis. Some examples of words with these suffixes are given in 4. Since the suffixes occur word-finally and both /l/ and /š/ are possible word-final codas, these suffixes show no evidence of alternations in voicing or manner and simply support the observations we have made regarding syllable structure up to this point.

(4) Codas in the lexicon: adverbial suffixes *-kel* and *-š*

a. *-kel* ‘somewhat, rather’

apȟé/apȟékel	‘to wait’/‘kind of waiting’
ečhá/ečhákel	‘naturally of such quality’/‘by nature, naturally’
naȟmá/naȟmákel	‘to hide sth., sb.’/‘hiding in a way’
pasí/pasíkel	‘to research’/‘kind of researching’
yasú/yasúkel	‘to judge’/‘passing judgment hastily’

b. *-š* ‘adversative; emphatic’

iyé/iyéš	‘he/she/it’/‘at least him/her/it’
naké/nakéš	‘finally now’/‘now at last’
miyé/miyéš	‘I, me, it is me’/‘at least I’
waná/wanáš	‘now, already’/‘now indeed, at last’

Another set of free morphemes that appear to be consonant-final are a small class of adverbs ending in *-b*, including: *itkób* ‘in the direction toward sb. who is approaching’, *óčib* ‘by degrees, slowly, step by step, little by little’ (red *óčibčib)*, and *sakhíb* ‘together’. Other *b*-final adverbs appear to be contracted forms of words ending in *-pȟa*: *akáb* ‘extra, overflowing, on top of’ from *akápȟa* ‘on the outside of, on top’; *hakáb* ‘afterward’ from *hakápȟa* ‘to be the following’; *hútawab*, *hútab* ‘downstream’ from *hútawapȟa* ‘somewhat farther downstream’; *ȟeyáb* from *ȟeyápȟa* ‘out of the way, removed’; *isáŋm* from *isáŋpȟa* ‘further than’; *ób* ‘with them (more than one), together with them’ from *ópȟa* ‘to join in something, to be a member of something’; and *watób* ‘by boat’ from *watópȟa* ‘to travel by boat’ (< *wáta* ‘boat’ + *opȟÁ* ‘to go by way of sth.’). On this basis, we hypothesize that adverbs like *itkób* historically derive from words ending in *-pȟa*, though synchronically, there is not always evidence of a longer form.[14]

In contrast to the coda consonants just mentioned, the majority of closed syllables in Lakota are the result of a process generally referred to as ‘truncation’. Under truncation,

[14] Further work is needed to determine whether truncation of forms ending in *…pȟV*, *…khȟV*, *…tȟa* is generally possible, limited to adverbs, or lexically determined. At present, examples with *…khȟV* and *…tȟa* are limited to adverbs *anúŋkȟa*/*anúŋg* ‘on both sides’ and *tókȟa*/*tog* ‘how, how is it?’, while examples with *…pȟV* include the adverbs mentioned in the text, as well as verbs *ičhápȟA*/*ičháb* ‘to get accidentally stabbed’, *mničhópȟA*/*mničhób* ‘to wade in water’.

a word (or stem) of a specific phonological shape undergoes final vowel loss, and the consonant that is rendered in final position may alternate predictably, depending on its quality and position. For example, the word *tópa* 'four' has a truncated form *tób*, which can occur as an independent word or as the first member of a derived word, as in *tóbkiya* 'in four ways, four places' and *tóbtopa* 'by fours' (RED). Truncation regularly occurs in two kinds of word-formation processes: prefixal reduplication, where the prefixed element can be viewed as a truncated base, and compounding, where the first element in the compound is a truncated base. With other word-formation processes, as detailed in Ullrich 2018, morphosyntactic properties may determine whether truncation takes place. Another important finding of Ullrich (2018) is that truncated forms can function as independent phonological words in various syntactic constructions. As a consequence, truncation may give rise to word-medial or word-final coda consonants. In the remainder of this section, we focus on the phonology associated with regular truncation, since truncation is the primary source of coda consonants in Lakota. We refer to morphological environments for truncation simply as 'complex words', including in this category nominal compounds and prefixal reduplication, as well as a host of derived verb forms.

Our understanding of truncation has four phonological components. First, the phonological conditioning of vowel loss (5a); second, the alternations in voicing, discussed earlier, that result when a consonant is in coda position (5b);[15] third, a dissimilatory process that applies to sequences of coronal consonants derived by reduplication (5c); and finally, an optional resyllabification of consonants into complex onsets that can result in derived ejectives, aspirates, or onset clusters (5d). This last process is supported by two facts: (i) the only consonant clusters that appear to show optional resyllabification are clusters that are allowed word-initially; (ii) the devoicing patterns can only be explained by resyllabification since a medial stop coda in V**C**.CV is typically voiced, as we show in §3.

(5) Understanding truncation as prosodic morphology
  a. TRUNCATION: If a Lakota form ends in /...VC_fV_f/, where C_f is a possible coda consonant, then:
    i. …VC_fV_f → …VC_f when it is the first member of a complex word.
    ii. VC_fV_f → …VC_f in isolation, provided that V_f is unstressed (optional).
  b. CODA VOICING CONSTRAINTS: In syllable coda position:
    i. Fricatives devoice: ǧ → ȟ, ž → š, z → s.
    ii. Oral stops and affricates voice: p → b, t → l, k → g, č → l.
  c. DISSIMILATION (in reduplication only/morphophonemic): Heterosyllabic lateral + coronal consonant clusters dissimilate:
    i. l.T → g.T, where T is a coronal consonant. (See §4 for further discussion.)
  d. OPTIONAL RESYLLABIFICATION (fast speech, variable): In $VC_1.C_2V$ where C_1C_2 is a possible syllable onset:
    i. $VC_1.C_2V$ → $V.C_1C_2V$ (with regressive devoicing as per Tables 8, 9).

The patterns described in 5 are illustrated in Table 11 with relevant reduplicated and compound forms. In this table 'n.a.' means 'not applicable', '—' indicates a predicted

[15] Truncation, voicing, and regressive and phrase-final devoicing (see §3.1) result in voiceless oral stop/nasal stop variants when the nuclear vowel is nasalized, since a voiced stop after a nasalized vowel is realized as a (voiced) nasal stop. For example, truncated variants of *núŋpa* 'two' are *núŋp* [nõp] and *núŋm* [nõm] < [nõb].

but unattested form, and '(?)' indicates a resyllabified form that may be indistinguishable from the original complex word, since no regressive voicing assimilation distinguishes the complex onset from the coda-onset CC cluster. Notice that, in word-medial position, triconsonantal clusters like /šmn/ and /ȟsn/ are found. Since only a single consonant is allowed in the coda, all medial CCC clusters must be syllabified as C.CC, with a simple coda followed by a complex onset.

BASE	5a.ii SIMPLE WD. TRUNCATION	5a.i COMPLEX WD. TRUNCATION	5b CODA VOICING	5c CLUSTER DISS.	5d RESYLLAB. (OPTIONAL)	2ND BASE OF COMPLEX WD.
okáspA 'to sink'	n.a. (…CCV#)	n.a. (…CCV#)	n.a.	n.a.	n.a.	n.a.
tȟağé 'saliva'	n.a. (…´V#)	*tȟaȟ.glá.tȟA* 'to chew cud'	ğ → ȟ	n.a.	n.a.	*gla.tȟÁ* 'to chew one's own'
šaké 'nail, claw'	n.a. (…´V#)	*šag.tȟúŋ* 'to have claws'	k → g	n.a.	*ša.ktȟúŋ*	*tȟúŋ* 'to have sth.'
sápA 'black'	*sáb*	*sab.sá.pA* 'black' (INAN.PL)	p → b	n.a.	*sa.psá.pA*	RED
čhápa 'beaver'	*čháb*	*čhab.síŋ.te* 'beaver tail'	p → b	n.a.	*čha.psíŋ.te*	*siŋté* 'tail'
ȟótA 'gray'	*ȟól*	*ȟol.ȟó.tA* 'gray' (INAN.PL)	t → l	n.a.	n.a.	RED
šókA 'thick'	*šóg*	*šog.šó.kA* 'thick' (INAN.PL)	k → g	n.a.	*šo.kšó.kA*	RED
tȟóka 'enemy, alien'	*tȟóg* 'different, foreign'	*tȟog.'í.yA* 'to speak a foreign language'	k → g	n.a.	*tȟo.k'í.yA*	*iyÁ* 'to speak'
lúta 'red, scarlet'	—	*lul.yÁ* 'to dye sth. red'	t → l	n.a.	n.a.	*-ya* CAUS
		lug.lú.ta 'red' (INAN.PL)	t → l	ll → gl	*lu.glú.ta* (?)	RED
šéča 'dry'	*šél*	*šeg.šé.ča* 'dry' (INAN.PL)	č → l	lš → gš	*še.kšé.ča*	RED
pȟéta 'fire'	*pȟél*	*pȟel.čhó.la* 'without a fire'	t → l	n.a.	n.a.	*čhóla* 'without'
wáŋčala 'only once'	—	*wáŋčagčana* 'only once each time'	n.a.	lč → gč	*wáŋ.ča.kčana*	RED
léžA 'to urinate'	*léš*	*leš.lé.žA* 'to urinate often'	ž → š	n.a.	*le.šlé.žA* (?)	RED
		leš.mná 'smell of urine'	ž → š	n.a.	n.a.	*mná* 'to smell'
máza 'metal'	—	*mas.kȟó.ka* 'can'	z → s	n.a.	n.a.	*kȟoká* 'keg'
čhážga 'ice'	*čháȟ*	*čhaȟ.sní.yaŋ* 'ice cream'	ğ → ȟ	n.a.	n.a.	*sniyÁŋ* 'to cool sth. off'
tȟaló 'meat'	n.a.	*tȟal.'á.gna.ke* 'rigid goldenrod'	n.a.	n.a.	n.a.	*agnákA* 'to lay sth. on'

TABLE 11. Illustrating the phonology of truncation.

Due to truncation in reduplication and compounding, triconsonantal sequences are not uncommon, and they support our view that voiced [b], [l], and [g] are codas, since word-initial clusters are limited to two consonants (see Table 8 and n. 13): *ağúya**b**-**sk**uyela* 'Danish, cake' from *ağúyapi* 'bread' + *skúyela* 'sweet'; *wetȟá**bsk**ala* 'white blood cells' from *wé* 'blood' + *tȟápa* 'ball' + *ská-la* 'white'; *tȟa**bšk**átA*, *tȟa**bšk**ál* 'playing basketball' from *tȟápa* 'ball' + *škátA* 'to play'; *pȟe**lmn**á* 'to smell of fire' from < *pȟéta* 'fire' + *mná* 'to smell of sth.'; *waší**lȟp**aya* 'garbage, trash' from *wa-* indef.obj + *šíčA* 'bad' + *ȟpáyA* 'to lie, be lying'.

It is worth stressing here that the phonological properties detailed in 5 for truncation are regular and, as far as we can tell, productive. New compounds—like those in Table 11 for *čhaȟ.sní.yaŋ* 'ice cream', *ağúyabskuyela* 'Danish (pastry)', and *wetȟábskala* 'white blood cells'—follow the same patterns as arguably older compounds that refer to indigenous culture items like *tȟal'ágnake* 'rigid goldenrod' (a plant used to lay meat on) and *čhabsíŋte* 'beaver tail' (used to comb hair). A further argument for the productivity of final obstruent voicing is that it occurs not only in the lexical truncation processes described here, but also in postlexical vowel dropping described as a feature of rapid speech by Rood and Taylor (1996:447): 'Also characteristic of rapid speech is the dropping of unstressed word-final vowels … In these examples, note that *p* and *k* are voiced to *b* and *g* when they come to stand before a consonant'.

Clearly Lakota is not a language like Czech or Basque in which all obstruents are devoiced in the coda. Devoicing is found for the fricatives /z/, /ž/, and /ğ/ (5b), but the oral stops /p/ and /k/ become voiced [b] and [g], respectively, while /t/ and the affricate /č/ are both pronounced as [l]. Since the existence of phonological coda obstruent voicing is debated, §3 of this article provides acoustic evidence for the patterns we have just described. More specifically, we demonstrate that the sounds transcribed as [b] and [g] are often voiced, that they have the closure duration, burst properties, and low energy values of oral stops, and that where they are voiced, their voicing is best viewed as a consequence of the voicing of coda stops. Before turning to this evidence, we briefly review previous analyses of Lakota voicing patterns and highlight details that distinguish our approach from previous ones.

2.5. Previous analyses of Lakota voicing patterns. The analysis presented above agrees in most respects with that presented by Rood and Taylor (1985, 1996). They describe /b/ as a marginal phoneme, and discuss [b] and [g] as positional variants of /p/ and /k/, respectively, when final vowels are dropped or words are reduplicated. In particular, they say that '[w]hen vowel dropping (of any origin except possibly the fast speech phenomena …) places /p t č k/ in word-final position or at an internal boundary between linguistic elements, these become [b], [l], [l], [g], respectively' (Rood & Taylor 1996:449).[16]

In contrast, Rankin (2001) and Rood (2016) take a very different view of the oral stop voicing process from historical and theoretical perspectives, respectively. Under both accounts, oral stops are lenited to sonorants in syllable-final position. Rankin (2001:5) suggests a sound law whereby syllable-final stops become sonorants by first becoming nasals, and then shifting to oral stops after oral vowels. Rood (2016) instead uses a

[16] Rood and Taylor (1996:449) continue: 'When a nasalized vowel precedes these sounds, they may further shift to a nasal consonant'. Recall that we are restricting our attention to oral syllables in this study, partly for reasons of time, but also because our preliminary findings are that stops are sometimes only partially nasalized in this context, and measuring their acoustic properties is more complex than examining those following oral vowels.

theoretical device, the feature [sonorant voice] (Rice 1993), which is assigned to oral stops in coda position. Rood's argument that voiced stops [b] and [g] are sonorants seems to have four parts. First, since /t/ and /č/ are realized as [l], a sonorant, in the coda, [b] and [g] should be sonorants too. Second, since [b] alternates with [m] in Lakota in nasal contexts, it should be a sonorant. Third, consonant lenition is common in coda position crosslinguistically, so '[s]ince our target sounds are in coda position, we should therefore look for a way to declare their voicing to be lenition' (Rood 2016:246). Finally, Rood claims that, like sonorants in many of the world's languages, voicing of [b] and [g] (but not [l]) is variable. Evidence for this variability 'comes from the informal observation that [b] and [g] in coda position often seem to match the voice phonetics of the following consonant' (Rood 2016:249). Since Rood treats [b] and [g] as sonorants, not obstruents, Lakota does not violate the universal markedness constraints discussed in §1 that give rise to common obstruent devoicing and prohibit obstruent voicing in the coda.

The most important difference between our account and all of the previous studies of Lakota voicing we are aware of, including those just mentioned, is that we provide acoustic evidence in §3 for the impressionistic descriptions of a range of voicing patterns in the language, including voicing of oral stops in coda position.

3. Phonetic analysis of lakota stop voicing patterns. Despite its endangered status, Lakota is well documented in comparison to most indigenous languages of North America. This is especially true where audio recordings are concerned. The third author of this article has made recordings of over 400 native speakers between 1992–2018, including hundreds of hours of narratives and dialogues, and has also collected recordings of several dozen speakers from other sources. Given that there are only about 2,000 speakers today, this corpus may represent the speech of 10–20% of the Lakota speech community. While we have listened to some of these recordings and analyzed voicing patterns in running speech from a handful of them, the central study of voicing in this article is based not on recordings of natural running speech, but on studio recordings of native speakers that form part of the database of the *NLD*, published by the Lakota Language Consortium (Ullrich 2011). Before saying more about this dictionary, a general comment is in order. Language documentation can take many forms, but a general piece of advice to those working on endangered languages is to create documentation that can be used for multiple purposes. Even if one has no interest in phonetics or phonology, creating high-quality audio recordings allows future scholars to do research in these areas. In this context, the *NLD*, a descriptive lexicographic work with multiple purposes, including language documentation, language pedagogy, and language revitalization, is exemplary. The high-quality recordings of the *NLD* are of great value to the scientific community, and without them, the current research would not be possible.

The *NLD* currently offers on its smartphone app approximately 52,000 sound files from eight native speakers representing 28,000 dictionary headwords, with about 95% of the audio files found in pairs, spoken by the same male and female native speakers. Though the audio files in the dictionary application are compressed OGG files, not ideal for acoustic analysis, the Lakota Language Consortium was extremely kind to offer us access to a subset of the original uncompressed audio WAV files. These recordings, made in a professional sound studio with control room constructed specifically for the Lakota Dictionary project, are of very high quality. The detailed phonetic analysis of Lakota stop voicing presented in this section is based entirely on these recordings. For almost every word examined, there are two tokens: one spoken by Ben Black Bear,

Jr. from the Rosebud Reservation, indicated by (M) after the token, and one spoken by Iris Eagle Chasing from the Cheyenne River Reservation, indicated by (F) after the token. Many people consider these two speakers to be the most competent and literate native speakers of Lakota.

The *NLD* recordings were made by prompting speakers with words as they are written in the dictionary. A possible complication for this study is that pronunciation of voiced vs. voiceless stops in the coda could have been influenced by spelling. While this possibility cannot be ruled out, two observations suggest that pronunciations of the two speakers are natural and not influenced by spelling. First, the patterns we discuss below do not always follow spelling conventions: for example, in *sabsápa* <b> is typically pronounced as voiceless though it is spelled with the symbol for a voiced stop (an alternative spelling is *sapsápa*). Second, both speakers have high levels of phonological awareness and commented when spellings did not fit their phonological intuitions. Both speakers were encouraged to say words in the way that was most natural to them. After recording sessions, if a speaker was uncomfortable with an audio file because it did not sound natural or did not feel right to them, that audio file was deleted.

Another issue that could influence pronunciation is that words recorded for the dictionary were spoken at a relatively slow rate, clearly and in isolation or as part of two-word phrases. Our results must be interpreted, then, as results relating to the phonology of clear speech.

The corpus compiled specifically for this study has a total of 611 words: 304 distinct words with two tokens each, spoken by the male and female speakers, plus two distinct words spoken only by the male speaker, plus one distinct word spoken only by the female speaker. From these 611 words, a database of oral stops was created including: 631 voiceless stops (<p> = 150, <t> = 196, <k> = 285); 584 voiced stops (<b> = 285, <g> = 299); and 14 ejectives (<p’> = 6, <t’> = 4, <k’> = 4). In addition, we included 111 instances of glottal stop, as we were particularly interested in the realization of coda stops before a glottal stop (see below). Of the 1,215 oral voiceless and voiced stops (excluding ejectives and glottal stops), 225 intervocalic tokens (between oral vowels) were used to establish voicing categories (see below). Stops in contact with a nasalized vowel were discarded from the analysis to avoid effects related to nasalization. The remaining 841 oral stops that were not intervocalic were subject to analysis based on the established voicing categories. Given our initial hypotheses that final coda voicing might be masked by phrase-final devoicing and possible assimilation to following voiceless obstruents, an attempt was made to include tokens where this masking effect might be absent, including word-final stops that were not phrase-final and syllable-final stops followed by glottal stop. All words were orthographically transcribed following the *NLD* spelling conventions. The transcriptions were then converted to SAMPA for automatic segmentation of the audio files using the WebMAUS application (Kisler et al. 2017) set for ‘Language independent (SAMPA)’, and subsequently hand-corrected as needed.

Given our central interest in determining whether Lakota shows final obstruent voicing of /p/ and /k/, the first part of the phonetic study, summarized in §3.1, was focused on the question of whether /p/ and /k/ show phonetic voicing word-finally and, more generally, in syllable-final position. With positive evidence of syllable-final voicing, we turned our attention to the question of whether these voiced segments still had properties of oral stops. This was necessary in order to rule out interpretations of voicing as a secondary feature of lenition, where voiced codas might be interpreted as fricatives or glides. In §3.2 we present evidence that coda [b] and [g] have acoustic properties of oral stops, includ-

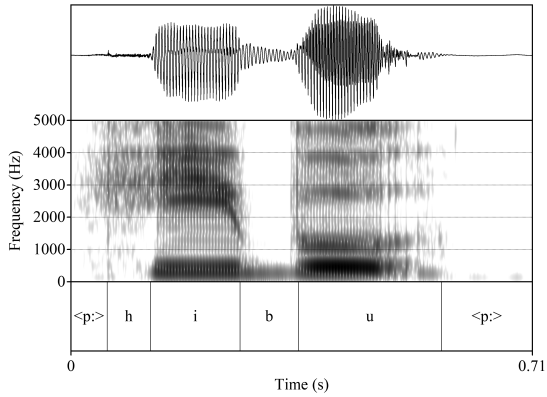
ing significant closure durations, absence of fricative noise, release bursts, and low energy levels typical of oral stops. Together, acoustic evidence for voiced [b] and [g] in the coda and acoustic evidence for oral stop production of these segments support the view that Lakota has a sound pattern of oral stop coda voicing for /p/ and /k/.

3.1. Evidence for voicing in oral stops: an analysis of auto-correlation coefficients. The corpus for this study was 876 oral stops extracted from the 611 word files from the original *NLD* audio WAV recordings. The central goal was to determine if there is a bimodal distribution of voicing in the data (voiceless vs. voiced), as described in phonological analyses, and, if so, to determine the distribution of each category.

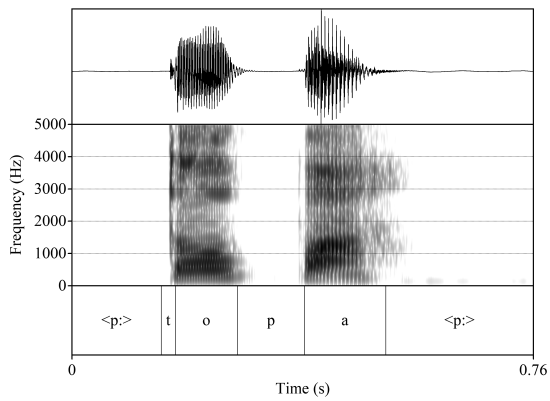
For each segment in the corpus, the auto-correlation (AC) peaks were calculated with the program EMU (Harrington 2010 for the legacy version, Winkelmann 2017 for the R package 'emuR'), using the ESPS method with a frame spacing of 10 ms, a window length of 7.5 ms, and pitch ranges of 60–400 Hz for the male speaker and 90–600 Hz for the female speaker. This yielded a total of 6,921 measurements with a voicing coefficient between 0 and 1 at each time point, 0 standing for no correlation (voiceless) and 1 for perfect correlation (voiced). Then, the median value of the AC coefficients of all measured time points within a given stop was calculated for all of the stops in the corpus.[17]

The statistical analysis was carried out in R (R Core Team 2019). We ran a binomial linear mixed-effects model that used the AC coefficients (continuous variable: probability of voicing) to predict the voicing label (binary categorical variable: voiceless or voiced) with the R function glmer() (from the R package 'lme4'; Bates et al. 2015). As random effects, we had intercepts for speakers. The intervocalic context was the only clear phonological context where the voiced vs. voiceless opposition appeared to be secure. Given this, we trained the binomial linear mixed-effects model on intervocalic stops (225 tokens). Figures 1a,b provide waveforms and spectrograms illustrating the intervocalic voiced vs. voiceless contrast for /b/ vs. /p/ (here and throughout, <p:> in the transcription line indicates pause). An AUROC test ('area under the receiver operating characteristic' curve; see Fawcett 2006) of the model predictions performed with the function auc() (R package 'pROC'; Robin et al. 2011) resulted in 0.8871 (0.5 being the chance threshold and 1 being the point of perfect prediction). A model with random intercepts for speakers and words showed a lower AUROC coefficient (0.8655) than the model with random effects only for 'speaker', and it was thus disregarded. The model showed no singularities or any other obvious deviations from the model assumptions. Based on the voicing values of intervocalic stops, we predicted response values between 0 and 1 (0 being fully voiceless and 1 fully voiced, with the categories having been divided by 0.5) for the full data set grouped by the phonological context of each of the analyzed stops.

[17] Regarding the preference for the metrics we selected (AC coefficients) over others (percentage of voicing into closure or absolute voicing duration), we chose to perform measurements of AC coefficients during the whole stop closure with the understanding that they provide information that is comparable to that provided by percentage of voicing into closure, with the additional benefit of the reduction of the degrees of freedom of the researcher, thus reducing the risk of a false positive. We selected the median over the mean, in part to avoid skewed results due to potential artifacts or subpar segment alignments. The results of this method should not differ greatly from those of methodologies in other works on voicing in endangered languages (e.g. Coetzee & Pretorius 2010), where closure duration and voicing into closure are measured (often by hand, which might yield complications regarding the marking of boundaries), and then the percentage of voicing into closure is calculated. Segments with a voicing-over-closure percentage of over 50% are considered voiced, and those below 50%, voiceless. In our study, the median of all measurements of a given stop (performed at each 10 ms) would mark 50% of the distribution: if that point is over 0.5, then the segment is considered voiced, while segments below 0.5 are considered voiceless.



a. *hibú* (F) ‘I am coming’.



b. *tópa* (F) ‘four’.

FIGURE 1. Lakota contrast between /b/ and /p/ in intervocalic position.

In addition to the intervocalic stops used to train the model, the analysis of phonetic voicing of Lakota stops was applied to three different general phonological contexts: word-initial stops, word-internal morpheme-final stops that constitute word-internal codas, and word-final stops.

Lakota word-initial stops /p/ vs. /b/ are considered to have a voicing contrast in the literature, and thus we analyzed word-initial voiced and voiceless stops separately by using annotation labels that follow the *NLD* transcription of each word. A Hartigans’ dip test (Hartigan & Hartigan 1985) performed on word-initial stops transcribed as voiceless (102 tokens in our data set) suggested a unimodal distribution (p-value = 0.8041), with results summarized in Figure 2a. However, applying the same test to word-initial stops that are written as <b> or <g> in the dictionary strongly suggested a bimodal distribution (p-value $< 2.2\times10^{-16}$).[18] Figure 2b shows word-initial stops that are written as voiced <b> or <g> in the dictionary: while some of these are clearly voiced, most were produced without voicing, according to our model. Since all word-initial stops in this corpus were

[18] Initial [b] and [g] are rare in Lakota. There were only thirty-eight tokens in our database, including words with initial <gl> clusters, where voicing of [g] is predictable and allophonic, as described in §2.

taken from words spoken in isolation, we interpret this result as evidence for optional devoicing of phrase-initial stops.

In contrast to word-initial stops, word-medial morpheme-final stops in Lakota function as codas and are thought to have predictable voicing, though there is debate as to if and where final obstruent voicing occurs (see §2). For the word-medial morpheme-final stops in our corpus that are analyzed as codas (361 tokens), Hartigans' dip test suggests a bimodal distribution ($p < 2.2\times10^{-16}$), as shown by the histogram in Figure 2c. However, dividing this single context into three subcontexts based on the nature of the following segment yields different distributions.

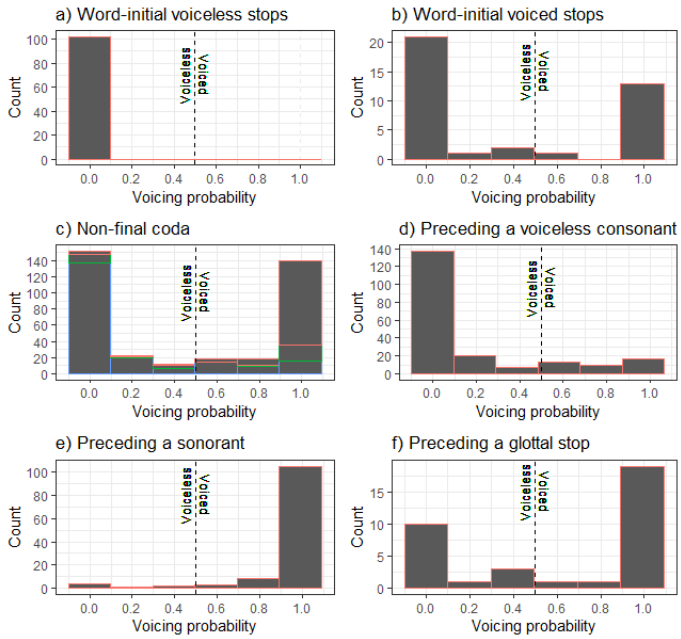
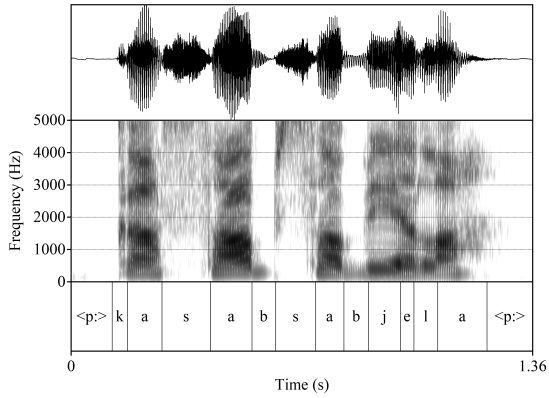


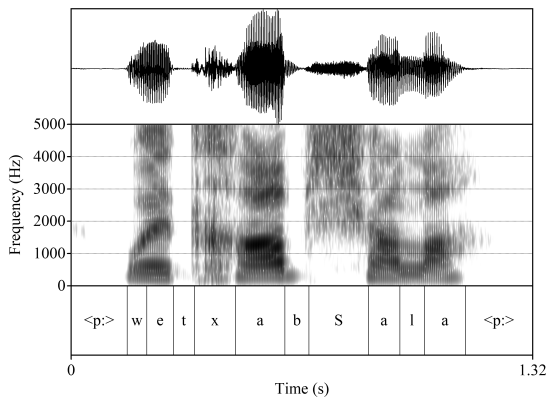
FIGURE 2. (a)–(f) Probability of stop voicing in different phonological contexts.

Recall from §2 that morpheme-final stops preceding [p, t, k, tʃ, s, ʃ, χ, h] can be resyllabified as complex onsets, in which case the entire cluster is expected to be voiceless (see example 5d). Figure 2d shows hypothesized resyllabified morpheme-final stops (203 tokens; $p = 0.3001$) with values close to 0, supporting the analysis of morpheme-final coda stops as typically voiceless in this context. At the same time, over forty tokens maintain voicing, suggesting that, although regressive devoicing (with resyllabification) is the norm in this context, it is not obligatory. (An alternative analysis without resyllabification is also possible: regressive devoicing optionally yields voiceless clusters in these contexts.) To highlight the possibility of a voiced coda preceding a voiceless consonant [p, t, k, tʃ, s, ʃ, χ, h], we offer the spectrograms in Figure 3.

Complementary contexts are those where morpheme-final stops in coda position precede [l, m, n, j, w] or glottal stop. Let us first consider codas followed by one of the voiced sonorants [l, m, n, j, w]. Under our analysis, these codas undergo coda stop voicing (5b.ii) and are expected to be voiced, since there is no contextual devoicing that occurs in this context. Figure 2e shows morpheme-final coda stops before sonorants (123 tokens; $p = 0.9697$), with values close to 1 and unimodal distribution, supporting an analysis of final stops as voiced in this context. Spectrograms in Figure 4 offer exam-



a. [b.s] in *kasábsabyela* (F) ‘(repeatedly falling) heavily or flat’.



b. [b.ʃ] in *wetx̌ábšala* (F) ‘red blood cells’.

FIGURE 3. Voiced coda stop preceding voiceless obstruent.

ples of this sound pattern. Though regressive voicing from following [l, m, n, j, w] might be suggested, the results summarized in Figure 2f above and Figure 2g below provide further support for a general process of coda stop voicing.

Figure 2f shows morpheme-final coda stops before glottal stop (thirty-five tokens; $p = 8.266\times10^{-5}$), suggesting bimodal distribution. While the majority of tokens are fully voiced, there are ten fully voiceless tokens, which we interpret as instances of the optional fusion/resyllabification of [b.ʔ], [g.ʔ] to onset [p’], [k’], respectively, included under 5d, and illustrated in Table 11 by variants *tx̌og.’í.yA*, *tx̌o.k’í.yA* ‘to speak a foreign language’. To illustrate the sequence of voiced stop in the coda followed by glottal stop, we offer the spectrograms in Figure 5. Since glottal stop is voiceless, the voicing of oral stops in the coda preceding glottal stop must have another source. We argue that the source is coda stop voicing (5b.ii).

The results tabulated in Figs. 2c–f support our analysis in §2. Oral stops /p/ and /k/ are voiced in the syllable coda: when followed by a voiceless obstruent or /h/, they are usually devoiced and resyllabified, forming complex onsets; when followed by glottal stop, optional fusion may give rise to voiceless ejectives.

Further support for the process of obstruent voicing in the coda is found in the last distributional category of word-final (and phrase-final) oral stops. The dip test performed on word-final stops (150 tokens) suggested a bimodal distribution ($p < 2.2\times10^{-16}$). Fur-

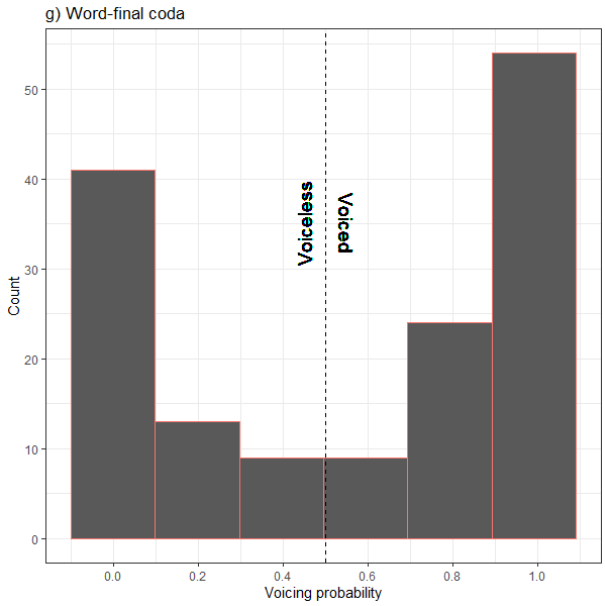
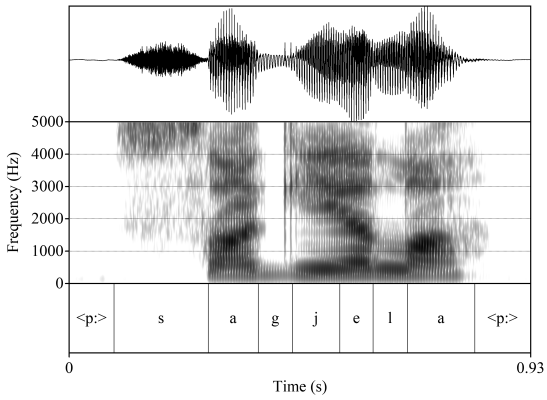
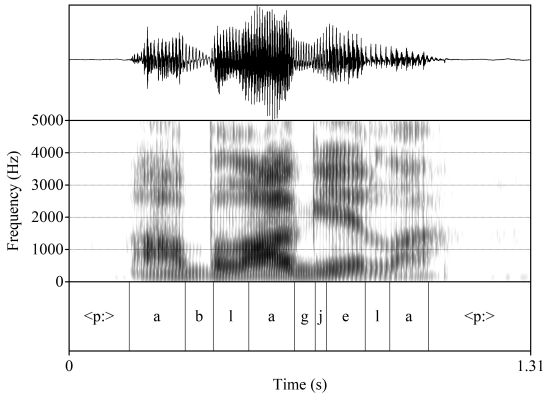


FIGURE 2. (g) Probability of stop voicing in word-final coda.

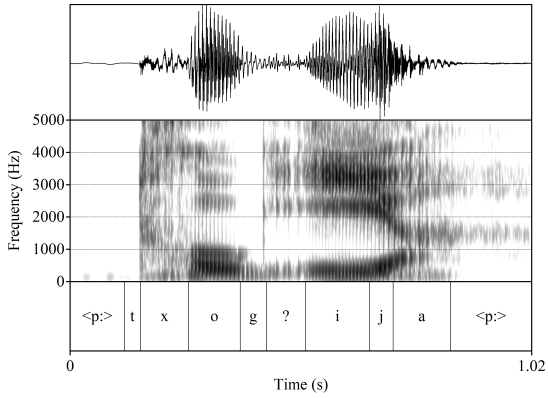


a. [g.j] in *sagyéla* (F) 'in a dried, hard, or stiff condition'.

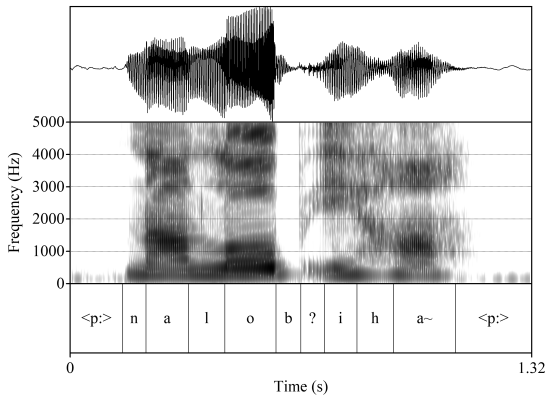


b. [b.l] and [g.j] in *ablágyela* (M) 'quietly, peacefully'.

FIGURE 4. Voiced coda stop preceding sonorant.



a. [g.ʔ] in *tȟog'íyA* (M) 'to speak a foreign language'.

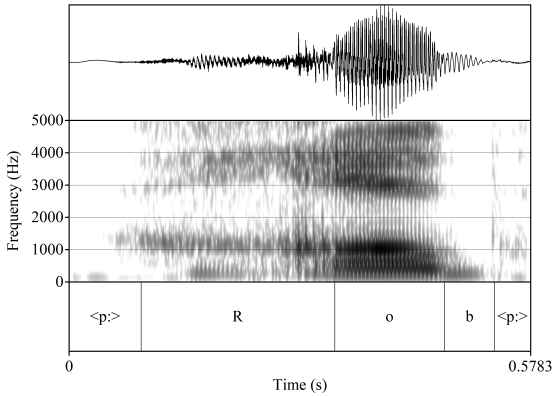


b. [b.ʔ] in *nalób ihÁŋ* (F) 'to step into sth. muddy or miry'.

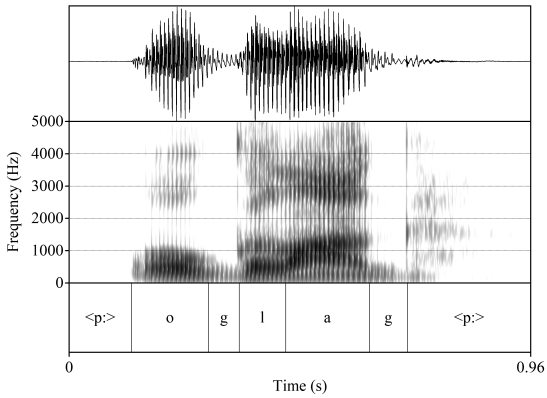
FIGURE 5. Voiced coda stop preceding (voiceless) glottal stop.

ther, Fig. 2g shows that, among these, there are more stops categorized as voiced than as voiceless. Since all stops in word-final position are analyzed as phonological instances of /p/ or /k/, voicing in this context can be interpreted as a consequence of coda voicing (5b.ii). Another interesting aspect of word-final stops visible in Fig. 2g is that there is a greater proportion of tokens in the intermediate prediction values. Given that all of our word-final tokens were also phrase-final, we interpret this as evidence of a gradient phrase-final devoicing process. In phrase-medial position, where a word-final stop is followed by something other than a voiceless obstruent, it is voiced, as predicted by 5b.ii. To highlight the common occurrence of voiced oral stops word-finally, we offer the spectrograms in Figure 6. Further discussion of the acoustic properties of coda [b] and [g] is offered in §3.2.

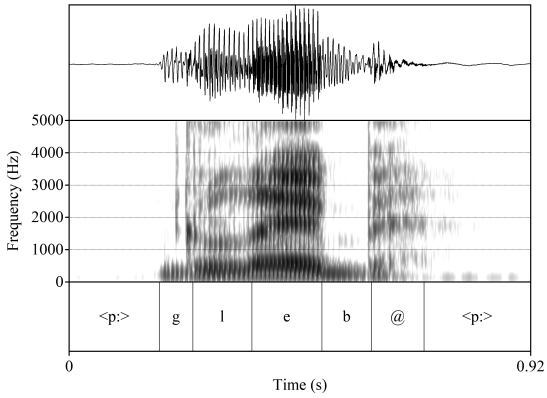
Overall, our acoustic analysis of voicing in Lakota stops suggests that stop voicing is contrastive intervocalically and word-initially, but is noncontrastive elsewhere. The distribution of phonetic voicing in Lakota oral stops supports an analysis of phonological syllable-final voicing of /p/ and /k/ to [b] and [g], respectively (5b.ii). In addition to coda voicing, four distinct devoicing processes are observable in the data. First, in phrase-initial position, there is often devoicing of (underlying) voiced stops (Fig. 2b). Second, in phrase-final position, coda voicing can be obscured by gradient phrase-final devoicing (Fig. 2g). Third, in phrase-medial position, a regular process of regressive



a. [b] in *ǧób* (F) ‘snoring’.



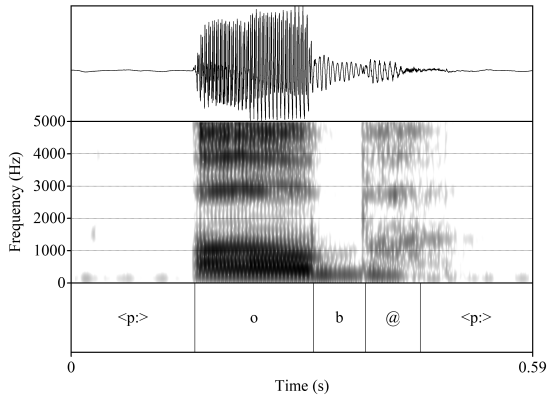
b. [g] in *oglág* (M) ‘telling one’s own, relating’.



c. [b] in *gléb* (M) ‘vomiting’.

(Figure 6. *Continues*)

obstruent devoicing is triggered by the class of voiceless sounds [p, t, k, tʃ, s, ʃ, χ, h], which, under our analysis, can be related to resyllabification (see example 5d). A final optional process that can yield voicelessness is fusion/resyllabification of [b.ʔ], [g.ʔ] to onset ejectives [p’], [k’], respectively (5d). While all of these processes (with the possi-



d. [b] in *ób* (F) 'with them, together with them'.

FIGURE 6. Word-final voiced stops.

ble exception of phrase-initial devoicing) have been described in the literature, this is the first study presenting acoustic evidence in support of syllable-final oral stop voicing together with these devoicing processes.

3.2. EVIDENCE THAT [b] AND [g] ARE ORAL STOPS, NOT FRICATIVES OR GLIDES. The general question of whether processes of final obstruent voicing exist is complicated by the existence of many sound patterns of final lenition or weakening where voicing is coupled with reduced stricture. For example, though Blevins (2006a) suggests that Tundra Nenets may show a pattern of final obstruent voicing, Kiparsky (2006:228–29) argues that Tundra Nenets /p, t, k/ vs. /b, d/ should be treated as a 'contrast of tenseness, not of voicing', noting that '/b/ and /d/ are lax, and articulated with various kinds of lenition'. In order to determine whether Lakota has a sound pattern of coda voicing for /p/ and /k/, then, it is necessary to show not only that these sounds are voiced, as we have in §3.1, but also that the voiced segments are oral stops. Here we present acoustic evidence of significant closure duration, absence of fricative noise, release bursts, and low energy levels that are all characteristic of oral stops in contrast to fricatives, taps/flaps, and glides.

Where hand-eye inspection of waveforms and spectrograms is mentioned, these waveforms and spectrograms are extracted from the *NLD* data. All waveforms, spectrograms, and annotations were plotted with the computer program Praat (Boersma & Weenink 2019) and were exported as 600 dpi PNG files.

CLOSURE DURATION. As detailed in §2, oral stops and oral fricatives contrast in Lakota, but due to coda stop voicing and coda fricative devoicing, the only environment where voiced and voiceless stops and fricatives contrast, and where accurate measures of duration can be made, is intervocalic position. Figure 7 shows durational measurements for all of the obstruents in our database in intervocalic position. Note that the horizontal line within each box marks the median, while the circle marks the mean. The durational measurements have been speaker-normalized by converting them to *z*-scores and back to milliseconds for easier comparison. Three outliers (> 3 *SD*s) have been removed under the assumption that they originated from hesitation or similar effects, resulting in a total of 348 intervocalic obstruents analyzed.

Voiced stops have the shortest closure durations of all obstruents, with averages ranging from 62 ms for [g] to 77 ms for [b]. Plain voiceless stops show longer closure durations, averaging 110–113 ms, with voiced fricatives slightly longer (99–120 ms), and

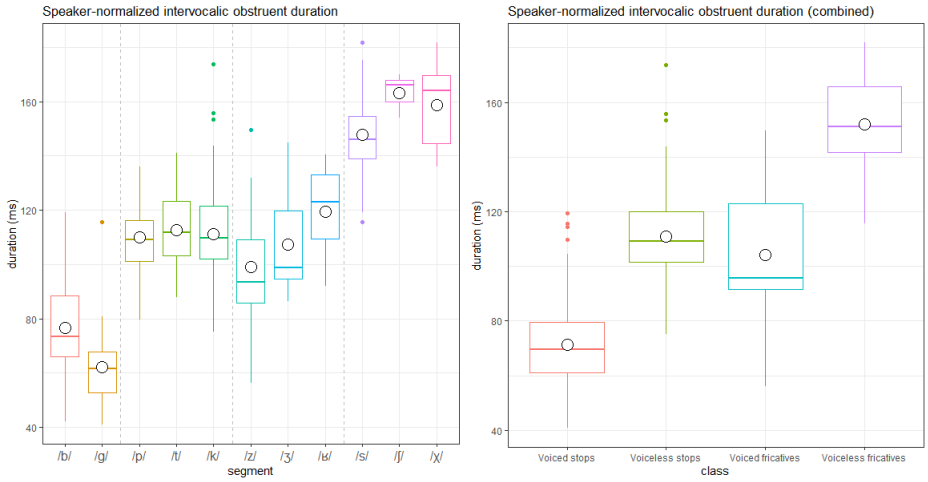


FIGURE 7. Speaker-normalized duration of intervocalic obstruents.

voiceless fricatives (excluding /h/) the longest of all (148–163 ms). These durational measurements are unremarkable and are consistent with interpreting sounds transcribed as [b] and [g] as voiced oral stops, since flaps or taps would be expected to show shorter closure durations averaging around 20 ms. We performed a linear mixed-effects analysis of the duration measurements of intervocalic consonants, with random intercepts by Speaker and Word, using the nonnormalized values to avoid singularity. A Tukey post-hoc comparison with Benjamini-Hochberg adjustment for multiple testing showed significant differences between [b] and [g] and any other given consonant at the 0.001 level, as shown by Table 12.

LINEAR HYPOTHESIS	EST	*SE*	*z*-VALUE	$Pr(>\|z\|)$
g − b == 0	−13.6356	3.9749	−3.430	< 0.001
k − b == 0	35.0904	3.0892	11.359	< 0.001
p − b == 0	32.5055	3.3951	9.574	< 0.001
R − b == 0	40.5085	9.1923	4.407	< 0.001
s − b == 0	72.6519	4.2175	17.226	< 0.001
S − b == 0	87.2562	10.5256	8.290	< 0.001
t − b == 0	35.7116	4.2953	8.314	< 0.001
x − b == 0	83.1465	5.7934	14.352	< 0.001
z − b == 0	24.7052	4.6610	5.300	< 0.001
Z − b == 0	31.0778	5.2792	5.887	< 0.001
k − g == 0	48.7260	3.6550	13.331	< 0.001
p − g == 0	46.1411	3.8593	11.956	< 0.001
R − g == 0	54.1441	9.3850	5.769	< 0.001
s − g == 0	86.2875	4.6902	18.398	< 0.001
S − g == 0	100.8918	10.6962	9.432	< 0.001
t − g == 0	49.3471	4.6953	10.510	< 0.001
x − g == 0	96.7821	6.0956	15.877	< 0.001
z − g == 0	38.3407	5.0317	7.620	< 0.001
Z − g == 0	44.7133	5.6074	7.974	< 0.001

TABLE 12. Comparison of the duration of different obstruents against [b] and [g].

However, since the measurements in Fig. 7 are taken only from intervocalic [b] and [g], checks are necessary to ensure that the same segments in syllable coda show similar closure durations. Since we could not automate this process due to the phonotactics,

we resorted to hand-eye-checking spectrograms of fully voiced tokens from data like that presented in Fig. 2e, Fig. 2f, and Fig. 2g. Waveforms and spectrograms in Figs. 4–6 exemplify our findings: voiced coda consonants transcribed as [b] and [g] in the *NLD* orthography that were measured as fully voiced by the acoustic analysis in §3.1 show closure durations falling within the central range of the distribution bars for voiced stops in Fig. 7. There is no evidence of significant temporal reduction of articulatory undershoot that might be associated with flapping/tapping or general lenition.

Absence of fricative noise during closure phase of voiced oral stops. To determine whether voiced codas were produced as stops vs. fricatives or glides, we performed two tasks. One was an automated measure of spectral energy, described in §3.2. The other was hand-eye-checking of spectrograms of fully voiced tokens from data presented in Figs. 2e,f and Fig. 2g, looking for noise above the voicing bar that would indicate incomplete stop closure. Spectrograms in Figs. 4–6 exemplify our findings: voiced coda consonants transcribed as [b] and [g] in the *NLD* orthography that were measured as fully voiced by the acoustic analysis in §3.1 show silent closure durations consistent with oral stop production. These stops do not show aperiodic components in the spectrum during the closure phase that would indicate frication, nor do they show formant structure indicative of vowel-like productions. There is no evidence of significant noise at frequencies above the fundamental that might be associated with incomplete oral closure, frication, or general lenition.

Presence of release bursts for voiced oral stops. A salient characteristic of (released) oral stops is their release bursts. Release bursts are not generally found in flaps/taps and are absent in fricatives and glides. To check for release bursts, we resorted to hand-eye-checking of spectrograms of fully voiced tokens from data presented in Figs. 2e,f and Fig. 2g, looking for a spike of noise at the end of the closure that would indicate release of intraoral air pressure. Spectrograms in Fig. 4 and Fig. 6 exemplify positive findings: voiced coda consonants transcribed as [b] and [g] in the *NLD* orthography that were measured as fully voiced by the acoustic analysis in §3.1 sometimes show visible release bursts before sonorants and in word-final position before pause. Note that in Fig. 4a, the [g] of *sagyéla* (F) ‘in a dried, hard, or stiff condition’ shows a double burst: this is not uncommon before sonorants, where preceding voiced stops are often described as being followed by a short open transition. Presence of release bursts in medial and final codas were taken as positive evidence that the sounds measured as fully voiced instances of [b] and [g] were indeed oral stops, and not fricatives or glides.

The finding of release bursts in phrase-final position, as illustrated in Fig. 6 for [b] in *ǧób* ‘snoring’, [g] in *oglág* ‘telling one’s own, relating’, [b] in *gléb* ‘vomiting’, and [b] in *ób* ‘with them’ was interesting for two reasons. First, as just mentioned, it supported the analysis of these sounds as voiced oral stops. Second, it focused our attention more on properties of the release. As detailed in §2, like the majority of Lakota words with final consonants, these words are contracted forms that have undergone final vowel loss: *ǧób* from *ǧópA* ‘to snore’, *oglág* from *oglákA* ‘to tell one’s own (as a story, decision, name)’, *gléb* from *glépA* ‘to vomit’, and *ób* from *ópȟa* ‘to take part’. Though our findings remain preliminary, formant structure in the final release burst of these tokens is consistent with production of a final voiceless or partially devoiced vowel. In §4, we discuss aspects of the diachrony of final stop voicing in Lakota and suggest an earlier stage (attested in Assiniboine) of intervocalic voicing concomitant with final vowel devoicing. Voiceless release of Lakota final voiced stops could continue this earlier hypothesized sound pattern.

Overall low spectral energy of voiced oral stops. If the coda consonants measured as fully voiced in §3.1 (Figs. 2e,f and Fig. 2g) are oral stops, as opposed to flaps/taps, fricatives, glides, or other sonorant sounds, then they will be expected to have lower overall energy envelopes than these other sound types. In order to assess this factor, we used quantitative measures across the entire data set. First, we applied a high-pass filter (350 Hz) to the audio data with the R package 'wrassp' (Winkelmann et al. 2017) to remove the lower frequencies, where most of the energy of voiced stops is expected to be concentrated. Then, we performed a root-mean-square analysis (rms), also with the R package wrassp, to the resulting filtered data. The rms values were speaker-normalized (*z*-normalization) and converted back to decibels. We excluded intervocalic voiced stops from the analysis, because there seems to be consensus on their status as stops, and because our focus was whether voiced consonants that result from the proposed process of stop coda voicing (5b.ii) were produced as oral stops or with some other manner of articulation. Our corpus of purported voiced coda stops included 299 tokens. These were compared with the entire set of approximants and voiced fricatives in our database, which included 219 tokens. Relevant results of our measurements are plotted in Figure 8.

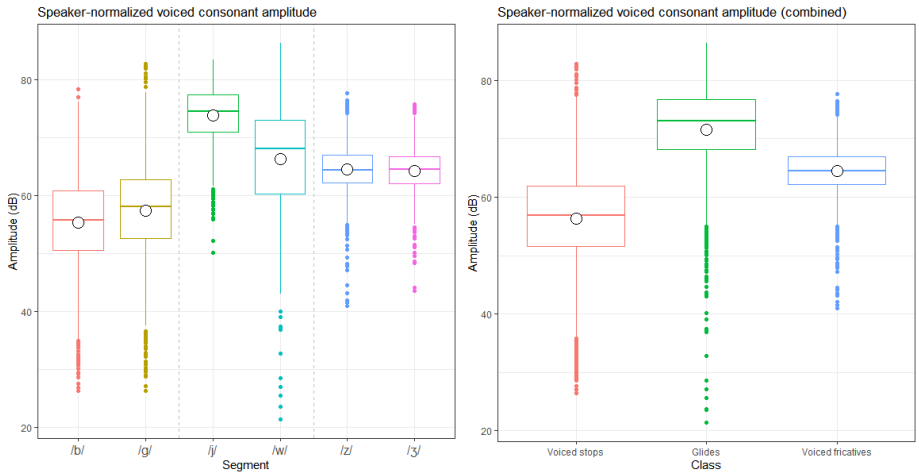


Figure 8. Comparison of the speaker-normalized spectral amplitude of some voiced segments.

Table 13 shows the results for the linear mixed-effects model constructed to test the intensity differences (in decibels; dB) between coda voiced stops and the approximants and voiced sibilants in the language. We added random slopes and random intercepts by speaker and word. Nonnormalized values were used in order to avoid singularity due to the incorporation of random effects by speaker to the model. Results shown are for Tukey post-hoc comparisons with α-level compensation for multiple comparisons using Benjamini-Hochberg adjustment.

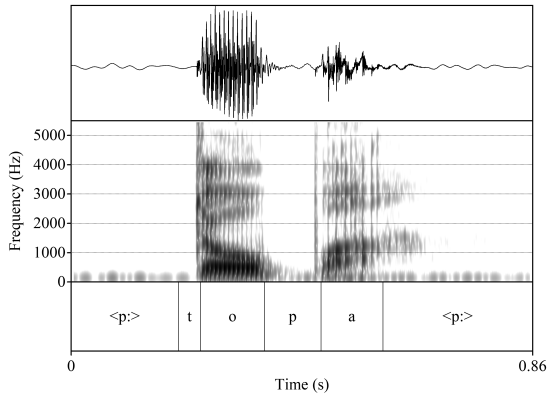
The comparison shows a categorical difference in amplitude between voiced stops in the coda and all voiced fricatives and approximants, and no statistically significant difference between [b] and [g]. Average spectral amplitude of voiced stops is below 60 dB (55 and 57 dB), while voiced fricatives have averages of 64 dB, and the averages for glides are 66 and 74 dB. These results suggest that [b] and [g] form a category of low-energy sounds, consistent with their production as oral (voiced) stops, distinct from the category of voiced fricatives and voiced approximants.

LINEAR HYPOTHESIS	EST	*SE*	*z*-VALUE	*Pr*(> |*z*|)
g − b == 0	2.1099	1.1627	1.815	0.075
j − b == 0	18.1264	1.6735	10.832	< 0.001
w − b == 0	12.0250	1.4124	8.514	< 0.001
z − b == 0	8.2392	0.8380	9.832	< 0.001
Z − b == 0	9.0625	1.1301	8.019	< 0.001
j − g == 0	16.0165	2.5464	6.290	< 0.001
w − g == 0	9.9151	2.0854	4.755	< 0.001
z − g == 0	6.1294	1.5069	4.067	< 0.001
Z − g == 0	6.9526	1.7378	4.001	< 0.001

TABLE 13. Comparison of the intensity of voiced consonants against [b] and [g].

3.3. FINAL VOICING AS ALTERNATION. As reviewed in §2, many researchers describe voicing alternations for Lakota when a stem ending in …VTV- (T a voiceless oral stop or affricate) is produced as consonant-final. For example, *sákA* 'to be dry, to be dried until hard or stiff' has medial [k], but is reduplicated as *sagsákA*, where truncation (5a) results in /k/ being pronounced as [g] in coda position (5b.ii). In this section, acoustic analysis of alternating consonants in morpheme-alternant pairs supports these descriptions: intervocalic instances of /k/ and /p/ are typically voiceless, while phrase-final and preconsonantal instances of the same consonants in the same morphemes are often voiced IN NONVOICING CONTEXTS, namely word-finally and before voiceless consonants and glottal stop. Since our tokens are taken from recordings made for dictionary purposes, we have limited numbers of these pairs, making a statistical analysis unsuitable. However, the acoustic measurements presented below demonstrate that the voicing alternations described for Lakota are attested. Spectrograms and waveforms of each word in Figures 9 and 10 are accompanied by a plot showing the details of the voicing of each relevant segment by means of AC coefficients as a function of time. As discussed in §3.1, AC coefficients go from 0 (completely voiceless) to 1 (completely voiced), with 0.5 as the voicing threshold. Time in the x-axis has been normalized to the duration of each segment to facilitate the visualization of the proportion of voicing. Note that the last time point in the plot might be affected by the voicing of the following segment.

In Figs. 9 and 10, the comparison between voiced and voiceless stops shows that, whereas prototypical intervocalic voiceless stops show a decrease in voicing that follows carryover voicing from the previous vowel (and spans most of the duration of the



(FIGURE 9. *Continues*)

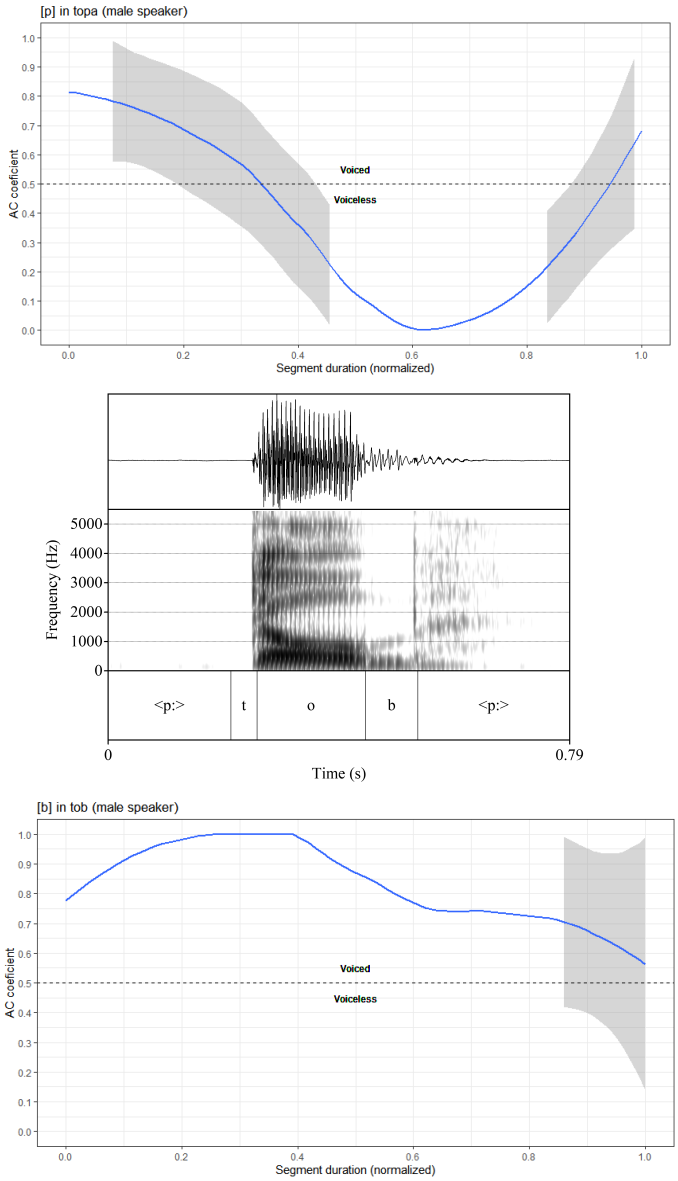
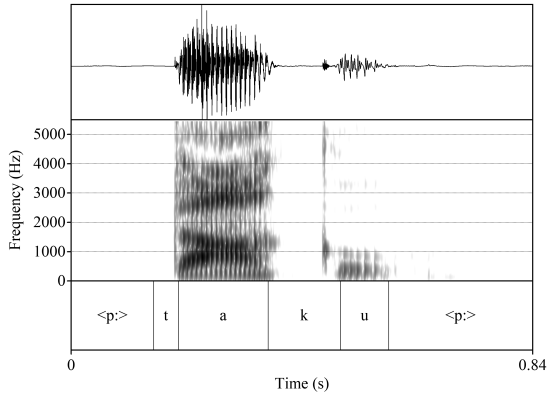
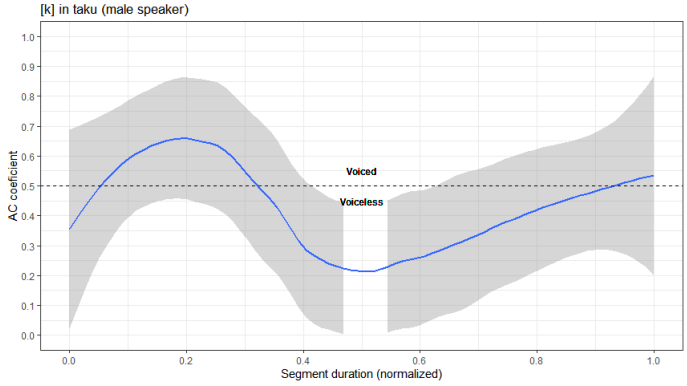


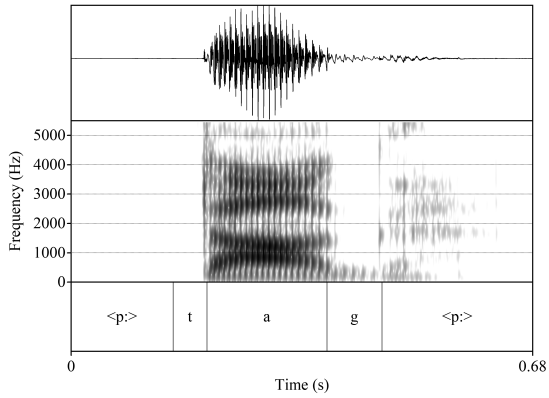
FIGURE 9. Comparison of alternating consonants in morpheme-alternant pair: *tópa* (M) 'four' vs. *tób* (M) 'four' (CONT).

stop), word-final voiced stops do not show any significant decrease in voicing until the end of the segment. This comparison can also be made in reduplicated forms, where the first morpheme appears in its reduced form (with coda stop voicing) preceding the full form (with an intervocalic voiceless stop). Figure 11 shows the reduplicated form *ȟobȟópA* 'to be extremely attractive' as produced by the female speaker.

3.4. SUMMARY OF ACOUSTIC ANALYSIS. Our acoustic analysis of voicing during oral stop closure presented in §3.1 is compatible with the view that Lakota oral stops /p/ and /k/ are voiced in coda position. This voicing is sometimes obscured by devoicing processes, including phrase-final gradient devoicing, assimilation to voicelessness







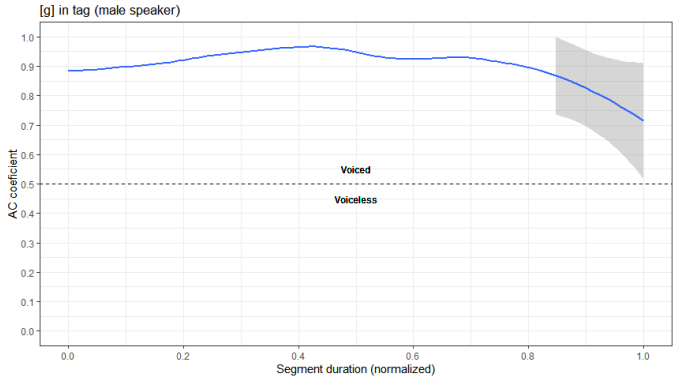


FIGURE 10. Comparison of alternating consonants in morpheme-alternant pair: *táku* (M) 'what' vs. *tág* (M) 'what' (CONT).

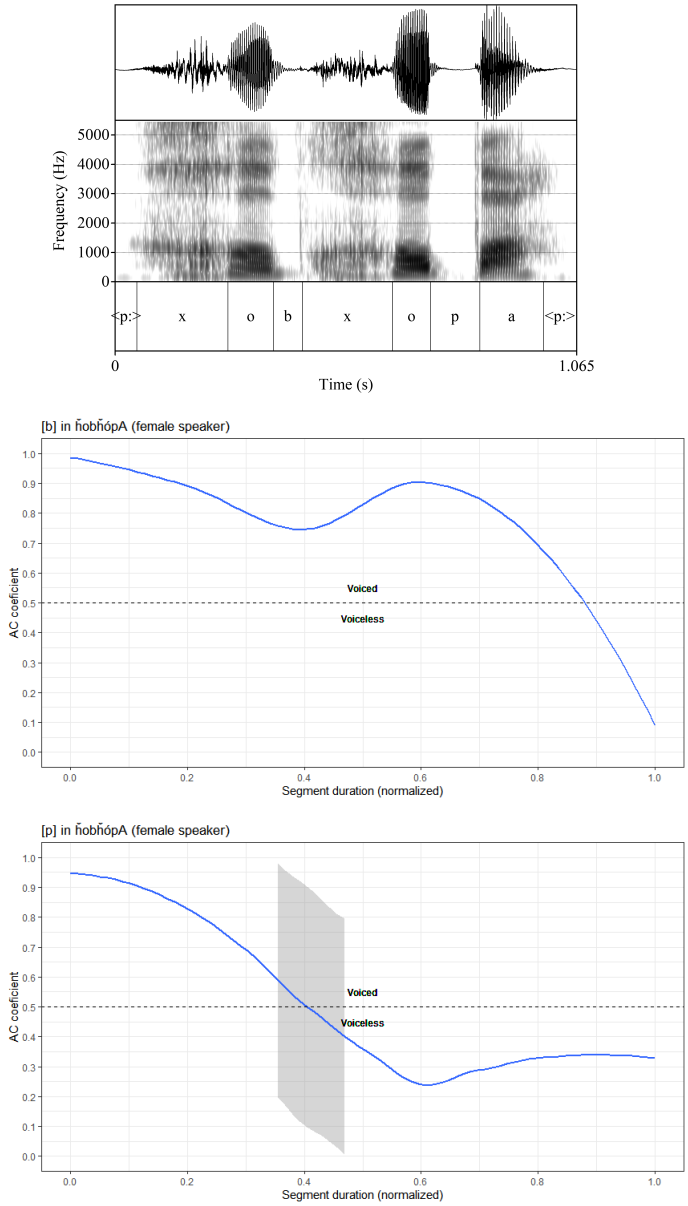


FIGURE 11. Comparison of alternating consonants in reduplicated words: *ȟobȟópA* (F) 'to be extremely attractive'.

before voiceless obstruents and /h/ (which may be coupled with resyllabification into onset), and devoicing that occurs with optional fusion with a following glottal stop, yielding an ejective. In order to support the view that voiced /p/ and /k/ are oral stops, §3.2 presented an array of acoustic properties consistent with stop production. Sounds that were categorized as fully voiced codas in §3.1 were shown in §3.2 to have significant closure durations in the range of normal for voiced stops, absence of fricative noise during closure, release bursts, and low energy levels that are all characteristic of oral stops in contrast to fricatives, taps/flaps, glides, or other sonorant sounds. Section 3.3 presented morphologically related word pairs with voiceless and voiced alternants,

highlighting the differences in voicing in a given stop with regard to its phonological context (intervocalic vs. word-final). In sum, there is acoustic evidence that Lakota has a sound pattern of oral stop coda voicing, supporting the impressionistic descriptions of earlier researchers.

4. Hypothesized origins of lakota obstruent voicing patterns. Recall from §1 that certain phonological markedness accounts like that of Kiparsky (2006, 2008) predict the nonexistence of synchronic sound patterns of final obstruent voicing. In contrast, evolutionary phonology (Blevins 2006a,b) suggests that such patterns may exist, but may be extremely rare due to the numerous phonetic factors that result in devoicing of obstruents in word-final position. One of Kiparsky's (2006) arguments against the evolutionary approach involves historical pathways of change. He suggests that, were there no grammatical markedness constraint against final obstruent voicing, it could easily evolve by the succession of two independently common sound changes: (i) intervocalic voicing: VTV > VDV, and (ii) final vowel loss: VDV# > VD#. Since, he argues, no clear cases of this kind are in evidence, the absence of synchronic final obstruent voicing supports the existence of phonological markedness constraints which demand that unmarked (voiceless) obstruents are preferred in positions of neutralization. We have shown above that there is acoustic support for a synchronic sound pattern of final voicing in Lakota. While this, in itself, is enough to call markedness accounts into question, we continue to be interested in the question of why sound patterns of this type are uncommon, and how they might evolve.

To this end, we offer below some hypotheses regarding the evolution of Lakota final obstruent voicing. After introducing the Proto-Siouan sound system in §4.1, we suggest Lakota /l/ < *d in §4.2. With this sound change established, the synchronic alternation of /p/, /t/, /k/ with [b], [l], [g] can be seen to reflect a uniform historical voicing of oral stops /p/, /t/, /k/ > [b], [d], [g] prior to the *d > l sound change. In §4.3, we suggest that this voicing process was similar to the historical pathway suggested by Kiparsky (2006): where he suggests intervocalic voicing followed by final vowel loss, we suggest intervocalic stop voicing concomitant with vowel reduction as a consequence of anticipatory coarticulation of the final vowel gesture.

4.1. The proto-siouan sound system. Proto-Siouan is reconstructed by Rankin, Carter, and Jones (1998) with the consonant inventory shown in Table 14, a system that also underlies the *Comparative Siouan dictionary* (Rankin et al. 2015), from which all reconstructions in this section are taken, unless noted otherwise. The Proto-Siouan consonant system in Table 14 consists minimally of a series of voiceless unaspirated stops and fricatives, *p, *t, *k, *s, *š, *x, laryngeals *ʔ and *h, three sonorants, *w, *r, and *y (which represent labio-velar, central rhotic, and palatal approximants, respectively), and two 'funny' resonants *W and *R, which are similar to *w and *r but more obstruent-like (Larson 2016).[19]

Given the numerous clusters in Proto-Siouan and the clear status of many postaspirates from heteromorphemic C+h clusters (Rankin et al. 1998:2), it is possible that all of the sounds in parentheses in Table 14 constitute historical clusters or, in the case of preaspirates *hp, *ht, *hk, allophonic variants of voiceless unaspirated stops intervocalically before accented vowels (Rankin et al. 1998:1, Larson 2016). Apart from *W and *R, which we return to below, the most notable feature of this inventory is its lack

[19] Larson (2016:66) states that *W is reconstructed no higher than Mississippi Valley Siouan, but the *Comparative Siouan dictionary* shows Proto-Siouan *Wa:si 'pine tree' and *Wá:te (?) 'boat'.

	LABIAL	DENTAL	PALATAL	VELAR	GLOTTAL
STOPS					
voiceless					
unaspirated	*p	*t		*k	
(postaspirated)	(*ph)	(*th)		(*kh)	
(preaspirated)	(*hp)	(*ht)		(*hk)	
(glottalized)	(*p’)	(*t’)		(*k’)	*’ = [ʔ]
FRICATIVES					
voiceless		*s	*š	*x	*h
(glottalized)		(*s’)	(*š’)	(*x’)	
RESONANTS					
sonorant	*w	*r	*y		
obstruent	*W = [b]?	*R = [d]?			

Table 14. Proto-Siouan consonants.

of nasal stops, though /n/ and /m/ are contrastive sounds in all Dakotan languages.[20] Nasalization in Proto-Siouan is hypothesized to be a feature of vowels only, with vowels *i, *e, *a, *o, *u, *į, *ą, *ų, and their long counterparts. By the time of Mississippi Valley Siouan (MVS), resonants preceding nasalized vowels were all pronounced as nasal stops, and subsequent to this point, were phonologized as nasal stops when vowel nasalization was lost.[21] While Proto-Siouan might seem odd in having nasalized vowels but no nasal stops, the same pattern exists in Mandan, and at least one non-Siouan language of North America had a similar pattern. Eyak, an extinct Na-Dené language, had no distinct nasal stops, but did have nasalized vowels: phonetic [m] occurred as a variant of /w/ before nasalized vowels, while phonetic [n] occurred as a variant of /l/ in the same contexts (Maddieson 2013).

With the possible exception of *W and *R, there is no contrastive voicing in Proto-Siouan. Though marginal, the protophonemes *W and *R are sounds thought to have been like *w and *r, but more obstruent-like and with different reflexes: where *w is usually continued as /w/ or /m/ in most daughter languages, *W is continued as /w/, /b/, /mb/, or /p/, showing more obstruent-like behavior; and where *r is usually continued as /r/, /k/, /n/, /ð/, or /y/, *R is continued as /r/, /l/, /d/, /nd/, or /t/, also showing more obstruent-like behavior. Given this, it would not be unreasonable to interpret Proto-Siouan *W as weak [b] (a voiced oral labial stop or tap with short closure duration) and Proto-Siouan *R as a weak [d] (a voiced oral dental stop or tap with short closure duration), which might suggest a limited incipient voicing contrast for the bilabial and dental stops alone.

Earlier researchers (e.g. Rankin 2001, Larson 2016, Rood 2016) have focused on the singleton consonant reflexes of *W and *R as evidence of their obstruent-like phonetics in the protolanguage. However, an additional piece of evidence for obstruent-like status of *W and *R is their behavior in clusters and, more specifically, the apparent assimilation in voicing that occurs when a voiceless obstruent preceded one of these sounds. Crosslinguistically, voice assimilation in obstruent clusters is common, while voice assimilation between a sonorant and a preceding obstruent is rare (Mielke 2008, 2013). If *W and *R are treated phonetically as voiced obstruents, we can better understand the source of the unusual process of presonorant voicing in Lakota.

[20] In contrast, Mandan appears to have inherited the Proto-Siouan pattern without change: it has phonetic [n] and [m] only as allophones of nonnasal consonants adjacent to nasalized vowels.

[21] See Michaud, Jacques, & Rankin 2012 for a detailed analysis of historical nasalization processes in the history of Siouan.

Recall that, in Lakota, /p/ and /k/ are pronounced as voiced [b] and [g], respectively, before sonorants /w, l, m, n/: [bl] from /pl/; and [gw], [gl], [gn], and [gm] from /kw/, /kl/, /kn/, and /km/, respectively (example 2 and Table 8). In the nonshaded rows of Table 15, these Lakota clusters are compared with cognate clusters in other dialects.

PROTO-SIOUAN	PREFORM	LAKȞÓTA	YANKTONAI	YANKTON	SISSETON-SANTEE	ASSINIBOINE	GLOSS
*i-Ró:te	***R**oté	**l**oté	**d**oté	**d**oté	**d**oté	**n**oté	'throat'
*sará:	*s**R**á	s**l**á	s**d**á	s**d**á	s**d**á	s**n**á	'greasy'
*parás-ka	***WR**aska	**bl**aská	**bd**aská	**bd**aská	**bd**aská, (†**md**aská)	**mn**aská	'to be flat & solid'
	*kWéza	**g**wéza	**g**béza	kbéza	hbéza	kméza	'rippled, ridged'
*kré:šE	*kReš-ka	**gl**ešká	**gd**ešká	kdešká	hdešká	kneška	'spotted, striped'
*karąškV	*knaška	**gn**aská	**gn**aská	knaská	hnaská	knaská	'vermin'/'frog'
	*kma	**gm**á	**gm**á	kmá	hmá	kmá	'walnut'

TABLE 15. Voiced obstruents across the Lakota-Dakota dialect continuum.[22]

The first shaded row illustrates *l*/*d*/*n* correspondences independent of cluster phonotactics, and the second shaded row shows that fricatives do not assimilate in voicing to a following consonant.[23] Of specific interest are the bolded singleton consonants and consonant clusters in Table 15. In the shaded cells, singleton *R and cluster *WR are continued as voiced singletons and clusters (whether sonorants or obstruents), respectively, in all dialects. In the nonshaded cells, bolded singletons are the result of dialectal voicing processes. In Lakota, as already discussed, the pattern is one of presonorant voicing, as in ***g**wéza* < *kWéza. However, in Yanktonai, the pattern appears to be one of regressive voice assimilation from a voiced oral stop to a preceding oral stop, as in ***g**béza* < *kbéza < *kWéza. (Sisseton-Santee and Assiniboine show no evidence of obstruent voicing in parallel contexts.) In cases where a preform contains an initial *WR cluster, voicing is continued in both consonants independently, as in *bdaská*, *mnaská* < *WRaska. Returning to the typological observation above, crosslinguistically, presonorant voicing in obstruent-sonorant (OR) consonant clusters is extremely rare. For example, in Indo-European, where OR clusters are reconstructed, there is no evidence in any of the 300+ languages of initial *kl > gl. In contrast, voicing assimilation between obstruents in tautosyllabic clusters is the norm crosslinguistically (Mielke 2008, 2013), with notable exceptions in languages like Hebrew, Khasi, and Tsou (Kreitman 2008, Blevins 2010). Given these facts, a view of *W and *R as phonetically voiced obstruents is indirectly supported by their role in triggering voice assimilation in word-initial clusters like those illustrated in Table 15. We conclude that one possible view of Proto-Dakota *r and *w (or, as written here, *R and *W) is that these sounds were voiced, oral dental and labial stops *d and *b, respectively, with short closure durations and weak bursts.

4.2. LAKOTA /l/ FROM *d. If Proto-Dakota *r/*R was pronounced something like [d], and Lakota /l/ is a regular reflex of this sound, then a sound change of *[d] > [l] is motivated.

[22] For the purposes of this discussion, we have replaced standard Proto-Dakota *r and *w with *R and *W to signal to the reader that these sounds may have had obstruent-like properties. Proto-Siouan *pr- and *wr- (from vowel syncope) merge as MVS *[b]r, and *wr merges with MVS *[b]r in Proto-Dakota, where [b] could be represented as *p, *b, *w, or *W, since there is no contrast in this context. We write these as *WR here.

[23] Since our interest is voiced obstruents, forms like Lakota-Dakota *mní* 'water' < *Wni < Proto-Siouan *wa-rį́:, where nasal assimilation takes *bnį > *mní* in Proto-Dakota, are not included in 5.

Some support for Lakota /l/ < *d can be found in data related to the process of coda voicing. Recall that in the modern language, /p/, /t/, /k/ are pronounced as [b], [l], [g], respectively, in the coda when phrase-final devoicing and/or assimilation to a following voiceless obstruent are absent. If voicing had begun as a natural phonetic process, the voiced counterparts of /p/, /t/, /k/ would be [b], [d], [g], not [b], [l], [g].

Evidence for earlier coda *[d] might be found in what could be interpreted as old compound forms. Consider the compound *lotkhú* 'the underjaw of animals' (Dakota *dotkhú*), from *loté* 'throat' + *khúl* 'down, below, underneath'. Recall that the expected coda form of /t/ is [l], so we expect ***lolkhúl*, not the attested *lotkhú*. The attested form is consistent with earlier /t/ > [d] in the coda, followed by local regressive devoicing under resyllabification (5d). Loss of final /l/ of /khul/ may also be an indicator of the age of the compound. Another word that may represent an old compound is the interjection *lotkȟúŋkešni* 'oh, by the way; incidentally; I forgot to mention; to go back to the subject', with a variant *lolkȟúŋkešni*. While the etymology of this term is unclear, final *-kešni* appears to be from /-kA-šni/ 'kind.of-not', while one might speculate that *-kȟuŋ-* could be the root of *kȟuŋyáŋ* 'quickly, promptly', an archaic term that usually begins a command. Here again, [t] can be interpreted as a devoiced/resyllabified instance of earlier *d, with the [l]-variant a continuation of coda *[d].[24] A final example of this kind is the pair of variants *pȟel'ížaŋžaŋ*, *pȟetížaŋžaŋ* 'lamp, light' from *pȟéta* 'fire' + *ižáŋžaŋ* 'to be lit, give light'. Here, [t] in *pȟetížaŋžaŋ* appears to be a reduction of [.t'] < [d.'], showing the stage prior to *d > l, while [l] in *pȟel'ížaŋžaŋ* is the modern (post *d > l) version of the same compound.

Another potential indicator of *d as the source of Lakota /l/ occurs in reduplicated forms. Recall that there is evidence for coronal dissimilation in reduplication: by rule 5c, l.T → g.T, where T is a coronal consonant: from *lúta* 'red', we have *luglúta* 'red' (INAN.PL). If we are correct in hypothesizing Lakota l < *d, the historic dissimilations (after coda voicing) are *d.d > g.d (> gl), *dš > gš, *ds > gs, and *dč > gč. In contrast, assuming l < *l (or any coronal sonorant) implies dissimilations like *l.l > g.l, *l.š > gš, *l.s > gs, and *l.č > gč. Since dissimilation of place is more likely when the target consonants share manner features, and since a simple change of place (under place dissimilation) is more natural than a change of place and manner, the first set of dissimilatory changes seems more likely than the second set. Since forms with reduplicated coda [g] for /l/ are arguably old (in some cases replaced with productive [l] forms), they are consistent with an earlier stage of the language where the voiced coda form of /t/ was [d]. Under this analysis, *sótA* 'clean, clear' has the old reduplicated form *soksóta* (< *sog-sóta < *sod-sóta), supported by the fact that the base *sótA* 'clean, clear' is obsolete in modern Lakota. In contrast, *yusótA* 'to use something up, expend' shows a productive reduplication *yusólsota*, where, in the synchronic phonology, coda /t/ is realized as [l].

If we are correct in hypothesizing an earlier stage of Lakota where /p/, /t/, /k/ were realized as [b], [d], [g], respectively, in the coda with evidence for the persistence of voiced stops [b] and [g] presented in §3, then a subsequent sound change *d > l is needed to account for the distribution of /l/ in the language. This sound change appears to be context-free and is supported by two kinds of evidence discussed above: com-

[24] These [l]/[t] coda variants in preconsonantal position should not be confused with [l']/[t] variants in compounds whose second members are vowel-initial. In arguably new compounds like *Lakȟóta-iyÀ* 'to speak Lakota' and *Lakȟóta-iyàpi* 'Lakota language', where the second member is vowel-initial, there are two common variants, one with intervocalic [l'] (often lenited to [l]) and one with [t], as in, for example, *Lakȟól'iyà*, *Lakȟótiyà* 'to speak Lakota'. In these cases, the variant with [l'] is considered older and more formal than that with [t].

pounds where first members have [t]- and [l]-final variants, and reduplicated forms where first members have [g]- and [l]-final variants. In both cases, obstruent-final variants reflect a stage of the language before the *d > l sound change: in compounds, *d > [t] in (resyllabified) complex onsets, and in reduplicated forms, *d > [g] under dissimilation with a following coronal obstruent. In sum, many instances of Lakota [l], including most of those in morpheme-initial and intervocalic position of inherited roots, are direct reflexes of Proto-Siouan *r or *R, which, we suggest, merged as *R, a voiced dental [d]-like obstruent, pre-Lakotan *d, which underwent *d > l.[25] Other instances of surface [l] are continuations of Proto-Dakota *t, pre-Lakota *t, which underwent regular voicing (see §4.3) to *d, and subsequent *d > l.

4.3. Precursors of final voicing in intervocalic coarticulatory voicing. If Lakota coda /l/ < *d, then there appears to be evidence of a historical obstruent voicing taking /p, t, k/ to [b, d, g]. Since final obstruent voicing is exceedingly rare crosslinguistically, and since we have evidence in all cases that these stops were historically medial, not final, it seems reasonable to investigate the possibility that synchronic coda voicing is somehow a transform of an earlier intervocalic voicing process.

As we demonstrated in §2, syllable-final voicing in Lakota is associated with truncation (5a). Since Proto-Dakota had only open syllables, the closed syllables created by truncation are new structures, not subject to any preexisting constraints on codas in the language. Nevertheless, there are conditions on truncation: in Lakota, truncation occurs in noncompound forms only when the final vowel is preceded by a single consonant (not a cluster) and when that single consonant is an obstruent or /l/ (not /n/, /w/, or /y/).[26] The picture becomes more complicated when truncated forms are compared across dialects, with representative data in Table 16, where '?' marks suspect forms from the published literature that require audio verification.

Truncated forms in other dialects show voiced codas, though, as in Lakota, the voicing can be allophonic (as for /g/), phonemic (for /p/ to [b]), or phonemic and coupled with a shift from obstruent to sonorant (for /t/ to [n]). Two notes are in order regarding the Assiniboine data. First, there is a regular rule of coda nasalization: /p/ and /t/ become [m] and [n], respectively, when they are word-final, or when followed by a sonorant consonant; /k/ is unaffected, since there is no velar nasal in the language (Cumberland 2005: 70–71). Second, there is some evidence that Assiniboine /n/ and /m/ were historically postploded nasals [n^d] and [m^b] when followed by oral vowels (Cumberland 2005:25–26). Where the (albeit functionally weak) contrast between intervocalic /t/ and intervocalic /l, d, n/ (< *R) is neutralized in Lakota and Assiniboine, the contrast is maintained in *d*-dialects by /d/ > [d] but /t/ > [n], where, perhaps, as in Assiniboine, [n] continues an earlier [n^d]. Though there have clearly been distinct paths of phonologization of voicing for coda consonants, we suggest a phonetic change of /p, t, k/ > [b, d, g] for Dakotan generally when intervocalic before final unstressed (voiceless) vowels, supported by modern Assiniboine sound patterns, as discussed below.

The phonetic realization of /p, t, k/ > [b, d, g] for Dakotan that we hypothesize is one that is grounded in coarticulation of a weak final vowel. We suggest that the weak articulation of the vowel is, in part, a manifestation of the anticipation of articulatory features from the vowel to the preceding consonant. Part of this anticipation yields voicing

[25] In a few cases, Dakota /l/ continues Proto-Dakota *y-.

[26] Here again we see /l/ patterning with the obstruents, suggesting historical *d. For prevocalic /m/, there is only one example where truncation may be observed: the proper noun *Iktómi* 'the trickster of Lakota myths' can be contracted to *Iktó*.

	DIAL	MEDIAL ONSET		TRUNCATION/ WORD-FINAL		TRUNCATION/MEDIAL PRECONSONANTAL	
p	L:	tó**p**a	'four'	tó**b**		tó**b**topa	'by fours'
< *p	Y:	tó**p**a		tó**b**		tó**b**topa	
	S:	tó**p**a		tó**b**		tó**b**topa	
	A:	tó**p**a		tó**m**		?tó**p**topa	(vs. num.nu.pa 'two')
t	L:	glusó**t**A	'to use one's own up'	glusó**l**		glusó**l**sotA	
< *t	Y:	gdusó**t**A		gdusó**n**		gdusó**n**sotA	
	S:	hdusó**t**A		hdusó**n**		hdusó**n**sotA	
	A:	knusó**t**A		knusó**n**		?knusó**n**sotA	
k	L:	ogláA	'to tell one's own'	oglá**g**		oglá**g**khiyA	'to let sb. tell his/her own'
< *k	Y:	ogdá**k**A		ogdá**g**		ogdá**g**khiyA	
	S:	ohdá**k**A		ohdá**g**		ohdá**g**khiyA	
	A:	okná**k**A		okná**[g]**		okná**k**khiyA	
l, d, n	L:	tȟa**l**ó	'meat'	buyáke**l**	'with thuds'	tȟa**l**'ágnake	'rigid goldenrod'
< *R	Y:	tȟa**d**ó		buyáke**d**		(tȟa**n**'ágnake)	
	S:	tha**d**ó		buyáke**d**		(tha**n**'áhnake)	
	A:	tȟa**n**ó		muyáke**n**		tha**n**'áknaŋke	
l, d, n	L:	lo**l**-, lo**t**-	'food'			lo**l**'óphiye	'leather bag for storing meat'
< *R	Y:	do**d**-				lo**t**óphiye	
	S:	do**d**-				do**t**óphiye	
	A:					do**t**óphiye	

TABLE 16. Truncation patterns across dialects.

of the consonant. In other words, the weaker the articulation of the final vowel, the more vowel-like (in terms of voicing/sonority) the articulation of the preceding consonant. Ultimately, the final vowel may be devoiced, as described below for Assiniboine, or lost altogether, as described above for Lakota. Under our analysis, it appears that something close to Kiparsky's sequence of sound changes involving intervocalic voicing followed by final vowel loss has occurred. What makes Lakota special, or unusual, is that when the final vowel is not reduced, there is no evidence of intervocalic voicing. It is only when final vowels are significantly reduced, or lost altogether, that voicing of the once-intervocalic derived coda is realized.

Our analysis of Lakota voicing as a modified version of intervocalic voicing is consistent with facts about the phonetics of closely related languages, and with our own data, where some voiceless stops show voicing between vowels. Many Siouan languages show intervocalic voicing of (voiceless) stops in running speech. One of these is Assiniboine (Nakota), as described by Cumberland (2005).[27] In Assiniboine, the voiceless unaspirated stops are voiced intervocalically, but voiceless elsewhere (Cumberland 2005:18). Of particular interest to this study is Cumberland's (op cit.) description of words like /t^{h}oka/ 'enemy', spoken in isolation. She transcribes this word as [tog^{a}] where the intervocalic /k/ of /t^{h}oka/ is voiced, and the final vowel of the same word is devoiced. Acoustic data is provided to support this analysis, and Cumberland is explicit in detailing a rule of word-final vowel devoicing, and in saying that 'voiceless vowels will still trigger intervocalic voicing so that, even when the vowel is virtually inaudible, evidence of its presence may be seen in a preceding obstruent' (Cumberland 2005:78).

[27] Recent work on the Stoney variant of Nakhóta by the third author of this article suggests a similar sound pattern, with regular voicing of intervocalic stops. There is at least one language, Gitksan, where an acoustic study demonstrates voiced allophones of voiceless unaspirated stops in prevocalic position (Rigsby & Ingram 1990).

Our phonetic study of Lakota revealed similar 'intermediate' stages in some tokens where final voiced obstruents were released into voiceless vowels. Spectrograms illustrating final voiceless vocalic release can be reviewed in Fig. 6. Though truncation typically eliminates final weak vowels in Lakota, we suggest that obstruent voicing, as a historical process, reflects an earlier stage of the language where voicing was noncontrastive and intervocalic stops were voiced before unstressed, reduced final vowels, a pattern extended to all intervocalic stops in Assiniboine. As noted above, what makes Lakota unusual is that when final vowels are not reduced, there is no evidence of intervocalic voicing. This is why we view historical obstruent voicing in Lakota as a phonetic process concomitant with final vowel reduction and loss.

While we have only touched the surface of the history of obstruent voicing in the Dakotan languages, the evidence in §§2 and 3 strongly suggests a synchronic sound pattern whereby Lakota /p/ and /k/ are voiced to [b], [g] in syllable-final position. We attribute the full set of alternations, including /t/ pronounced as [l], to an earlier sound change of intervocalic /p, t, k/ > [b, d, g] concomitant with reduction of final unstressed vowels, including vowel devoicing and loss. This sound change was followed by a context-free change of *d > l in Lakota, supported by data presented in §4.2.

5. Summary and implications for phonological theory. In §2 we described Lakota sound patterns that appear to involve a synchronic process whereby stems ending in …VTV (T a voiceless unaspirated stop or affricate) can be pronounced as …VD (D a voiced consonant). Under this process, /p, t, k/ are pronounced as [b, l, g] in syllable-final position. Our laboratory phonology study in §3 provides phonetic evidence that /p/ and /k/ undergo voicing and maintain their obstruent quality. The segments reported as [b] and [g] in syllable-final position in the Lakota literature are not lenited segments: they are typically produced with significant closure duration, are voiced for the duration of closure when not in a devoicing environment, and have bursts consistent with the production of oral stops. In cases where voicing was partial or absent, it was explained by context-sensitive devoicing: phrase-finally there is gradient devoicing, while before voiceless obstruents and /h/, voiced obstruents tend to be devoiced. The Lakota voicing process is an unusual one from a phonological perspective, as it is neither wholly neutralizing nor wholly allophonic. In the case of /t/ to [l], a robust contrast exists, so final voicing can be seen as neutralizing, though, of course, this alternation can be viewed in terms other than final voicing. In the case of /p/ to [b], a weak contrast exists between /p/ and /b/; here voicing is neutralizing, but barely. In contrast, for /k/ to [g], voicing is purely allophonic: Lakota, like other Dakotan languages, has /k/, but no /g/. Given the nonuniformity of the phonological alternations involved, the psychological reality of the process of final voicing may also be nonuniform. In this context, it is interesting to note that most orthographies represent [b] as <b> and [g] as <g> (e.g. *tópa*, *tób* 'four'; *íyotakA* 'sit down', *íyotagkhiyA* 'to cause someone to sit down') (Rood & Taylor 1985:7) and, further, that 'the most widely used transcription systems do not consistently write these instances of /p/ and /k/ with *b* and *g* when a voiceless obstruent follows', suggesting that speakers could be aware of the contextual devoicing described in §3 as well. Our historical discussion in §4 provided support for a historical pattern of intervocalic /p, t, k/ > [b, d, g] before final unstressed vowels, concomitant with devoicing and/or loss of these vowels, as still found in modern Assiniboine and some tokens in our Lakota database. Under this analysis, Lakota underwent a later *d > l change, distinguishing it from the *d*-dialects.

The historical source of Lakota final obstruent voicing in an earlier intervocalic stop voicing process, followed by final vowel loss, is of theoretical import. Kiparsky (2006)

adopts a general (violable) universal constraint prohibiting obstruent voicing in syllable codas on the basis of the following argument: if this general universal constraint did not exist, how could we explain the fact that, of the many languages with intervocalic voicing and the many languages with final weak vowel loss, not one shows a progression whereby intervocalic voicing of /p, t, k/ to [b, d, g] is followed by final vowel loss, yielding a sound pattern where only voiced stops are found word-finally? In earlier work, Blevins (2006a,b) suggested Somali as such a case modulo more recent final devoicing, but the argument from Lakota is stronger. In Lakota, we have clear phonetic data supporting voicing of /p/ to [b] and /k/ to [g] in the coda (see §3), with comparative evidence from Assiniboine (Cumberland 2005) attesting intermediate stages of intervocalic voicing with final vowels intact, with final vowels devoiced, and with final vowels lost. Further, we can see why final voicing is precarious and rarely attested. In Lakota, the process was rendered nonuniform by the *d > l change and is currently evidenced only where later processes of phrase-final devoicing and regressive devoicing are not found. We are led to conclude that Kiparsky's (2006) argument is flawed. Consistent with evolutionary phonology (Blevins 2004, 2006a,b, 2015), Lakota illustrates a possible but fleeting instance of syllable-final obstruent voicing. In languages where intervocalic voicing of plain voiceless stops is the norm (Kakadelis 2018), loss of final vowels can yield final-voicing sound patterns as well. However, synchronic sound patterns of this kind may be rare due to the high frequency of final obstruent devoicing and voice assimilation, two processes with well-studied phonetic bases.

We conclude that the Lakota language, as currently spoken, shows evidence of a rare pattern of syllable-final obstruent voicing. The oral stops /p/ and /k/ are voiced to [b] and [g] in syllable-final position, while /t/ is pronounced as [l] in the same position. Since we have examined oral stops only in the context of oral vowels, sound patterns in the context of nasalized vowels may differ and are deserving of further study. Obstruent devoicing is also in evidence. Fricatives /z, ž, ǧ/ are devoiced to [s, ʃ, χ] in syllable-final position. In addition, there is evidence of variable initial stop devoicing, gradient phrase-final devoicing, and fairly regular regressive devoicing of [b] and [g] before voiceless obstruents and /h/. Our conclusion is consistent with earlier descriptions of Lakota phonology (e.g. Rood & Taylor 1985, 1996, Ullrich 2008, 2011, 2018, 2019, Ullrich & Black Bear 2016), building on these in three ways. First, we support the sound pattern of syllable-final stop voicing with acoustic analyses of syllable-final stops in the language. Second, we use the same acoustic evidence to argue for gradient phrase-final devoicing and regressive devoicing before voiceless obstruents. Third, we offer a preliminary historical explanation for the final voicing process and for its rarity: final voicing is a consequence of earlier, conditioned intervocalic voicing, preserved only when the final vowel was reduced or lost. Finally, we highlight the importance of the Lakota sound patterns to phonological theory. Traditional markedness accounts predict that such sound patterns do not and should not exist. In contrast, phonetic-historical accounts like evolutionary phonology explain skewed patterns of voicing in terms of common phonetically based voicing and devoicing tendencies, allowing for rare cases of final-obstruent voicing like that found in Lakota.

Like many indigenous languages of the Americas, Lakota is endangered. The acoustic component of this study serves to highlight the central role of endangered language documentation in phonological description and theory (Blevins 2007) and the continued importance of producing high-quality recordings in language documentation, as exemplified by the *NLD*. Independent of its scientific merit, we hope that this study of one indigenous language of the Great Plains will inspire other researchers to bring important data from indigenous languages to bear on central issues in linguistic theory

and encourage colleagues around the world to continue their excellent work in high-quality language documentation.

REFERENCES

Bates, Douglas; Martin Maechler; Ben Bolker; and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. DOI: 10.18637/jss.v067.i01.

Beckman, Jill; Michael Jessen; and Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49.259–84. DOI: 10.1017/S0022226712000424.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006a. A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics* 32.117–65. DOI: 10.1515/TL.2006.009.

Blevins, Juliette. 2006b. Reply to commentaries. *Theoretical Linguistics* 32.245–56. DOI: 10.1515/TL.2006.017.

Blevins, Juliette. 2007. Endangered sound patterns: Three perspectives on theory and description. *Language Documentation and Conservation* 1.1–16. Online: http://hdl.handle.net/10125/1721.

Blevins, Juliette. 2008. Natural and unnatural sound patterns: A pocket field guide. *Naturalness and iconicity in language*, ed. by Klaas Willems and Ludovic De Cuypere, 121–48. Amsterdam: John Benjamins.

Blevins, Juliette. 2010. Phonetically-based sound patterns: Typological tendencies or phonological universals? *Laboratory phonology 10*, ed. by Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio, and Nathalie Vallée, 201–24. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110224917.2.201.

Blevins, Juliette. 2015. Evolutionary phonology: A holistic approach to sound change typology. *The Oxford handbook of historical phonology*, ed. by Patrick Honeybone and Joseph Salmons, 485–500. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199232819.013.006.

Blevins, Juliette. 2017. Areal sound patterns: From perceptual magnets to stone soup. *The Cambridge handbook of areal linguistics*, ed. by Raymond Hickey, 88–121. Cambridge: Cambridge University Press.

Boas, Franz, and Ella Cara Deloria. 1941. *Dakota grammar*. (Memoirs of the National Academy of Sciences 23.) Washington, DC: National Academy of Sciences.

Boersma, Paul, and David Weenink. 2019. Praat: Doing phonetics by computer. Online: http://www.praat.org/.

Brockhaus, Wiebke. 1995. *Final devoicing in the phonology of German*. Tübingen: Max Niemeyer.

Buechel, Eugene S. J. 1939. *A grammar of Lakota: The language of the Teton Sioux Indians*. Saint Francis, SD: Saint Francis Mission.

Coetzee, Andries W., and Rigardt Pretorius. 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38.404–21. DOI: 10.1016/j.wocn.2010.03.004.

Cresci, Michela. 2014. *The sound patterns of Camuno: Description and explanation in evolutionary phonology*. New York: The Graduate Center, CUNY dissertation. Online: https://academicworks.cuny.edu/gc_etds/191.

Cumberland, Linda A. 2005. *A grammar of Assiniboine: A Siouan language of the Northern Plains*. Bloomington: Indiana University dissertation.

de Jong Boudreault, Lynda J. 2009. *A grammar of Sierra Popoluca (Soteapanec, a Mixe-Zoquean language)*. Austin: University of Texas dissertation. Online: http://hdl.handle.net/2152/ETD-UT-2009-12-523.

Egurtzegi, Ander. 2013. Phonetics and phonology. *Basque and Proto-Basque: Language-internal and typological approaches to linguistic reconstruction*, ed. by Mikel Martínez-Areta, 119–72. Frankfurt am Main: Peter Lang.

Elson, Ben. 1947. Sierra Popoluca syllable structure. *International Journal of American Linguistics* 13.13–17. DOI: 10.1086/463925.

Fawcett, Tom. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27. 861–74. DOI: 10.1016/j.patrec.2005.10.010.

Harrington, Jonathan. 2010. *Phonetic analysis of speech corpora*. Chichester: Wiley-Blackwell.

Harris, John G. 2001. States of the glottis of Thai voiceless stops and affricates. *Essays in Tai linguistics*, ed. by Kalaya Tingsabadh and Arthur S. Abramson, 3–11. Bangkok: Chulalongkorn University Press.

Hartigan, J. A., and P. M. Hartigan. 1985. The dip test of unimodality. *The Annals of Statistics* 13.70–84. Online: https://www.jstor.org/stable/2241144.

Henderson, Eugenie J. A. 1964. Marginalia in Siamese phonetic studies. *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*, ed. by David Abercrombie, Dennis B. Fry, Peter A. D. MacCarthy, Norman C. Scott, and J. L. M. Trim, 415–24. London: Longmans, Creen and Co.

Houlihan, Kathleen, and Gregory K. Iverson. 1979. Functionally constrained phonology. *Current approaches to phonological theory*, ed. by Daniel A. Dinnsen, 51–73. Bloomington: Indiana University Press.

Ingham, Bruce. 2003. *Lakota*. (Languages of the world: Materials 426.) Munich: LINCOM Europa.

Iverson, Gregory K., and Joseph C. Salmons. 2011. Final devoicing and final laryngeal neutralization. *The Blackwell companion to phonology, vol. 3: Phonological processes*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 1622–43. Chichester: Wiley-Blackwell. DOI: 10.1002/9781444335262.wbctp0069.

Jansen, Wouter. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian and Dutch*. Groningen: University of Groningen dissertation. Online: http://hdl.handle.net/11370/c53fd622-8670-4b8e-89f9-770e59adb816.

Kager, René. 1999. *Optimality theory*. Cambridge: Cambridge University Press.

Kakadelis, Stephanie. 2018. *Phonetic properties of oral stops in three languages with no voicing distinction.* New York: The Graduate Center, CUNY dissertation. Online: https://academicworks.cuny.edu/gc_etds/2871.

Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32.217–36. DOI: 10.1515/TL.2006.015.

Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. *Linguistic universals and language change*, ed. by Jeff Good, 23–53. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199298495.003.0002.

Kisler, Thomas; Uwe D. Reichel; and Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45.326–47. DOI: 10.1016/j.csl.2017.01.005. WebApp: https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface.

Kreitman, Rina. 2008. *The phonetics and phonology of onset clusters: The case of Modern Hebrew*. Ithaca, NY: Cornell University dissertation. Online: https://hdl.handle.net/1813/9456.

Larson, Rory. 2016. Regular sound shifts in the history of Siouan. *Advances in the study of Siouan languages and linguistics*, ed. by Catherin Rudin and Bryan J. Gordon, 63–83. Berlin: Language Science. DOI: 10.17169/langsci.b94.118.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in optimality theory. *Natural Language and Linguistic Theory* 17.267–302. DOI: 10.1023/A:1006182130229.

Maddieson, Ian. 2013. Absence of common consonants. *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: http://wals.info/chapter/18.

Michaud, Alexis; Guillaume Jacques; and Robert L. Rankin. 2012. Historical transfer of nasality between consonantal onset and vowel: From C to V or from V to C? *Diachronica* 29.201–30. DOI: 10.1075/dia.29.2.04mic.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mielke, Jeff. 2013. Phonologization and the typology of feature behavior. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 165–80. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199573745.003.0008.

Nichols, Johanna. 2011. *Ingush grammar*. Berkeley: University of California Press. Online: https://escholarship.org/uc/item/3nn7z6w5.

Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17(2).178–203.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, ed. by Peter F. MacNeilage, 189–216. New York: Springer. DOI: 10.1007/978-1-4613-8202-7_9.

Parks, Douglas R., and Raymond J. DeMallie. 1992. Sioux, Assiniboine, and Stoney dialects: A classification. *Anthropological Linguistics* 34.233–55. Online: https://www.jstor.org/stable/30028376.

R Core Team. 2019. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org.

Rankin, Robert L. 2001. On Dakotan pre-boundary and cluster phonology. Paper presented at the Siouan and Caddoan Languages Conference, University of Chicago, June 2001.

Rankin, Robert L.; Richard T. Carter; and A. Wesley Jones. 1998. Proto-Siouan phonology and grammar. *Papers from the 1997 Mid-America Linguistics Conference*, ed. by Xingzhong Li, Luis López, and Tom Stroik, 366–75. Columbia: University of Missouri Press.

Rankin, Robert L.; Richard T. Carter; A. Wesley Jones; John E. Koontz; David S. Rood; and Iren Hartmann (eds.) 2015. *Comparative Siouan dictionary*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: http://csd.clld.org.

Rice, Keren. 1993. A reexamination of the feature [sonorant]: The status of 'sonorant obstruents'. *Language* 69.308–44. DOI: 10.2307/416536.

Rigsby, Bruce, and John Ingram. 1990. Obstruent voicing and glottalic obstruents in Gitksan. *International Journal of American Linguistics* 56.251–63. DOI: 10.1086/466152.

Robin, Xavier; Natacha Turck; Alexandre Hainard; Natalia Tiberti; Frédérique Lisacek; Jean-Charles Sanchez; and Markus Müller. 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. DOI: 10.1186/1471-2105-12-77.

Rood, David S. 2016. The phonology of Lakota voiced stops. *Advances in the study of Siouan languages and linguistics*, ed. by Catherine Rudin and Bryan J. Gordon, 233–55. Berlin: Language Science. DOI: 10.17169/langsci.b94.118.

Rood, David S., and Allan R. Taylor. 1976. *Beginning Lakhota, vol. 1 and 2*. Boulder: University of Colorado Lakota Project.

Rood, David S., and Allan R. Taylor. 1985. Boundaries and Lakhota orthography. *Siouan and Caddoan Linguistics* 5.5–10.

Rood, David S., and Allan R. Taylor. 1996. Sketch of Lakhota, a Siouan language, Pt. I. *Handbook of North American Indians, vol. 17: Languages*, 440–82. Washington, DC: Smithsonian Institution.

Scarborough, Rebecca; Georgia Zellou; Armik Mirzayan; and David S. Rood. 2015. Phonetic and phonological patterns of nasality in Lakota vowels. *Journal of the International Phonetic Association* 45.289–309. DOI: 10.1017/S0025100315000171.

Shaw, Patricia A. 1980. *Theoretical issues in Dakota phonology and morphology*. New York: Garland.

Smolders, Joshua. 2016. A phonology of Ganza (Gwami Nana). *Linguistic Discovery* 14(1). DOI: 10.1349/PS1.1537-0852.A.470.

Šimáčková, Šárka; Václav Jonáš Podlipský; and Kateřina Chládková. 2012. Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association* 42.225–32. DOI: 10.1017/S0025100312000102.

Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie*. Göttingen: Vandenhoeck and Ruprecht.

Ullrich, Jan F. (ed.) 2008. *New Lakota dictionary: Incorporating the Dakota dialects of Santee-Sisseton and Yankton-Yanktonai.* Bloomington, IN: Lakota Language Consortium.

Ullrich, Jan F. (ed.) 2011. *New Lakota dictionary*. 2nd edn. Bloomington, IN: Lakota Language Consortium.

Ullrich, Jan F. 2018. *Modification, secondary predication and multi-verb constructions in Lakota*. Düsseldorf: Heinrich Heine University dissertation.
Ullrich, Jan F. (ed.) 2019. *New Lakota dictionary online*. Bloomington, IN: Lakota Language Consortium. Online: https://www.lakotadictionary.org.
Ullrich, Jan F., with Ben Black Bear, Jr. 2016. *Lakota grammar handbook*: *A pedagogically orientated self-study reference and practice book for beginner to upper-intermediate students*. Bloomington, IN: Lakota Language Consortium.
Wetzels, W. Leo, and Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77.207–44. DOI: 10.1353/lan.2001.0123.
Winkelmann, Raphael. 2017. *The EMU-SDMS*. Munich: Ludwig Maximilian University dissertation. Computer software program available online: https://cran.r-project.org/package=emuR.
Winkelmann, Raphael; Lasse Bombien; and Michel Scheffers. 2017. wrassp: Interface to the 'ASSP' library. Online: https://cran.r-project.org/package=wrassp.
Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80.73–97. DOI: 10.1353/lan.2004.0049.

Blevins
Linguistics Program
The Graduate Center, CUNY
365 Fifth Avenue
New York, NY 10016
[jblevins@gc.cuny.edu]

Egurtzegi
IKER (UMR5478)
Campus de la Nive
Château-Neuf, 15 Place Paul Bert
64100 Bayonne, France
[ander.egurtzegi@iker.cnrs.fr]

Ullrich
The Language Conservancy
2620 N Walnut St Ste. 810
Bloomington, IN 47404
[jfu@languageconservancy.org]

[Received 28 April 2019;
revision invited 3 October 2019;
revision received 18 November 2019;
accepted 21 December 2019]

Phonetically conditioned sound change

Contact induced /u/-fronting in Zuberoan Basque

Ander Egurtzegi
Institute of Phonetics and Speech Processing, Ludwig-Maximilians University of Munich

All modern Basque dialects have at least 5 contrastive vowels /i, e, a, o, u/. One Basque dialect, Zuberoan, has developed a contrastive sixth vowel, the front rounded high vowel /y/. This development is arguably due to sustained contact with neighboring Gallo-Romance languages. This paper supports empirically the historical development of the /u/ vs. /y/ contrast and provides a detailed analysis of the contexts that inhibited the /u/ > /y/ sound change. Fronting was inhibited when the vowel was followed by an apical sibilant, a tap /ɾ/ or an rT cluster (where r is a rhotic, and T an alveolar obstruent), arguably due to coarticulatory effects. Fronting occurred when /s̪/, /r/ or non-coronal rhotic-obstruent clusters followed /u/. Zuberoan /u/-fronting illustrates the importance of language contact and phonetics in the phonological analysis of historical developments.

Keywords: u-fronting, sound change, coarticulatory effects, language contact, Basque, Zuberoan

1. Introduction[1]

All modern Basque dialects have at least 5 vowels /i, e, a, o, u/. Of these dialects, only Zuberoan (also known as Souletin) and Low Navarrese from Mixe (cf. Lafon 1999 [1962b]: 105–106; Camino 2009a, 2009b), two varieties in close contact

1. I am greatly indebted to Juliette Blevins, Gorka Elordieta, Jonathan Harrington, José Ignacio Hualde, Joseba Lakarra, two anonymous reviewers and the editor of the journal for their comments and suggestions to previous versions of this paper. They have considerably improved the paper. Markus Jochim and Beatriz Umbon helped me with the German and French translations of the abstract, respectively. All remaining errors are mine. This research has been partially funded by the Spanish MINECO (FFI2016-76032-P; FFI2015-63981-C3-2; FJCI-2014-21348), the Basque Government (IT769/13) and the Alexander von Humboldt Foundation.

Diachronica 34:3 (2017), **331–367**. DOI 10.1075/dia.34.3.02egu
ISSN 0176-4225 / E-ISSN 1569-9714 © John Benjamins Publishing Company

with Bearnese Gascon, have developed a contrastive sixth vowel quality, the front rounded high vowel /y/. This vowel systematically corresponds to /u/ in the other dialects and can be traced back to that segment by means of cross-dialectal comparison. Nevertheless, /y/ now contrasts with /u/, since certain phonological contexts inhibited fronting (Michelena 2011 [1977]: 42; see below) and since there are other sources of /u/ in this dialect (cf. Egurtzegi 2014a: 171–188).

This study has two goals. On the empirical side, Zuberoan evidence is supported by an analysis of the earliest texts in this dialect. On the theoretical side, I tackle three aspects of the /u/ > /y/ shift with implications for theories of sound change. First, while contact has clearly played a role, the evolution of /y/ in Zuberoan Basque is complex: it is not the simple result of the influx of Romance loanwords with /y/, nor is it a borrowed sound change. Instead, it is best treated as a consequence of contact-induced sound change, where the change in question shows clear phonetic conditioning not present in the contact language. Second, this phonetic conditioning involves the set of contexts where /u/ > /y/ was inhibited; interestingly, this environment is problematic in terms of traditional phonological natural classes. A final problem for the general account is why the /u/ of inherited /au/ and /eu/ sequences fails to undergo the /u/ > /y/ shift. I suggest that this is because the vowel devocalized (or was non-syllabic) prior to the /u/ > /y/ shift. This detail is of general interest, since it is rare to find arguments for true glides in Basque phonological systems.

Before turning to historical data, I offer a few remarks on modern Zuberoan phonetics and phonology. Zuberoan has six contrastive vowels /i, y, u, e, o, a/ as well as contrasting high and low nasalized vowels /ĩ, ỹ, ũ, ã/. According to Michelena (2011 [1977]: 41), the realization of Zuberoan /y/ is closer to French [ø] than [y]. Hualde (1993: 290) also describes this segment as being intermediate between French [y] and [ø]. Following Larrasquet (1932) and Lafon (1999 [1958]), he notes that this is consistent with the evidence that Zuberoan /i/ and /u/ are more open than they are in French (Hualde 1993: 290–291). Gavel also describes the three Zuberoan high vowels as lower than their French counterparts and states that this is also the case in many Bearnese varieties of Gascon (Larrebat 1926: 29 fn. 1, cf. Michelena 2011 [1977]: 41 fn. 6).

A contrast between front high rounded /y/ and back /u/ is evident in 17th century documents in the Zuberoan dialect based on the interpretation of distinct orthographic symbols for the two vowels. The contrast is present in the authors Bela (Michelena 2011 [1964]: 168–171) and Etxart (in his letters from 1616–1617; edited by Lakarra et al. forthcoming). Oihenart (2003 [1657]) describes the vowel /y/ and

© 2017. John Benjamins Publishing Company
All rights reserved

writes it as <u>, as opposed to /u/, which he writes as <ou>.[2] Michelena mentions Tartas (1995 [1666]) as the first book written in Zuberoan and says the author is consistent with this tradition (cf. Michelena 2011 [1977]: 41).[3] Other 17th century documents that distinguish the two high rounded vowels include Belapeire (1997 [1696]) and *Pronus* (Agirre Sarasola 1998 [ca. 1676]).

Following French written tradition, /y/ is written as <u> and /u/ is written as <ou>. However /y/ is consistently written as <ü> in modern texts, and that is the grapheme that will be used in the examples given here. In some old texts, <eu> is used instead of <u> (cf. Oyharçabal 1991, where the word *lür* /lyr/ "Earth, land" is written <leur> more than ten times). In French, <eu> represents /ø/ or /œ/: *heureux* /øʁø/ "happy", *feu* /fø/ "fire" and *peu* /pø/ "a little", but also *peur* /pœʁ/ "fear", *jeune* /ʒœn/ "young" and *leur* /lœʁ/ "their, them". This is consistent with the aforementioned observation that Zuberoan high vowels are more open than their French counterparts. Also consistent with this height difference is the pattern of some loanwords involving /œ/ borrowed with /y/ in Zuberoan: French *tracteur* /tʁaktœʁ/ >> Zuberoan *traktür* /tɾaktyr/ "tractor", French *beurre* /bœʁ/ >> Zuberoan *bürra* /byra/ "butter" (Coyos 2002: 207).

The high front rounded vowel is not completely unknown in the other two northern dialects, Lapurdian and Low Navarrese. However, in these dialects, it is limited to French borrowings such as *faktüra* [faktyɾa] "receipt", *kandidatüra* [kan̪diðatyɾa] "candidacy", *nümero* [nymeɾo] "number" (cf. Zuazo 2008: 43). There is no evidence for a /u/ > /y/ shift in these varieties.

Zuberoan has other sound patterns in common with neighboring Romance languages as well. One of these is the contrast between oral and nasalized vowels (not only present in Zuberoan but also in western Basque dialects such as 16th century Bizkaian; cf. Egurtzegi 2015). Nasalized vowels in Basque are due to the loss of nasalized /h̃/, which developed from intervocalic *n (cf. Michelena 2011 [1950]: 8–9, 2011 [1977]: 171; Hualde 1993: 294; Igartua 2015; Egurtzegi 2014a: 68–72, forthcoming). Both the loss of intervocalic *n and the development of nasalized

2. Peillen (1992: 252) claims to have seen <u> used this way in manuscripts from the 14th century, without providing details on the source. However, there is, to date, no evidence of a consistent orthographic convention distinguishing /u/ vs. /y/ in Zuberoan prior to the 17th century. The oldest texts I examine in §2.2 from the end of the 14th century do not show this contrast in a clear way (the manuscript Peillen referred to might be the *Censier* examined there).

3. Nevertheless, the classification of Tartas' variety as Zuberoan is highly debatable today (Ricardo Gómez, personal communication). In addition, in most work, Tartas does not distinguish /u/ from /y/ orthographically and consistently uses <u> for all high rounded vowels (cf. Tartas 1995 [1666], 1996 [1672]). However, when he does distinguish /y/ from /u/, he also writes the first as <u> and the second as <ou>.

© 2017. John Benjamins Publishing Company
All rights reserved

vowels are common to Bearnese Gascon and Zuberoan Basque. Both languages have similar accentual systems and share other sound patterns including /o/-raising and /u/-fronting (Egurtzegi 2014a).

There has been extensive contact between Zuberoan and Bearnese Gascon for centuries, as described, for example, by Haase (1992: 19–22) who speaks of intense bilingualism, mentioning close connections between the counties of Zuberoa and Béarn since the early Middle Ages. Zuberoa was part of Béarn from the 12th century (Haase 1992: 19), and Bearnese Gascon was the only written language in use in Zuberoa until the 17th century (Séguy 1952: 385). In addition, Gascon was an official language and the language for scholarly communication in Zuberoa during that period (Gavel 1920: 47). Although French is the main language now, the presence of Gascon has been mantained in the region. In sum, Basque-Gascon bilingualism has existed for centuries (Lafon 1999 [1962a]: 108).

2. Describing the u > y change

This section describes in detail the Zuberoan sound change of /u/ > /y/. First, I summarize earlier descriptions of this process. Then, two new studies are shown to be consistent with the historical evidence. The first (cf. Egurtzegi 2014a, 2014b) randomly surveys more than 500 pages of the *General Basque Dictionary* (*Orotariko Euskal Hiztegia*; Michelena & Sarasola 1987–2005), in addition to a sample from the *Dictionnaire basque-français* of Lhande (1926–1938) and the glossary in *Le Basque de la Basse-Soule Orientale* by Larrasquet (1939). These examples are used to describe the distribution of /y/ in the language. A second study surveys the oldest documents written in Zuberoan from the 14th–17th centuries. The results are compared to those of the first survey in order to highlight differences in the distribution of /y/ in the attested stages of the dialect and details of the sound change and its inhibition not noted in the earlier literature.

2.1 Previous descriptions of the sound change

The fronting of u > y in Zuberoan has been described in earlier literature (cf. Uhlenbeck 1903; Gavel 1920; Lafon 1999 [1937], 1999 [1958], 1999 [1962a], 1999 [1962b]; Michelena 2011 [1977]) with a range of commonly agreed upon findings. As in various Gallo-Romance languages such as French (Harris 1988: 211) and Occitan (Wheeler 1988: 247), Zuberoan Basque developed a sound pattern in which most instances of the high back vowel /u/ were fronted to /y/. This process appears to have been context-free, as shown by the examples in (1), though a subclass of

© 2017. John Benjamins Publishing Company
All rights reserved

coronals inhibits fronting, as discussed below. Modern Zuberoan lost the tap /ɾ/ during the 19th century – see Michelena (2011 [1977]: 272), who mentions that Gèze (1873: 2) already shows this loss (cf. also Camino 2009a: 167); for this reason, Literary Zuberoan, which maintains the tap, is used in all the Zuberoan examples.

The general pattern is shown in (1): (1a) shows word-final contexts; (1b) shows instances of fronting before some coronals; (1c–d) show instances of /u/ fronting in pre-velar and pre-labial contexts respectively; and (1e) are loans that are arguably old, predating the fronting sound change and regularly undergoing it.

(1) /u/-fronting in Zuberoan Basque

a. Word-final position

Standard Basque	Literary Zuberoan	IPA	Gloss
su	*sü*	[s̪y]	"fire"
zu	*zü*	[s̪̺y]	"you (SG)"
sagu	*sagü*	[s̪aɣ̟y]	"mouse"

b. Before a coronal consonant

Standard Basque	Literary Zuberoan	IPA	Gloss
egun	*egün*	[eɣ̟yn]	"day"
uda	*üda*	[yð̞a]	"summer"
dut	*düt*	[dyt]	"I have"
mutur	*müthür*	[myt^{h}yr]	"snout"
guzti	*güzi*	[gys̪̺i]	"all"
buztan	*büztan*	[bys̪̺tan]	"tail"
lur	*lür*	[lyr]	"ground"
egur	*egür*	[eɣ̟yr]	"wood"
zuzen	*xüxen*	[ʃyʃen]	"direct"

c. Before a velar consonant

Standard Basque	Literary Zuberoan	IPA	Gloss
suge	*süge*	[s̪yɣ̟e]	"snake"
sukalde	*sükhalte*	[s̪yk̟^{h}alte]	"kitchen" (with root *sü* "fire")
ukan	*ükhen*	[yk^{h}en]	"to have"
uko	*ükho*	[yk^{h}o]	"refusal"
nauk	*nük*	[nyk]	"I am (ALLOC.M)"

d. Before a labial

Standard Basque	Literary Zuberoan	IPA	Gloss
zubi	*zübü*	[s̪̺yβ̞y]	"bridge"
ume	*hüme*	[hyme]	"child"

© 2017. John Benjamins Publishing Company
All rights reserved

e. In loanwords

Standard Basque	Literary Zuberoan	IPA	Gloss
mundu	*mündü*	[myn̪dy]	"world"
duda	*düda*	[dyð̞a]	"doubt"
lukainka	*lükhainka*	[lykʰa͡iŋka]	"spicy sausage" (cf. Vulgar Latin *lucanicia*)
zeru	*zelü*	[s̬ely]	"sky"
kutxa	*hütxa*	[hyʧa]	"box"
uxatu	*üxütü*	[yʃyty]	"to shoo"

Based on this data, the /u/-fronting sound change in Zuberoan can be stated as in (2).

(2) Definition of /u/-fronting in Zuberoan Basque
/u/ > /y/

Two details of the sound change are remarkable and are reviewed in §2.2 in the context of the oldest written Zuberoan texts. First, there are specific contexts where the sound change is inhibited, as detailed below. Inhibitory effects are always the result of a following coronal, though not all coronals inhibit the shift. A second observation is that Common Basque /u/ does not front to /y/ when it is part of an inherited /au/ or /eu/ sequence. For example, from Common Basque *gau* "night", we do not find Zuberoan **gaü, but instead, *gai*. Older texts allow us to refine and ultimately explain these details.

The contexts hindering fronting involve a following coronal segment, but, as should be clear from the examples in (1b), not all coronals inhibit fronting. The contexts where fronting is inhibited have been described in earlier literature (cf. Uhlenbeck 1903: 29–31; Gavel 1920: 39–69; Lafon 1999 [1937], 1999 [1958], 1999 [1962a], 1999 [1962b]; Michelena 2011 [1977]) and are also repeated in more recent papers (cf. Zarabozo 1972; Hualde 1993; Zuazo 2008; Oñederra 2009a, 2009b; Egurtzegi 2013: 129, 2014b). Michelena (2011 [1977]: 41–42) described three different contexts where fronting was inhibited: before /s̬/; before /ɾ/; and before /rtʰ, rt, rd/. Examples of maintained /u/ are shown in (3): in (3a) fronting is inhibited before an apical sibilant; in (3b) before a tap; and in (3c) before rhotic-coronal clusters, where the first element is a neutralized rhotic (probably a light trill, as in most modern varieties) and the second element is a coronal oral stop.

© 2017. John Benjamins Publishing Company
All rights reserved

(3) Inhition of /u/-fronting in Zuberoan Basque

a. Before an apico-alveolar sibilant /s̪/

Standard Basque	Literary Zuberoan	IPA	Gloss
ikusi	*ikhusi*	[ik^{h}us̪i]	"see"
itsusi	*itxusi*, *itsusi*	[itʃus̪i], [its̪us̪i]	"ugly"
uste	*uste*	[us̪te]	"thought"
busti	*busta*, *busti*	[bus̪ta], [bus̪ti]	"to wet, wet"

b. Before a tap /ɾ/

Standard Basque	Literary Zuberoan	IPA	Gloss
ur	*hur*	/huɾ/	"water"
zur	*zur*	/s̪uɾ/	"wood"
hura	*hura*	[huɾa]	"he/she/it"
gure	*gure*	[guɾe]	"our" (but *gü* "we (ABS)", *gük* "we (ERG)", etc.)
zuri	*xuri*	[ʃuɾi]	"white"
urin	*urin*	[uɾin]	"animal fat"
hiru	*hirur*	/hiɾuɾ/	"three"
barau	*barur*	/baɾuɾ/	"fasting"

c. Before a (heterosyllabic) rhotic-dental cluster /rt, rt^{h}, rd/

Standard Basque	Literary Zuberoan	IPA	Gloss
urte	*urthe*	[urt^{h}e]	"year"
urde	*urde*	[urð̪e]	"pig, animal"
urdin	*urdin*	[urð̪iɲ]	"blue"

Based on the data in (3), the statement of /u/-fronting in (2) must be revised. A preliminary revision is given in (4).

(4) Definition of /u/-fronting in Zuberoan Basque

/u/ > /y/ except /_X

Where X is either:

i. An apical fricative /s̪/,

ii. A tap /ɾ/,

or

iii. A rhotic + coronal-oral-stop cluster /rt, rt^{h}, rd/

The problem posed by the blocking environments is a simple one. What do the environments in (4i–iii) have in common that is not shared by the laminal series /s̪, ts̪/, the pre-palatals /ʃ, ʧ/, the trill /r/, plain /t, d, n, l/ or other rhotic-stop clusters like /rk/ and /rp/? Though it may be tempting to invoke an apical/laminal contrast, both the trilled /r/ and the tapped /ɾ/ are apical sounds. Attempting to define the

© 2017. John Benjamins Publishing Company
All rights reserved

class of blockers in terms of place of articulation also seems misguided. While all of the inhibitors could be classified as [−anterior] (assuming [−anterior] assimilation in rt/rd clusters), the natural class of [−anterior, +coronal] would include the palatals, which are not inhibiting segments. In §3, a phonetic explanation is proposed for the class of blocking environments, grounded in theories of coarticulation with special reference to the tongue tip and tongue body.

In addition to Zuberoan, the northern subvarieties of Mixean Low Navarrese Basque also show /u/-fronting (cf. Camino 2009a, 2009b: 69). The only notable difference in these two varieties is the group of segments that inhibit the process. In Mixean, the fronting of /u/ was inhibited before /s̻/, /ɾ/ and the clusters /rt^h, rt, rd/, as in Zuberoan. However, in addition, fronting did not occur before the two velar obstruents /k/ and /g/ (cf. Lafon 1999 [1962b]: 105–106; Camino 2009b: 70–71), where /g/ is phonetically realized as [ɣ̞] after a vowel. Examples like those in (1c), which were systematically fronted in Zuberoan, do not show fronting in Mixean Low Navarrese, as illustrated in (5).

(5) Maintenance of /u/ before a velar consonant in Mixean Low Navarrese

Standard Basque	**Literary Zuberoan**	**Mixean Low Navarrese**	**IPA**	**Gloss**
suge	*süge*	*suge*	[s̻uɣ̞e]	"snake"
sukalde	*sükhalte*	*sukhalde*	[s̻uk^{h}alde]	"kitchen" (with root *sü* "fire")
ukan	*ükhen*	*ukhan*	[uk^{h}an]	"to have"
uko	*ükho*	*ukho*	[uk^{h}o]	"refusal"
nauk	*nük*	*nuk*	[nuk]	"I am (ALLOC.M)"
lukainka	*lükhainka*	*lukhainka*	[luk^{h}a͡iŋka]	"spicy sausage" (cf. Vulgar Latin *lucanicia*)

As a consequence, in addition to (4), I also propose an explanation for why Mixean has the pattern of /u/-fronting described in (6).

(6) Definition of /u/-fronting in Mixean Low Navarrese Basque

/u/ > /y/ except /_X

Where X is either:

i. An apical fricative /s̻/,
ii. A tap /ɾ/,
iii. A rhotic + coronal-oral-stop cluster /rt, rt^h, rd/,
or
iv. A velar oral stop /k, g/

© 2017. John Benjamins Publishing Company
All rights reserved

Two additional aspects of Zuberoan sound patterns directly related to the /u/-fronting sound change have obscured the distribution of /u/ vs. /y/: phonological alternations that result from the sound change and apparent exceptions to the patterns discussed thus far. First, as a consequence of the sound change in (4), morpheme alternants have evolved. For example, the Common Basque 2nd person pronoun *zu* and its possessive form *zure* and dative form *zuri* are continued as *zü* /s̬y/, *zure* /s̬uɾe/, *zuri* /s̬uɾi/ in Literary Zuberoan, with an alternation between /y/ and /u/ in the pronominal root. Alternations due to old inherited compounds exist as well. For example, compare Literary Zuberoan *hur* /huɾ/ "water" and *zur* /s̬uɾ/ "wood" with compounds *ühaitz* /yhaits̬/ "river" (< *hur-han-i-tz "water-big") and *zühañ* /s̬yhaɲ/ "tree" (< *zur-han-i "wood-big"), where the tap which inhibits /u/-fronting in the root has been lost under compound formation, allowing the fronting to take place.

At the same time, Romance loans, as well as analogical leveling, have given rise to clear exceptions to the general patterns of fronting and fronting-inhibition described above. The borrowed Romance suffix /-yɾa/ (cf. Gascon /-yɾa/) shows a front /y/ before a tap, as in Zuberoan /berdyɾa/ "verdure". Internal to Zuberoan, a stem like *larrü* /lary/ "skin, leather", from Common Basque *larru*, maintains the final /y/ in derived forms such as in *larrüsta* "cover with leather", where the suffix-initial segment is expected to inhibit fronting. Note, however, that the form of the suffix may have been *-zta* – with a non-inhibiting sibilant, as attested in other varieties – when the fronting occurred.

2.2 A revised description based on the oldest Zuberoan texts

In order to understand the distribution of /y/ in the oldest attested stages of Zuberoan Basque, I have examined the oldest texts written in this dialect. The results of this second study are compared to those of the first survey in §2.1 in order to highlight differences in the distribution of /y/ in the attested stages of Zuberoan and details of the sound change and its inhibition which were not noted in the earlier literature. Most of the documents for this second study date from the 17th century, although the 14th century *Censier gothique* is a rare witness of an older stage. The full set of documents is: *Le censier gothique de Soule* (14th century, edited by Cierbide 1994); Etxart (from 1616–1617, edited by Lakarra et al. forthcoming); Bela (17th century, included in Michelena 2011 [1964]: 168–171); Zalgize (17th century, included in Sarasola 1983: 173–178); Oihenart (2003 [1657]); Tartas (1995 [1666], 1996 [1672]); *Pronus* (Agirre Sarasola 1998 [ca. 1676]); and Belapeire (1997

© 2017. John Benjamins Publishing Company
All rights reserved

[1696]). Texts from the 18th and 19th century have also been used for comparison and are noted where relevant.[4]

Le censier gothique de Soule (Cierbide 1994) is the oldest known text which includes words in the Zuberoan dialect. It is a long text written in Gascon from the end of the Middle Ages and it contains many Basque house names from the region of Zuberoa (French *Soule*). The document does not have a date, and it is only known thanks to an attested copy (or vidimus) made in 1690. According to Jaurgain (Cierbide 1994: 16), the census was due to a survey ordered by the King of England in 1377. Orpustan (1990: 147) places it in the 15th century but later proposes the end of the 14th century instead (Orpustan 1999: 31, 75), presumably following Jaurgain's observation.

The writer of the *Censier* does not seem to have any clear differentiation of /u/ and /y/. There are some instances of the digraph <ou> in the document, but these were probably added by the 17th century copyist, given that the use of <ou> for /u/ in Gascon began in the 16th century (cf. Cierbide 1994: 48). Nevertheless, Orpustan (1999: 74–76) suggests that the appearance of assimilated vowels in variants and derivatives of Standard Basque *iturri* "source, fountain" may point to the front rounded vowel being already present in Zuberoan by the 14th century. The forms in question include: <uthurria> /yt^{h}yria/, <uthurraltea> /yt^{h}yraltea/; and, especially, the variants with high vowel assimilation together with palatalization of the intervening stop as in <utchurrie> and <utchurry>, /yc^{h}yria/ and /yc^{h}yri/.[5] In addition to the *Censier*, Orpustan (1999: 75) discusses the form <uthurburie> /yt^{h}yrbyɾia/, attested in Béarn in 1385. This form shows boundary unrounding of /y/ > /i/ (*ithürburü- > *üthürbürü- + -a* > *üthürbüria*), which is necessarily more recent than u-fronting.

A recently discovered early attestation of the Zuberoan dialect is the correspondence between the Roncalese Ros and the Zuberoan Etxart from 1616–1617 (Lakarra et al. forthcoming). Etxart clearly differentiates /y/ from /u/ in the contexts of blocking discussed in §2.1: compare <goure> "our", <gourequi> "with us", <hirour> "three", <hirourec> "the three (ERG)", <irourac> "the three (ABS)", <hamairo-urguerren> "thirteenth", <ikoussi> "to see", <ikoussiric> "having seen", <ikoussiren> "to see (FUT)", <ouste> "opinion, idea", <eraccoutsi> "to show", <ourthe> "year", <ourthez> "for years", <ourthian> "in the year", <ourthiaren> "of the year", all with inhibition of fronting, to <uda> "summer", <guciac> "all", <hartu> "take", <çu> "you

4. I'm grateful to Joseba Lakarra and Manuel Padilla for helping me find certain old Zuberoan texts.

5. Nevertheless, assimilation of a back rounded vowel /u/ with previous palatalization of the stop due to the etymological /i/ (cf. <itchurie>, also in the *Censier*) cannot be ruled out for these examples.

© 2017. John Benjamins Publishing Company
All rights reserved

(SG)", where <u> corresponds to the fronted vowel /y/. A few words show variation, including *zure* "your (SG)", written as <çure> 17 times but as <çoure> 3 times. Recall that the uninflected form of this pronoun is *zü* "you (SG)", so, here, it is possible that the alternation in vowels has played a role. On the other hand, the inflected forms of *gü* "we" do not show variation, and this author may have accommodated to his interlocutor, so the reason for this variation is unclear.

The Zuberoan writer Bela compiled a total of 45 proverbs around the beginning of the 17th century (Michelena 2011 [1964]: 168–171). Bela systematically differentiates /u/ <ou> from /y/ <u>. The contexts of blocking are shown by words such as <çouretic> "from the wood", <goure> "our", <houretaric> "from the waters", <eztakoussa> "s/he doesn't see", <dakoussanac> "the one who sees", <oustegabearen> "of the unexpected" and <bousti> "wet".

The Catechism written by Belapeire in 1696 provides additional evidence for the distribution of /y/ vs. /u/ in the 17th century. The examples of inhibition of u-fronting in this text include <houra> "that, she/he", <hour> "water", <barour> "fasting", <barourtu> "to fast", <hirour> "three", <hamirour> "thirteen", <goure> "our", <gouri> "to us", <çoure> "your", <çourequi> "with you", <çourez> "of wood", <iracourt> "to read", <iracourçaler> "to the readers", <iſſouri> "to spill", <itchouran> "in the appearance", <itçour> "to escape", <chouritarçuna> "excuse", <ourthe> "year", <ouste> "opinion, idea", <ikhouſſi> "to see", <eracoux> "to show", <eracouſten> "showing", <houx> /hut̪s/ "mistake, empty". This last example is important as it suggests that the apical affricate /t̪s/ was also a segment that inhibited /u/-fronting, a topic discussed below.

Another text consistently distinguishing /u/ and /y/ is the anonymous Zuberoan translation of the *Pronus singulis diebus Dominicis clarè et distinctè immediatè post Evangelium populo legendus* (c. 1676) thought to have been written by Belapeire (Agirre Sarasola 1998: 22). In this text the author consistently uses <u> for fronted /y/ and <ou> for the cases of inhibition of the fronting, as in <houra> "that", <goure> "our", <gouri> "to us", <çoure> "your", <çouri> "to you", <çourequi> "with you", <hirour> "three", <barour> "fasting", <ourthia> "the year", <houx> "only", <houxic> "empty".

However, not all old texts provide clear evidence of the sound changes discussed in this paper. One of the classic works in the Zuberoan tradition is the collection of proverbs and poems written by Arnaud Oihenart. Although Oihenart specifically mentions the use of /y/ in the regions of Zuberoa and Mixe in the book's preface[6] (Oihenart 2003 [1657]), giving us an unusually concrete picture

6. "Vocalic V is pronounced in Basque like 'ou', as in Italian and Spanish, except in the region of Soule and a few areas in Lower Navarre, where it is pronounced as French 'u'" ("L' V voyelle se prononce en Basque comme, ou, tout ainsi qu'ès Langes Italienne & Espanole, excepté au païs

© 2017. John Benjamins Publishing Company
All rights reserved

of the dialectal distribution of this vowel, Oihenart did not write his texts aiming for a Zuberoan public but rather for a wider Basque-speaking public. As a consequence, Oihenart did not use his native Zuberoan dialect in his work but looked for forms closer to the Lower Navarrese dialect (cf. Michelena 2011 [1964]: 171). Thus, Oihenart shows no graphic difference between /u/ and /y/ in his proverbs nor in his poems, and he uses <u> (or the variant <v>) for both vowels. For example, where /y/ is expected, one finds <ahuns> "goat", <ehun> "hundred" and <leku> "place", and the same where /u/ is expected: <vstea> "opinion, idea", <buſti> "wet", <barur> "fasting", <hirur> "three", <urde> "pig" or <urte> "year". The use of <ou>, the sequence used to represent /u/ by his contemporaries, is kept to a minimum in his texts. Most instances of <ou> are found in his poetry and are consistently used to represent the vowel /u/, with examples such as <ourdin> "blue" and <ehoula> "in no way". There is one case in which Oihenart seems to make an orthographic distinction between /u/ and /y/, probably due to the proximity of the minimal pair in the text: <ehoun> /ehun/ "nowhere" (Standard Basque *inon*) vs. <ehun> /ehyn/ "hundred" (Standard Basque *ehun*), both in §IV, ¶7, p. 13 of his poetry (see Altuna & Mujika's introduction to Oihenart 2003 [1657]: 38).

Sauguis or Zalgize (Sauguis 1908–1909; Sarasola 1983: 173–178) is another Zuberoan author who seemed to avoid representing dialect differences in his orthography (cf. Sarasola 1983: 182). Zalgize compiled a total of 205 proverbs, none of which shows a difference between /u/ and /y/. He represents all instances of /u/ and /y/ with <u>, the only exception being the word <ourde> "pig" (Standard Basque *urde*), in which he (probably unintentionally) shows the non-fronted nature of the first vowel by using <ou>.

Tartas is yet another classic Zuberoan author who does not distinguish the two rounded high vowels orthographically. *Onsa hilceco bidia* (1666) does not show any graphic difference between the non-fronted vowels in the introduction, in words such as <çure> "your (SG)", <çuri> "to you (SG)", <hura> "that" or <iccuſtera> "to see" vs. <çu> "you (SG)", <mundu> "world" or <lurrian> "in the Earth", and the same is true in most of his book. Nevertheless, in the first chapter, the author intersperses forms with <u> with forms with <ou> for words with the vowel /u/. As an example, page 12 shows five instances of the word *gure* "our", three of them written as <goure> and two as <gure>. Tartas only uses <ou> in the beginning of his first book, where it seems to be a consistent writing of /u/. In the following chapters, as well as in his second book *Arima penitentaren occvpatione devotaq* (1672), he consistently uses <u> for all examples of /u/ and /y/ where expected under this analysis. Words written with <ou> in inhibition contexts include: <goure> "our", <icouſſi>

de Soule, & en quelques endroits de la Basse-Navarre, où il se prononce comme l'u François"; Oihenart 2003 [1657], my translation).

© 2017. John Benjamins Publishing Company
All rights reserved

"see", <icouſteco> "to see", <irakhourtçen> "reading", <houra> "that", <eracouſten> "showing". Each of Tartas' books has some introductory or final short texts, some of them written in clear Zuberoan dialect, and these distinguish the high rounded vowels consistently. The dedication by P. Darhetz (Tartas 1995 [1666]) includes the words <çouré> "your" and <iracourtzen> "reading", the text by Bonnecasse (Tartas 1995 [1666]) includes <IRACOVRTV> "to read", and the second text by Bonnecasse (Tartas 1996 [1672]) includes <iracourtu> "to read" and <ourthian> "in the year".

Examination of these texts allows a more comprehensive description of /u/-fronting with respect to the class of inhibiting consonants and the evolution of diphthongs. Most of the examples of inhibited fronting in the oldest texts correlate with those in the initial survey. These words include words derived from *hura* "that", *hur* "water", *gure* "our", *zure* "your", *zuri* "white", *zura* "wood", *hirur* "three", *barur* "(to) fast", *axuri* "lamb", *isuri* "to spill", *itzuri* "to escape", *irakurtu* "to read", *urte* "year", *urde* "pig", *urdin* "blue", *ikhusi* "to see", *itsusi* "ugly", *uste* "opinion", *busti* "wet" and *gustü* "flavor".

The tap /ɾ/ is a consistent inhibitor from the earliest texts showing a /u/ vs. /y/ contrast. One word that shows a non-fronted vowel in the oldest texts but fronting in the modern language is *itxura* "appearance", *ütxüa* in modern Zuberoan. Modern Zuberoan *ütxüa* appears as *itxura* in Belapeire (1997 [1696]) and in Maister (1757). The intermediate form *itxüra*, with fronting but no assimilation of the initial /i/ to /y/, appears at least in Maister (1757), Eguiateguy (1983 [1785]) and Archu (1848). The form *ütxüra*, with fronting and assimilation of the high front vowels, can be found in later texts such as the 19th century Inchauspe (1852) and *Catuchuma españoul* (1899, in Padilla-Moyano 2017). The irregular *itxura* > *itxüra* > modern *ütxüa* might be due to analogy with the suffix *-üra*. Recall that this suffix has been borrowed directly from Romance with a fronted vowel; furthermore, it is a high frequency suffix in both Zuberoan and Gascon, and it is expected to exert an analogical influence on similar wordforms.

Several lexical items suggest that a new segment must be added to the list of inhibiting consonants, namely the apical affricate /t̪s/. In the oldest texts, *huts* "empty; void; fault; mistake" and *erakutsi* "to show" both have non-fronted vowels, though the former is continued as *hüts* in modern Zuberoan. Confirmation of earlier *huts* is found in other texts: *huts* occurs at least once in *Pronus* and seven times in Belapeire. In contrast, *hüts* appears systematically in the 18th century writer Maister, as well as in later authors. Peillen's (1992: 252) statement about the change of *huts* to *hüts* occurring at the end of the 18th century can be refined: 18th century authors such as Maister and others used the fronted variant *hüts* systematically, but there was already variation between *huts* and *hüts* in the 17th century *Pronus*. The shift from *huts* to *hüts* during this period is difficult to explain but could be due to

© 2017. John Benjamins Publishing Company
All rights reserved

infection from a distinct lexical item, *ütz* "leave". Additional evidence confirming /t̪s̪/ as an inhibitor comes from Mixean Low Navarrese, where the modern dialect still shows *huts* and *erakutsi*. In sum, this additional evidence suggests that the sound change of /u/-fronting in Zuberoan should be modified as in (7).

(7) Definition of /u/-fronting in Zuberoan Basque
/u/ > /y/ except /_X
Where X is either:
i. An apical fricative or affricate /s̪/, /t̪s̪/
ii. A tap /ɾ/,
or
iii. A rhotic + coronal-oral-stop cluster /rt, rt^{h}, rd/,

A phonetic explanation for this set of inhibitory segments is offered in §3.

Other descriptive details gleaned from the oldest texts inform our general understanding of the reflexes of Common Basque *u*. Here I comment on an assimilatory process eliminating earlier /y/ from the stem-determiner boundary. This study allows us to confirm an intermediate stage of glide epenthesis and to date the completed sound change to the 17th century. I also suggest a stage of Zuberoan with incomplete fronting of diphthongs /au/ and /eu/ at the beginning of the 17th century.

The oldest texts provide a clearer picture of the development of processes triggered by the addition of a vowel to a word ending in a vowel.[7] Modern Zuberoan has three prevocalic boundary processes: -e+V > -iV, -o+V > -uV and -ü+V > -iV. Compare *bele* "crow" with *belia* "the crow", *beso* "arm" with *besua* "the arm" and *negü* "winter" with *negia* "the winter".

Boundary e-raising (-e+V > -iV) is nearly systematic from the first texts of the 17th century, such as Etxart (Lakarra et al. forthcoming) from 1616–1617 (<baquia> "the peace", from *bake*, <doblia> "double (DET)", from *doble*),[8] so it may be assumed that this process is the oldest of the three. Boundary raising of -o+V to -uV is variable in the first authors of the 17th century (Bela's <escoüa> /eṣ̪kua/ "the wax" from *ezko* but <bilhoa> "the hair", <otsoa> "the wolf" and <Elçoa> "mosquito", from *bilho*, *otso* and *eltxo*, respectively; cf. Michelena 2011 [1964]: 168–171), while later authors consistently show raising.

The case of boundary /y/ > /i/ (-ü+V > -iV) seems to be more complex. Bela shows maintenance of /y/ and glide epenthesis. Examples of /y/ maintenance are

7. These processes are also found in other Basque dialects, cf. Michelena 2011 [1977]: 89–101.

8. Bela (Michelena 2011 [1964]: 168–171) shows e-raising: <Gossia> "the hunger", <umia> "the kid", etc., but also <luzea> "long (DET)" (cf. the bare forms *gose*, *üme* and *lüze*).

© 2017. John Benjamins Publishing Company
All rights reserved

<cençua> "the sense", stem *zentzü*, <eskua> "the hand", stem *eskü* and examples of glide epenthesis include <lekuya> "the place", stem *lekü*. Etxart (1616–1617; in Lakarra et al. forthcoming) probably shows a more developed state of the evolution: he shows boundary epenthesis alongside cases of boundary unrounding: <lekuyetan> "in the places", <abizuia> "the warning", <manuia> "the command", <Parlemenduian> "in the Parliament", <Conseilhuiac> "the Council" vs. <ordian> "then, in time", stem *ordü* "hour", <governamendian> "in the government", stem *gobernamendü*, <abentiaren> "of December", stem *abentü*. Boundary glide epenthesis may have been the stage preceding the pattern found in Modern Zuberoan: *mündü – mündüa* > *mündü – mündüya* > *mündü – mündia* "(the) world". This intermediate stage was first observed by Lafon (1999 [1937]: 87), who noticed that monosyllabic words maintained glide epenthesis in their determined forms for a longer period of time: *sü – süya* "(the) fire", *blü – blüya* "(the) blue" and *thü – thüya* "(the) saliva". The intermediate pattern of older stages of Zuberoan is similar to that found in the neighboring Low Navarrese dialect (where -u+V > -ujV) but involving /y/ instead of /u/. The neighboring Roncalese dialect shows determined forms such as *esku* "hand" + *-a* > *eskiua* "the hand", *buru* "head" + *-a* > *buriua* "the head" (Azkue 1931: 254–255). I suggest that, in Roncalese, a similar pattern of glide insertion is followed by a glide metathesis: -u+V > -ujV > -juV. This would be consistent with the dialectal distribution of the glide insertion, and glide metatheses are attested in Roncalese as well as in other Basque dialects (Egurtzegi 2014a: 195, Example 8.5). Later texts such as Tartas (1995 [1666]), *Pronus* (Sarasola 1998 [1676]) and Belapeire (1997 [1696]) show systematic boundary /y/ > /i/ (<Deebriac Infernian> "the demons in Hell", <ſaintia> "the saint", <eſpiritia> "the spirit", <maniac> "the commands", etc.), with very few exceptions such as <suya> "the fire", from *su*, a form used by both Tartas and Belapeire.

This apparent evolution during the 17th century contrasts with the pattern of <uthurburie> /yt^{h}yrbyɾia/ (Béarn, 1385), the example used by Orpustan (1999: 75) to argue for the presence of /y/ in Zuberoan in the 14th century. If this early attestation is to be taken as a first instance of Zuberoan boundary /y/ > /i/ (instead of a mistake or a later adaptation of the place name), then we could assume that either the process spread very slowly from one variety of the dialect to another or that many of the authors from the 17th century tried to accommodate to a different dialect. While this accommodation is clear in the texts of Oihenart, Zalgize and Tartas and may be a possibility in cases such as Etxart accommodating to his Roncalese interlocutor, this does not seem to be the case for other authors such as Bela or Belapeire. Taking into account that the evolution seems to have a path that can be inferred from the texts (/-yV/ > /-yjV/ > /-iV/), this last possibility can probably be dismissed.

© 2017. John Benjamins Publishing Company
All rights reserved

The oldest text available, the *Censier,* was examined in order to determine whether boundary /y/ > /i/ occurred as early as the 14th century. In this list of house names, many items that end in the determiner *-a* can be found, providing the context for prevocalic boundary processes. Although there are nearly two hundred items with potential contexts for boundary processes, only half a dozen show boundary e-raising.[9] Thus, the oldest attestations point towards an early development of -e+V > -iV. However, at this point in time, there is no evidence of the other two boundary processes.[10]

Recall from §2.1 that the /u/ of Common Basque vowel sequences /au/ and /eu/ does not undergo fronting to /y/. Instead of yielding a front rounded glide, most /Vu̯/ diphthongs were fronted and unrounded to /Vi̯/ in both Zuberoan and Roncalese (cf. Michelena 2011 [1977]: 76). Examples of fronted /Vu̯/ diphthongs from Literary Zuberoan are shown in (8).

(8) Fronting of /Vu̯/ to /Vi̯/

Standard Basque	**Literary Zuberoan**	**IPA**	**Gloss**
gau	*gai*	[gai̯]	"night"
gauza	*gaiza*	[gai̯ʂa]	"thing"
auzo	*aizo*	[ai̯ʂo]	"neighbor"
laudatu	*laidatü*	[lai̯ð̞aty]	"to laud, to praise"
belaun	*belhaiñ*	[belɦai̯ɲ]	"knee"
leun	*lein*	[lei̯ɲ]	"soft"

However, this fronting was contextually limited, in a manner reminiscent of the inhibitory contexts for /u/-fronting discussed above, as illustrated by the examples in (9).

9. The words identified are *Aguerrie*, *Albistia*, *Barnechia*, *Guaricotchie*, *Iratzssitie*(?) and *Ibarrexia*.

10. There is one potential instance of boundary /y/ > /i/ in *Aramburrie*, but only if the second word was intended to be written <burie>, with a single <r>. The basis of this suggestion is another housename, *Aramburue* (cf. Standard Basque *haran* "valley", *buru* "head"), and the fact that *burri, burru* are not recognizable words in Zuberoan, while *bürü* "head" is. Nevertheless, it is also possible that this is a scribal error and that the second <r> and the <i> were originally an <u>, so that the original word was <burue> /buɾua/, the expected unaltered form.

© 2017. John Benjamins Publishing Company
All rights reserved

(9) Diphthongs with maintenance of /u̯/

a. Before a rhotic /ɾ, r/ or an apico-alveolar sibilant /s̪, ts̪/

Standard Basque	Literary Zuberoan	IPA	Gloss
lau	*laur*	/lau̯ɾ/ [lau̯r]	"four"
hau	*haur*	/hau̯ɾ/ [hau̯r]	"this"
zauri	*zauri*	[s̪au̯ɾi]	"wound"
euri	*euri*	[eu̯ɾi]	"rain"
ausartu	*ausartü*	[au̯s̪arty]	"dare"
euskara	*euskara*	[eu̯s̪kaɾa]	"Basque"
haur	*haur*	/hau̯r/	"child"
hauts	*hauts*	[hau̯ts̪]	"dust"

b. After a palatal glide/fricative /ʒ/ (< *j)

Standard Basque	Literary Zuberoan	IPA	Gloss
jaun	*jaun*	/ʒau̯n/	"sir"[11]
jauki	*jauki*	/ʒau̯ki/	"to commit"
jantzi	*jauntsi*	/ʒau̯nts̪i/	"to dress"
jauzi	*jauzi*	/ʒau̯s̪i/	"to fall, to jump"

As shown by (9a), fronting was inhibited when the semivowel was followed by /s̪/, /ts̪/, /ɾ/ and /r/. Examples in (9b) show inhibition when the palatal glide/fricative precedes the diphthong. Two main differences, then, between this fronting and the general u > y sound change are the inhibitory status of trilled /r/ in the process and the inhibitory role of the syllable-initial palatal.[12]

The fronting of diphthongs is illuminated by the correspondence between Etxart and Ros (1616–1617). Neither Zuberoan nor Roncalese show a consistent pattern until the 17th century. For instance, both use *gauza* "thing" instead of *gaiza*, the form found in the later stages of both dialects. In addition, Etxart uses *gauaz* "at night" instead of *gai-* as well as consistently using verbal forms with the root *-erau-*: *igorten derauziet* "I send (it) to you", *gomendatzen derauzut* "I recommend (it) to you", *respuestu eginen derautazu* "You will give me a response", *eskribaturen*

11. See also Roncalese *jein* "sir", with fronting of the glide.

12. Concerning the forms in (9b), the inhibition of the fronting has been attributed to avoidance of *jau-* > **jai- in the literature, although Michelena (2011 [1977]: 77) expresses doubts about this statement and adds that some instances of *-au-* in verbal radicals may be secondary. Note, however, that had the words in (9b) not diphthongized by the time the general Zuberoan u-fronting occurred, they would show a fronted /y/ instead of a velar semivowel. Word-initial yod is maintained in some Basque dialects, although modern Zuberoan shows [ʒ] instead. Rising diphthongs involving /w/ were unaffected by the process, as they probably evolved after this sound pattern was developed (cf. Egurtzegi 2014a: 126).

© 2017. John Benjamins Publishing Company
All rights reserved

derautazula "that you will write to me", etc. Nevertheless, Etxart systematically writes *aizo* "neighbor", showing that fronting of diphthongs was occurring in some words in his variety. Ros does not show any instance of diphthong fronting in the few potential cases where it could occur (e.g., *gauza* "thing"). Tartas (1995 [1666]) does not show fronting in diphthongs either in his most accurately written first chapter (*gauza*, *gauaz* "at night"), nor is it observed in the short texts and dedications written in Zuberoan by other authors in his books (e.g., *lauda* "to laud"), but there are not as many potential examples as in Etxart and Ros' correspondence. Other 17th century works such as *Pronus* (c. 1676), Belapeire (1997 [1696]) and the undated text of Bela (who died in 1667), regularly show fronted diphthongs.

In sum, the oldest texts reveal details of the evolution of vowel sequences that have not been remarked on before. At the word boundary, unrounding of final /y/ followed by the determiner occurred after glide-insertion, and this process was preceded by the raising of /e/ to /i/ at the same boundary. Word-internally, diphthongs /au/ and /eu/ shifted to /ai/ and /ei/ respectively, with inhibitory contexts similar to but distinct from those detailed above and summarized in (7) for /u/-fronting in Zuberoan. More importantly, if the diphthong fronting is as recent as suggested by the texts, beginning only in the 17th century, they must be distinct from the earlier process of /u/ fronting, which appears to have started at least several hundred years before then. With this review of the oldest texts complete, and the description of Zuberoan /u/-fronting stated as in (7), I attempt to explain aspects of this process. How and why did /u/ undergo fronting, and why was this fronting inhibited in certain contexts?

3. Explaining the u > y sound change

Examination of the oldest texts has allowed us to offer a more detailed description of Zuberoan /u/-fronting and the contexts where fronting did not apply. I now discuss previous approaches to Zuberoan /u/-fronting before presenting a proposal aiming to integrate contact as a causal factor and phonetic factors determining the inhibitory contexts.

3.1 Previous approaches to Basque /u/-fronting

Since Zuberoan /u/-fronting was first described more than a century ago, different authors have refined its description and proposed a range of analyses, briefly reviewed here in chronological order.

© 2017. John Benjamins Publishing Company
All rights reserved

Uhlenbeck's treatment of Zuberoan describes the distribution of /u/ vs. /y/ in relation to the segment that follows the vowel. Uhlenbeck (1903: 29–31) describes a general /u/ > /y/ process with maintenance of /u/ before /ɾ/ and /s̪/ but does not mention other blocking contexts.

Gavel (1920: 46–69) argues generally for a Bearnese origin of Zuberoan /u/-fronting. He describes a distribution of /y/ similar to that of Uhlenbeck (1903) but adds some cases of maintenance before "preconsonantal r", followed by some seemingly similar cases that show fronting instead. In order to account for presence or absence of fronting in the same contexts, he hypothesizes that unfronted /u/ may have been followed by a tap in older times, while the rest were followed by a trill (1920: 45). Gavel treats some of the exceptions to the general pattern in terms of analogy and others as due to the influence of Bearnese Gascon (1920: 42–44).

Lafon dedicated several papers to Basque /u/-fronting (cf. Lafon 1999 [1933]; 1999 [1937]; 1999 [1962a]; 1999 [1962b]). In addition to the already established inhibitory effects of /ɾ/ and /s̪/, Lafon (1999 [1962a]: 97) proposes that /u/ was also maintained before /nk/ clusters, although he only offers one word as evidence (*hunki* "to touch"). Lafon (1999 [1933]) also discusses the maintenance of /u/ before rC clusters, as first described by Gavel (1920: 45): According to Lafon (1999 [1933]: 75), /u/ was only maintained when followed by a cluster formed by a rhotic and a dental stop /d, t, tʰ/ or a rhotic and an affricate /t̪s̪, t̪s̪, ʧ/. Nevertheless, the only evidence Lafon offers for maintenance of /u/ before rhotic + affricate clusters is *murtxatü* "to suck". Lafon went farther than Gavel in his discussion of contact-induced change. He proposed that contact and bilingualism with Gascon triggered Basque /u/-fronting (1999 [1962a]: 108–111; 1999 [1962b]: 139–142) but that the different "articulatory habits" of Basque speakers conditioned it (1999 [1962b]: 142). He assumed that /u/-fronting was borrowed either at the same time or shortly after it developed in Gascon, giving preference to the former possibility (1999 [1962a]: 111). Lafon also assumed that /u/ > /y/ preceded /o/ > /u/, given that late /u/ was unaffected by /u/-fronting (1999 [1962a]: 102). Lafon (1999 [1937]: 87) was the first to propose the intermediate stage /y.ja/ in the evolution of boundary unrounding of /y/ (/y.V/ > /y.jV/ > /i.V/), after observing that monosyllabic words showed glide epenthesis in their definite forms: *sü – süya* "(the) fire", *blü – blüya* "(the) blue" and *thü – thüya* "(the) saliva". My analysis of the oldest Zuberoan texts in §2.2 provides further support for these developments.

Zarabozo (1972) was the first to try to propose a unified account of the inhibitory contexts of Zuberoan /u/-fronting. Zarabozo defines the natural class of /ɾ/ and /s̪/) in terms of the phonological feature set [+cor, +cont, –tense, –distr, –lat] (Zarabozo 1972: 176), which he later redefines as [+cor, +back, –tense, –lat] (1972: 177). His use of [–tense] is meant to include the tap and exclude the trill. Zarabozo does not discuss rT clusters, although these clusters were already

© 2017. John Benjamins Publishing Company
All rights reserved

established as inhibitory contexts by Lafon (1999 [1933]) and the first edition of Michelena (2011 [1977]) in 1961. In addition to rT clusters, Zarabozo (1972) does not treat /t̪s/ as a blocking segment. Finally, the natural class Zarabozo defines will include palatals, however, as noted earlier, fronting does occur before palatals in Zuberoan.

Michelena (2011 [1977]: 41–42) places /u/-fronting prior to the oldest Zuberoan documents and describes the contexts where it did not develop as before /ɾ/, /s̪/ and before a cluster formed by a rhotic and an apical stop rT. Michelena discards Lafon's (1999 [1962a]) proposal that /u/-fronting was inhibited in *hunki* "to touch" due to the /nk/ cluster, since /u/ in this case derives from older *o. In addition, Michelena argues against /rʧ/ as a blocking cluster, first by stating that Lafon (1999 [1933]) does not give enough evidence for it (2011 [1977]: 42) and later by finding an etymology for Lafon's only example *murtxatü* "to suck" in Latin *mors-* (*mordere*) "to bite" (2011 [1977]: 360). Michelena (2011 [1977]) refines the contexts of maintenance of /u/ that Uhlenbeck, Gavel and Lafon previously described but does not propose an analysis, leaving the description in the form of a list. Nevertheless, he presents a hypothesis based on reconstructed allophones of /o/ ([ɔ, o]) and /u/ ([ʊ, u]) that, after a new distribution, give rise to a third phoneme (2011 [1977]: 44): [ɔ] > /o/, [o, ʊ] > /u/ and [u] > /y/. Michelena's description of the inhibitory contexts differs from the description in (7) in not including /t̪s/ as a blocking segment.

After Michelena's work, the discussion on Zuberoan /u/-fronting did not advance until Oñederra (2009a, 2009b). Most authors mentioning this process just echoed previous descriptions. In his book on contact and sound change in continental Basque, Haase (1992: 41–42) briefly discusses /u/-fronting by describing it and linking it to Gascon. He probably follows Lafon in his description of the inhibitory contexts, given that he includes /rʧ/ clusters among the environments where /u/ was maintained (Lafon 1999 [1933]: 75). Hualde (1993: 291) discusses /u/-fronting briefly and mentions only /ɾ/ and /s̪/ as inhibitory contexts, while Trask (1997: 150) lists the environments of the change as previously described by Michelena (2011 [1977]). More recently, in his overview of Basque dialects, Zuazo (2008: 44) echoes the inhibitory contexts listed by Michelena (2011 [1977]) and states that Zuberoan /u/-fronting is due to influence from Gascon.

Oñederra (2009a) presents /u/-fronting in Zuberoan as an example of a process which resists explanation in terms of its motivation, stating that "A clear phonetic explanation seems unreachable in light of the available data" ("Una explicación fonética clara parece inalcanzable a la luz de los datos disponibles", my translation) (2009a: 56). Oñederra (2009a: 58, 2009b: 671) follows Michelena (2011 [1977]: 41–42) in not considering /t̪s/ an inhibitory segment. Following Bichakjian's (1974) proposal for French /u/-fronting, Oñederra (2009a: 62) reconstructs a multi-stage pathway involving diphthongization and monophthongization (*u* > *uw* > *iw* > *y*)

© 2017. John Benjamins Publishing Company
All rights reserved

for Zuberoan /u/ > /y/. Though she suggests that it was diphthongization that was inhibited, not fronting, invoking opaque prosodic factors as potential triggers, she presents no evidence to support this proposal (Oñederra 2009a: 63, 2009b: 676). Oñederra's (2009a) proposal is similar to earlier accounts in invoking contact with Gallo-Romance; however, in a more recent paper, Oñederra (2009b: 675) leaves contact aside.

The present analysis builds on the description of Zuberoan /u/-fronting first presented by Uhlenbeck and Gavel and later refined by Lafon and Michelena. It adds another inhibitory segment (/t͡s/) as observed in the oldest attestations. In contrast to Oñederra (2009b), a central role for contact is acknowledged (§3.2), and a phonetic explanation grounded in typological and articulatory studies is proposed (§3.3).

3.2 The role of contact

Our analysis of Zuberoan involves a general shift of /u/ > /y/ with a closed class of post-vocalic consonants or consonant clusters inhibiting the shift (7). Languages with a context-free high back vowel fronting include Gallo-Romance languages such as French (Harris 1988: 210) and Occitan (Wheeler 1988: 247), the Gallo-Italic languages (Piedmontese, Lombard, Emilian, Romagnol, etc.) and Arpitan (Franco-Provençal). Outside of Romance, a similar sound change is described for Greek (Fortson 2010: 253), Albanian and the Lolo-Burmese language Akha (Labov et al. 1972). Some English dialects such as Standard Southern British English (Hawkeys & Midgley 2005; Henton 1983; McDougall & Nolan 2007; Harrington et al. 2008; Harrington 2012), Australian English (Cox & Palethorpe 2001), New Zealand English (Gordon et al. 2004) and Southern and General American English (Bailey 1997; Fridland 2008) have fronting of /u/, but it gave rise to /ʉ/ instead of /y/ and it did not create a new phonological contrast. /u/ > /ʉ/ fronting is also found in Swedish, Açorean and some varieties of European Portuguese (José Ignacio Hualde p.c.). /u/-fronting is one of the sound changes involved in chain shifts as analyzed by Labov (1994: 116).

Though front rounded vowels are quite common in Western Europe, they are less so in other parts of the world, which might suggest that a catalyst for the Zuberoan u > y shift was contact with surrounding Romance languages with /y/. Only 6.6% of languages (37 out of 562) in the WALS database (Maddieson 2013) have front rounded vowels. In addition, 78% of the languages with front rounded vowels in the survey (29 out of 37) are found in the North-Central area of the Eurasian continent. Outside of this area, languages with front rounded vowels are scarce (cf. Blevins 2017).

© 2017. John Benjamins Publishing Company
All rights reserved

As noted earlier, Basque front rounded vowels have not evolved in isolation. Zuberoan is spoken in an area adjacent to where Romance languages that also have front rounded vowels are spoken, and, indeed, this is part of a larger region where front rounded /y/ may be seen as an areal feature. Given this, it seems likely that contact has played a role in the development of a cross-linguistically uncommon sound pattern. However, determining to what extent this contact has influenced the evolution of areal /y/ is non-trivial. If /y/ evolves from /u/ in native vocabulary, how does the shift occur?

Evolutionary Phonology (Blevins 2004, 2015, 2017) highlights the evolution of regular sound change in the context of areal contact and may be useful in understanding how sound patterns like the /y/ vs. /u/ contrast may spread. Blevins (2017) suggests that language experience in a multilingual environment alters phonetic perception via the 'perceptual magnet effect' (cf. Kuhl 1991, 2000; Kuhl & Iverson 1995). The central idea is that when first acquiring a language, proto-categories act as magnets, drawing nearby perceptual stimuli into them. In language contact situations, continuous exposure to a second language may result in a warping of perceived distances of phonetic tokens in the first. Blevins (2017) argues that in situations of long-term bilingualism, an external phonetic prototype may be internalized and act as a perceptual magnet in the first (or other) language of the infant.

Crucial to this model are the notions that the establishment of a phonetic prototype requires perceptual saliency of the segment involved, as well as intense language contact spanning multiple generations. An interesting prediction of the model is that "sound change will appear to be natural and phonetically motivated and indistinguishable from internal developments" (Blevins 2017). In general terms, the model demonstrates how sound patterns can spread via contact independent of lexical borrowing, phoneme borrowing or borrowing of specific sound patterns or sound changes.

Turning back to Zuberoan, we see that all of the ingredients for a /u/ > /y/ shift are present. Zuberoan has /u/, and speakers of the language have had long-term intense contact with Gascon, a language with a /u/ vs. /y/ contrast. The vowel /y/, like other front rounded vowels, is, arguably, perceptually salient. If it were to act as a perceptual magnet as understood by Blevins (2017), it would draw tokens of the phonetically close /u/ even closer to its prototype when speakers of a language which lack /y/ are in close and continued contact with speakers of a language that has /y/ in its phonological inventory.[13]

This is the scenario I propose for speakers of Zuberoan and Mixean Basque in contact with speakers of Bearnese Gascon. Recall that Zuberoa was part of Béarn

13. The model developed in Harrington & Schiel (2017) makes the same prediction based on different assumptions.

© 2017. John Benjamins Publishing Company
All rights reserved

in the Middle Ages (Haase 1992: 19) and Bearnese Gascon was the only written language in use in Zuberoa until the 17th century (Séguy 1952: 385), as well as being the official language of the region and the language of scholarly communication (Gavel 1920: 47). Thus, many Zuberoan speakers were bilingual in Gascon for centuries (Lafon 1999 [1962a]: 108). This has resulted in a great number of Gascon loanwords in the two dialects (for instance, see the long list of borrowings with a stressed nasalized vowel in Egurtzegi 2014a: 156, Example 6.9) and many shared sound patterns including loss of intervocalic *n, development of nasalized vowels, similar stress systems, /o/-raising and /u/-fronting (cf. Egurtzegi 2014a). However, it should be clear that /u/-fronting is not the result of Romance loan words coming into the language. Fronting of /u/ takes place in inherited Basque words and is sensitive to properties of inherited Basque consonants that follow the target vowels. Loanwords or borrowed morphemes (e.g., Romance /-yra/) may constitute exceptions to the general pattern. The inhibiting segments described in (7) are specific to Zuberoan, and there is no known language in the world that has a sound change with these conditions from which the Basque pattern could be directly borrowed. In sum, Zuberoan fronting is a unique sound change integrating an external stimulus with internal phonetic and phonological conditions. In the next section, I suggest phonetic factors underlying the unique inhibitory contexts.

3.3 The phonetic basis of inhibited fronting

Why do seemingly context-free vowel shifts occur so often in the world's languages? The pronunciation of a vowel is variable, and the extent of this variability may be limited by the division of the vocalic perceptual space in each language (Bradlow 1995). The perception of a given vowel is not categorical. Instead, some exemplars (prototypes) are judged to be more typical than others. Prototypes make the perceptual distance between them and the exemplars that surround them smaller than psychophysically expected, warping these exemplars into the same category. This is known as the 'perceptual magnet effect' (Kuhl 1991, 1995; Kuhl & Iverson 1995; see also Blevins 2004: 286).

In addition, according to the exemplar theory of speech (Johnson 1997; Pierrehumbert 2001), "a new token which is well positioned with respect to a category can actually provide a better example of that category (in being recognized quickly and rated highly) than any actual example of that category that has been previously experienced" (Pierrehumbert 2001: 143). If no phonological category is assigned to a neighboring psychoacoustic space, prototypes may move over time, effectively dragging all non-prototypical exemplars with them. Thus, context-free vowel shifts may be expected when acoustic space is available.

© 2017. John Benjamins Publishing Company
All rights reserved

In Gallo-Romance, systematic /u/-fronting has been attributed to a push-chain (cf. Labov 1994: 116). The raising of the mid back vowel /o/ to /u/ would have overcrowded the high back vowel space, and that would have triggered the fronting of /u/ to /y/ (Haudricourt & Juilland 1970: 114).

Harrington (2012) argues that context-free fronting of /u/ starts from coarticulatory contexts. Coarticulation can be a source of various kinds of sound changes (Ohala 1981, 1993), like vowel harmony (Beddor et al. 2002) or vowel nasalization (Hajek 1993).

According to Harrington (2012: 104), perception-production relationships tend to be aligned in coarticulatory patterns (Fowler 2005), and it is only during a sound change in progress that production and perception are misaligned. In this scenario, the actual sound change would occur as the context-dependent and context-independent phonetic variants come closer together and the perceptual compensation for coarticulatory effects is reduced, giving rise to a new production-perception alignment (Harrington 2012: 104).

Under this account, both perception and production may be involved in the source of the change: contextual coarticulation is no longer compensated by the listener (a case of hypocorrection, cf. Ohala 1993) and the outcome of fronting environments is then phonologized in other phonological contexts (Harrington 2012: 116), given a shift of the variants in non-fronting contexts toward the variants found in fronting contexts. This analysis is consistent with Harrington's observation that the context-less /u/ in younger generations of speakers of English is similar to the /u/ produced in fronting contexts by older generations of speakers of the same variety.

The most usual consonantal coarticulatory situations where a back vowel is fronted are environments involving alveolar consonants (Flemming 2003; Öhman 1966). Harrington et al. (2011) looked for the predisposition towards the fronting of /u/ in German in T_1uT_1 context (where T is a voiceless stop) and found that both the onset and offset of /u/ in /tut/ as well as the onset of /u/ in /kuk/ were well into the /y/ space (Harrington 2012: 106). Further, high back vowels are more prone to diachronic fronting than high front vowels are to retraction, as empirically tested by Harrington (2012: 115–116).[14] Given that coronals seem to trigger /u/-fronting, the class of inhibiting segments specified in (7), all coronals, is surprising. However, when articulatory properties of these sequences are considered, a unified explanation may be possible.

From the typological parallels mentioned above, the case of the context-free fronting of American English /u/ is of special interest. This was a southern feature

14. Centralization is known in lax vowels: /ɪ/ in New Zealand English (Maclagan & Hay 2007; see also Moon & Lindblom 1994).

© 2017. John Benjamins Publishing Company
All rights reserved

but now is described as covering most of the North American continent (Labov 2008: 27). In the American case, a following dark [ɫ] inhibits the process when following the target /u/, parallel to inhibition of fronting before /ɾ, s̪, t̪s̪, rt^{h}, rt, rd/ in Zuberoan Basque. However, this constraint does not occur in Southern American varieties, although it does exist in most non-Southern varieties (cf. Labov et al. 2006: 152).

It may be hypothesized that the inhibition before [ɫ] was active when the sound pattern began to spread northwards, but it is not active anymore. Alternatively, both sound patterns, although clearly areal, may have developed independently. Koops (2010: 113) weighs both possibilities and concludes that these sound patterns represent two processes.[15] Inhibition before [ɫ] is still active in non-Southern dialects, where the degree of fronting is even greater than in the Southern dialects (cf. Labov et al. 2006: 153), although this sound change is described as being already complete there (cf. Baranowski 2008).

The inhibition of English /u/ > /ʉ/ by a following dark [ɫ] may be attributed to the degree of shared gesture of the consonant and the affected vowel. The degree of coarticulation in tautosyllabic /ul/ sequences is extreme, and this, coupled with the 'back' specification for the lateral, is ultimately inhibitory.

The consonants that inhibit /u/-fronting in Zuberoan do not appear to share gross articulatory gestures with the preceding vowel. However, Recasens & Pallarès (2001) suggest reasons for the resistance to coarticulatory processes in a combination of place and manner that may illuminate the nature of inhibition of fronting. As Recasens & Pallarès (2001: 274) point out, tongue dorsum coarticulation data show that highly constrained consonants have large coarticulatory effects in contiguous vowels and can inhibit vowel dependent effects (Fowler & Saltzman 1993). According to them, some consonants involving apical activity of the tongue also require a concrete dorsal placement of the tongue, and this may inhibit neighboring vowels from articulatory processes towards the palatal zone:

> It thus appears that consonants involving demanding manner requirements (and little dorsopalatal contact) block consonantal and vocalic effects at the palatal zone, i.e., apical vibration for r, frication for s and, less so, laterality and the formation of a secondary lingual constriction for dark ɫ. (Recasens & Pallarès 2001: 288)

Thus, this secondary placement of the tongue dorsum may play a role in inhibiting /u/ > /y/ in Zuberoan Basque. While the shift from [u] to [y] involves a fronting in the placement of the tongue, the consonants "involving demanding manner

15. Koops discerns two different kinds of 'fronted u' in the speech of Houston Anglos, which "differ in a number of fine phonetic details" (2010: 119). According to him, these two types of /u/ show the properties of Southern fronted /u/ and the general American palatalized back vowel.

© 2017. John Benjamins Publishing Company
All rights reserved

requirements and little dorsopalatal contact" crucially require a lowering and back placement of the tongue dorsum and thus inhibit /u/-fronting. Recall that the group of segments or clusters that inhibit the process include precisely the expected segments: the apical rhotic /ɾ/ and the apical sibilants /s̪/ and /ts̪/ but not the laminal sibilants /s̺/ and /ts̺/, which deviate from the other sibilants precisely in being produced with the blade of the tongue instead of the tip and thus do not require a back placement of the tongue dorsum. Interestingly, the only rhotic-obstruent clusters included /rt^{h}, rt, rd/ are those involving activity of the tip of the tongue in both consonants. The production of [t] in neutral contexts like /ata/ is lamino-dental, i.e., [at̪a], the realization of the /t/ after a rhotic is more retracted than in intervocalic contexts and may also involve less of the tongue blade. This can shed light on why these clusters inhibit fronting, while sequences like /tut/ favor it. The other rhotic-obstruent clusters do not inhibit fronting, since they involve non-coronal segments as second members.

The segments that inhibit fronting in Mixean Low Navarrese, namely /k/ and /g/, are also produced with a back placement of the tongue dorsum, so that they could potentially create the same coarticulatory effect that the segments that inhibit the process in Zuberoan Basque produce. It is worth mentioning that, as stated above, productions of the sequence /kuk/ by speakers of German show instances of /u/ into the /y/ space but, crucially, only in the onset, while in the case of /tut/ the coarticulation occurs in the offset as well as the onset (Harrington et al. 2011; Harrington 2012: 106). Thus, the VC coarticulation of the sequence /uk/ seems to result in a back vowel, which is consistent with its status as an inhibiting segment in Mixean Low Navarrese.

In short, inhibition of u > y as analyzed here is a consequence of coarticulatory effects due to the tongue dorsum lowering and backing required for the production of the set of coronal segments and clusters /ɾ, s̪, ts̪, rt^{h}, rt, rd/, which demand precise movements of the tip of the tongue.

4. A revised Zuberoan chronology based on /u/-fronting in neighboring Romance languages

In this section I briefly describe the /u/-fronting processes found in the Romance languages that have been in contact with Zuberoan Basque, Gascon and French, before discussing the chronology of the processes in these languages in order to propose an approximate chronology for Zuberoan /u/-fronting. Vulgar Latin long *ū* (/uː/) was regularly fronted to /y/ in French (Bourciez 1967: 94) and Gascon (Rohlfs

© 2017. John Benjamins Publishing Company
All rights reserved

1977: 124).[16] Examples in (10) show this fronting in both Gascon and French with Latin forms for comparison. The examples of Bearnese Gascon are taken from Lespy & Raymond (1998 [1887]) and Palay (1980 [1932–1934]). The transcriptions of Donzacese Gascon are taken from Kelly (1973).[17]

(10) /u/ fronting in Gascon and French

Bearnese Gascon	Donzacese Gascon	French	IPA	Latin	Gloss
utile	/yˈtile/	*utile*	/ytil/	*ūtĭlis*	"useful, helpful"
lue, *lu*[18]	/ˈlyno/[19]	*lune*	/lyn/	*lūnam*	"moon"
ue, *u*	/ˈyno/	*une*	/yn/	*ūnam*	"a, one (F)"
lèytugue	/lɛjˈtygo/	*laitue*	/lety/	*lactūcam*	"lettuce"
dur	/dyr/	*dur*	/dyʁ/	*durum*	"hard"
cuu, *cu*	/kyw/	*cul*	/ky/	*cūlum*	"ass"

Given the widespread presence of /u/-fronting in the Gallo-Romance languages, one might think that this sound pattern developed during an older, common period of these languages, as in Gallo-Romance, for instance, which is usually placed from the end of the 5th century until the middle of the 9th century. However, Hall (1976) does not reflect this sound change in his transcriptions of Early Old French.

The /u/-fronting sound change was systematic in French and has been accounted for by means of a push-chain (cf. Labov 1994: 116). It has been proposed that the raising of the mid back vowel /o/ to /u/ overcrowded the high back vowel space, which subsequently triggered the second step in the chain-shift (Haudricourt & Juilland 1970: 114). This explanation separates the fronting of /y/ from that of /æ/ (/'a/, /aː/ > /aː/ > /æː/ > /ɛ/), which is one of the vowels triggering the second (or Gallo-Romance) palatalization. In fact, /u/ > /y/ fronting took place after the so-called second palatalization, which is taken as the crucial process in the separation

16. Gascon is usually considered an Occitan dialect (although not necessarily; cf. Chambon & Greub 2002) but deviates from the other Occitan dialects to some extent.

17. Examples of the fronting in Occitan include *un* /yn/ "one (M)", *tu* /ty/ "you (SG)", *segur* /segyr/ "safe, sure".

18. Intervocalic /n/ is dropped in Gascon (Anglade 1921: 185) as in Gascon *lüa* /ˈlya/ < Latin *lūnam* "moon", Gascon *üa* /ˈya/ < Latin *ūnam* "a, one (F)", Gascon *prüa* /ˈpɾya/ < Latin *prūnam* "coal" or Gascon *fiestra* /fiˈestɾa/ < Latin *fĕnestram* "window".

19. Latin word-final *-a* is maintained as /a/ in Bearnese, raised to /o/ in most Gascon dialects, including the Donzacese variety in Example (10) – as well as in Occitan – and raised to /e/ in Landes and in the Bayonne-Orthez region (cf. Rohlfs 1977: 125). In addition, the Bayonnese variety drops this final *-e* after a stressed /y/ or /i/: *lu* /ly/ "moon" (< Latin *lūnam*), *u* /y/ "a, one (F)" (< Latin *ūnam*), *hari* /haˈɾi/ "flour" (< Latin *farīnam*), *gari* /gaˈɾi/ "hen" (< Latin *gallīnam*).

© 2017. John Benjamins Publishing Company
All rights reserved

between northern and southern Gallo-Romance, i.e., Old French and Old Occitan (cf. Buckley 2009: 32).[20] The second palatalization began around the end of the 5th century or the beginning of the 6th (apud Meyer-Lübke 1890: 354ff.), but /k/-palatalization did not develop until much later, being usually placed around the 7th century (Fouché 1958: 203ff.; Bourciez 1967: 94–95; Matte 1982: 102; Buckley 2009: 38–39). Examples such as Latin *curam* > Old French *cure* /kyr/ "cure" (and not **/ʧyr/) or Latin *cūpam* > French *cuve* /kyv/ "barrel" (and not **/ʧyv/) provide the evidence for placing French /u/-fronting after Gallo-Romance palatalization (Bourciez 1967: 95; Buckley 2009: 39). Thus, /u/-fronting in French may be placed in the period when the first Old French texts were written, i.e., around the 9th century (Bourciez 1967: 94–95).

Given that /u/-fronting developed after the division of the languages of Oïl and Oc (Old French and Old Occitan), the time frame in Occitan might be different. Latin /uː/ being maintained orthographically after /u/-fronting makes establishing a chronology in Old Occitan more complicated than in Old French (cf. Anglade 1921: 82). Nevertheless, Anglade (1921: 84) finds instances of palatalization of /l/ before front vowels in medieval texts from the end of the 13th century, and these include fronting before graphic <u>. This evidences the fronted status of /y/ in Occitan by the end of the 13th century but does not give us a clear date.

Dating the fronting is even more complicated within Gascon. Rohlfs (1977: 124) acknowledges this and proposes a rather recent time frame, without further temporal specification. Rohlfs (1977) proposes that the sound pattern extended gradually from one region to another and that it was already completed by the time /o/ raised to /u/ before a nasal, given that the latter was unaffected by u-fronting. According to Chambon & Greub (2002: 489), the processes that made Gascon distinct from Occitan began around the 5th century and were completed by ca. 600, so that Gascon was an independent language by the time u-fronting developed within French.

The fronting of Vulgar Latin /u/ (< Latin /uː/) seems to have spread geographically from one Romance language to another and ultimately into Zuberoan; it does not appear to be an inherited innovation of all three Romance languages that Zuberoan took from them afterward. According to Buckley (2009: 39), /uː/ > /y/ can be placed around the 9th century for Old French, but, due to the Occitan writing system not reflecting this sound change, the literature does not give any clear date for Occitan – it only states that /u/ produced consonant palatalization by the end of

20. Northern Occitan – which is usually taken as a transition zone (Lafont 1971: 107) – shows Gallo-Romance fronting: *chevra*/*chavra* "goat", *cheira* "dear", *chen*/*chin* "dog", *chas*/*chies* "('s) house" (Buckley 2009: 57). In any case, Buckley (2009: 59) concludes that this sound-pattern is borrowed from French.

© 2017. John Benjamins Publishing Company
All rights reserved

the 13th century – and even less for Gascon. If Old French were the source of the process, which seems plausible given the direction of the isoglosses, a somewhat later date for the fronting could be assumed in southern languages such as Occitan and Gascon, and even later for Zuberoan Basque. Thus, the only proposal that can be made is that Zuberoan developed the fronting some time after the 9th century, possibly closer in time to the first attestations of u-fronting in Occitan by the end of the 13th century. This makes the potential attestation of Zuberoan u-fronting mentioned by Orpustan (1999: 75) especially important, if it were indeed to be seen as an early attestation of this process in Basque.

5. Conclusions

I have analyzed the fronting of Common Basque /u/ to /y/ in Zuberoan. To that end, I have surveyed two corpora: one based on historical dictionaries and word lists and a second consisting of the oldest texts written in Zuberoan Basque. I have confirmed that the flap /ɾ/ and the rhotic-dental clusters /rt^{h}, rt, rd/ systematically inhibited fronting, as did the apical sibilant fricative /s̪/. I have found old cases of inhibition of the fronting preceding the apical sibilant affricate /ts̪/, not listed by most previous authors, in words such as *huts* "mistake, empty", pronounced *hüts* in the modern language. I have also noted that palatals, which are often included in the class of coronal non-anterior segments, do not inhibit /u/-fronting. Taking everything into account, it is clear that not all coronal segments inhibited fronting of a preceding /u/: fronting of /u/ occurred systematically before /t/, /d/, /l/, /ʎ/, /s̪/, /ts̪/, /ʃ/, /tʃ/ and /r/.

I have proposed that the fronting of /u/ to /y/ was inhibited in these contexts due to coarticulatory effects: maintenance of *u was a consequence of the coarticulation caused by consonants requiring active tongue dorsum lowering and backing. This tongue dorsum placement is required to perform the fine movements of the tongue tip involved in the production of inhibitory segments and clusters /ɾ, s̪, ts̪, rt^{h}, rt, rd/.

The fronting of /u/ not only occurs in Zuberoan but also in the neighboring Mixean variety of Low Navarrese Basque. In most northern subvarieties within this variety, the fronting process has developed similarly to that of Zuberoan. The only difference is found in the group of segments that inhibit the process, which includes the velar obstruents /k/ and /g/, in addition to those found in Zuberoan. Articulatorily, these segments are produced with a back placement of the tongue dorsum, so that they could create the same coarticulatory effect as the segments that inhibit the process in Zuberoan Basque.

© 2017. John Benjamins Publishing Company
All rights reserved

The potential importance of contact between Zuberoan and Bearnese Gascon in the spread of this sound pattern has been addressed. Blevins (2017) predicts that areal sound patterns may develop in one language when listeners are exposed to perceptually salient segments (or features) through significant, continued exposure to a second language. The model predicts this sound change to be similar to other phonetically motivated, natural sound changes, as is the case of /u/-fronting in Zuberoan.

French, Occitan and Gascon show evidence of a sound change where Latin /uː/ gave rise to /y/ in several modern Romance languages. In contrast to what is found in Zuberoan, the fronting has been described as context free and exceptionless in all Romance languages that developed the process. In the Gallo-Romance languages, the back vowel fronting affected all instances of Latin /uː/, which became Vulgar Latin /u/ around the 2nd century AD. In Zuberoan, in contrast, fronting of Common Basque /u/ was inhibited in specific phonological environments, conditions which do not seem to have a parallel in any of the Romance languages.

In addition to this, Zuberoan – as well as Roncalese – shows a different pattern of fronting for the falling diphthongs, /V͡u/ to /V͡i/. This fronting which, interestingly enough, shares some phonological conditions of the fronting of syllabic /u/, may have developed more recently. If Zuberoan texts such as Etxart (Lakarra et al. forthcoming) show an early stage in the development of this process, the fronting of back semivowels may have been a late sound pattern, unrelated to developments in Gascon, and later spreading to Roncalese.

The development of /y/ in Zuberoan is unique and highlights the potential complexity of contact-induced sound patterns. Existing texts allow for a more detailed description of the phenomenon, in particular of inhibitory contexts. Overall, Zuberoan /u/-fronting illustrates the importance of perception, production and contact in the analysis of historical developments and the importance of typological comparison in the study of language isolates such as Basque.

List of abbreviations

ABS	absolutive
ALLOC	allocutive
DET	determined
ERG	ergative
F	feminine
FUT	future
M	masculine
SG	singular

© 2017. John Benjamins Publishing Company
All rights reserved

References

Agirre Sarasola, Pello (ed.). 1998 [ca. 1676]. Pronus singulis diebus Dominicis clarè & distinctè immediatè post Evangelium populo legendus. *International Journal of Basque Linguistics and Philology [ASJU]* 32. 1–46.

Anglade, Joseph. 1921. *Grammaire de l'ancien provençal ou ancienne langue d'Oc*. Paris: Klincksieck.

Archu, Jean-Baptiste. 1848. *La Fontainaren aleghia berheziak, neurt-hitzez franzesetik uskarara itzuliak*. La Réole: Pasquier.

Azkue, Resurrección M. de. 1931. Particularidades del dialecto roncalés. *Euskera* 12(2–4). 207–406.

Bailey, Guy. 1997. When did Southern American English begin? In Edgar W. Schneider (ed.), *Englishes around the world: Studies in honour of Manfred Görlach*, vol. 1: *General studies, British Isles, North America*, 255–275. Amsterdam: John Benjamins. doi:10.1075/veaw.g18.24bai

Baranowski, Maciej. 2008. The fronting of the back upgliding vowels in Charleston, South Carolina. *Language Variation and Change* 20. 527–551. doi:10.1017/S0954394508000136

Beddor, Patrice S., James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30. 591–627. doi:10.1006/jpho.2002.0177

Belapeire, Athanase. 1997 [1696]. *Catechima laburra* [*Athanase Belapeire: Catechima laburra (1696). Autorearen garaia, nortasuna eta idazlanak. Grafiak eta fonologia. Edizio kritikoa eta hiztegia*, ed. by Pello Agirre Sarasola. Vitoria-Gasteiz, Spain: University of the Basque Country dissertation].

Bichakjian, Bernard H. 1974. The evolution of French /y/: An integrated change. In John M. Anderson & Charles Jones (eds.), *Historical linguistics: Proceedings of the first international conference on historical linguistics*, vol. 2: *Theory and description in phonology*, 71–88. Amsterdam: North-Holland.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486357

Blevins, Juliette. 2015. Evolutionary phonology: A holistic approach to sound change typology. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 485–500. Oxford: Oxford University Press.

Blevins, Juliette. 2017. Areal sound patterns: From perceptual magnets to stone soup. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press. doi:10.1017/9781107279872.006

Bourciez, Édouard. 1967. *Phonétique française: Étude historique*. Paris: Klincksieck.

Bradlow, Ann R. 1995. A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97. 1916–1924. doi:10.1121/1.412064

Buckley, Eugene. 2009. Phonetics and phonology in Gallo-Romance palatalization. *Transactions of the Philological Society* 107. 31–65. doi:10.1111/j.1467-968X.2008.00212.x

Camino, Iñaki. 2009a. Mugako hiztun eta aldaerak ipar-mendebaleko Zuberoan. *Fontes Linguae Vasconum* 111. 153–218.

Camino, Iñaki. 2009b. Hizkuntza-aldakortasunari Amiküzetik so. *Lapurdum* 13. 61–80. doi:10.4000/lapurdum.1979

© 2017. John Benjamins Publishing Company
All rights reserved

Chambon, Jean-Pierre & Yan Greub. 2002. Note sur l'âge du (proto)gascon. *Revue de linguistique romane* 66. 473–495.

Cierbide, Ricardo. 1994. *Le censier gothique de Soule*. Tarnos: Éditions Izpegi Baigorri.

Cox, Felicity & Sallyanne Palethorpe. 2001. The changing face of Australian English vowels. In David C. Blair & Peter Collins (eds.), *Varieties of English around the world: English in Australia*, 17–44. Amsterdam: John Benjamins.

Coyos, Jean – Baptiste. 2002. Zubereraren idazkeraz: ortografia, fonetika eta fonología. *Lapurdum* 7. 201–219. doi:10.4000/lapurdum.1088

Eguiateguy, Jusef. 1983 [1785]. *Lehen liburia edo filosofo huskaldunaren ekheia*. Edited by Txomin Peillen. Bilbao: Euskaltzaindia.

Egurtzegi, Ander. 2013. Phonetics and phonology. In Mikel Martínez-Areta (ed.), *Basque and Proto-Basque: Language-internal and typological approaches to linguistic reconstruction* (*Mikroglottika* 5), 119–172. Frankfurt am Main: Peter Lang.

Egurtzegi, Ander. 2014a. *Towards a phonetically grounded diachronic phonology of Basque*. Vitoria-Gasteiz, Spain: University of the Basque Country dissertation.

Egurtzegi, Ander. 2014b. Zubereraren /y/: Bere sorrera eta hedapenaren inguruan. In María José Ezeizabarrena & Ricardo Gómez (eds.), *Eridenen du zerzaz kontenta: Sailkideen omenaldia Henrike Knörr irakasleari (1947–2008)*, 149–162. Bilbao: University of the Basque Country.

Egurtzegi, Ander. 2015. Different distributions of contrastive vowel nasalization in Basque. *Loquens* 2(1). e017. doi:http://dx.doi.org/10.3989/loquens.2015.017.

Egurtzegi, Ander. Forthcoming. On the phonemic status of nasalized /ñ/ in Modern Zuberoan Basque. To appear in *Linguistics* 56.

Flemming, Edward. 2003. The relationship between coronal place and vowel backness. *Phonology* 20. 335–373. doi:10.1017/S0952675704000041

Fortson, Benjamin W. 2010. *Indo-European language and culture: An introduction*, 2nd edn. Oxford: Wiley-Blackwell.

Fouché, Pierre. 1958. *Phonétique historique du français*, vol. 2: *Les voyelles*. Paris: Klincksieck.

Fowler, Carol A. 2005. Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics* 33. 199–213. doi:10.1016/j.wocn.2004.10.003

Fowler, Carol A. & Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36. 171–195.

Fridland, Valerie. 2008. Patterns of /uw/, /ʊ/, and /ow/ fronting in Reno, Nevada. *American Speech* 83. 432–454. doi:10.1215/00031283-2008-030

Gavel, Henri. 1920. *Éléments de phonétique basque*. Paris: Édouard Champion.

Gèze, Louis. 1873. *Éléments de grammaire basque, dialecte souletin, suivis d'un vocabulaire basque-français et français-basque*. Bayonne: Imprimerie de Veuve Lamaignère.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486678

Haase, Martin. 1992. *Sprachkontakt und Sprachwandel im Baskenland: Die Einflüsse des Gascognischen und Französischen auf das Baskische*. Hamburg: Helmut Buske.

Hajek, John. 1993. Old French nasalization and the universals of sound change. *Journal of French Language Studies* 3. 145–164. doi:10.1017/S0959269500001721

Hall, Robert A. 1976. *Proto-Romance phonology*. New York: Elsevier.

Harrington, Jonathan. 2012. The coarticulatory basis of diachronic high back vowel fronting. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production and social factors*, 103–122. Amsterdam: John Benjamins. doi:10.1075/cilt.323.10har

© 2017. John Benjamins Publishing Company
All rights reserved

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America* 123. 2825–2835. doi:10.1121/1.2897042

Harrington, Jonathan, Phil Hoole, Felicitas Kleber & Ulrich Reubold. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics* 39. 121–131. doi:10.1016/j.wocn.2010.12.006

Harrington, Jonathan & Florian Schiel. 2017. /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language* 93(2). 414–445. doi:10.1353/lan.2017.0019

Harris, Martin. 1988. French. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 209–245. London: Routledge.

Haudricourt, André & Alphonse Juilland. 1970. *Essai pour une histoire structurale du phonétisme français*, 2nd edn. The Hague: Mouton.

Hawkeys, Sarah & Jonathan Midgley. 2005. Formant frequency of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35. 183–199. doi:10.1017/S0025100305002124

Henton, Caroline. 1983. Changes in the vowels of received pronunciation. *Journal of Phonetics* 11. 353–371.

Hualde, José Ignacio. 1993. Topics in Souletin phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *Generative studies in Basque linguistics*, 289–327. Amsterdam: John Benjamins. doi:10.1075/cilt.105.12hua

Igartua, Iván. 2015. Diachronic effects of rhinoglottophilia, symmetries in sound change, and the curious case of Basque. *Studies in Language* 39(3). 635–663. doi:10.1075/sl.39.3.04iga

Inchauspe, Emmanuel T. 1852. *Kurutchiaren bidiaren debocionia*. Bayonne: Foré & Lasserre.

Johnson, Keith. 1997. Speech perception without speaker normalization. In Keith Johnson & John W. Mullennix (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Kelly, Reine C. 1973. *A descriptive analysis of Gascon*. The Hague: Mouton.

Koops, Christian. 2010. /u/-Fronting is not monolithic: Two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers in Linguistics* 16(2). 113–122.

Kuhl, Patricia K. 1991. Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50. 93–197. doi:10.3758/BF03212211

Kuhl, Patricia K. 1995. Mechanisms of developmental change in speech and language. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 2, 132–139. Stockholm: KTH & Stockholm University.

Kuhl, Patricia K. 2000. Language, mind, and brain: Experience alters perception. In Michael S. Gazzaniga (ed.), *The new cognitive neurosciences*, 2nd edn., 99–115. Cambridge, MA: MIT Press.

Kuhl, Patricia K. & Paul Iverson. 1995. Linguistic experience and the 'perceptual magnet effect'. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 121–154. Timonium: York Press.

Labov, William. 1994. *Principles of linguistic change,* vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, William. 2008. Triggering events. In Susan M. Fitzmaurice & Donka Minkova (eds.), *Empirical and analytical advances in the study of English language change*, 11–54. Berlin: Mouton de Gruyter.

© 2017. John Benjamins Publishing Company
All rights reserved

Labov, William, Malcah Yaeger & Richard C. Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. New York: Mouton de Gruyter. doi:10.1515/9783110167467

Lafon, René. 1999 [1933]. Sur le traitement de *u* devant le groupe « r plus consonne » en Souletin. *Revista Internacional de Estudios Vascos* 24. 173–175 [reissued in Lafon (1999), 75–77].

Lafon, René. 1999 [1937]. Tendance à la palatalisation de la sonante *u* dans les parlers basques du Nord-Est. *Revue de Linguistique Romane* 13. 73–82 [reissued in Lafon (1999), 85–93].

Lafon, René. 1999 [1958]. Contribution a l'étude phonologique du parler basque de Larrau (Haute-Soule). In Diego Catalán (ed.), *Estructuralismo e historia. Miscelánea homenaje a André Martinet* 2, 77–106. La Laguna: Universidad de La Laguna [reissued in Lafon (1999), 113–133].

Lafon, René. 1999 [1962a]. Contact de langues et apparition d'une nouvelle voyelle: *u* et *ü* en basque. *Actes du X*e Congrès International de Linguistique et Philologie Romanes III. Strasbourg, 901–909 [reissued in Lafon (1999), 135–143].

Lafon, René. 1999 [1962b]. Sur la voyelle *ü* en basque. *Bulletin de la Société de Linguistique de Paris* 57. 83–102 [reissued in Lafon (1999), 95–111].

Lafon, René. 1999, *Vasconiana* (Iker 11), Bilbao: Euskaltzaindia [collected writings, Jean Haritschelhar & Piarres Charritton (eds.)].

Lafont, Robert. 1971. Deux types ethniques. In Robert Lafont (ed.), *Le sud et le nord, dialectique de la France*, 101–136. Toulouse: Privat.

Lakarra, Joseba A. & Íñigo Ruiz Arzalluz (eds.). 2011a. *Obras completas VI: Fonética histórica vasca*. Donostia-San Sebastián, Vitoria-Gasteiz: Diputación Foral de Guipúzcoa, University of the Basque Country.

Lakarra, Joseba A. & Íñigo Ruiz Arzalluz (eds.). 2011b. *Obras completas VII: Fonética y fonología. Morfosintaxis. Dialectología*. Donostia-San Sebastián, Vitoria-Gasteiz: Diputación Foral de Guipúzcoa, University of the Basque Country.

Lakarra, Joseba A. & Íñigo Ruiz Arzalluz (eds.). 2011c. *Obras completas XII: Textos arcaicos vascos. N. Landuchio, Dictionarium Linguae Cantabricae (1562)*. Donostia-San Sebastián, Vitoria-Gasteiz: Diputación Foral de Guipúzcoa, University of the Basque Country.

Lakarra, Joseba A., Gidor Bilbao & Céline Mounole. Forthcoming. *Edición crítica de la correspondencia entre Etxart y Ros*.

Larrasquet, Jean. 1932. Phonétique du basque de Larrajá (quartier de Barcus). *Revista Internacional de los Estudios Vascos* 23. 153–191.

Larrasquet, Jean. 1939. *Le basque de la Basse-Soule Orientale*. Paris: Klincksieck.

Larrebat, Justin. 1926. *Poésies gasconnes. Nouvelle édition annotée avec deux portraits de l'auteur, une notice biographique et une étude sur le gascon de Bayonne, par H. Gavel*. Edited by Henri Gavel. Paris: Société des sciences, lettres, arts et d'études régionales de Bayonne.

Lespy, Vastin J.D. & Paul L. Raymond. 1998 [1887]. *Dictionnaire béarnais ancien et moderne*. Pau: Marrimpouey.

Lhande, Pierre. 1926–1938. *Dictionnaire basque-français (Dialectes labourdin, bas-navarrais et souletin)*. Paris: Gabriel Beauchesne.

Maclagan, Margaret & Jennifer Hay. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English vowel shift. *Language Variation and Change* 19. 1–25. doi:10.1017/S0954394507070020

© 2017. John Benjamins Publishing Company
All rights reserved

Maddieson, Ian. 2013. Front rounded vowels. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, feature 7A. Munich: Max Planck Digital Library. http://wals.info/chapter/11 (25 September, 2013.)

Maister, Martin. 1757. *Jesu-Kristen imitacionia. Cuberouaco uscarala, herri beraurtaco. Apheç bateç, bere jaun apheçcupiaren baimentouareki utçulia*. Pau: G. Dugué & J. Desbaratz.

Matte, Edouard J. 1982. *Histoire des modes phonétiques du français*. Geneva: Librairie Droz.

McDougall, Kirsty & Francis Nolan. 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1825–1828. Saarbrucken: Universität des Saarlandes.

Meyer-Lübke, Wilhelm. 1890. *Grammaire des langues romanes*. Paris: E. Welter.

Michelena, Luis [Mitxelena, Koldo]. 2011 [1950]. De fonética vasca. La aspiración intervocálica. *Boletín de la Real Sociedad Vascongada de los Amigos del País* 6. 443–459 [reissued in Joseba A. Lakarra & Íñigo Ruiz Arzalluz (eds.), 2011b, 3–20].

Michelena, Luis [Mitxelena, Koldo]. 2011 [1964]. *Textos arcaicos vascos*. Madrid: Minotauro [reissued in Joseba A. Lakarra & Íñigo Ruiz Arzalluz (eds.), 2011c, 1–179].

Michelena, Luis [Mitxelena, Koldo]. 2011 [1977]. *Fonética histórica vasca*. Donostia-San Sebastián: Diputación Foral de Guipúzcoa [1961, 2nd edn. 1977; reissued in Joseba A. Lakarra & Íñigo Ruiz Arzalluz (eds.), 2011a].

Michelena, Luis [Mitxelena, Koldo] & Ibon Sarasola. 1987–2005. *Orotariko euskal hiztegia* [*General Basque dictionary*] (16 vols.). Bilbao: Euskaltzaindia. http://www.euskaltzaindia.net/oeh/ (25 September, 2013).

Moon, Seung-Jae & Björn Lindblom. 1994. Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96. 40–55. doi:10.1121/1.410492

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman.

Oihenart, Arnaud d'. 2003 [1657]. *Les proverbes basques recueillis par le Sr d'Oihenart, plus les poesies basques du même auteur*. Paris: Bibliotheca Colbertina [Patxi Altuna & José María Mujika (eds.), *Euskal atsotitzak eta neurtitzak* (*IKER* 15), Bilbao: Euskaltzaindia].

Oñederra, M. Lourdes. 2009a. Diacronía y motivación fonética. In Fernando Sánchez Miret (ed.), *Romanística sin complejos. Homenaje a Carmen Pensado*, 47–67. Bern: Peter Lang.

Oñederra, M. Lourdes. 2009b. Ohar batzuk bokal batez: Zubereraren [y]. *International Journal of Basque Linguistics and Philology [ASJU]* 43. 663–678.

Orpustan, Jean-Baptiste. 1990. *Toponymie basque*. Bordeaux: Presses Universitaires de Bordeaux.

Orpustan, Jean-Baptiste. 1999. *La langue basque au Moyen Age (IX*e*–XV*e siècles). Tarnos: Éditions Izpegi Baigorri.

Oyharçabal, Beñat [Oihartzabal, Beñat]. 1991. *La pastorale souletine. Édition critique de Charlemagne (Supplements of the International Journal of Basque Linguistics and Philology [ASJU] 16, Corpus Pastoralium Vasconicarum 1)*. Donostia-San Sebastián: Diputación Foral de Guipúzcoa, University of the Basque Country.

Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39. 151–168. doi:10.1121/1.1909864

© 2017. John Benjamins Publishing Company
All rights reserved

Padilla-Moyano, Manuel. 2017. *Analyse diachronique du dialecte souletin (XVIe–XIXe siEdes) – Zuberoako euskararen azterketa diakronikoa (XVI–XIX. mendeak)*. Vitoria-Gasteiz, Spain: University of the Basque Country, Université Bordeaux Montaigne dissertation.

Palay, Simin. 1980 [1932–1934]. *Dictionnaire du béarnais et du gascon modernes*. Pau: Marrimpouey.

Peillen, Txomin. 1992. Zubereraren bilakaeraz ohar batzuk. In Euskaltzaindia (ed.), *Nazioarteko Dialektologia Biltzarra. Agiriak* (*Iker* 7), 247–272. Bilbao: Euskaltzaindia.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–158. Amsterdam: John Benjamins. doi:10.1075/tsl.45.08pie

Recasens, Daniel & Maria Dolors Pallarès. 2001. Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics* 29. 273–301. doi:10.1006/jpho.2001.0139

Rohlfs, Gerhard. 1977. *Le gascon: Études de philologie pyrénéenne*, 3rd edn. Tübingen: Niemeyer. doi:10.1515/9783111330969

Sarasola, Ibon. 1983. Contribución al estudio y edición de textos antiguos vascos. *International Journal of Basque Linguistics and Philology [ASJU]* 17. 69–212.

Sauguis, Bertrand de [Zalgize, Bertrand]. 1908–1909. Los refranes vascos de Sauguis traducidos y anotados [Julio de Urquijo (ed.)]. *Revista Internacional de los Estudios Vascos* 2/3. 677–724/144–157.

Séguy, Jean. 1952. Basque et gascon dans l'Atlas linguistique de la Gascogne. *Orbis* 1. 385–391.

Tartas, Juan de. 1995 [1666]. *Iván de Tartas. Onsa hilceco bidia. Edizio kritikoa Patxi Altunak paratua* [Patxi Altuna (ed.)]. Bilbao: Universidad de Deusto.

Tartas, Juan de. 1996 [1672]. *Iván de Tartas. Arima penitentaren occupatione devotaq* [Patxi Altuna (ed.)]. Bilbao: Mensajero.

Trask, R. Larry. 1997. *The history of Basque*. London: Routledge.

Uhlenbeck, Christianus C. 1903. *Beiträge zu einer vergleichenden Lautlehre der baskischen Dialecte*. Amsterdam: Johannes Müller.

Wheeler, Max. 1988. Occitan. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 246–278. London: Routledge.

Zarabozo, Carlos. 1972. Two borrowed sound changes in Basque. *International Journal of Basque Linguistics and Philology [ASJU]* 6. 174–182.

Zuazo, Koldo. 2008. *Euskalkiak. Euskararen dialektoak*. Donostia-San: Elkar. Sebastián.

Résumé

Tous les dialectes basques actuels ont au moins cinq voyelles contrastives: /i, e, a, o, u/. Le souletin, un dialecte basque, a développé une sixième voyelle contrastive: la voyelle haute antérieure arrondie /y/. On peut soutenir que ce développement est dû au contact prolongé avec les langues gallo-romanes voisines. Cet article défend empiriquement le développement historique de la distinction entre /u/ et /y/ et propose une analyse détaillée des contextes qui ont empêché le changement de /u/ en /y/. L'antériorisation a été empêchée quand la voyelle était suivie par la sibilante apicale, la battue /ɾ/, ou le groupe consonantique ɾT (r une rhotique, T une occlusive alvéolaire), probablement en raison des effets co-articulatoires. L'antériorisation apparaît quand /s̬/, /r/ ou les groupes rhotique-occlusive non-coronale suivent /u/. Cette antériorisation de la voyelle /u/ démontre l'importance du contact entre langues et de la phonétique dans l'analyse phonologique des développements historiques.

© 2017. John Benjamins Publishing Company
All rights reserved

Zusammenfassung

Alle modernen baskischen Dialekte besitzenhaben mindestens 5 kontrastive Vokale, /i, e, a, o, u/. Der baskische Dialekt Zuberoa (bzw. Suletinisch), ein baskischer Dialekt, hat mit dem hohen, gerundeten Vorderzungenvokal /y/ einen sechsten kontrastiven Vokal entwickelt. Dessen Diese EntstehungEntwicklung beruht vermutlich auf dem anhaltendenm Kontakt mit den benachbarten gallo-romanischen Sprachen. Dieser Artikel präsentiert empirische Belege für die historische Entwicklung des /u/- vs. /y/-Kontrasts des /u/ vs. /y/ Kontrasts und bietet eine detaillierte Analyse der Kontexte, die den Lautwandel /u/ > /y/ Lautwandel unterbunden haben. Frontierung wurde verhindertunterbunden, wenn der Vokal von einem apikalen Sibilanten, einem Tap-/ɾ/ Tap /ɾ/ oder einem rT-Cluster (r ist ein rhotischer, t ein r: ein Rhotic, T: ein alveolarer Obstruent) gefolgt wurde, vermutlich was wohl auf Koartikulation zurückzuführen istaufgrund von Koartikulation. Frontierung fand statt, wenn /s̪/, /r/ oder ein nicht-koronales rhotisches Rhotic-Obstruenten-Cluster auf /u/ folgte. Die /u/-Frontierung im Zuberoanischen zeigtillustriert, wie wichtig Sprachkontakt und Phonetik in der phonologischen Analyse historischer Entwicklungen sind.

Author's address

Ander Egurtzegi
Institut für Phonetik und Sprachverarbeitung (Ludwig-Maximilians-Universität München)
Schellingstr. 3, Büro 249
München, Bayern 80799
Germany
egurtzegi@phonetik.uni-muenchen.de

© 2017. John Benjamins Publishing Company
All rights reserved

DE GRUYTER MOUTON Linguistics 2018; 56(6): 1353–1367

Ander Egurtzegi*

On the phonemic status of nasalized /h̃/ in Modern Zuberoan Basque

https://doi.org/10.1515/ling-2018-0024

Abstract: Modern eastern Basque dialects have several conservative features, including the maintenance of historical /h/, which is lost in other dialects. Zuberoan, the easternmost dialect of Basque still spoken today, shows both this /h/ as well as a phonetically nasalized segment [h̃] which is a reflex of intervocalic *n. In this paper I argue that these two segments contrast in Zuberoan. Evidence for the contrast comes from both a newly described process of assimilation of /h/ to /h̃/ in nasal environments which then serves as a basis of the analogical extension of the nasalized aspirate in a context where it cannot be phonologically derived, and from neighboring Mixean Low Navarrese where the nasalized [h̃] has no other obvious source. Since a contrast between oral and nasalized aspirates is rare crosslinguistically, the Zuberoan and Mixean sound patterns discussed here should be of interest to typologists and phonologists alike.

Keywords: laryngeals, nasalized aspirates, rhinoglottophilia, nasalization, Basque, Zuberoan

1 Introduction

More than half of the world's languages have a phoneme /h/, a voiceless aspirate, often classified as an approximant, glide or voiceless laryngeal.[1] In contrast, few languages if any have a clear contrast between /h/ and its nasalized counterpart /h̃/, though, as argued by Walker and Pullum (1999), [nasal] must be a possible phonological feature of glottals given that /h̃/ has been

1 88 % of the languages in the UPSID database (279/317) have /h/ (Maddieson 1984). For a detailed discussion of the frequency of /h/ in contrast to voiceless sonorants, see Blevins (2018). Classification of this sound as an approximant or glide is based on the absence of significant constriction and/or air pressure build-up in the supralaryngeal cavity (cf. the definition of [consonantal] in Halle and Clements 1983).

***Corresponding author: Ander Egurtzegi,** Institute of Phonetics and Speech Processing (IPS), Ludwig Maximilian University of Munich, Schellingstr. 3, 80799 Munich, Germany, E-mail: egurtzegi@phonetik.uni-muenchen.de

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

described to trigger phonological nasal spreading. The phonological spread of nasalization triggered by /h̃/ requires this segment to be phonologically specified for nasality, yet an opposition between /h/ and /h̃/ is scarcely attested in the world's languages. Ohala (1975) suggested early on that cues for aspiration and nasality were similar, attempting to explain the absence of this contrast crosslinguistically in terms of perceptual similarity (see also Ohala 1990). Since then, at least two languages have been argued to have a contrast between oral and nasalized /h/: Kwangali, a Bantu language (Ladefoged and Maddieson 1996) and Seimat, an Austronesian language of the Admiralty Islands (Blust 1997, 1998). At least two other languages, Aguaruna, a Jivaroan language, and Arabela, a Zaparoan language, are analyzable with the oral vs. nasalized /h/ contrast, or with the nasalized aspirate viewed as a predictable allophone of a velar nasal consonant (Walker and Pullum 1999: 768–770; Durvasula 2009: 52–55). Given the typological rarity of a contrast between oral and nasalized /h/, any language showing a possible contrast between these two segment types should be of interest.

One language with a possible contrast of this kind is Zuberoan (Souletin) Basque, a modern Basque dialect spoken in the northeasternmost corner of the Basque Country. In this variety of Basque nasalized and non-nasalized aspirates exist, as shown by minimal pairs such as *ehe* 'wash water' vs. *ẽh̃ẽ* 'no' (Lafon 1958). Hualde (1993: 294) states the hypothesis clearly: "...our proposal is that Souletin has two aspirated phonemes: oral /h/ and nasal /h̃/". While many adopt Hualde's analysis, an alternative, first proposed by Larrasquet (1932: 168), views nasalization of aspirates as allophonic, with nasalization spreading from adjacent nasalized vowels to /h/ (Michelena 2011b [1977]). The issue has not been seriously revisited in modern studies of Zuberoan phonology, and the central role of this paper is to fill this gap.[2]

In this paper, I begin with the basic observation of minimal and near-minimal pairs contrasting oral vs. nasal aspirates. In addition, I offer two different kinds of evidence for this as a contrast between oral vs. nasal aspirates, as opposed to allophonic nasalization of the aspirate in nasalized vowel contexts. The first piece of evidence is a newly identified analogical spread of the nasalized aspirate discussed in Section 2.2, and the second, offered in Section 2.3, is comparative dialect evidence where vowel nasalization does not appear to be a possible source for aspirate nasalization. In Section 2.1, I offer an overview of Basque dialect phonology, and of the Zuberoan phoneme inventory. Before turning to this, a brief summary of the diachronic phonology that has given rise

2 For a recent treatment of rhinoglottophilia in Basque historical phonology, including the evolution of nasalized aspirates, see Igartua (2015).

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

to the contrastive aspirates in earlier stages of Basque and in modern Zuberoan is presented. Throughout, transcriptions enclosed in square brackets make use of standard IPA symbols.[3] The IPA refers to /h/ as a "glottal fricative"; in this paper, I use the terms "aspirate" and "nasalized aspirate" to refer to these sounds, since intense fricative noise, of the kind associated with Basque sibilants, is typically weak or absent.

Proto-Basque, as reconstructed by Michelena (2011b [1977]: 171–172), has a contrast between intervocalic *h, intervocalic *n, and zero: *zahar 'old', *seni 'boy', *gau 'night'.[4] Intervocalic /h/ and /n/ are both attested in Aquitanian, the oldest attested Euskarian language (Gorrochategui 1984; Martínez-Areta 2013), with /h/ persisting into Michelena's Common Basque (ca. 600 CE), and later lost in western varieties, but preserved in Zuberoan and other eastern dialects. Intervocalic *n was also weakened. While many instances of Proto-Basque *h have been maintained by the modern eastern dialects (compare Aquitanian (*Umme*)*sahar* to modern Basque *zahar* 'old'), evidence for the reconstruction of intervocalic *n can be found in Aquitanian attestations such as *Seni-* (in the names *Senicco, Seniponnis* and *Senitennis*, modern Basque *sehi, sei*(*n*) 'boy, servant'), names of Basque origin with early introduction into Romance (cf. *Anuncibai*, formed by modern Basque *ahuntz* 'goat' and (*h*)*ibai* 'river') or old compounds that lost the phonological context that triggered the change (cf. *mingain* 'tongue', formed by the words *mihi* < **bini* 'tongue' and *gain* < **gane* 'top'). Michelena (2011a [1950]: 8–9, 2011b [1977]: 171) proposes the sound change of *VnV > ṼhṼ; by contrast, evidence from Zuberoan has led Hualde (1993), Igartua (2008, 2015) and Egurtzegi (2014) to formulate the weakening as *VnV > Vh̃V, giving rise to a nasalized aspirate phoneme.[5] In (1a) this development is shown for inherited vocabulary, where reconstructions are from Michelena (2011b [1977]; Arbelaiz 1978), and examples in (1b) show the same sound change in Latin loans.

3 Phonetic nasalization of both /h̃/ and vowels is transcribed throughout, so as not to bias interpretation of the source of nasalization.

4 Trask (1997) does not reconstruct *h for Common Basque. Among problems for his account are the many attested cases of /h/ in Aquitanian, continued faithfully in Medieval Basque, Zuberoan and other northern dialects. For a critique of his proposal for a synchronic /h/ autosegment, see Igartua (2001) or Lakarra (2009); see also papers in Martínez-Areta (2013), which all assume *h for Proto-Basque, and its continuation in Aquitanian, Medieval Basque and northern dialects.

5 The weakening of a voiced nasal to /h/ may seem phonetically unnatural, but has been analyzed as a case of rhinoglottophilia, recognizing the similar acoustic/perceptual properties shared by nasalization and aspiration. See Igartua (2008, 2015) for general references, and for a detailed treatment of the Basque case.

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

(1) *n > /h̃/ /V_V

(a) Inherited lexicon:

Zuberoan	IPA	Reconstruction	Gloss
ahai	/ãh̃ãĩ/	**anari*	'ram'
ihi	/ĩh̃ĩ/	**ini*	'rush, reed'
ihes	/ĩh̃ẽs̻/	**enes*	'to escape, run away'
sehi	/s̻ẽh̃ĩ/	**seni*	'servant' (Bizkaian Basque *sein* /s̻e̯in/ 'boy')
mihi	/mĩh̃ĩ/	**bini*	'tongue' (cf. *mingain* 'tip of the tongue')

(b) Latin loanwords:

Zuberoan	IPA	Latin	Gloss
uhue	/ũh̃ũẽ/	*honōre*(*m*)	'honor' (Standard Basque *ohore*)
ahate	/ãh̃ãte/	*anăte*(*m*)	'duck'
mahuka	/mãh̃ũka/	*manica*, **manuca*	'sleeve'

Since the contrastive status of the synchronic nasalized aspirate in Zuberoan is the subject of this paper, we do not use this diachronic development as evidence for any aspect of the synchronic phonology. It is mentioned only to give the reader an overview of the sound patterns in earlier stages of Basque that are widely agreed to have given rise to the aspirate contrasts discussed in the remainder of this paper.

2 The Zuberoan /h/ vs./h̃/ contrast

2.1 Segmental contrasts in major Basque dialects

Major dialects of Basque at the end of the nineteenth century are illustrated in Figure 1. Zuberoan, the dialect of central interest in this study, is located at the far northeastern corner of the Basque speaking region, surrounded on its northeastern borders by Romance languages (Gascon and French), and in past centuries, by other dialects of Basque to the south and west, though now, these are regions dominated by Castilian.

Table 1 illustrates the segmental inventory of Standard Basque, and Table 2 illustrates the contrastive segments of Zuberoan as analyzed by Hualde (2003).

Zuberoan has a range of contrasts that are absent in Standard Basque (and the central and western varieties on which this standard is based), including: /h/

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

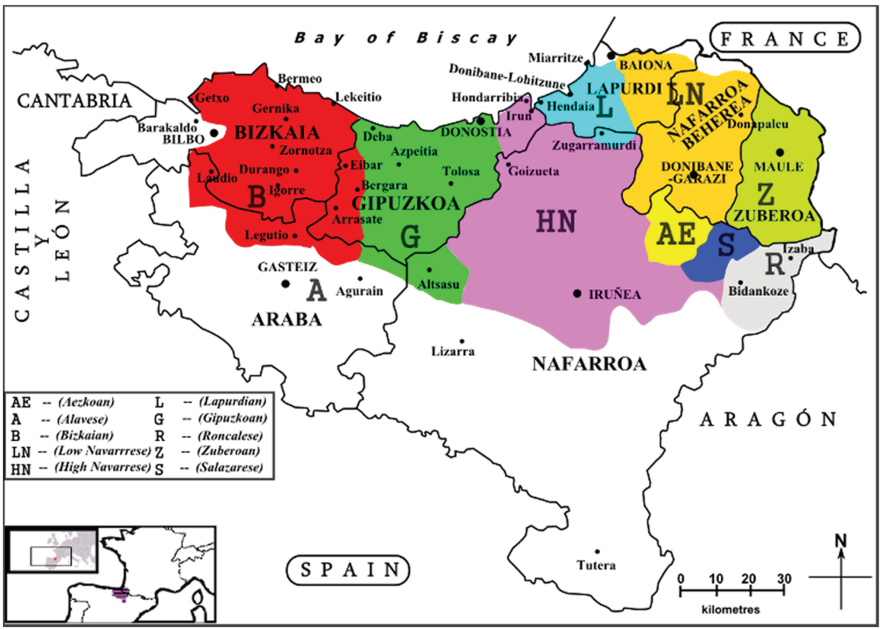


Figure 1: The historical dialectal division of Basque (Egurtzegi 2014).

Table 1: Segment inventory of Standard Basque (after Hualde 2003).

		labial		apical		laminal	predorsal		postdorsal
		bilabial	labio-dental	apico-dental	apico-alveolar		palato-alveolar	palatal	
stop	voiceless	p		t				c	k
	voiced	b		d				ɟ	g
fricative			f		s̪	s̪	ʃ		x
affricate					ts̪	ts̪	tʃ		
nasal		m			n			ɲ	
lateral					l			ʎ	
tap					ɾ				
trill					r				

	front	central	back
high	i		u
mid	e		o
low		a	

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

Table 2: Segment inventory of Zuberoan (after Hualde 2003; Egurtzegi 2014).

		laminal								
		labial		apical		(alveolar)	predorsal		postdorsal	laryngeal
		bilabial	labio-dental	apico-dental	apico-alveolar		palato-alveolar	palatal		
stop	voiceless	p		t				c	k	
	aspirated	p^h		t^h				c^h	k^h	
	voiced	b		d				ɟ	g	
Fricative	voiceless		f		s̪	s̪	ʃ			h/h̃
	voiced			z̪	z̪	ʒ				
affricate	voiceless				t̪s̪	t̪s̪	tʃ			
	voiced			d̪z̪						
nasal		m			n			ɲ		
lateral					l			ʎ		
tap					ɾ					
trill					r					
glides								j	w	

	front	central	back
high	i/ĩ y/ỹ		u/ũ
mid	e		o
low		a/ã	

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

vs. /h̃/; oral vs. nasalized vowels, and an /u/ vs. /y/ contrast.[6] While the latter two features, namely vowel nasalization and presence of a front rounded vowel, can arguably be attributed to Romance influence (Egurtzegi 2014, Egurtzegi 2017), the aspirate contrast under investigation here cannot be attributed to external influence, and, indeed, is typologically unusual, as noted earlier.

In Table 3, the presence (✓) or absence (**X**) of these three features is tabulated for selected dialects and earlier stages of Basque, following Michelena (2011b [1977]), Hualde (2003) and Egurtzegi (2013, 2014).

Table 3: Contrastive aspirates and nasalization in Basque, past and present.

	h	h̃	Ṽ	y
Aquitanian	✓	X	X	X
Old Common Basque	✓	✓[a]	X	X
Bizkaian	X	X	X	X
Gipuzkoan	X	X	X	X
High Navarrese	X	X	X	X
Roncalese	X	X	✓	X
Lapurdian	✓	X	X	X
Low Navarrese	✓	X	X	X
Mixean (LN)	✓	✓	X	✓
Zuberoan	✓	✓	✓	✓

[a]There is no evidence that contradicts this.

Today, /h/ is contrastive in the three northern dialects of the French territory, Lapurdian, Low Navarrese and Zuberoan (L, LN and Z in the map), while /h̃/ has been claimed to be contrastive only in Zuberoan. Nasalized vowels were present in Archaic Bizkaian and Roncalese, but today are found only in Zuberoan. Front rounded /y/ is found in Zuberoan and Mixean Low Navarrese.

Note that the earliest stages of Basque, represented here by Aquitanian, show intervocalic and word-initial /h/ vs. /n/ vs. zero, while later stages of the language appear to show a shift of intervocalic /n/ to nasalized /h̃/ or /h/, and western dialects show a loss of /h/ altogether. Some common lexical items illustrating dialect correspondences for historical initial and intervocalic /h/ and /n/ are provided in Table 4.

6 Some spectrograms illustrating the /h/ vs. /h̃/ contrast can be found in Egurtzegi (2014).

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

Table 4: Dialect correspondences for Proto-Basque *h and *n.

Pre-Basque	Zuberoan	Low Navarrese	Western	Gloss
*#hV	[hari]	[hari]	[ari]	'stone'
	[hil]	[hil]	[il]	'dead'
*#nV	[naɣ̞ u̪i]	[naɣ̞ u̪i]	[naɣ̞ u̪i]	'chief'
	[ni]	[ni]	[ni]	'I' (1sg.abs)
*VhV	[s̪ahar]	[s̪ahar]	[s̪aar], [s̪ar]	'old'
	[bihi]	[bihi]	[bii], [bi]	'grain'
*VnV	[ãh̃ãrdi]	[ahardi]	[aardi]	'sow'
	[ĩh̃ẽs̪i]	[iheṣ̪i]	[ieṣ̪i], [iɲeṣ̪i]	'to flee'

Throughout this paper, we use the voiceless symbols [h] and [h̃] for the two sounds in question, though in intervocalic contexts, both aspirates are typically voiced.

Though /h/ is phonemic in Zuberoan, Low Navarrese and Lapurdian, it has different lexical frequencies in these varieties, decreasing significantly as one moves from east to west. Zuberoan, the easternmost dialect, appears to represent the most conservative stage, where /h/ has the highest functional load. Loss of /h/ has moved areally from western varieties into Lapurdi, and is now complete on the Lapurdian coast, extending into Low Navarre.[7] In Zuberoan, however, /h/ has been preserved in most of its historic contexts. In sum, Zuberoan appears to be the only modern variety of Basque in which /h/ does not show any critical trace of recession, and, apart from one variety of Low Navarrese discussed in Section 2.3, it is also the only dialect that maintains the distinction between oral /h/ and nasalized intervocalic /h̃/ (from *VnV).

2.2 New evidence for Zuberoan /h/ vs. /h̃/: Analogical extension of /h̃/

Hualde's (1993) proposal for an /h/ vs. /h̃/ contrast in Zuberoan is based on minimal and near-minimal pairs like those shown in (2).

7 Bonaparte's (1991[1869]) dialect descriptions allow us to see that /h/ was already lost in coastal areas of Lapurdi, and was being lost in the eastern varieties of the Lapurdian dialect in the nineteenth century. The first context where /h/ was lost in Lapurdian is in post-sonorant position: compare Lapurdian *alaba* 'daughter', *alargun* 'widow', *eraztun* 'ring vs. Zuberoan *alhaba, alhargun, erhaztun.*

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

(2) Surface /h/ vs. /h̃/ in Zuberoan

zahar	[s̪ahar]	'old'	vs.	*ahardi*	[ãh̃ãrdi]	'sow'
ehe	[ehe]	'wash water'	vs.	*ehe*	[ẽh̃ẽ]	'no! (emphatic)'
bihi	[bihi]	'grain'	vs.	*mihi*	[mĩh̃ĩ]	'tongue'
bühürt	[byhyrt]	'twisted'	vs.	*zühür*	[s̪ỹh̃ỹr]	'wise'

There is also a modern minimal quadruplet contrast intervocallically in Modern Zuberoan between /h/, /h̃/, /n/ and zero: *ehi* ['ehi] 'finger' (< Literary Zuberoan *erhi*) vs. *ehi* ['ẽh̃ĩ] 'easy' vs. *eni* ['eni] 'to me' vs. *ei* [e̯i] (< Literary Zuberoan *eri*) 'ill' (cf. Lafon 1958).

Under Hualde's analysis, surface nasalized vowels are created when nasalization of the underlying nasalized aspirate spreads to adjacent segments (Hualde 1993: 294–295). The alternative analysis, originally proposed by Larrasquet (1932: 168), and later adopted by Michelena (2011b [1977]), is that nasalization of aspirates is allophonic, with nasalization spreading from adjacent nasalized vowels to /h/. Since contrastive nasalized vowels are independently necessary in all analyses of Zuberoan based on data like that in (3), the analysis where nasalization of /h/ is allophonic appears simpler. In addition, given the rarity of /h/ vs. /h̃/ contrasts crosslinguistically, Hualde's proposal could be questioned on typological grounds.

(3) Contrastive vowel nasalization in Zuberoan

Oral vowel		Nasalized vowel	
biga	'two'	*bigã*	'two-year heifer'
dügü	'we have'	*hügũ*	'repugnance'
hi	'you'	*fĩ*	'fine, prudent'
lili	'flower'	*lĩ*	'linen'

However, as shown below, there is additional evidence that nasalized /h̃/ is a contrastive segment in Zuberoan. Following Hualde (1993: 294–295, 2003: 31), I will argue that /h̃/ is a contrastive nasalized segment which induces phonetic nasalization in neighboring vowels. Contextual nasalization is a consequence of these vowels being in a nasal environment similar to that found in sequences such as *ona* [õnã] 'the good' or *ama* [ãmã] 'mother' (cf. Hualde 1993: 294).

One argument against the spread of nasalization from vowels to /h/ involves the distribution of vowel nasalization in Zuberoan. As illustrated in (3) above, contrastively nasalized vowels in modern Zuberoan are (i) limited to /ĩ ỹ ũ ã/ (Egurtzegi 2015: 4–5); and (ii) found only in word-final stressed syllables (Hualde 1993: 295; Egurtzegi 2014: 151–152). If we follow Michelena (2011b

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

[1977]) in deriving /h/ nasalization from surrounding vowels, the vowels in question would be the only unstressed contrastively nasalized vowels in Zuberoan. In addition, adopting the /ṼhṼ/ analysis with spread of nasalization to /h/ would require adding more nasalized vowels to the phonemic inventory of the language. Words like *ehe* [ẽh̃ẽ] 'no! (emphatic)', *mehe* [mẽh̃ẽ] 'thin' and *Johanne* [ʒõ'h̃ãɲe] 'John' require adding at least /ẽ/, and possibly /õ/ as well. Thus, the general distribution of nasalized vowels in Zuberoan must be unnecessarily complicated for nasalization to spread from vowels to /h/. There is on the other hand a more general principle based on contrastive /h̃/ by which allophonically nasalized unstressed vowels, and allophonically nasalized mid vowels like [ẽ] and [õ] can be derived as a consequence of nasal coarticulation. If this were the only argument in favor of the contrastive nasalized aspirate, one might still lean towards the alternative due to the extreme rarity of the /h/ vs. /h̃/ contrast crosslinguistically, but other evidence exists.

In Zuberoan, /h/ is assimilated to [h̃] in the context of a nasal consonant (Egurtzegi 2014: 79–81). Zuberoan examples of /h/ surfacing as [h̃] include: *nihau*(*r*) [nĩ'h̃ãu̯] 'me myself' from *ni* 'me' and *hau*(*r*) 'this'; *Johanne* [ʒõ'h̃ãɲe] 'John', with a known Biblical origin; *ahin* [ãh̃ĩn] 'light' and *ahan* [ãh̃ãn] 'plum' (< Literary Zuberoan *arhin* and *arhan*). All of these examples involve etymological /h/. In these cases, nasalization in the aspirate is due to assimilation to a nasal consonant elsewhere in the word, and not a consequence of intervocalic *VnV > VH̃V. This local assimilation in and of itself has little bearing on the status of /h̃/ in words like those in (2). However, a subsequent development strongly suggests that at the time this local assimilation occurred, nasalized /h̃/ was a contrastive segment in Zuberoan.

The subsequent development in question is an analogical change within the intensive pronominal paradigm. As illustrated above, secondary [h̃] in *nihau* [nĩ'h̃ãu̯] 'me myself' is a consequence of a local assimilation of nasality from the initial /n/ of /ni/ 'I' to the second element of the compound *hau* /hau̯/ 'this' (< Literary Zuberoan *haur* /hau̯ɾ/). The facts of interest involve analogical extension of this second compound element with initial nasalized [h̃] to the rest of the paradigm, despite the absence of a nasal consonant that could trigger /h/ nasalization. The full paradigm is shown in (3). On the left are the early Zuberoan forms, where nasalization of the initial segment of /hau(ɾ)/ was found only in the first person singular, triggered by the /n/ of /ni/. In the second column are the modern Zuberoan forms showing the extension of the phonological form /h̃au̯(ɾ)/ 'self' to the entire pronominal paradigm.

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

(4) Analogical extension of /h̃au̯(ɾ)/ 'self' within the pronominal paradigm

Early Zuberoan	Later Zuberoan	
ni-haur [nĩ'h̃ãũ̯ɾ]	/nih̃au̯/ [nĩ'h̃ãũ̯]	'me myself'
hi-haur [i'hau̯ɾ]	/hih̃au̯/ [ĩ'h̃ãũ̯]	'you yourself'
zi-haur [ʂi'hau̯ɾ]	/ʂih̃au̯/ [ʂĩ'h̃ãũ̯]	'you yourself' (formal)
gi-haur [gi'hau̯ɾ]	/gih̃au̯/ [gĩh̃ãũ̯]	'we ourselves'

In all but the first person singular, there is no historical or synchronic contextual source for the nasalization of /h/. In order for the [h̃] in [nĩ'h̃ãũ̯] to extend to the rest of the paradigm, the contextually conditioned phonetic variant [h̃] must have been phonologized (i.e. form part of the phonologically contrastive inventory in speech communication; cf. Janda 2003: 409). We conclude that at the time of this analogical extension, the opposition between /h/ and /h̃/ was contrastive in Zuberoan, and has remained so to this day. At the same time, these examples cannot be easily accounted for under an analysis where /h/ nasalization is allophonic due to spread from adjacent nasalized vowels. Under such an account, an abstract nasalized vowel would need to be posited for the morpheme /hau/, though the vowel would only trigger nasalization within this pronominal subparadigm and not in the bare demonstrative *hau* [hau]. In sum, analogical extension of a pronominal form beginning with /h̃/ suggests that this segment was contrastive at the time the paradigmatic shift took place.

2.3 Comparative evidence for Zuberoan /h̃/: Mixean low navarrese /h̃/

Beyond Zuberoan, one other modern dialect appears to preserve and extend a contrast involving /h/ (from historical /h/) versus nasalized /h̃/ from historical intervocalic *n. The variety in question is Mixean Low Navarrese as spoken in the region of Mixe (Basque Amiküze) and described by Camino (2016). Unlike Zuberoan, where contrastive nasalized vowels occur in final stressed syllables (3), there is no evidence in modern Low Navarrese for contrastively nasalized vowels. Nasalized vowels are found only in words containing an aspirate, and, for the most part, these words are cognate with Zuberoan words with nasalized /h̃/. Cognate words in the two eastern varieties involving both aspirates are shown in (5) with Standard Basque forms for comparison.

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

(5) Contrastive /h̃/ in Mixean Low Navarrese

Mixean	Zuberoan	Standard Basque	Gloss
hasi	*hasi*	*hasi*	'begin'
hirü/hiü	*hiü*	*hiru*	'three'
ahür	*ahür*	*ahur*	'palm (of the hand)'
behi	*behi*	*behi*	'cow'
zahar	*zahar*	*zahar*	'old'
ãh̃ãrdi	*ãh̃ãrdi*	*ahardi*	'sow'
ãh̃ãtiak	*ãh̃ãtiak*	*ahateak*	'ducks'
lẽh̃ẽn	*lẽh̃ẽn*	*lehen*	'first'
ĩh̃ĩzin	*ĩh̃ĩzin*	*ehizan*	'hunting'
ũh̃ũre	*ũh̃ũe*	*ohore*	'honor'

Since there is general agreement on the sound change of *VnV > VhV in the history of Basque, and good evidence for this in Latin loans (e.g. Zuberoan *ũh̃ũe* 'honor' < **onore* ≪ Lat. *honore*), Mixean Low Navarrese appears to be conservative in retaining nasalized /h̃/ in these contexts. However, in Mixean, one cannot derive synchronic nasalization of aspirates from nasalized vowels, since contrastively nasalized vowels do not exist. The Mixean data in (5), then, not only lends further support to a contrast between /h/ and /h̃/ in Zuberoan, but also argues for the same contrast in Mixean Low Navarrese, and, more generally, for this as a conservative inherited feature continued from an earlier stage of Basque.

3 Conclusions

In addition to languages such as Kwangali, Seimat, Aguaruna and Arabela, there is an arguable contrast between oral and nasalized /h/ in two modern varieties of Basque: Zuberoan and Mixean Low Navarrese. Where earlier authors suggest that the nasalization of h in Zuberoan be derived from spread of nasalization in adjacent nasalized vowels, Hualde (1993) suggests an alternative where nasalization of the aspirate is contrastive, and spreads to neighboring vowels. Building on Hualde's analysis, I argue above that the distribution of nasalized aspirates in Zuberoan cannot be adequately handled via assimilation from neighboring nasalized vowels without increasing the complexity of the nasalized vowel system. A stronger argument for a contrastively nasalized /h̃/ is the analogical extension of this segment in initial position of a morpheme, /h̃au̯(ɾ)/, within a subpart of the pronominal paradigm. Since no nasalized

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

vowels are present where this form has been extended, there is no phonological source of nasalization, and nasalized /h̃/ should be analyzed as basic and contrastive. Finally, comparative evidence from the Mixean variety of Low Navarrese shows cognate words with nasalized aspirates to those found in Zuberoan. However, Low Navarrese lacks contrastively nasalized vowels. Therefore, the alternative analysis, where nasalization spreads from vowels to /h/ is not possible in this variety. I conclude that /h̃/ is contrastive in both Zuberoan and Mixean Low Navarrese, and that this is a conservative feature of these dialects, directly inherited from earlier stages of Basque. Contrastive /h̃/ evolved from *n intervocalically, and the category was apparently strengthened by contextual nasalization of /h/ in nasal domains.

While the study of Basque phonology has not yielded many typological rarities, this detailed reconsideration of Zuberoan aspirates demonstrates the importance of holistic approaches to grammar incorporating phonetic, phonological and morphological detail. Without evidence from the pronominal paradigm, the arguments for a nasalized aspirate would be equivocal. At the same time, this study also illustrates how small endangered varieties of languages can be the repository of linguistic information of great value. In this case, the Mixean Low Navarrese words with nasalized aspirates allow us to confirm an analysis where the /h/ vs. /h̃/ contrast is basic, and not derived through vowel nasalization. At the same time, Zuberoan itself, a small and endangered variety of Basque, reveals what, phonetically, was thought to be a difficult or impossible contrast (Ohala 1975), instantiates the predictions of phonological theory (Walker and Pullum 1999).

Acknowledgments: I am greatly indebted to Juliette Blevins, José Ignacio Hualde, Joseba Lakarra and two anonymous reviewers of *Linguistics* for their insightful comments and suggestions that have considerably improved this paper. All remaining errors are mine.

Funding: This research has been partially funded by the Spanish Ministry of Economy and Competitiveness [FFI2016-76032-P; FFI2015-63981-C3-2; FJCI-2014-21348] and the Alexander von Humboldt Foundation.

References

Arbelaiz, Juan José. 1978. *Las etimologías vascas en la obra de Luis Michelena*. Tolosa: Kardaberaz.

Blevins, Juliette. 2018. Evolutionary phonology and the life cycle of voiceless sonorants. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and*

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

diachrony (Typological Studies in Language 121), 29–58. Amsterdam & Philadelphia: John Benjamins.
Blust, Robert. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36(1). 149–179.
Blust, Robert. 1998. Semiat vowel nasality: A typological anomaly. *Oceanic Linguistics* 37(2). 298–322.
Bonaparte, Louis-Lucien. 1991 [1869]. Le verbe basque en tableaux. In Euskaltzaindia (ed.), *Opera omnia vasconice*, vol. 1, 175–442. Bilbao: Euskaltzaindia.
Camino, Iñaki. 2016. *Amiküze eskualdeko heskuara*. [The Basque of the region of Amiküze (Mixe)] (Mendaur 11). Bilbao: Euskaltzaindia.
Durvasula, Karthik. 2009. *Understanding nasality*. Newark, DE: University of Delaware dissertation.
Egurtzegi, Ander. 2013. Phonetics and phonology. In Mikel Martínez-Areta (ed.), *Basque and Proto-Basque: Language-internal and typological approaches to linguistic reconstruction* (Mikroglottika 5), 119–172. Frankfurt am Main: Peter Lang.
Egurtzegi, Ander. 2014. *Towards a phonetically grounded diachronic phonology of Basque*. Vitoria-Gasteiz: University of the Basque Country dissertation.
Egurtzegi, Ander. 2015. Different distributions of contrastive vowel nasalization in Basque. *Loquens*, 2(1). e017.
Egurtzegi, Ander. 2017. Phonetically conditioned sound change: Contact induced /u/ -fronting in Zuberoan Basque. *Diachronica* 34(3). 331–367.
Gorrochategui, Joaquín. 1984. *Estudio sobre la onomástica indígena aquitana*. Bilbao: University of the Basque Country, University of Salamanca.
Halle, Morris & George N. Clements. 1983. *Problem book in phonology: A workbook for introductory courses in linguistics and in modern phonology*. Cambridge, MA: MIT Press.
Hualde, José Ignacio. 1993. Topics in Souletin phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *Generative studies in Basque linguistics*, 289–327. Amsterdam & Philadelphia: John Benjamins.
Hualde, José Ignacio. 2003. Segmental phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 15–65. Berlin & New York: Mouton de Gruyter.
Igartua, Iván. 2001. La aspiración en vasco: Ensayo tipológico y diacrónico. *International Journal of Basque Linguistics and Philology/Anuario del Seminario de Filología Vasca "Julio de Urquijo"* 35(1). 185–213.
Igartua, Iván. 2008. La aspiración de origen nasal en la evolución fonética del euskera: Un caso de rhinoglottophilia. *International Journal of Basque Linguistics and Philology/Anuario del Seminario de Filología Vasca "Julio de Urquijo"* 42(1). 171–189.
Igartua, Iván. 2015. Diachronic effects of rhinoglottophilia, symmetries in sound change, and the curious case of Basque. *Studies in Language* 39(3). 635–663.
Janda, Richard D. 2003. "Phonologization" as the start of dephoneticization – Or, on sound-change and its aftermath: Of extension, generalization, lexicalization, and morphologization. In Brain D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 401–422. Oxford: Blackwell.
Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
Lafon, René. 1958. Contribution a l'étude phonologique du parler basque de Larrau (Haute-Soule). In Diego Catalán (ed.), *Estructuralismo e historia: Miscelánea homenaje a André Martinet*, vol. 2, 77–106. La Laguna: Universidad de La Laguna.

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

Lakarra, Joseba A. 2009. Adabakiak /h/ -aren balio etimologikoaz [Amendments on the etymological value of /h/]. *International Journal of Basque Linguistics and Philology/Anuario del Seminario de Filología Vasca "Julio de Urquijo"* 43(1–2). 565–596.
Larrasquet, Jean. 1932. Phonétique du basque de Larrajá (quartier de Barcus). *International Journal on Basque Studies/Revista Internacional de Estudios Vascos* 23(1). 153–191.
Maddieson, Ian. 1984. *Patterns of sounds*. New York: Cambridge University Press.
Martínez-Areta, Mikel (ed.). 2013. *Basque and Proto-Basque. Language-internal and typological approaches to linguistic reconstruction* (Mikroglottika 5). Frankfurt am Main: Peter Lang.
Michelena, Luis. 2011a [1950]. De fonética vasca. La aspiración intervocálica. In Joseba A. Lakarra & Íñigo Ruiz Arzalluz (eds.), *Obras completas 7*, 3–20. Donostia-San Sebastián & Vitoria-Gasteiz: Diputación Foral de Guipúzcoa & University of the Basque Country.
Michelena, Luis. 2011b [1977]. Fonética histórica vasca. In Joseba A. Lakarra & Íñigo Ruiz Arzalluz (eds.), *Obras completas 6*. Donostia-San Sebastián & Vitoria-Gasteiz: Diputación Foral de Guipúzcoa & University of the Basque Country.
Ohala, John J. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 289–316. Stanford, CA: Stanford University Press.
Ohala, John J. 1990. There is no interface between phonetics and phonology: A personal view. *Journal of Phonetics* 18(2). 153–171.
Trask, R. Larry. 1997. *The history of Basque*. London: Routledge.
Walker, Rachel & Geoffrey K. Pullum. 1999. Possible and impossible segments. *Language* 75(4). 764–780.

Authenticated | egurtzegi@phonetik.uni-muenchen.de author's copy
Download Date | 11/5/18 10:06 AM

JASA ARTICLE

An acoustic description of Mixean Basque

Ander Egurtzegi[1,a),b)] and Christopher Carignan[2,a),c)]

[1]*Centre National de la Recherche Scientifique (CNRS)–IKER (UMR 5478), Campus de la Nive, Château-Neuf 15, place Paul Bert 64100, Bayonne, France*

[2]*Speech Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom*

ABSTRACT:
This paper presents an acoustic analysis of Mixean Low Navarrese, an endangered variety of Basque. The manuscript includes an overview of previous acoustic studies performed on different Basque varieties in order to synthesize the sparse acoustic descriptions of the language that are available. This synthesis serves as a basis for the acoustic analysis performed in the current study, in which the various acoustic analyses given in previous studies are replicated in a single, cohesive general acoustic description of Mixean Basque. The analyses include formant and duration measurements for the six-vowel system, voice onset time measurements for the three-way stop system, spectral center of gravity for the sibilants, and number of lingual contacts in the alveolar rhotic tap and trill. Important findings include: a centralized realization ([ʉ]) of the high-front rounded vowel usually described as /y/; a data-driven confirmation of the three-way laryngeal opposition in the stop system; evidence in support of an alveolo-palatal to apical sibilant merger; and the discovery of a possible incipient merger of rhotics. These results show how using experimental acoustic methods to study under-represented linguistic varieties can result in revelations of sound patterns otherwise undescribed in more commonly studied varieties of the same language.
© *2020 Acoustical Society of America*. https://doi.org/10.1121/10.0000996

(Received 3 May 2019; revised 20 September 2019; accepted 7 October 2019; published online 30 April 2020)

[Editor: Benjamin V. Tucker] Pages: 2791–2802

I. INTRODUCTION

Basque is a well-known language in terms of general linguistic description. A number of reference grammars can be consulted for particular details on its syntax or its phonology (especially Hualde and Ortiz de Urbina, 2003). Nevertheless, there are no general acoustic descriptions of the standard language, but rather a limited number of papers that include acoustic characterizations of specific segments as produced in particular varieties of Basque. In this paper, we present a general acoustic description of an endangered and under-documented variety of Basque, namely Mixean Low Navarrese—henceforth, "Mixean Basque" or "Mixean." Most of the work on the linguistic description of this variety has been carried out by the Basque dialectologist Iñaki Camino, who started performing fieldwork in the Mixe region in the early 1980s. His work involves general phonological, morpho-syntactic, and lexical descriptions (Camino, 2016). Given that the acoustics of Mixean Basque have never been discussed in previous work, we will compare our results with the acoustical description of each group of segments as described in different varieties of the language, insofar as they can be found in the literature.

A. General background

The region of Mixe (*Amiküze* in Mixean Basque) is situated in the South-West of France, in the Pyrénées-Atlantiques department. Within the Basque Country, it is located in the northern part of the historical province of Low Navarre or *Nafarroa Beherea*. The population of Mixe was 7856 as of 2015 (L'Institut national de la statistique et des études économiques, 2019), although the number of speakers of Basque in the region is much lower (Camino, 2016). The Mixe region is formed by 32 towns, the main city being Donapaleu (*Dnaplü* in Mixean Basque and *Saint-Palais* in French), with a population of 1842 in 2016 (L'Institut national de la statistique et des études économiques, 2019). The variety of Basque spoken in the region, Mixean Basque or *amiküzera*, is usually classified as part of the Low Navarrese dialect (Michelena, 2011) or as a transition variety (Zuazo, 2008). Nevertheless, Mixean Basque has more in common, phonologically, with the neighboring Zuberoan dialect, usually considered the most deviant variety of Basque, than with Low Navarrese.

Mixean Basque has historically been in contact with Bearnese Gascon to the North (Mixe being South of the region of Béarn) and Zuberoan Basque to the East.[1] Due to an increasing French influence at all levels of the society, the number of speakers of Basque and Gascon has steadily decreased in the region during the last century (Camino, 2016, pp. 48–49). Although no study regarding the current language use in Mixe has been performed recently, Camino

[a)]Also at: Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Munich, Germany.
[b)]Electronic mail: ander.egurtzegi@iker.cnrs.fr, ORCID: 0000-0002-3451-323X.
[c)]ORCID: 0000-0002-1565-3691.

(2016, p. 51) describes the language as being on its way to disappearing, with children currently raised and schooled in French. According to a 2016 study by the association *Zabalik* (Camino, 2016, p. 55), only 9% of children (75 individuals) are currently schooled in a Basque-speaking model (Zabalik, 2016), where children will be taught the standard language instead of their local varieties.

In short, Mixean is one of the varieties of Basque that is most different from the rest (alongside Zuberoan), but is understudied due to being underrepresented in the literature and not being considered a dialect of its own by most Basque dialectologists. Given that it is an endangered variety and that its speakers are among the oldest population, failure to study this variety now may result in the inability to study it at all.

B. Mixean phonology

The phonemic inventory of Mixean Basque has not been clearly established in the previous literature. Nevertheless, a provisional description might be inferred from the study by Camino (2016) and the work on neighboring varieties (Egurtzegi, 2018b; Hualde, 2003). Thereby, Mixean ostensibly includes at least 34 contrastive consonants—12 stops /p, t, c, k, pʰ, tʰ, cʰ, kʰ, b, d, ɟ, ɡ/, ten sibilants /s̪, s, ʃ, t̪s̪, tʃ, z̪, z, ʒ, d̪z/, a labiodental fricative /f/, nine sonorants /m, n, ɲ, l, ʎ, ɾ, r, j, w/, two laryngeals /h, h̃ / (which are sonorant-like except word-initially)—and six contrastive vowels /a, e, o, i, y, u/, although not all segments are equally frequent.[1] In addition, some non-nativized French loanwords are produced with /ʁ/, /v/, /ɛ/, /œ/, or nasalized vowels (phonemic segments in French), although these are rare and generally not considered part of the Mixean Basque inventory. Excluding recent loanwords, there are no phonemic nasal vowels in Mixean Basque (Camino, 2016, p. 200), but coarticulatory vowel nasalization can be found in contact to any of the nasal segments /m, n, ɲ, h̃/.

Some segments are only contrastive in particular phonological environments. Laryngeals and aspirated stops can only be found in the onset of the first two syllables of the word; stops are absent from word-medial codas and their laryngeal configuration is neutralized in word-final position (to the plain voiceless series). Rhotics only contrast intervocallically; they are absent from word-initial position and they neutralize in tautosyllabic obstruent-rhotic onset clusters and in coda position to a segment that is not clearly described in the literature on this or any close variety. In his general description of Basque, Hualde (2003, p. 30) describes this neutralized rhotic as a trill realized with fewer vibrations. Regarding sibilants, /ʧ/ is the only affricate found word-initially, and voiced sibilants are mostly present in loanwords and compound boundaries such as *deuse* /deuz̪e/ "nothing (at all)" (*deus* "nothing" + *ere* "as well"). Sibilants preceding a voiced obstruent are also voiced. The sonorants /m, ɲ, ʎ/ do not occur word-finally. See Hualde (2003) for a description of the phonotactic restrictions of most Basque dialects.

The Mixean variety of Low Navarrese Basque shows a more complex phonemic inventory than general Low Navarrese or most other varieties of Basque. The only variety of Basque that has a more complex inventory is Zuberoan, which has a phonemic inventory similar to that of Mixean Basque but also includes four contrastive nasalized vowels (Egurtzegi, 2015). The most notable differences between Mixean and Zuberoan Basque, on the one hand, and the other Basque varieties, on the other, are the set of aspirated stops /pʰ, tʰ, cʰ, kʰ/, the nasalized laryngeal /h̃/, and the rounded high front vowel /y/. All other Basque varieties only have five vowels /a, e, o, i, u/, two sets of stops (voiced and voiceless unaspirated), and either one (/h/) or no laryngeal approximants. Aspiration—both as a phonological segment (Egurtzegi, 2018b; Hualde, 2018) and as a feature of a two-way stop distinction (Egurtzegi, 2018a)—was arguably part of an earlier, common stage of the language. However, contrastive aspiration is limited to the Easternmost varieties today, and it has been reported to show recession even in the dialects where it is most present, namely Zuberoan and Mixean (Camino, 2016). Although both Zuberoan and Mixean have been reported to have a front rounded vowel, Mixean /y/ has been auditorily described as different from Zuberoan /y/, and closer to /ø/, especially in pre-pausal position (Haase, 1992, p. 29). Another difference between Mixean and Zuberoan is that Zuberoan lost the tap in the 19th century, while Mixean still preserves two rhotics (/ɾ, r/) intervocalically.

II. PREVIOUS ACOUSTIC DESCRIPTIONS

Most acoustic descriptions of particular varieties of the Basque language focus either on sibilants or on the vocalic inventory. In addition to these, there are some studies that focus on rhotics and a couple of studies on the stop system of Zuberoan Basque. We are not aware of any acoustical study that analyzes all segments in any variety of the Basque language. The current study will focus on the sets of segments that have been analyzed in other varieties of Basque: vowels, stops, sibilants, and rhotics. Although prosody has been widely studied in Basque, accentuation is not contrastive in Mixean,[2] so it will not be analyzed in this paper.

A. Previous findings: Vowels

The acoustics of vowels have received the most attention in the literature on Basque. The most general and complete work on the acoustics of vowel systems is Urrutia *et al.* (1995a), which summarizes previous works and presents descriptions and formant charts of multiple dialects of the language. Among the varieties studied in this book, we can find Eastern Low Navarrese (ELN) (Urrutia *et al.*, 1995a, pp. 151–175; results for ELN were also published in Urrutia *et al.*, 1995b) which includes data from a sample of three male speakers from Donibane Garazi (Cizean Low

Navarrese), Donapaleu (Mixean Low Navarrese), and the Salazar Valley (Salazarese). This somewhat heterogeneous group is the closest reference to our data that can be found in the literature.

Urrutia *et al.* (1995a) follow the dialectal classification of Yrizar (1981, pp. 39–45), who classify very different varieties of Basque under ELN. The study involved three speakers, each representing a different sub-variety. The results are not presented by speaker, but only as aggregates, so we cannot extract the particular information about the Mixean informant. In addition, there is no information about /y/, a vowel which has been claimed to be contrastive in Mixean Low Navarrese (Michelena, 2011), but is not present in the varieties of the other two informants (Cizean Low Navarrese and Salazarese). We could find no information on whether the instances of /y/ were discarded from the analysis or aggregated to the /u/ tokens, the segment to which it corresponds in the varieties of the two speakers who do not have /y/ in their inventory.

The only acoustic study of a variety of Basque that has a six-vowel inventory, namely Zuberoan Basque, is found in the same book (Urrutia *et al.*, 1995a, pp. 203–234). It presents the aggregate results of two informants of (Northern) Zuberoan Basque, with no information on the number of tokens under analysis. The authors mention [ø] and specify that it is not a contrastive vowel (Urrutia *et al.*, 1995a, p. 206), but an allophonic variant of /y/. However, we could not find a phonological context for this realization in their work, and this proposal is not found in any other work that we are aware of—only Haase (1992) proposes [ø] as an allophone of /y/ in prepausal position, but in Mixean instead of Zuberoan. The study concludes that stress does not play a significant role in the acoustic realization of Zuberoan vowels (Urrutia *et al.*, 1995a, p. 206). The results of the two relevant studies in Urrutia *et al.* (1995a) are summarized in Table I.

All studies,[3] with the exception of the study of Zuberoan discussed above, describe five-vowel systems with two back non-low vowels, two front non-low vowels, and a low central vowel. In most cases (all in the case of back vowels), mid vowels are mid-closed: they are closer to high vowels than low vowels. [i] is consistently the highest and most anterior vowel, and [o] tends to be less retracted than [u], although multiple studies report no meaningful differences in $F2$ for back vowels (Hualde *et al.*, 2010; Salaburu, 1984, etc.).

TABLE I. $F1$ and $F2$ values (Hz) and duration values (ms) in ELN and Zuberoan (Z) (Urrutia *et al.*, 1995a; Urrutia *et al.*, 1995b).

Vowel	Variety	$F1$	$F2$	Dur.
/i/	ELN	348	2277	57.0
	Z	353	2442	68.0
/e/	ELN	504	1879	59.1
	Z	509	2002	59.1
/a/	ELN	730	1469	65.0
	Z	821	1449	71.0
/o/	ELN	521	1058	64.7
	Z	505	1024	76.7
/u/	ELN	383	1036	58.5
	Z	390	949	82.1
/y/	ELN	-	-	-
	Z	379	1812	80.8
[ø]	ELN	-	-	-
	Z	416	1755	78.5

B. Previous findings: Stops

Only a handful of studies can be found in the literature that analyze the acoustic realization of Basque stops. Two of these studies investigated Zuberoan Basque, which makes them relevant as a means of comparison for our study. Gaminde *et al.* (2002) analyzed the recordings of four male speakers of Zuberoan Basque (from the region of Pettarra), including a total of 302 analyzed tokens. The study was restricted to word-initial stops in a stressed syllable. The authors describe a three-way contrast involving prevoiced ("negative" voice onset time, VOT; i.e., voicing starts during the closure phase of the stop), plain voiceless (short-lag positive VOT; i.e., voicing starts shortly after the release of the stop), and voiceless aspirated (long-lag VOT; i.e., voicing starts long after the release) stops. Mounole (2004) performed a comparable study of the same dialect, also involving four male participants (from Larrain), including a total of 861 tokens. She analyzed word-initial voiceless stops and found a difference between plain voiceless and voiceless aspirated stops, with no significant differences due to stress (Mounole, 2004, p. 222).

The few other studies on Basque stops found in the literature investigated intervocalic stop lenition. They demonstrated that plain voiceless stops showed a tendency towards voicing and lenition intervocalically (Nadeu and Hualde, 2015; Saadah, 2011), which was strongest for word-final voiceless stops when preceding a word-initial vowel (Hualde *et al.*, 2019).

The studies discussed above show that Zuberoan has a three-way stop opposition based on VOT (pre-voicing or voicing during closure, short-lag, and long-lag VOT) that is not present in most Basque varieties. VOT values increase with increased posteriority for voiceless and voiceless aspirated stops but are roughly similar for all voiced stops. Stops are measured word-initially due to intervocalic lenition, and stress does not seem to be a factor affecting VOT.

C. Previous findings: Sibilants

There are six sibilants—represented orthographically as <s, z, x, ts, tz, tx>—that have traditionally been recognized as common to all Basque dialects. Most authors (Egurtzegi, 2013; Hualde, 2003; Michelena, 2011, inter alia) describe these segments as voiceless apico-alveolar (transcribed as /s̪, ts̪/), dorso- or lamino-alveolar (transcribed as /s̻, ts̻/), and post-alveolar (transcribed as /ʃ, tʃ/). A number of studies (Larrasquet, 1934; N'Diaye, 1970; Txillardegi, 1982) have described the pronunciation of written <s> as retroflex

(instead of apico-alveolar), usually referring to Eastern varieties of Basque, restricting the apico-alveolar realizations to varieties in contact with Spanish (N'Diaye, 1970, p. 15). Yárnoz (2002a,b) described the six sibilants in the Basque variety of Bortziri (Northern High Navarrese) as flat post-alveolar (transcribed by the author as /s̻, ts̻/), denti-alveolar (/s̪, ts̪/), and palatalized post-alveolar (/ɕ, tɕ/) instead, although only Jurado Noriega (2011) followed this description for other varieties. In addition, some Eastern varieties have developed voiced sibilants that are mostly found in loanwords from Gascon and French and in *liaison* (Michelena, 2011, p. 230). Some dialects (Bizkaian and Gipuzkoan, in particular) have developed mergers, resulting in a reduction of the size of the sibilant inventory (see Muxika-Loitzate, 2017, and Beristain, 2018b, for recent acoustic studies on sibilant merger in Bizkaian and Gipuzkoan). In Bizkaian and some neighboring varieties, the merger of alveolar sibilants /s̪, ts̪, s̻, ts̻/ results in an apico-alveolar fricative and a lamino-alveolar affricate realization [s̻, ts̻] (maintaining the post-alveolar sibilants), whereas in Gipuzkoan /s̪/ and /s̻/ merge in a lamino-alveolar realization ([s̻]). In some varieties in contact with French, the apico-alveolar fricative /s̪/ merges with the post-alveolar fricative /ʃ/ instead (see Hualde, 2010, for a comprehensive discussion of sibilant mergers). Some authors have even concluded that the sibilant merger is a completed phonological process in Basque (Urrutia *et al.*, 1991, p. 331), but this assertion is only widely accepted for western dialects (Yárnoz, 2002b, p. 14).

As in the case of the vowels, the study that is geographically most relevant for comparison with our analysis is that by Urrutia *et al.* (1991), in which some acoustic parameters of the sibilants of different eastern varieties of Basque are described. However, their study of ELN sibilants shares the problematic dialectal classification followed in their aforementioned study on vowels (see Sec. II A) and, most importantly, they report the lower energy cut-off frequencies of the sibilants instead of their spectral center of gravity (CoG), reducing the comparability with more recent studies. The recent studies that analyze sibilant CoGs differ from our study in various aspects: Iglesias *et al.* (2016) used nonce words instead of lexical items, Gandarias *et al.* (2014) included three tokens for each sibilant, Beristain (2018b) and Muxika-Loitzate (2017) analyzed fricative sibilants (rather than both fricatives and affricates), Hualde (2010) and Iglesias *et al.* (2016) reported results of a single speaker, and Gandarias *et al.* (2014) only provided complete CoG values from one speaker (female from Lekeitio) even though the speech of three speakers was analyzed. All of these studies analyzed speakers from central or western dialects, while Mixean is one of the most eastern varieties of Basque.

Studies[4] on western varieties have consistently reported mergers (as described above), while studies on High Navarrese have reported maintenance of all six (fricative and affricate) voiceless sibilant phonemes. Voiced sibilants tend to be linked to eastern varieties (although they can also be found in some western varieties, see Gandarias *et al.*, 2014), but no study measuring the CoG of an eastern variety can be found in the literature. In varieties with no sibilant mergers, lamino-alveolar sibilants have the highest CoGs, followed by apico-alveolars and then by post-alveolars.

D. Previous findings: Rhotics

Most, if not all of the previous studies on the realization of Basque rhotics have been performed by Gaminde and colleagues (Gaminde, 2006; Gaminde *et al.*, 2017; Gaminde *et al.*, 2016). Gaminde *et al.* (2016) studied the different realizations of intervocalic taps in 15 speakers (20–23 years old) of Bizkaian Basque. They measured duration, formants, and acoustic energy in a total of 330 tokens of read speech. They observed realizations of five allophones of the intervocalic tap in Basque—[ɾ, ɾ͡ɹ, ɹ͡ɹ, ɹ̝, r]—classified by whether the rhotic showed lingual contact and whether it showed fricative noise (partial frication, frication throughout the rhotic, or no frication at all). Gaminde *et al.* (2017) followed this work with a larger study on the realization of intervocalic trills, involving 155 young speakers (23–36 years old) from the whole Basque-speaking territory. Their results showed that the majority (50%) of canonical trills were realized with two contacts, with far less occurrence of three (8.54%) and four (0.4%) contacts. Among the trills realized with incomplete occlusions (Lindau, 1985), most showed a tap followed by an approximant (27.13%) or fricative (10.67%).

Gaminde *et al.* (2017) showed that speakers of the central and western Basque dialects (in contact with Spanish) tend to use the alveolar trill, while speakers of eastern Basque dialects (in contact with French) use voiced uvular rhotics [ʁ, ʀ] instead. Interestingly, they reported that all of the speakers of eastern dialects in their study consistently used uvular rhotics. All studies have reported a wide range of variation in the production of both taps (Gaminde *et al.*, 2016) and trills (Gaminde *et al.*, 2017).

III. METHODOLOGY

The data used for the acoustic analyses presented in the current study include audio recordings from Camino (2016), compiled after intensive fieldwork in the Mixe region during the last four decades. Given the difficulty of accessing this population, these might likely be the only available recordings of Mixean. In addition, the advanced age of the participants makes any later recording problematic. Camino (2016) recorded more than an hour of audio data from each of 15 speakers of Mixean Basque. All speakers lived in different villages from the Mixe region. From this data set, we have selected ten audio recordings that were made between 2005 and 2015. The selected audio files were recorded in rural environments, with a portable minidisc recorder (SONY MZ-R30), at a sampling rate of 44.1 kHz. Recordings of speakers older than 85 years were excluded from the analysis to avoid age-related phonetic biases, and recordings made in the 1980s with a DAT recorder were excluded in order to obtain

homogeneous data regarding the state of the language and the recording method used. The resulting corpus includes recordings from seven male speakers and three female speakers from ten different villages of the region of Mixe (Donapaleu, Uhartehiri, Sorhapürü, Arrüeta, Martxüeta, Labetze, Amendüze, Gamue, Zohota, and Arberatze), with ages ranging from 80 to 85 years [$\mu = 83$, standard deviation (SD) = 1.7]. The length of the recordings was 5.5 min on average, with a range of 3.5 min to 8.5 min (SD = 1.7). All audio recordings were force-aligned using the WebMAUS application (Kisler *et al.*, 2017) set to Basque (FR); the resulting Praat TextGrids were subsequently hand-corrected by either the first author, the second author, or a graduate student in phonetics. All acoustic analyses were performed in Praat (Boersma and Weenink, 2019) using custom-written functions created by the second author, following the protocols outlined below.

A. Methods: Vowels

We analyzed a total of 2221 vowel tokens, ranging between 112 and 370 tokens per speaker ($\mu = 222.1$, SD = 72.2). Formant estimations were made using the Burg LPC method in Praat. The formant estimator was optimized for each speaker by manually adjusting the maximum formant parameter (five formant estimation) until the formant trajectories aligned consistently with visible formant bands in a broad-band spectrogram. Average $F1$, $F2$, and $F3$ measurements were then obtained within the middle 10% of each vowel interval (i.e., a window equal to 10% of the vowel duration, centered on the vowel midpoint). Additionally, the duration of the entire vowel interval (in ms) was measured and logged. The formant measurements were then imported into R (R Core Team, 2019), where speaker-normalized values were computed using Lobanov normalization (i.e., z-score transformation). The resulting z-scores were then converted back to the Hz scale using the average standard deviation and grand mean of all ten speakers. These normalized formant values maintain the interpretability of the Hz scale, while also retaining the speaker-specific normalized structure of the z-scores.[5] For the characterization of the acoustic vowel space, the speaker-normalized $F1$ and $F2$ values were combined for each of the six vowels /i, e, a, o, u, y/ and vowel-specific z-scores were computed. Outliers were removed from each vowel category by excluding observations associated with vowel-specific $F1$ or $F2$ z-scores with absolute values greater than 3.

B. Methods: Stops

Only stops in utterance-initial position were included in the analysis, since stop lenition occurs intervocalically (Nadeu and Hualde, 2015), even across word boundaries (Hualde *et al.*, 2019). In this environment, the palatal stop series /ɟ, c, c^h/ was produced rarely in the data—/ɟ/ was produced twice by one speaker and one time each by three speakers; /c/ was produced four times by one speaker, three times by another speaker, and only once by a third speaker; /c^h/ was produced only one time by a single speaker. This stop series was therefore excluded from the final analysis. We analyzed a total of 717 stop consonant tokens, ranging between 41 and 114 tokens per speaker ($\mu = 71.7$, SD = 25.2). Measurements of VOT were made for the nine stop consonants /b, d, ɡ, p, t, k, p^h, t^h, k^h/. VOT was first estimated automatically using AutoVOT (Keshet *et al.*, 2014) in Python (Python Software Foundation, 2019), trained on a subset of manually-annotated Praat TextGrids. The onsets and offsets of the estimated VOT segments were then manually verified by the second author and hand-corrected as needed. Before the final analysis, speaker-normalized z-scores were computed, and outliers were removed by excluding observations associated with z-scores with absolute values greater than 3.

C. Methods: Sibilants

We analyzed a total of 1494 sibilant consonant tokens, ranging between 97 and 214 tokens per speaker ($\mu = 149.4$, $SD = 37.8$). In the production of sibilants, the sound source is located in the front cavity, and the tight vocal tract constriction results in a weakening of the acoustic coupling between the front and back cavities; thus, the resonant frequencies are primarily associated with the length of the front cavity (Johnson, 2003, pp. 124–125). As such, in the absence of articulatory measurements (e.g., electropalatography), spectral CoG can be used as an approximation of place of articulation for sibilant consonants. Accordingly, CoG measurements for the sibilants /ʃ, ʧ, s̪, ts̪, s̺, ts̺, ʒ, z̪, z̺/ were obtained in Praat. CoG measurements were calculated from the absolute spectrum after band-pass filtering the audio (from 300 Hz to 19 kHz). For fricative consonants, average CoG measurements were obtained within the middle 10% of the consonant interval (i.e., the same analysis window used for formant measurements of the vowels). For affricate consonants, the same 10% window was used for analysis; however, instead of centering the window on the midpoint of the consonant (i.e., 50% of the consonant interval) the window was centered on 75% of the consonant interval. This helped to ensure that average CoG values were obtained in the fricative portion of the affricate and excluded the closure and/or release.

Although CoG measurements are not necessarily expected to vary between speakers due to the same factors that condition inter-speaker differences in formant frequencies of vowels (e.g., differences in vocal tract length between males and females), it is expected that some degree of inter-speaker variation in CoG may be observed due to differences in speaker morphology (e.g., height of the palate, formation of the teeth, etc.). Because of this, CoG values were Lobanov normalized before conversion back to Hz, in the same manner as previously described for formant values. Before the final analysis, outliers were removed by excluding observations associated with z-scores with absolute values greater than 3.

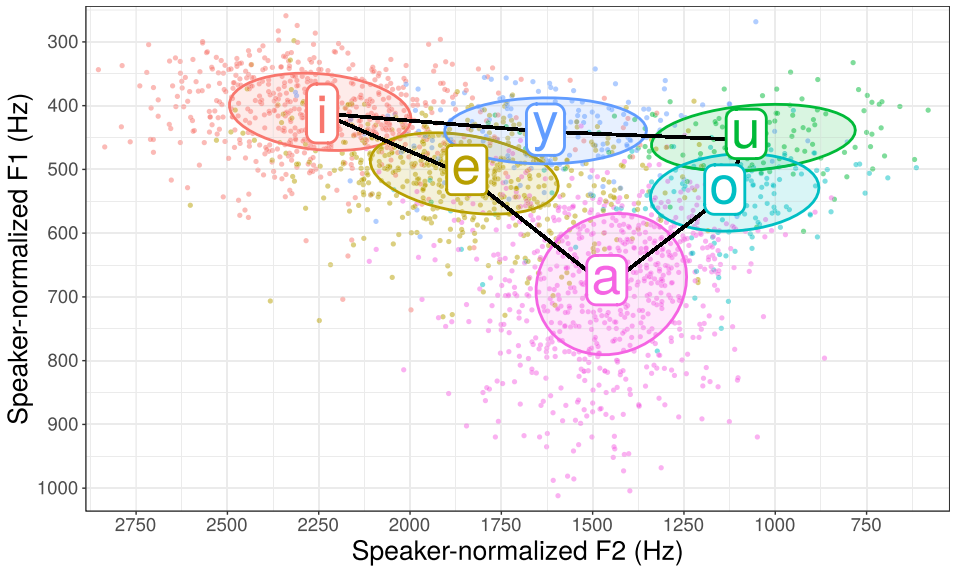


FIG. 1. (Color online) Acoustic vowel space of speaker-normalized $F1/F2$ values (Hz) for Mixean Basque. Ellipses represent coverage of 50% of the data in each vowel category.

D. Methods: Rhotics

We analyzed a total of 171 rhotic tokens, ranging between 9 and 29 tokens per speaker ($\mu = 17.1$, SD = 6.2). The number of lingual contacts made during the productions of /r/ and /ɾ/ was estimated programmatically in Praat according to the following method. The audio signal was first low-pass filtered at 2000 Hz, and an intensity/amplitude (dB) trajectory was generated from the filtered signal. Second, the intensity minima within the interval of each /r, ɾ/ token were identified, and the total number of minima within each interval was logged. We interpret each intensity minimum as a lingual contact, according to the understanding that any sort of intra-oral constriction will result in a loss of overall energy, due to the increased airflow impedance; a rapid constriction associated with a lingual tap is expected to result in a corresponding rapid loss of acoustic energy, thus producing an intensity minimum. Three environments were investigated: intervocalic rhotics, rhotics in onset clusters, and rhotics in coda position. As previously described in Sec. I B, rhotics only contrast intervocalically in Basque: they are neutralized in onset clusters as well as in coda position. Thus, separate analyses were carried out for /r/ and /ɾ/ in intervocalic position, but /r/ and /ɾ/ items were combined for the analyses of onset clusters and codas.

IV. RESULTS

A. Results: Vowels

The speaker-normalized $F1/F2$ acoustic vowel space is shown in Fig. 1. The ellipses shown denote 50% of the data variation for each vowel category (i.e., data ellipses, not confidence interval ellipses). The colored vowel symbols denote the $F1/F2$ mean for each category; black bars connect the category means in order to help visualize the overall shape of the acoustic vowel space. The average $F1$, $F2$, $F3$, and duration values for the six vowel categories are given in Table II.

With regard to $F1$ (i.e., acoustic vowel height), a three-way distinction can be observed: /i, y, u/ are realized as high vowels with similar $F1$ values, /e, o/ are realized as mid vowels with similar $F1$ values, and /a/ is realized as a low vowel. However, it should be noted that the $F1$ distinction between the high vowels /i, y, u/ and the mid vowels /e, o/ is not as great as the $F1$ distinction between the mid vowels /e, o/ and the low vowel /a/, as observed for other Basque varieties in previous works. With regard to $F2$ (i.e., acoustic vowel anteriority/posteriority), the results are somewhat more varied. Among the three high vowels, there is a clear three-way distinction in which /i/ and /u/ are realized as front and back vowels, respectively, while /y/ seems to be realized as a central (rather than front) vowel. In order to test for differences among the high vowels—and, thus, properly characterize the acoustic realization of /y/—linear mixed effects (LME) models were constructed using the *lme4* R package (Bates *et al.*, 2015) with either $F2$ or $F3$ as the response, *Phone* as the fixed factor, and random intercepts by *Speaker* and *Word*. Tukey *post hoc* comparisons were constructed using the *multcomp* package (Hothorn *et al.*, 2008), with α level compensation for multiple comparisons using Benjamini-Hochberg adjustment; significant differences are reported for adjusted p-level 0.05. With regard to $F2$, each of the pairwise differences are significant, and are summarized as: /i/ > /y/ > /u/. With regard to $F3$, each of the pairwise differences are

TABLE II. Average $F1$, $F2$, and $F3$ values (Hz) and duration values (ms) of vowels in Mixean Basque.

Vowel	$F1$	$F2$	$F3$	Dur.
/i/	412	2241	2946	81
/e/	510	1845	2724	76
/a/	683	1461	2714	88
/o/	541	1138	2758	98
/u/	454	1079	2804	85
/y/	441	1629	2711	79

https://doi.org/10.1121/10.0000996

significant, and are summarized as: /i/ > /u/ > /y/. In the absence of articulatory data, these acoustic results suggest that /y/ is rounded (i.e., low *F*3) but centralized (i.e., *F*2 that is midway between /i/ and /u/), and would thus more appropriately be characterized as /ʉ/ in this variety. The mid-vowels /e, o/ display a clear front-back distinction (although /e/ is much more retracted in comparison to /i/), while the low vowel /a/ is realized as a central vowel along the *F*2 dimension. These results suggest that the Mixean Basque vowel system is characterized by two front vowels (/i/ and /e/), two central vowels (/ʉ/ and /a/), and two back vowels (/u/ and /o/).

B. Results: Stops

VOT values for voiced stops /b, d, ɡ/, unaspirated stops /p, t, k/, and aspirated stops /p^h, t^h, k^h/ are displayed in box plots in Fig. 2. In this figure and all similar box plot figures in the paper, mean values are displayed as white circles, median values are displayed as horizontal black bars, inward-facing notches display the standard error around the median (if the notches of two boxes do not overlap along the *y* axis, this suggests that the medians are significantly different), the boxes denote the inter-quartile range (IQR; the middle 50% of the data), and the whisker lines denote 1.5 × the IQR; outliers have been removed to aid visualization. While the results suggest that there are differences between the three stop categories in Mixean Basque that are in the expected direction—voiced stops are produced with voicing during the closure phase, unaspirated and aspirated stops are produced with positive VOT, and aspirated stops are produced with greater positive VOT values than unaspirated voiceless stops—the range of VOT values is substantially smaller in the Mixean variety compared to the Zuberoan variety. Note that VOT values referring to voicing during closure are denoted by negative values, following conventional use in studies on VOT.

Both Gaminde *et al.* (2002) and Mounole (2004) report a distinction between unaspirated and aspirated voiceless stops in the Zuberoan variety, indicating a three-way voicing distinction. In our results for the Mixean variety, although the voiceless consonants that are (ostensibly) aspirated are indeed realized with overall greater positive VOT values than their unaspirated counterparts, the distinction between the two groups is not as large as has been reported for the Zuberoan variety. Recall from the introduction that stop aspiration has been lost in most Basque dialects and has been reported to show recession in the dialects that maintain it. Therefore, it is of interest for our current study to determine in an objective, data-driven manner whether a three-way contrast does exist in Mixean Basque. To this end, we performed clustering of speaker-normalized VOT values (*z*-scores) with the *mclust* R package (Fraley *et al.*, 2019), using the Bayesian Information Criterion (BIC) based on finite Gaussian mixture modeling to determine the optimal number of clusters present in the data. The results suggest that the overall distribution is indeed best described by three clusters/groups. The proportions of items belonging to these groups are shown in Fig. 3. In this figure, the three panels correspond to the three groups identified by the Gaussian clustering, and bars are shown in each panel for each of the nine consonant categories that are suggested by the orthographic representations. A given bar displays the percentage of the total number of items of the given phone that are identified as belonging to the given cluster. The average VOT for the given consonant belonging to the cluster is displayed above each bar. By way of example, we take the clustering results for /ɡ/: 81% of /ɡ/ items were identified as belonging to Group 1, with an average VOT of -40 ms for these items; 12% of /ɡ/ items were identified as belonging to Group 2, with an average VOT of 15 ms for these items; and 7% of /ɡ/ items were identified as belonging to Group 3, with an average VOT of 34 ms for these items.

These results suggest that a three-way voicing contrast is indeed present in the variety. However, the classification of the observations based on these groups/clusters does not clearly delineate categories comprised solely of the three stop consonant groups /b, d, ɡ/, /p, t, k/, and /p^h, t^h, k^h/. Group 1 consists of consonants with average pre-voicing values in the range of -44 ms to -40 ms (i.e., from 40 to 44 ms of voicing during the closure). Only /b, d, ɡ/ items are included in this group, and the majority (81%–94%) of the

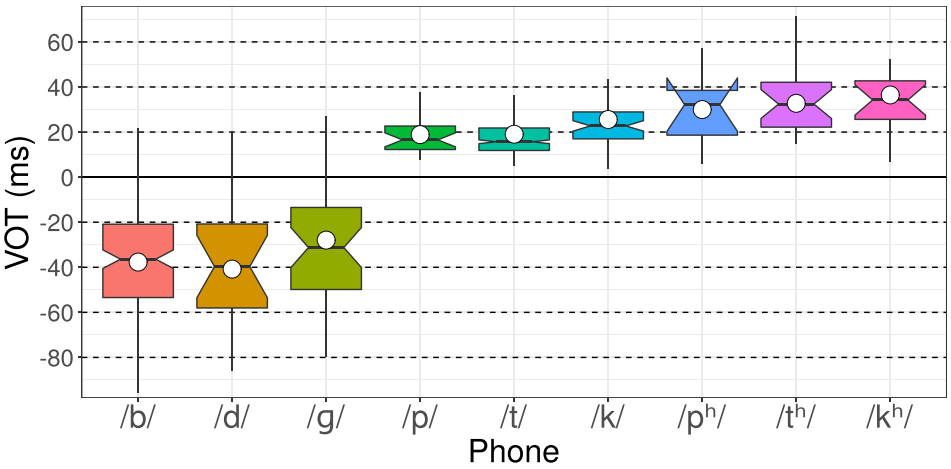


FIG. 2. (Color online) VOT (ms) values of stop consonants in Mixean Basque.

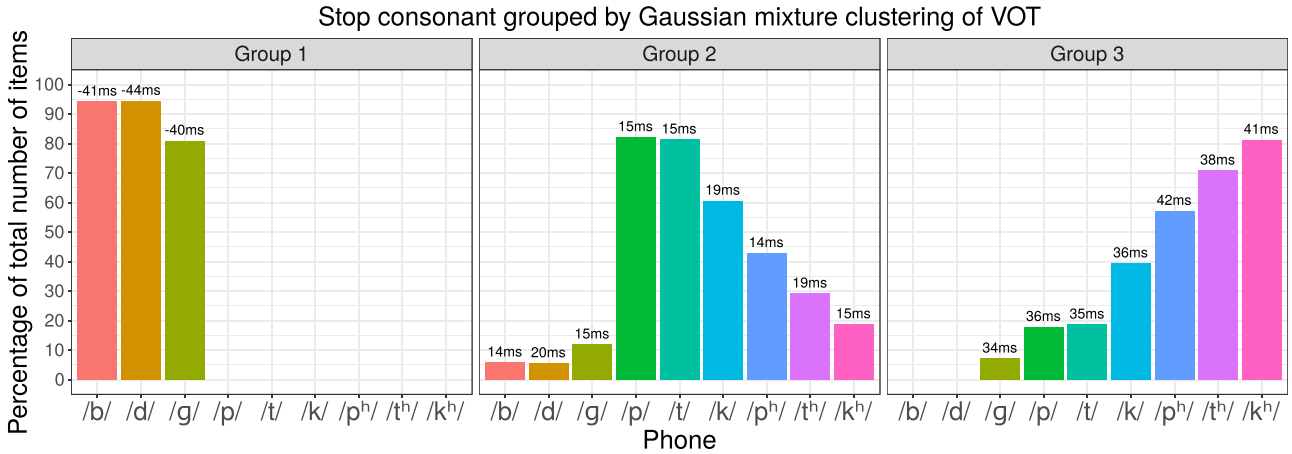


FIG. 3. (Color online) VOT (ms) values of stop consonants in Mixean Basque, separated into three clusters identified by finite Gaussian mixture modeling.

items for these consonants are included in this group for each of the three consonants. This suggests that Group 1 represents voiced consonants, comprised solely of /b, d, g/ items, and that these /b, d, g/ items are nearly categorically realized with voicing during the closure (i.e., pre-voicing).

However, considerably more variation can be observed for Groups 2 and 3. Group 2 consists of consonants with average positive VOT values in the range of 14–20 ms. Interestingly, items from all nine stop consonants are realized with VOT values in this range. While the most prominent group of consonants in this cluster is indeed the unaspirated voiceless /p, t, k/ triad—82%, 81%, and 61% of the total number of items for these consonants, respectively—a small percentage of each of the voiced consonants /b, d, g/, as well as a much larger percentage of the aspirated voiceless /pʰ, tʰ, kʰ/ are also produced with VOT values in this 14–20 ms range. Finally, Group 3 consists of consonants with average positive VOT values in the range of 34–42 ms. While there are no instances of voiced /b, d/ in this group, 7% of the total number of /g/ items are produced with VOT values in this range, as well as a higher percentage of unaspirated voiceless /p, t, k/—18%, 19%, and 39% of the total number of times for these consonants, respectively. However, the most prominent group of consonants in this third cluster is indeed the aspirated voiceless /pʰ, tʰ, kʰ/ triad—57%, 71%, and 81% of the total number of items for these consonants, respectively.

Overall, the VOT results for the stop consonants suggest that there is a clear categorical distinction between voiced (negative VOT values, i.e., voicing starts during the closure) and voiceless (positive VOT values, i.e., voicing starts after the release) consonants, but that the sub-distinction of aspiration vs non-aspiration among the voiceless consonants displays a substantial degree of overlapping phonetic realizations. It has previously been shown that phonological stop aspiration can vary greatly from town to town within the Zuberoan variety, including recession of this feature (Michelena, 2011). This intra-dialectal variation, if also present in Mixean Basque, may provide a possible explanation for the phonetic variation observed here.

C. Results: Sibilants

Box plots of speaker-normalized CoG values for the sibilant consonants are displayed in Fig. 4. The average values among all nine consonants range from 3365 Hz (for /ʒ/) to 5088 Hz (for /ʦ/). In order to test for significant differences between the phones, an LME model was constructed with *CoG* as the response, *Phone* as the fixed factor, random slopes and intercepts by *Speaker*, and random intercepts by

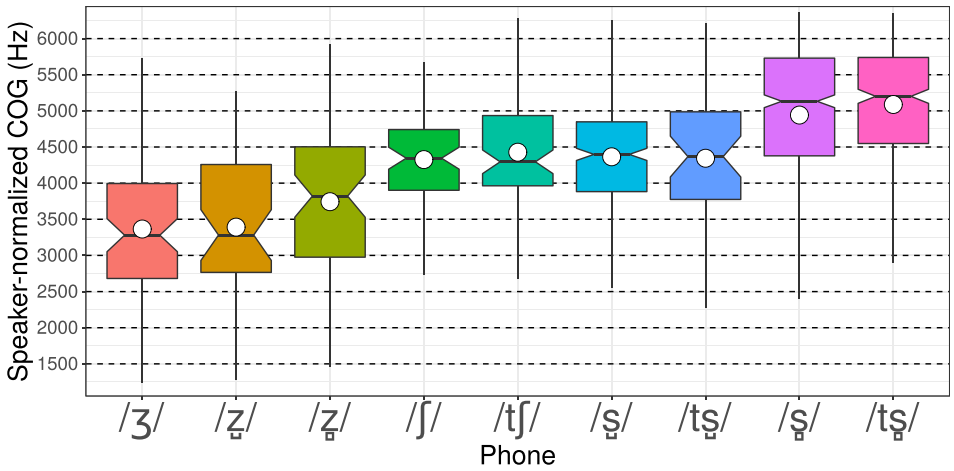


FIG. 4. (Color online) CoG (Hz) of sibilant consonants in Mixean Basque.

Word. Tukey *post hoc* comparisons were constructed with α level compensation for multiple comparisons using Benjamini-Hochberg adjustment; significant differences are reported for adjusted p-level 0.05.

The group of voiced sibilants /ʒ, s̪, z̪/ is realized with the lowest CoG values;[6] these three consonants have similar CoG values, except for a marginally significant difference ($p = 0.027$) between /ʒ/ and /z̪/ due to the slightly raised CoG for /z̪/. The group of voiceless laminal sibilants /s̪, ts̪/ is realized with the highest CoG values, with no significant difference between them. It is of interest to note that both fricative and affricate apical sibilants have similar CoG values to their palatal counterparts (i.e., no significant differences among /ʃ, tʃ, s̪, ts̪/), suggesting a merger in the place of articulation of the apical and palatal categories; this pattern is consistent for both voiced and voiceless sibilants. Regarding CoG values for fricative and affricate counterparts, there are no significant intra-pair differences among any of the three pairs /ʃ, ʧ/, /s̪, ts̪/, and /s̪, ts̪/. This suggests that, in each of the three cases, the place of articulation is consistent for the plain fricative and the fricative portion of its corresponding affricate, unlike other varieties of Basque.[7]

D. Results: Rhotics

As mentioned in Sec. I, the opposition between the two Basque rhotics (a tap and a trill) is only realized intervocalically. Rhotics neither occur word-initially nor do they contrast in onset clusters or in codas. Thus, it is important to differentiate these three contexts for their analysis, as outlined in the current section.

The results for the number of lingual contacts produced in the tap /ɾ/ and in the trill /r/ are shown in Table III. For the intervocalic tap /ɾ/, the majority of items (82.4%) are produced with a single lingual contact, as expected for the phonetic realization of a tap. A small portion of the items (10.5%) are produced without any measured lingual contact; these realizations may represent occurrences of allophones such as [ɹ̪͡ɹ̪, ɹ̪], reported by Gaminde *et al.* (2016) for Bizkaian Basque. An even smaller portion of the items (7.1%) are produced with multiple contacts, suggesting that the tap /ɾ/ is sometimes (but infrequently) produced as a trill in the Mixean variety of Basque. For the intervocalic trill /r/, there is a nearly equal proportion of items realized with a single contact (43%) as of items realized with two contacts (40.9%), with slightly over half of the total number of /r/ items (56.5%) produced with multiple lingual contacts (i.e., two or more taps) and a single item (0.5% of the total data) with no measured lingual contacts. The high proportion of single-contact items is surprising, given the fact that trills are characterized by multiple articulator contacts; this result suggests a possible merger in progress of /r/ to /ɾ/ in Mixean Basque. Nevertheless, it could also be the case that rhotics form a stable non-robust opposition in Mixean, with a gradient difference and no clear boundary between categories. If so, the opposition would be conditioned by the mostly predictable distribution of the two sounds, which are only contrastive in a small subset of phonological contexts (see Currie Hall, 2009; Hualde, 2004; Renwick and Ladd, 2016; Renwick, 2014; Simonet, 2005).

TABLE III. Percentage of total number of taps realized in alveolar rhotics of Mixean Basque in onset clusters, intervocalic position, and syllable codas.

	Intervocalic		Cluster	Coda
Taps	/ɾ/	/r/	/ɾ/ ∼/r/	/ɾ/ ∼/r/
0	10.5	0.5	35.7	54.8
1	82.4	43.0	60.7	34.7
2	5.7	40.9	3.6	6.5
3	1.4	13.4	-	2.4
4	-	2.2	-	1.6

In comparison to the intervocalic environment, the occurrence of alveolar rhotic items produced without any lingual contact is much greater in the onset cluster and coda environments. 35.7% of /ɾ, r/ items are produced without any lingual contact when they occur in onset clusters, while the majority of alveolar rhotics (54.8%) are produced without any lingual contact when they occur in coda position. However, in both environments, some of the alveolar rhotic items are produced with at least one tap. In onset clusters, the majority of items (60.7%) are produced with a single tap, and a small proportion of items (3.6%) are produced with two taps; no items are produced with three or more lingual taps. In coda position, while the majority of items (54.8%) are produced without any lingual contact, a substantial proportion of alveolar rhotics are produced with one (34.7%) or multiple (10.5%) taps, including rare cases of three and four taps.

V. DISCUSSION

Based on the previous literature, we expected to find six vowel categories in Mixean Basque: the common Basque vowels /a, e, o, i, u/, and a high front rounded vowel /y/, as found in Zuberoan. Our acoustic analysis confirms that there are six vowels in Mixean. Nevertheless, our results suggest that the Mixean vowel system is characterized by two front vowels (/i/ and /e/), two central vowels (/y/ ∼[ʉ] and /a/), and two back vowels (/u/ and /o/). The retraction of Mixean /y/ has been auditorily described as closer to /ø/ (Haase, 1992, p. 29), although our results suggest that the sixth vowel found in Mixean is better represented as /ʉ/, a segment not found in any other variety of the Basque language. In addition, all high vowels in Mixean seem to be comparatively lower than their equivalents in other Basque dialects. The observation that Zuberoan high vowels are lower than those found, for instance, in French has been made by other authors as early as Larrasquet (1932), but our results for Mixean point to even lower high vowels than these described by Urrutia *et al.* (1991) for Zuberoan Basque. Thus, Mixean high vowels /i, y, u/ are actually articulated with mean formants close to the values expected of [e, ɵ, o].

Regarding Mixean stops, our study has found evidence that there is still a three-way laryngeal distinction in the language (pre-voiced, plain voiceless, and voiceless aspirated). Nevertheless, the mean VOT values that resulted from our study are less extreme than these reported by previous authors for the Zuberoan dialect. Although the finite Gaussian mixture modeling applied to our data suggested the presence of three stop series, the clusters that resulted from this classification did not consistently reflect the orthographic notation. These two observations suggest that the three-way distinction is not as strong in Mixean Basque as has been reported for Zuberoan Basque, and that distinctive aspiration is perhaps being lost in this variety, as has arguably occurred in other Basque varieties (Egurtzegi, 2018a).

Within sibilants, we did find a categorical difference between apical and laminal sibilants, but our results are not consistent with the most general description of the place of articulation of sibilant segments (namely, apico-alveolar, lamino-alveolar, and post-alveolar/alveolo-palatal). As the rest of the studies outlined in this paper, our study found lower CoG values for the apical sibilants than for the laminal sibilants, which is in line with the observation that apical sibilants may be articulated with a post-alveolar (or even retroflex) place of articulation, as suggested by a number of previous descriptions. However, no differences in CoG were observed between the apical and alveolo-palatal sibilants, suggesting a merger in place of articulation between the two sets of segments. This process could involve one of three scenarios: (1) a merger of apical to alveolo-palatal, as described in Hualde (2010); (2) a merger of alveolo-palatal to apical; or (3) a merger of both categories towards an intermediate place of articulation. In comparing our CoG measurements to the values reported in other studies, the CoG values for the apical sibilants are similar to those reported by Hualde (2010) for High Navarrese; however, the values for the alveolo-palatal sibilants are lower in High Navarrese than those observed in the current study. We believe that these results suggest that the alveolo-palatal sibilants have merged to apico-(post)alveolar in the Mixean variety, i.e., scenario 2 above. Finally, unlike previous reports, we did not find CoG differences between any given fricative and its affricate counterpart, suggesting that fricative and affricate counterparts are produced with the same place of articulation in Mixean Basque.

We expected to find an opposition between a tap and a trill in intervocalic position, as in other Basque dialects. The results from our study suggest that, in Mixean, this opposition is not as strong as in other Basque dialects. While the number of intervocalic productions of the phonological tap with more than one lingual contact was low (7.1%), the percentage of productions of the phonological intervocalic trill with a single tap was much higher than expected (43%). Although the majority of the intervocalic trills were actually trilled (56.5%), the unexpectedly high number of trills realized as a tap intervocalically points in the direction of an incipient merger of the two rhotics in Mixean Basque. Following the more general descriptions of Basque rhotics, we may have expected the rhotics in neutralizing contexts (i.e., onset clusters and codas) to be articulated as shorter trills (i.e., two lingual contacts). However, our study suggests that the most common realization of onset-cluster rhotics involves one tap (60.7% of all items, and 35.7% with no taps), while coda rhotics are most frequently produced with no taps (54.8% of all items, and 34.7% with one tap). These results suggest that neutralized rhotics are produced as taps or approximants in Mixean, and that coda position is the most frequent context for the realization of approximant/fricative rhotics, followed by onset clusters. Finally, regarding the place of articulation of Mixean rhotics, while some of the speakers produced a number of uvular articulations, none of them limited their rhotics to uvular segments. This observation is based on manual transcription and concomitant perception by the first author (a native speaker of Basque). This finding contrasts with observations by Gaminde *et al.* (2017), who consistently found uvular articulations in the speakers from eastern Basque dialects. However, it is worth mentioning that, while the speakers in their study were of young age, the participants in our study encompassed speakers from a much older generation, so that the spread of uvular rhotics in eastern Basque dialects can potentially be viewed as a recent innovation.

VI. CONCLUSION

This study has presented a general description of an endangered variety of Basque, namely Mixean Low Navarrese, via acoustic analyses of most segments in its phonological inventory. This paper has underlined the uneven nature of the acoustic studies on the Basque language: while studies on vocalic inventories and sibilants are fairly common, the rest of the segments of the language are understudied, and no general acoustic description of any variety (or the standard language) can be found in the literature. Important results of this study include the first proposal of a centralized rounded vowel in any Basque variety, a data-driven confirmation of the maintenance of the three-way stop distinction in Mixean, the description of a merger of the series of alveolo-palatal sibilants to the apico-alveolar series, evidence in support of an incipient merger of rhotics, and the realization of rhotics in neutralizing positions with one or even no lingual contacts. These results would have remained unknown had we focused on a more accessible dialect instead of an endangered variety.

ACKNOWLEDGMENTS

This research was supported by the Alexander von Humboldt Foundation, the Spanish Ministry of Economy and Competitiveness (FFI2016-76032-P; FFI2015-63981-C3-2), and ERC Advanced Grant No. 295573 *Human interaction and the evolution of spoken accent* (J. Harrington). We thank Iñaki Camino for letting us use his recordings, Rosa Franzke for helping us with the annotation process, Martxel Martínez for his help in editing the map, and José Ignacio Hualde for reading and commenting on an early version of this paper. All errors are ours.

[1]See supplementary material at https://doi.org/10.1121/10.0000996 for a dialectal map of the Basque varieties and for a provisional segmental inventory of the Mixean variety.

[2]Hualde (1997, p. 106) describes the accent in Low Navarrese as non-contrastive, with no clear stress in connected speech and paroxytonic in isolated words, a description that Camino (2016, pp. 190–191) confirms for Mixean.

[3]A nearly-exhaustive list of acoustic studies of Basque vowels would also include Salaburu (1984) with five speakers from Baztan (High Navarrese); Pagola (1992), with 18 speakers of (Northern) High Navarrese (from Baztan, Bortziri, and Ultzama; and a speaker of Lapurdian from Zugarramurdi); Hualde *et al.* (2010) with two speakers from Goizueta (High Navarrese); Etxebarria Ayesta (1991, p. 48) with one speaker from Zeberio (Bizkaian); Gaminde (1992, 1995), with speakers from Urduliz and Gatika (Bizkaian); Etxebarria Arostegi (1995) also from Bizkaian; and Etxeberria (1990, 1991) with speakers of Zaldibia (Gipuzkoan). Etxebarria Arostegui (1991) and Iribar Ibabe and Túrrez Aguirrezabal (2001) do not specify the varieties under study.

[4]A nearly-exhaustive list of acoustic studies involving Basque sibilants would include Yárnoz (2002b), Hualde (2010), Jurado Noriega (2011), Gaminde *et al.* (2013), Gandarias *et al.* (2014), Iglesias *et al.* (2016), Muxika-Loitzate (2017), Beristain (2018a,b), Txillardegi (1982), Isasi Martínez *et al.* (2009), Urrutia *et al.* (1988, 1989, 1991), and Txillardegi (1982). Classic articulatory descriptions of Basque sibilants include Navarro Tomás (1923, 1925) and Alonso (1923).

[5]In comparison with the raw Hz values, the formant normalization resulted in an average absolute difference of 4.5 Hz for $F1$ (SD = 2.2), 9.8 Hz for $F2$ (SD = 5.1), and 12.5 Hz for $F3$ (SD = 11.4).

[6]Given that the speech signals were band-pass filtered from 300 Hz to 19 kHz before spectral measurements were made, the lower CoG values for the voiced sibilants are not likely due directly to $F0$ energy. However, it is reasonable to speculate that energy associated with lower harmonics (energy that arises due to voicing) causes the overall spectral energy to shift towards lower frequencies.

[7]A study of the contrast between fricative and affricate sibilants is beyond the purposes of the present study. Such a study would require additional acoustic metrics such as total duration or rise time (Johnson, 2003, pp. 144–145).

Alonso, A. (**1923**). "Consonantes de timbre sibilante en el dialecto vasco-baztanés" ("Sibilant consonants in the Basque dialect of Baztan"), in *Tercer Congreso de Estudios Vascos* (*Third Conference of Basque Studies*), edited by Eusko Ikaskuntza (Diputación de Guipuzcoa, San Sebastián-Donostia), pp. 57–64.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Beristain, A. (**2018a**). "The acoustic realization of /s̪/ and /ts̪/ by L1 and L2 Basque speakers," in *Linguistic Variation in the Basque Language and Education—III*, edited by A. Iglesias (University of the Basque Country, Bilbao, Spain).

Beristain, A. (**2018b**). "Basque dialectal substrate in the realization of /s/ in L2 Spanish," M.A. thesis, University of Illinois at Urbana-Champaign, Champaign, IL.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer [computer program]," http://www.praat.org/ (Last viewed 01/19/2019).

Camino, I. (**2016**). *Amiküze Eskualdeko Heskuara* (*The Basque of the Region of Mixe*) *(Mendaur 11)* (Euskaltzaindia, Bilbao, Spain).

Currie Hall, K. (**2009**). "A probabilistic model of phonological relationships from contrast to allophony," Ph.D. thesis, The Ohio State University, Columbus, OH.

Egurtzegi, A. (**2013**). "Phonetics and phonology," in *Basque and Proto-Basque: Language-Internal and Typological Approaches to Linguistic Reconstruction*, edited by M. Martínez-Areta (Peter Lang, Frankfurt am Main, Germany), pp. 119–172.

Egurtzegi, A. (**2015**). "Different distributions of contrastive vowel nasalization in Basque," Loquens **2**(1), e017.

Egurtzegi, A. (**2018a**). "Herskarien ustezko ahoskabetasun asimilazioa eta euskal herskari zaharren gauzatzea" ("The alleged voicelessness assimilation and the realization of old Basque stops"), Int. J. Basque Ling. Philol. (ASJU) **52**(1–2), 189–206.

Egurtzegi, A. (**2018b**). "On the phonemic status of nasalized /h̃/ in Modern Zuberoan Basque," Linguistics **56**, 1353–1367.

Etxebarria Arostegui, M. (**1991**). "En torno al vocalismo vasco" ("On Basque vowel patterns"), in *Memoriae L. Michelena Magistri Sacrum. Anejos de ASJU*, pp. 635–655.

Etxebarria Arostegi, M. (**1995**). "Análisis acústico de las vocales del dialecto vizcaíno" ("An acoustic analysis of the vowels of the Bizkaian dialect"), Euskal Dialektologiako Kongresua (1991. Donostia) Anejos del ASJU **28**, 519–537.

Etxebarria Ayesta, J. M. (**1991**). *Euskal Fonetika eta Fonologia: Euskera Biziaren Azterbidean* (*Basque Phonetics and Phonology: Toward Studying Living Basque*) (Labayru Ikastegia, Bilbao, Spain).

Etxeberria, P. (**1990**). *Zaldibiako Bokalen Azterketa Akustikoa* (An acoustic study of the vowels of Zaldibia) (Mundaiz, Donostia, Spain).

Etxeberria, P. (**1991**). "Zaldibiako eta beste zenbait hizkuntzatako triangelu akustikoen alderaketa" ("Comparison of the acoustic triangles of Zaldibia and other languages"), in *Memoriae L. Mitxelena magistri sacrum. Anejos del ASJU*, Vol. 14, pp. 657–675.

Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B., and Fop, M. (**2019**). mclust: Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation, [computer program], https://CRAN.R-project.org/package=mclust (Last viewed 4/15/2019).

Gaminde, I. (**1992**). "Urduliz eta Gatikako herri hizkuntza azterketa linguistikoa" ("Linguistic study of the local language of Urduliz and Gatika"), Ph.D. thesis, Deustuko Unibertsitatea / Universidad de Deusto, Bilbao, Spain.

Gaminde, I. (**1995**). "Gatika eta Urdulizko bokalen azterketa akustikoa" ("Acoustic study of the vowels of Gatika and Urduliz"), **11**, 125–149.

Gaminde, I. (**2006**). "Dardarkarien ezaugarri akustikoez" ("On the acoustic features of rhotics"), Euskalingua **8**, 75–82.

Gaminde, I., Etxebarria, A., Eguskiza, N., Romero, A., and Iglesias, A. (**2017**). "Dardarkari anizkunaren aldakortasunaz" ("On the variability of the trill"), Fontes Linguae Vasconum **123**, 29–64.

Gaminde, I., Hualde, J. I., and Salaberria, J. (**2002**). "Zubereraren herskariak: Azterketa akustikoa" ("The stops of Zuberoan: An acoustic study"), Lapurdum **7**, 221–236.

Gaminde, I., Romero, A., Eguskiza, N., and Etxebarria, A. (**2016**). "Dardarkari bakunaren alofonoez testu irakurrietan" ("On the tap allophones of read speech"), Fontes Linguae Vasconum **121**, 121–140.

Gaminde, I., Unamuno, L., and Iglesias, A. (**2013**). "Bizkaiko neska gazteen kontsonante afrikatuen izari akustikoez" ("On the acoustic cues of affricate consonants of Bizkaian young girls"), Euskalingua **23**, 6–13.

Gandarias, L., Plaza, J., and Gaminde, I. (**2014**). "Lekeitioko txistukariez: Frikariak eta afrikatuak" ("Sibilants of Lekeitio: Fricatives and affricates"), Euskalingua **24**, 6–21.

Haase, M. (**1992**). *Sprachkontakt und Sprachwandel im Baskenland: Die Einflüsse des Gaskognischen und Französischen auf das Baskische* (*Linguistic Contact and Linguistic Change in the Basque Country: The Influence of Gascon and French on Basque*) (Helmut Buske, Hamburg, Germany).

Hothorn, T., Bretz, F., and Westfall, P. (**2008**). "Simultaneous inference in general parametric models," Biometr. J. **50**(3), 346–363.

Hualde, J. I. (**1997**). *Euskararen Azentuerak* (*Basque Stress Patterns*) (Euskal Herriko Unibertsitatea / Universidad del País Vasco, Bilbao—Donostia).

Hualde, J. I. (**2003**). "Segmental phonology," in *A Grammar of Basque*, edited by J. I. Hualde and J. Ortiz de Urbina (Mouton de Gruyter, Berlin, Germany), pp. 16–65.

Hualde, J. I. (**2004**). "Quasi-phonemic contrasts in Spanish," in *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*, April 23–25, Davis, CA, pp. 374–398.

Hualde, J. I. (**2010**). "Neutralización de sibilantes vascas y seseo en castellano" ("Neutralization of Basque sibilants and seseo in Spanish"), Oihenart **25**, 89–116.

Hualde, J. I. (**2018**). "Aspiration in Basque," Papers Hist. Phonol. **3**, 1–27.

Hualde, J. I., Beristain, A., Icardo Isasa, A., and Zhang, J. (**2019**). "Lenition of word-final plosives in Basque," in *Proceedings of International Congress of Phonetic Sciences 2019*, August 5–9, Melbourne, Australia.

Hualde, J. I., Lujanbio, O., and Zubiri, J. J. (**2010**). "Goizueta Basque," J. Int. Phon. Assoc. **40**, 113–127.

Hualde, J. I., and Ortiz de Urbina, J. (**2003**). *A Grammar of Basque* (Mouton de Gruyter, Berlin, Germany).

Iglesias, A., Gaminde, I., Gandarias, L., and Unamuno, L. (**2016**). "Euskararen txisturkariak aztertzeko indize akustikoez" ("On the acoustic cues to study Basque sibilants"), Euskalingua **28**, 6–18.

Iribar Ibabe, A., and Túrrez Aguirrezabal, I. (**2001**). "Algunos factores de variación vocálica" ("Some factors of vocalic variation"), Letras de Deusto **92**, 229–241.

Isasi Martínez, C., Iribar Ibabe, A., and Moral del Hoyo, C. (**2009**). "Una transferencia vasca: el seseo de hablantes vizcaínos y guipuzcoanos" ("A Basque transfer: The seseo of Bizkaian and Gipuzkoan speakers"), Oihenart **24**, 201–235.

Johnson, K. (**2003**). *Acoustic and Auditory Phonetics* (Blackwell, Oxford, UK).

Jurado Noriega, M. (**2011**). "Caracterización de sibilantes fricativas vascas y su percepción en el sistema fonético español" ("Characterization of Basque fricative sibilants and their perception in the Spanish phonetic system"), Int. J. Basque Ling. Philol. (ASJU) **45**(1), 81–137.

Keshet, J., Sonderegger, M., and Knowles, T. (**2014**). "AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [computer program]," https://github.com/mlml/autovot/ (Last viewed: 2/12/2019).

Kisler, T., Reichel, U. D., and Schiel, F. (**2017**). "Multilingual processing of speech via web services," Comput. Speech Lang. **45**, 326–347.

Larrasquet, J. (**1932**). "Phonétique du basque de Larrajá (Quartier de barcus)" ("The phonetics of the Basque of Larrajá"), Revista Internacional de Estudios Vascos **23**, 153–191.

Larrasquet, J. (**1934**). *Le Basque Souletin Nord-Oriental* (Floch, Paris).

Lindau, M. (**1985**). "The story of /r/," in *Phonetic Linguistics: Essays in Honor of P. Ladefoged*, edited by V. Fromkin (Academic Press, New York), pp. 157–167.

L'Institut National de la Statistique et des Études Économiques (**2019**). "Populations légales 2016" ("Legal populations 2016"), https://www.insee.fr/fr/statistiques/3681328 (Last viewed April 29, 2019).

Michelena, L. (**2011**). "Fonética histórica vasca" ("Basque historical phonetics"), in *Obras Completas 6*, edited by J. A. Lakarra and I. Ruiz Arzalluz (Diputación Foral de Guipúzcoa & University of the Basque Country, Donostia-San Sebastián and Vitoria-Gasteiz).

Mounole, C. (**2004**). "Zubererazko herskarien azterketa akustikoa" ("An acoustic study of Zuberoan stops"), Int. J. Basque Ling. Philol. (ASJU) **38**(1), 207–248.

Muxika-Loitzate, O. (**2017**). "Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano," Languages **2**(4), 1–19.

Nadeu, M., and Hualde, J. I. (**2015**). "Biomechanically conditioned variation at the origin of diachronic intervocalic voicing," Lang. Speech **58**, 351–370.

Navarro Tomás, T. (**1923**). "Observaciones fonéticas sobre el vascuence de Guernica" ("Phonetic remarks on the Basque of Gernika"), in *Tercer Congreso de Estudios Vascos* (Third Conference of Basque Studies), edited by E. Ikaskuntza (Diputación de Guipuzcoa, San Sebastián-Donostia) pp. 49–56.

Navarro Tomás, T. (**1925**). "Pronunciación guipuzcoana" ("Gipuzkoan pronunciation"), in *Homenaje a Menéndez Pidal*, Vol. III, pp. 593–653.

N'Diaye, G. (**1970**). *Structure du Dialecte Basque de Maya* (*Structure of the Basque Dialect of Maia*) (Mouton, The Hague).

Pagola, R. M. (**1992**). *Euskal Fonetika Nafarroan* (*Basque Phonetics in Navarre*) (Nafarroako Gobernua / Gobierno de Navarra, Iruñea, Spain).

Python Software Foundation. (**2019**). "Python Language Reference, version 2.7 [computer program]," http://www.python.org (Last viewed 2/12/2019).

R Core Team (**2019**). "R: A language and environment for statistical computing [computer program]" http://www.R-project.org (Last viewed 3/22/2019).

Renwick, M. E. L. (**2014**). *The Phonetics and Phonology of Contrast: The Case of the Romanian Vowel System* (de Gruyter Mouton, Berlin, Germany).

Renwick, M. E., and Ladd, D. R. (**2016**). "Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts," Lab. Phonol. **7**(1), 1–29.

Saadah, E. (**2011**). "Towards quantifying lenition in Ondarroan Basque," Stud. Ling. Sci. Illinois Working Papers **2011**, 89–107.

Salaburu, P. (**1984**). *Hizkuntzaren Soinu Egitura. Hizkuntz Teoria eta Baztango Euskalkia. Fonetika eta Fonologia* (*The Sound Pattern of Language. Linguistic Theory and the Dialect of Baztan. Phonetics and Phonology*) (Euskal Herriko Unibertsitatea / Universidad del País Vasco, Bilbao, Spain).

Simonet, M. (**2005**). "Prosody and syllabification intuitions of [CiV] sequences in Catalan and Spanish," in *Prosodies*, edited by S. Frota, M. Freitas, and M. Vigário (Mouton de Gruyter, Berlin, Germany), pp. 247–267.

Txillardegi (**1982**). "Some acoustic data about the three Basque sibilants," in *Proceedings of the First International Basque Conference in North America*, Fresno, CA, pp. 18–34.

Urrutia, H., Etxebarria, M., Túrrez, I., and Duque, J. C. (**1988**). *Fonética Vasca 1: Las Sibilantes en el Vizcaíno* (*Basque Phonetics I: The Sibilants in Bizkaian*) (Deustuko Unibertsitatea/ Universidad de Deusto, Bilbao, Spain).

Urrutia, H., Etxebarria, M., Túrrez, I., and Duque, J. C. (**1989**). *Fonética Vasca 2: las Sibilantes en el Guipuzcoano* (*Basque Phonetics II: The Sibilants in Gipuzkoan*) (Deustuko Unibertsitatea / Universidad de Deusto, Bilbao, Spain).

Urrutia, H., Etxebarria, M., Túrrez, I., and Duque, J. C. (**1991**). *Fonética Vasca 3: Las Sibilantes en los Dialectos Orientales* (*Basque Phonetics III: The Sibilants in the Eastern Dialects*) (Deustuko Unibertsitatea / Universidad de Deusto, Bilbao, Spain).

Urrutia, H., Etxebarria, M., Túrrez, I., and Duque, J. C. (**1995a**). *Fonética Vasca 4: Las Vocales en Euskera* (*Basque Phonetics: The Vowels in Basque*) (Deustuko Unibertsitatea/ Universidad de Deusto, Bilbao, Spain).

Urrutia, H., Etxebarria, M., Túrrez, I., and Iribar, A. (**1995b**). "El dialecto bajo-navarro oriental: Vocalismo" ("The Eastern Low Navarrese dialect: Vowel patterns"), in *Phonetica. Trabajos de Fonética Experimental* (*Phonetica. Works on Experimental Phonetics*), edited by A. Elejabeitia and A. Iribar (Deustuko Unibertsitatea / Universidad de Deusto, Bilbao, Spain), pp. 61–99.

Yárnoz, B. (**2002a**). "Descripción de las sibilantes vascas mediante el parámetro Tongue shape" ("Description of the Basque sibilants using the *tongue shape* parameter"), Euskalingua **1**, 25–31.

Yárnoz, B. (**2002b**). *Sibilants in the Basque Dialect of Bortziri: An Acoustic and Perceptual Study* (Nafarroako Gobernua / Gobierno de Navarra, Iruñea, Spain).

Yrizar, P. D. (**1981**). *Contribución a la Dialectología de la Lengua Vasca* (*Contribution to the Dialectology of the Basque Language*) (Caja de Ahorros Provincial de Guipúzcoa, Zarauz, Spain).

Zabalik (**2016**). "Euskara gure eskoletan: dena eta ez dena" ("Basque in our schools: Fact and fiction") http://zabalik-amikuze.eus/euskara-gure-eskoletan-dena-eta-ez-dena (Last viewed September 9, 2019).

Zuazo, K. (**2008**). *Euskalkiak. Euskararen Dialektoak* (*Euskalkiak. The Dialects of Basque*) (Elkar, Donostia-San Sebastián, Spain).

Phonological Data & Analysis Volume 4, Article 5: 1–31 (2022)
https://doi.org/10.3765/pda.v4art5.67

Received 31 July 2021; Revised 01 July 2022; accepted 01 August 2022. © 2022 Dorota Krajewska et al.
Published by the Linguistic Society of America with permission of the authors under a CC BY 3.0 license.

Sibilant mergers in 18th-century Basque: A quantitative study

Dorota Krajewska[a], Eneko Zuloaga[b] & Ander Egurtzegi[c*]

[a]University of the Basque Country UPV/EHU – dorota.krajewska@ehu.eus
[b]University of the Basque Country UPV/EHU – eneko.zuloaga@ehu.eus
[c]Centre National de la Recherche Scientifique – IKER (UMR5478) – ander.egurtzegi@iker.cnrs.fr

Conservative Basque dialects distinguish apical (/s̪/ and /t͡s̪/) and laminal (/s̻/ and /t͡s̻/) alveolar sibilants in the fricative and affricate series. This paper analyses the changes this system was undergoing in the Central Basque variety of San Sebastián in the 18th century: (1) the "Western merger": neutralisation of the laminal and apical fricative sibilants in favour of the latter and the neutralisation of the laminal and apical alveolar affricates in favour of the former, which started in Western Basque and spread to some Central varieties, and (2) the "Central merger", a more recent development, limited to some central dialects, where both fricative and affricate alveolar sibilants are realised as laminals. A generalised linear mixed-effects model was fitted to the data extracted from an early-18th-century manuscript which shows evidence of both mergers. We propose that sibilant mergers were still in progress in the variety and time period under study and that they are interrelated processes. The Western merger started as a phonetically-conditioned sound change due to coarticulation to a following consonant. As this neutralisation extended to other positions, a hypercorrective change was initiated in some Central varieties, which eventually resulted in a mirror-image process, namely a change from apical to laminal fricatives.

Keywords: sibilants; Basque; merger; corpus study

1 Introduction

In Basque historical phonology, the study of consonants has prevailed over that of vowels (Michelena 1990; Lakarra 1995; Lakarra 2013; Trask 1997; Egurtzegi 2013). Among processes affecting consonants, the changes that have altered the phonological opposition between pairs of fricative and affricate sibilants with different places of articulation have aroused considerable interest, with renewed discussion in recent times (Muxika-Loitzate 2017; Egurtzegi & Carignan 2020; Beristain 2021). The interest of the Basque sibilant system lies in its relative complexity: it includes voiceless fricative-affricate pairs with three contrasting places of articulation – lamino-alveolar (/s̻/ and /t͡s̻/), apico-alveolar (/s̪/ and /t͡s̪/) and postalveolar (/ʃ/ and /t͡ʃ/) – and voiced counterparts of these in some eastern varieties (cf. Michelena 1990; Hualde 2003; Egurtzegi & Carignan 2020).

The best-understood development in the evolution of the phonological contrasts involving sibilants in Basque is what we call here the "Western merger" (Figure 1). This merger dates back to the 17th century

* The research reported in this paper was funded by the following projects: "Monumenta Linguae Vasconum VI" (PID2020-118445GB-I00), "Modern approaches to diachronic phonology applied to Basque (MADPAB)" (ANR-20-CE27-0007) and "Diachronic Linguistics, Typology and the History of Basque" (IT1534-22). Data and code used in this paper as well as supplementary materials are available at https://osf.io/t6a7w/.

and has been proposed as the main consonantal development differentiating the Western dialect of Basque from the rest (Zuloaga 2020: 273). It involves the neutralisation of the laminal and apical alveolar fricative sibilants in favour of the latter and the neutralisation of the laminal and apical alveolar affricates in favour of the former.

Though geographically less widespread and more recent, other sibilant neutralisation processes also occurred in different Basque varieties (see Hualde 2010 for an overview). In addition to the Western merger, we also analyse the "Central merger" (Figure 2), an innovation which developed during the 17th-19th centuries in some central varieties. In this merger, fricative and affricate alveolar sibilants are realised as fricative and affricate laminals (see Beristain 2018; Beristain 2019; Beristain 2021 for analyses of modern data).

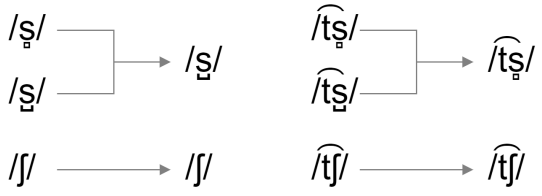


Figure 1: The Western merger

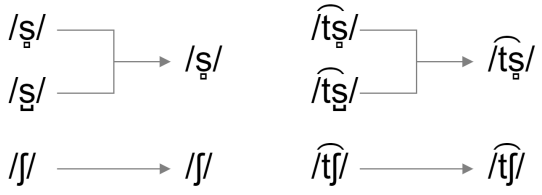


Figure 2: The Central merger

This paper presents two main novelties. First, it focuses on the intermediate phase of the Central merger, which has barely been analysed. Second, it involves a quantitative approach. Research on the history of sibilant neutralisation in Basque has generally been qualitative, mainly because the Basque pre-19th-century corpus does not always allow quantitative analyses due its limited size. In this paper, we take advantage of the recent discovery of a relatively extensive manuscript: "Lubieta's dictionary", dated 1728. It represents the variety of the town of San Sebastián (located in the province of Guipuscoa, in the centre of the Basque Country). The text shows evidence of two patterns of merger, namely the Western and Central mergers.

We intend to specify the degree of development of the Western and Central mergers in the variety of San Sebastián in the middle of the 18th century, whether a potential connection between the two mergers may have existed, and the details of the process that led to the generalisation of the Central merger in that area. We also explore a number of general issues, such as how to establish whether a phonemic opposition was lost in quantitative diachronic data, how to study a merger in progress, and how quantitative data analysis can inform qualitative diachronic phonological studies. Most importantly, this study is particularly interesting for general theories of sound change because 18th century San Sebastián Basque provides an excellent example of two competing sound changes at an early stage of development.

The paper is structured as follows: In Section 2, we present the background of the present research: (a) the definition of a merger and problems related to the study of mergers in diachrony, (b) an overview of the Basque sibilants, and (c) a description of our source, Lubieta's dictionary. We conclude Section 2 by presenting the objectives of our study. Section 3 presents our methodology with an introduction to the statistical model we constructed. This is followed by Section 4, with descriptive statistics of the data, and Section 5, detailing the model's results, diagnostics and effects. We conclude the paper with a general discussion of our results in Section 6.

2 Background and objectives

2.1 The concept and study of mergers

A merger is usually understood as the identification (and gradual equation) of all instances of one (or more) phoneme(s) A with those of a different phoneme B, with a concomitant loss of the phonological features distinguishing them (and thus their phonemic contrast), such that A, B > A. When the loss of a phonemic contrast is underway or when it is limited to a particular phonological context, we speak of neutralisation instead.

Three types of mergers are distinguished in the literature. In a "merger by approximation", the phonetic targets of two phonemes gradually converge and the outcome might be an intermediate sound (Labov 1994: 321). A "merger by transfer", as explained by Labov (1994: 321), is a unidirectional process in which one phoneme is replaced with another in a word-by-word manner. Finally, in a "merger by extension", first described by Herold (1990), "the lexical constraints on the distribution of the two former phonemes are removed, and the range that was previously divided between the two phonemes is used for the new phoneme, with allophonic distributions in appropriate areas of the new range" (Labov 1994: 323–324).

Several procedures can be used to study an ongoing merger, the simplest being producing and perceiving contrasts in minimal pairs (Gordon 2015). Acoustic analysis might also be used. The situation is different for diachronic data (Minkova 2015), where the analysis depends mainly on spelling and its interpretation (though "indirect" evidence, such as poetry, can be used (cf. Minkova 2015)). The analysis of the gradual neutralisation requires a detailed study of the spelling in testimonies, since pronunciation is not the only criterion for the establishment of orthography (Martínez Alcaide 2010: 14).[1]

The study of the neutralisation of Basque sibilants poses two general problems. First, the brevity of most texts does not allow for extensive quantitative analysis. Traditionally (see Michelena 1990; Hualde 2010; Zuloaga 2020), a sufficiently consistent graphic distinction, even if there is some deviation, has been taken as a reflection of a phonological contrast. Thus, graphic confusion is considered a reflection of the loss of contrast, especially when it occurs repeatedly in the same positions in different texts of the same period and area. One of the objectives of the present study is to apply a more quantitatively oriented methodology to study merger phenomena in longer texts.

The second problem is the lack of a univocal orthographic rule for pre-20th-century Basque, which makes it difficult to distinguish between the sibilants (both between fricatives and affricates and between laminals and apicals). Basque writers had to resort to the spelling systems of other languages, and in the Peninsular Basque Country, Spanish orthography was used (Echenique 2013: 71). As a result, authors had to work with defective spelling, since they had to express sounds that did not exist in Spanish by means of the Spanish spelling.

The situation in Basque bears certain similarity with that of Middle English before the standardisation of English orthography (Stenroos 2002): during that period, there was not a single model for spelling, and thus it is not easy to distinguish between phenomena which reflect different aspects of phonology from those related to clashes between various models (see also Oyosa 2015 for a discussion of the situation in Romance languages). Thus, even though our study is concerned with Basque, the methodology we apply might be useful to study other languages as well.

2.2 Sibilants in Basque

The "moderately complex" (Igartua & Zabaltza 2012: 11) sibilant system reconstructed for Proto-Basque (Michelena 1990; Hualde 2010) remained stable in all Basque dialects until the Archaic Basque period, i.e.

[1] A distinction should be made between spelling and orthography: "Spelling involves the graphic realisations of all spoken items, whereas orthography is limited to a more or less binding norm that can lead to criticism in case of non-compliance" (Rutkowska & Rössler 2012: 214).

ca. 1400-1600 (Mounole & Lakarra 2018: 418).[2] The system is still preserved in some central-eastern dialects. It includes three fricative consonants (lamino-alveolar /s̻/, apico-alveolar /s̺/ and postalveolar /ʃ/, nowadays expressed by the graphemes <z>, <s> and <x>), and three affricates with analogous place of articulation (lamino-alveolar /t͡s̻/, apico-alveolar /t͡s̺/ and postalveolar /t͡ʃ/, nowadays expressed by the digraphs <tz>, <ts> and <tx>).[3]

Regarding their phonotactic distribution, alveolar fricative sibilants occur word-initially (*zu* 'you (sg.)', *sagu* 'mouse'), medially – both intervocalically (*hezi* 'to educate', *hesi* 'fence') and preconsonantally (*gazte* 'young', *asko* 'a lot') –, as well as in post-vocalic word-final position (*ez* 'no', *ados* 'OK'). Alveolar affricates are absent from word-initial position –where only the postalveolar /t͡ʃ/ is possible (*txakur* 'dog') – and only occur in medial position – intervocalically (*atzo* 'yesterday', *atso* 'elderly woman') or after a sonorant (*mintzatu* 'to talk', *hartza* 'the bear') – or final position (*hotz* 'cold', *hots* 'sound'), either after a vowel or after a sonorant. Affricates are found in word-final coda clusters (*beltz* 'black', *hortz* 'tooth'), while fricatives occur in non-final coda clusters (*belztu* 'to become black', *horzkari* 'dental'). Almost all such tautosyllabic coda clusters show the laminal sibilant. No sibilant can appear in tautosyllabic onset clusters.

The most important change in this system is the Western merger, initiated in the Western Basque (see Figure 3). It has led to the loss of the distinction between the laminal and apical alveolar fricatives in favour of the latter (Michelena 1990; Hualde 2010; Zuloaga 2020). Conversely, in the series of affricates, neutralisation occurred in favour of the laminal (see Figure 1).[4]

Zuloaga (2020) shows that initially (between Archaic and Old Basque) the neutralisation of laminal to apical fricatives was sporadic and limited to the preconsonantal (especially before stops, e.g. *guzti* 'all' or *zazpi* 'seven') and word-final position (e.g. with the instrumental suffix *-z*). It was not systematic in any phonological context, nor did it extend to any prevocalic position (i.e., intervocalic or word-initial). Examples are found in most varieties attested at that time. Zuloaga does not specify whether the process could be explained as assimilatory or dissimilatory. Here, we will argue that the preconsonantal position was crucial: the merger could have started due to coarticulation to the following consonant, especially before /t/. During the Old Basque period, the neutralisation became more systematic in preconsonantal and word-final positions in the Central and Western varieties. Moreover, neutralisation started to extend to new contexts, i.e. to word-initial and intervocalic position. As regards affricates, the first examples of apical to laminal neutralisation are attested in the early Old Basque period. From the 18th century, the Western neutralisation in fricatives and affricates extended to all possible phonological contexts in the Western area (Biscay, Alava and western Guipuscoa). Moreover, throughout the 18th and 19th centuries, it spread to the whole peninsular Basque coast and was even attested in some coastal villages in Labourd (Zuloaga 2020).

[2] The history of Basque is commonly divided into the following periods: Old Proto-Basque, Modern Proto-Basque (last cent. BC), Aquitanian (first cent. BC), Old Common Basque (6–9th cent.), Medieval Basque (10–15th cent.), Archaic Basque (1400–1600), Old Basque (1600–1745), Early Modern Basque (1745–1890), Late Modern Basque (1891–1968) and Unified Basque (1968–) (Lakarra, Manterola & Segurola 2019).

[3] For the sake of simplicity, we will use *apical* and *laminal* to refer to these sounds, focusing on the shape of the tongue, but note that the main constriction occurs in the alveolar region for these segments.

[4] Postalveolar sibilants have a special status in Basque, often patterning with palatal sounds, which have affective connotations in the language. These will be left out of our analysis, given that they remained unchanged in most varieties and show no obvious variation in our data.

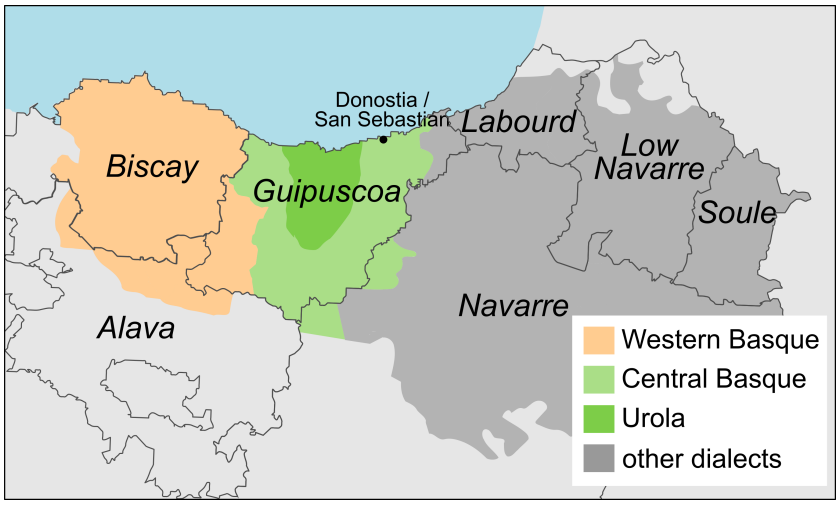


Figure 3: Varieties of Basque

The second change relevant for this paper emerged in Central Basque, and it involved the neutralisation of alveolar fricative and affricate sibilants in favour of the laminal (Figure 2). It is first attested between the Archaic and Old Basque periods, especially in intervocalic and word-initial position (though there are a few examples in preconsonantal position as well). In the 18th and 19th centuries, it is always documented alongside the Western neutralisation, i.e. in the fricative series, and both patterns co-occur in the same text. From the 19th century onwards, the Central merger became the prevailing pattern in the Guipuscoan coastal areas and in the Urola valley (see, e.g. Camino 2000; Sagarzazu 2005; Beristain 2018; Beristain 2019).

Finally, it is worth mentioning that in the affricate series no variety shows a laminal to apical merger. Though unusual, some examples of an apical realisation of the laminal affricate, best understood as hypercorrection, are found in texts and in eastern coastal villages of Guipuscoa nowadays, such as Hondarribia (see Sagarzazu 2005: 67).

This paper aims to reassess and build on the descriptions and proposals by Zuloaga (2020) from a quantitative and qualitative point of view. In particular, we would like to evaluate Zuloaga's hypothesis that the Central merger developed through generalised hypercorrection.

The concept of "hypercorrection" has been widely used in sociolinguistics since Labov (1972), where it is usually defined as a "process and result of an exaggerated attempt on the part of a speaker to adopt or imitate linguistic forms or a linguistic variety that he/she considers to be particularly prestigious" (Bussmann, Kazzazi & Trauth 1998: 525). In this paper, however, we follow a more general definition: "A pronunciation or grammatical construction often arising from a mistaken analogy with standard usage and produced out of the wish to be 'correct', but usually resulting in incorrect or stilted forms" (Brown & Miller 2013).

2.3 Lubieta's dictionary

Diccionario en Castellano y Basquenze que Sirve para la Enseñanza de la Vascongada, commonly known as "Lubieta's dictionary", is a bilingual Spanish-Basque manuscript written by Joseph Domingo Lubieta, dated from 1728 and preserved in the private library of the Sociedad Bilbaina. The existence of the manuscript went practically unnoticed until 2006 (see Bilbao 2012 for more details).

At least two features make Lubieta's dictionary particularly interesting. Firstly, it is one of the most extensive texts in the Central dialect preceding the 19th century. It is also one of the few texts documenting the Basque of San Sebastián (*Donostia* in Basque), an important coastal town that became the capital of Guipuscoa in the early 19th century. Lubieta has 6316 instances of sibilants, which allows for a quantitative analysis.

Secondly, the textual genre is worth mentioning. The text was prepared for Juan Francisco de Lullier, a merchant from San Sebastián, so that he could learn Basque. The text includes, as was customary in the manuals of the time, general notions of grammar, lists of words and a series of practical examples, such as dialogues, verses, sayings and parts of the Christian doctrine. The data's reliability is to some extent limited due to the genre and to the fact that it is based on materials written in other languages. At the same time, however, the text offers an interesting advantage: it contains a much wider range of vocabulary than religious texts, which prevail in the Basque historical corpus.

The preliminary analysis in Zuloaga (2020), based on a fragment of the text, has shown that Lubieta's dictionary exhibits a pervasive graphic confusion in the representation of alveolar sibilants: <s>, expected to represent apical sibilants, is found in words with an etymological laminal, and <z> and <c>, used to represent laminals at that time, are used with etymologically apical sibilants. There are also instances of confusion in the affricate alveolar sibilants: what is expected to be an apical affricate in other varieties is often represented by means of <z>, and there are a few instances etymologically laminal sibilants spelled with <s>.[5] Thus, the text shows both patterns of sibilant merger presented in Section 2.2.

2.4 Research questions

The main objective of this paper is to answer the following questions:

Q1. Was the contrast between apical and laminal alveolar sibilants already lost in 18th-century Basque of San Sebastián, either in the fricative or affricate class (or in both)? Or was the neutralising change still in progress?

Q2. What was the role of the phonological context on one hand, and lexical and frequency effects on the other hand, in the observed changes? Which had a bigger effect?

Q3. How can we account for the observed distribution of innovative spellings?

3 Methodology

3.1 Generalised linear mixed-effects models

Logistic regression models quantify the effects of predictors on a binary dependent variable. Generalised linear mixed-effects models, apart from the usual predictors ("fixed effects"), also include "random effects", which allow models to better deal with unbalanced and hierarchically organised data by taking into account the variation related to author, speaker or specific item. Thus, these models are particularly adequate for the analysis of various topics in linguistics (Coupé 2018; Winter 2020), e.g. in corpus studies (Gries & Hilpert 2010; Gries 2015a; Gries 2015b). They have proven fruitful in diachronic corpus studies, especially in the morphosyntax (see, for example, Gries & Hilpert 2010; Wolk et al. 2013; De Smet & Van de Velde 2020), and have also been used to study changes in spelling in German (Barteld, Hartmann & Szczepaniak 2016; Dücker, Hartmann & Szczepaniak 2020).

We use the top-down model selection strategy to construct our model (Zuur et al. 2009: 121–122; Gries 2013: 259–261). First we fit a "beyond optimal" model, with all independent variables and as many interactions as possible, inasmuch as they are theoretically justified. The next step is to evaluate whether the model should be simplified. The decision on whether to discard a predictor, an interaction, or a random effect will be based on significance testing (in that an element should be discarded if the deletion does not make the model significantly worse). The effect of this procedure is the minimal adequate model.

We use R (R Core Team 2021) for our statistical analyses as well as the libraries *lme4* (Bates et al. 2015) and *ggeffects* (Lüdecke 2018). For plotting, we use *ggplot2* (Wickham 2016), *sjPlot* (Lüdecke 2020) and *ggstatsplot* (Patil 2018).

[5] Note that, as usual at that time, fricative and affricate sibilants are not graphically distinguished in the text.

3.2 Variables

The dependent variable is SPELLING with two levels, *ConSpell* (conservative spelling) or *InnSpell* (innovative spelling). We use the label ConSpell when an etymologically laminal sibilant is spelled with <z> or <c> or when an apical sibilant is spelled with <s>. Conversely, InnSpell is used when a laminal sibilant is written with <s> or an apical one with <z> or <c>.[6] As far as the interpretation of spelling is concerned, we infer the place of articulation of a given segment by directly translating the spelling conventions to their phonological equivalent, under the assumption that variation would reflect the merger in progress.

Independent variables are the following:

(1) Etymological place of articulation (PLACE_ETYM): *apical* vs *laminal*.
(2) Manner of articulation (MANNER): *fricative* vs *affricate*.
(3) CONTEXT: prevocalic word-medial position (_*V*),[7] word-initial position (#_), word-final position (_#) and preceding a consonant (_*C*).
(4) LOAN: *yes* / *old* (loanwords showing historical sound change) / *no*.
(5) CATEGORY: *grammatical* (affixes + function words) vs *content*.
(6) FREQUENCY_CLASS of the lexeme: *high* / *mid* / *low*.

Thus, we take into account the sibilant's features (variables 1 and 2), the phonological context in which it appears (variable 3) and a number of factors related to lexical diffusion and frequency effects (variables 4-6).

The variable MANNER is mainly used to evaluate potential differences in the direction of the change in fricative and affricate series, given that, unlike in the case of fricatives, both mergers result in a laminal affricate. The value was established on the basis of the written tradition and modern data, as fricatives and affricates are not graphically distinguished in the text. Most Basque sibilants are unambiguously fricative or affricate. Nonetheless, Basque shows a systematic process of post-sonorant affrication (Michelena 1990: sec. 14.6), but the fact that sibilant manner of articulation was rarely reflected in spelling until later times makes dating the affrication process challenging. Given that recent research suggests that post-sonorant affrication developed during the preceding Old Basque period (Mounole & Lakarra 2018: 498; Mujika 2002: 221; Zuloaga 2020: 471–472), we decided to treat sibilants following a sonorant as affricates.

Lexemes' frequency values have been extrapolated from modern data and only approximate the language at the time of our text. Lubieta's dictionary, because of its nature, cannot be used to establish frequencies, and there are no other adequate texts for the time period and language variety under study. Thus, we used the Goenkale Corpus (Euskara Institutua 2011), a corpus of 13.3 million words which contains scripts of the Basque soap opera "Goenkale", broadcasted in the Basque television from 1994 to 2015. Currently, this is the only readily available source of frequency data from (nearly-)conversational language. Our working assumption is that the frequency of a word as used in the oral language (and not in a literary register) could condition a sound change in progress.

The number of occurrences per 10,000 words for each lexeme was extracted from the corpus. The data was available for 499 lexemes out of 873 (57%). Most of the lexemes for which frequency data was not available are unadapted loanwords (339/373), which are not common in the modern language. To avoid undefined values, the lexemes not found in the reference corpus were given the lowest value found for other lexemes, which is 0.01. The values were then log transformed (Gries 2010; Baayen 2001: 32), scaled and centred (using the *scale* function in R). We then binned the data into three classes: high, mid and low as

[6] The etymological place of articulation was determined by taking into account data from older texts from the same variety or from dialects that do not show any evidence of merger. Innovations are only considered with regard to place of articulation. Potential innovative spellings related to manner of articulation are not expected.

[7] Fricatives are intervocalic and affricates can follow a vowel or a sonorant consonant (see 2.2 for an overview of phonotactics).

shown in Figure 4.[8] The cuts are -1.3 and 0.2 (which correspond to 0.06 and 27 per 10.000 words), and they are meant to respect the form of the distribution and allow apical sounds to have items in all three frequency levels. Although data-driven, this decision is necessarily arbitrary given the lack of comparable studies on Basque.[9]

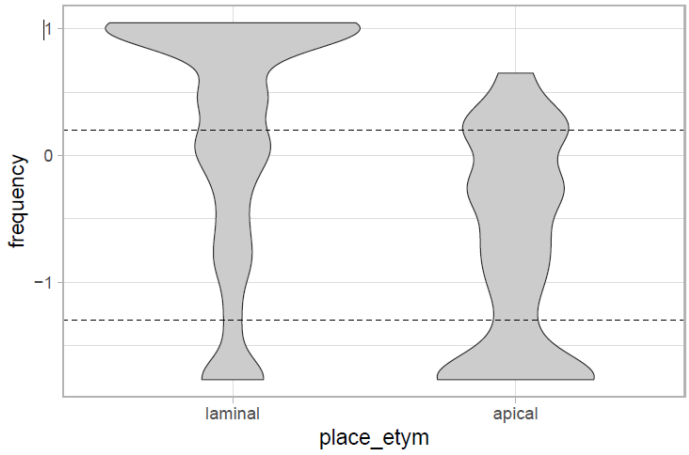


Figure 4: Violin plot for logged frequency and etymological place of articulation. Dashed lines represent the thresholds used for binning the data

The choice of the variables PLACE_ETYM and MANNER also requires some comments. An alternative approach would be to have a variable representing the etymological sound with four levels (/s̪/, /s̪/, /ts̪/, /ts̪/). However, this approach is not feasible as the low number of affricate sibilants (and especially of apico-alveolar affricates) in our dataset would cause some interactions to have very few tokens. Having two separate variables, PLACE_ETYM and MANNER, is a compromise which enables us to model both features whenever possible and relevant.

Taking these considerations into account, the interactions specified in the model are the following:

- PLACE_ETYM – MANNER: to find out whether there is a difference between the four sounds of interest, or rather the processes are better described in terms of place or manner only.
- FREQUENCY_CLASS – PLACE_ETYM, FREQUENCY_CLASS – MANNER: frequency might interact differently with different sounds.
- FREQUENCY_CLASS – CATEGORY: frequency might interact differently with different categories.
- CONTEXT – PLACE_ETYM: the phonetic environment might interact differently with different sounds.
- LOAN – PLACE_ETYM: words of different origin might show different behaviour.
- CATEGORY – PLACE_ETYM, CATEGORY – MANNER: the lexeme's category might interact differently with different sounds.

Furthermore, the model includes the following random effects:

[8] We decided not to use a numeric variable because the frequency values come from modern data, and thus they are only an approximation, which is better reflected in broader frequency classes. Furthermore, as we will show, the relation between frequency and other variables does not appear to be linear. Adding frequency as a factor is the simplest way to reflect this in the model.

[9] For English, the frequency of 35 per million words is sometimes used as the cut-off point between the high-frequency words and the rest, and this is motivated by psycholinguistic studies as explained by Bybee (2007: 203). An alternative approach would be to use, for example, k-means to cluster values, as done in Kang, Yoon and Han (2015). However, this would not solve the problem of different top frequency values for laminal and apical sibilants (apical sibilants would not have lexemes in the highest frequency class).

- PAGE: page on which the example is found. Lubieta's text is a manuscript, and the effect of PAGE is treated as a proxy of (writing) session (or trial in experimental studies), under the assumption that the different parts/pages of the text were written in different moments/days.
- LEXEME: different lexemes might show different tendencies.

3.3 Predictions

3.3.1 General predictions

Taking into account what we know about the chronology and geographic extension of the sibilant merger in Basque, both patterns of merger are expected to be quite advanced in Lubieta's variety but not necessarily completed. Less variability is expected in affricate consonants than in fricatives because for the affricates, the direction of change is apical to laminal in both the Western and the Central merger.

3.3.2 Phonetic environment

As explained in Section 2.2, in the Western merger, sibilants are more likely to neutralise before a consonant and in word-final position than before a vowel, where the contrast is preserved until later (Zuloaga 2020). Thus, if the merger was complete in Lubieta's variety, the phonetic environment would not determine the choice of spelling. However, if there was a conditional merger (with neutralisation of the alveolar sibilants restricted to a particular phonological context), innovative spelling would be more common in some environments than others. Furthermore, if hypercorrection was responsible for the change from apical to laminal in the Central pattern, we hypothesise that the change would start from the contexts where the Western merger was not systematic yet: speakers are unlikely to start to hypercorrect in contexts where the opposition between the two fricatives was already completely lost, but would rather do so in contexts where the contrast is partially maintained.

3.3.3 Lexical effects and frequency

The relation between lexical frequency and the generalisation of the sibilant neutralisation or the potential role of word class have not been discussed in the literature on Basque yet (among other reasons, because of the usually small number of examples and the limited range of vocabulary found in texts). Therefore, we cannot formulate precise predictions regarding these factors. Our results might be important to determine the nature of the change that resulted in the loss of contrast in sibilants. In particular, they can shed light on the role of hypercorrection: following Bybee (2001, 2007, 2012), we hypothesise that a change initiated as hypercorrective would start from low-frequency lexemes and affect most frequent words last (Bybee 2001: 81).

However, if the change from laminal to apical was initiated due to articulatory reasons (e.g. if it followed from contextual coarticulatory patterns), the data might show the opposite: frequent words showing innovative behaviour contrary to infrequent words (cf. Bybee 2001: sec. 4.4; Bybee 2007: 201; Bybee 2012).

Because of the sociolinguistic situation of Basque, another potentially relevant lexical factor is the distinction between native vocabulary, adapted loanwords and unadapted (recent) loanwords. The lexeme's origin might influence spelling: writers are likely to borrow a given word alongside its spelling. It is important to note that spelling was much more normalised for Spanish than it was for Basque at the time the text was written and that writers were more proficient in writing Spanish than they were in writing Basque. Therefore, we expect a low occurrence of innovative spelling in recent loanwords.

4 Data exploration

4.1 Overview

6316 tokens of alveolar sibilants, spelled with <z>, <c> or <s>, were gathered from the text.[10] Taking into account their etymological value, /s̻/ is the most frequent (64% of examples), followed by /s̪/ (20%) and /ts̻/ (9%). /ts̪/ appears only in 2% of the examples. Thus, laminal sibilants have much higher token frequency than apical sibilants. However, the type frequency (the number of distinct lexemes in which the sound appears) is only slightly higher for laminals (Table 1). The lexemes' mean frequency (based on an external corpus, as discussed earlier) is also higher for laminal consonants, as is the maximum frequency value. This means that, generally speaking, laminal sibilants occur in high-frequency lexemes, while apical sibilants occur in many infrequent ones.

Table 1: Number of examples, distinct lexemes and lexemes' frequency (per 10 000 words) for apical and laminal consonants

PLACE_ETYM	Tokens	Distinct lexemes	Mean frequency	SD	Max frequency
Laminal	4907	521	13.8	81.9	805
Apical	1409	426	1.6	9.2	164

4.2 Manner and place of articulation

We observe innovative spelling in 29% of the tokens. InnSpell is found for all sounds under study, but with varying proportions (Figure 5). Fricative sibilants are represented with InnSpell in around 30% of the cases, and the difference in place of articulation is small (34% laminal vs 28% apical). In affricates, InnSpell is more common in apical sibilants than in laminal sibilants (39% vs 7%).

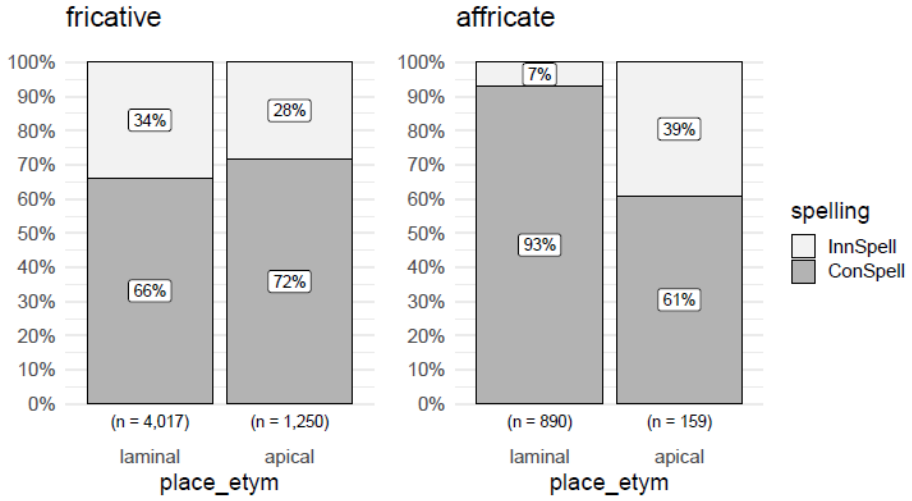


Figure 5: Proportions of InnSpell and ConSpell for fricative and affricate sibilants

4.3 Context

For laminal sibilants, InnSpell appears most often in preconsonantal position (73%). It is also frequent in word-initial and word-final position (53% and 41%, respectively; Figure 6). The results for word-initial position are somewhat skewed because of two frequent lexemes, *zu* 'you (sg.)' and *zuek* 'you (pl.)', which

[10] More details on spelling can be found in the supplementary materials.

are consistently spelled in an innovative manner. The proportion of word-initial InnSpell drops to 26% when these two lexemes are not taken into account (see Figure S8 in the supplementary materials). Prevocalic laminal sibilants are represented with InnSpell only in 6% of the cases.

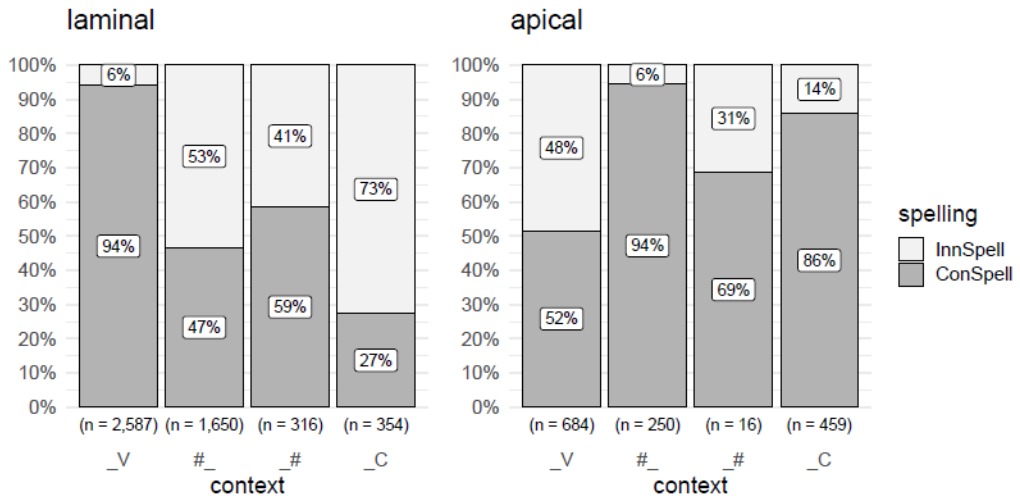


Figure 6: Proportions of InnSpell and ConSpell in the different contexts for laminal and apical sibilants

The situation is different for apical sibilants, where the prevocalic context favours the appearance of InnSpell most (48%) followed by the word-final context (31%). Apical sibilants which appear at the beginning of a word or before a consonant are spelled conservatively most of the time (94% and 86% respectively).

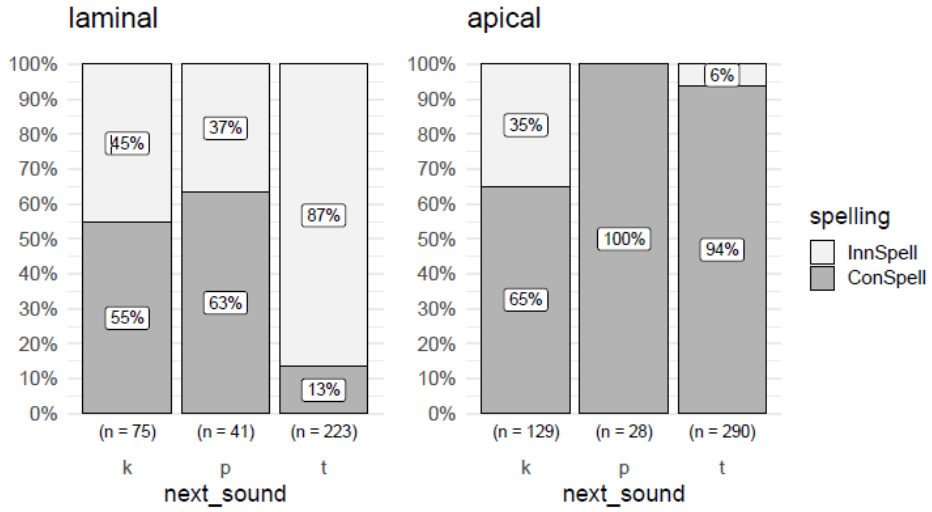


Figure 7: Proportions of InnSpell and ConSpell in preconsonantal position

With regards to voiceless stops (the most common consonants after a sibilant in Basque, Figure 7), sibilants preceding /t/ and /p/ exhibit different patterns: InnSpell is found in etymologically laminal sibilants (37% for [p] and as much as 87% for /t/), but not in apical sibilants (0% for /p/ and 6% for /t/). The most common examples for laminals include verbal forms (e.g. the imperative *zaitezte*) and negated verbs (*ezta* 'it is not'). A following /k/ triggers InnSpell in both apical and laminal sibilants (35% and 45%, respectively). The most frequent examples include *asko* 'a lot', *esku* 'hand', *eskatu* 'ask' for apicals, and *ezkondu* 'marry', *bizkotxo* 'cake' and various verbal forms for laminals.

4.4 Category

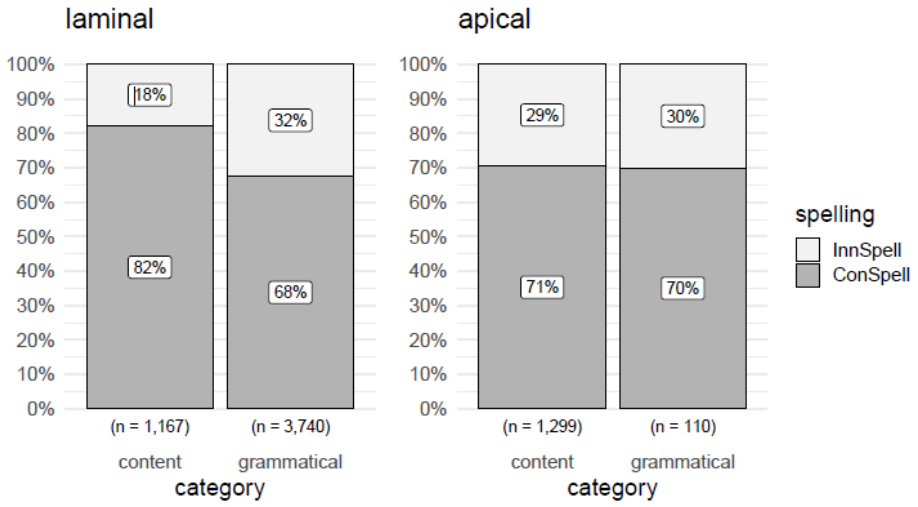


Figure 8: Proportions of InnSpell and ConSpell in content and grammatical items for laminal and apical sibilants

As regards laminal sibilants, InnSpell is more common in grammatical items (32%) than it is in content words (18%) (Figure 8). However, not all grammatical items behave in the same way. For example, forms of the verbs *izan* 'to be' or **edun* 'to have' are only rarely spelled innovatively, but the negative particle *ez* 'no', personal pronouns *zu* 'you (sg.)' and *zuek* 'you (pl.)', or the interrogatives *zer* 'what' and *zein* 'which', and the indefinite *ezer* 'something' often show InnSpell.

The proportion of InnSpell in apical sibilants is similar for content words and grammatical words and affixes. Nevertheless, it is important to mention that apical sibilants are very infrequent in grammatical words. Some examples from the text analysed here are *oso* 'very', *asko* 'a lot', *ainbeste* 'so much', *beste* 'other' and adjectives with the suffix *-so* (borrowed from Spanish) such as *odioso* 'hateful'. Among them, *asko* tends to be spelled with InnSpell. There are also a few examples of the suffix *-so* spelled with <z>.

4.5 Loanwords

There are many loanwords in our text: 522 lexemes out of 871 are recent (barely adapted) loanwords (60%), 6% are adapted loanwords and 34% are native words. This distribution is, to some extent, related to the nature of the text: it is partly a dictionary, which is likely to contain words for which the translator could not find Basque equivalents. Loanwords have low token frequency, though: taking into account word forms instead of lexemes, 76% of the examples (4802/6316) pertain to the native lexicon, while 20% involve recent loanwords and 3% older borrowings.

In general, as shown in Figure 9, InnSpell is more common in the native lexicon (32% for laminal sibilants and 50% for apical sibilants) than in recent or older loanwords (7-16%). The lowest proportion of InnSpell is for borrowings with an apical sibilant (7%). Generally speaking, older and more recent loanwords appear to behave in a similar way.

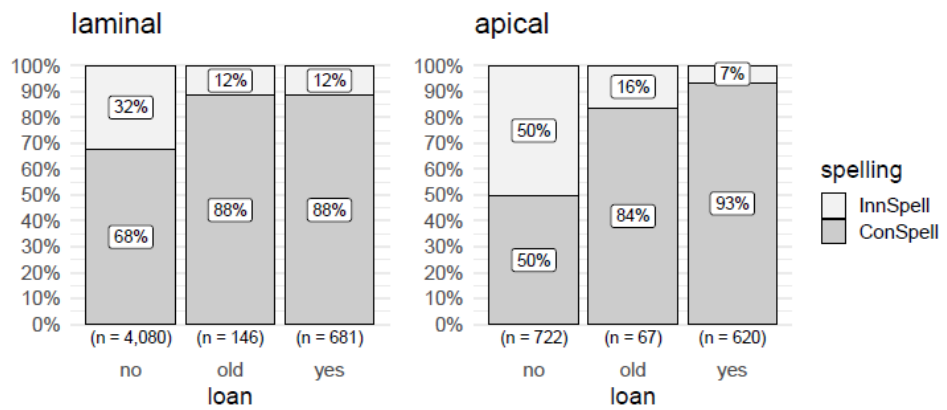


Figure 9: Proportions of InnSpell and ConSpell in lexemes of different origins for laminal and apical sibilants

4.6 Frequency

There are 871 lexemes in the corpus. Mean number of tokens is 7.25, but there are huge differences between lexemes: standard deviation is 42.9. 425 lexemes have only one occurrence, which is not surprising in a text which is, in part, a dictionary. Figure 10 plots log number of tokens found in the text and log frequency for lexemes which have more than 10 occurrences. For lexemes with a lower number of occurrences, the frequency values assigned from the external corpus vary greatly, as expected considering the type of text. Most recurrent lexemes in the text tend to have high frequency values, but there are exceptions: for example, the word *aborrezitu* 'abhor' appears so many times because it is used in a verbal paradigm, but it is an infrequent word.

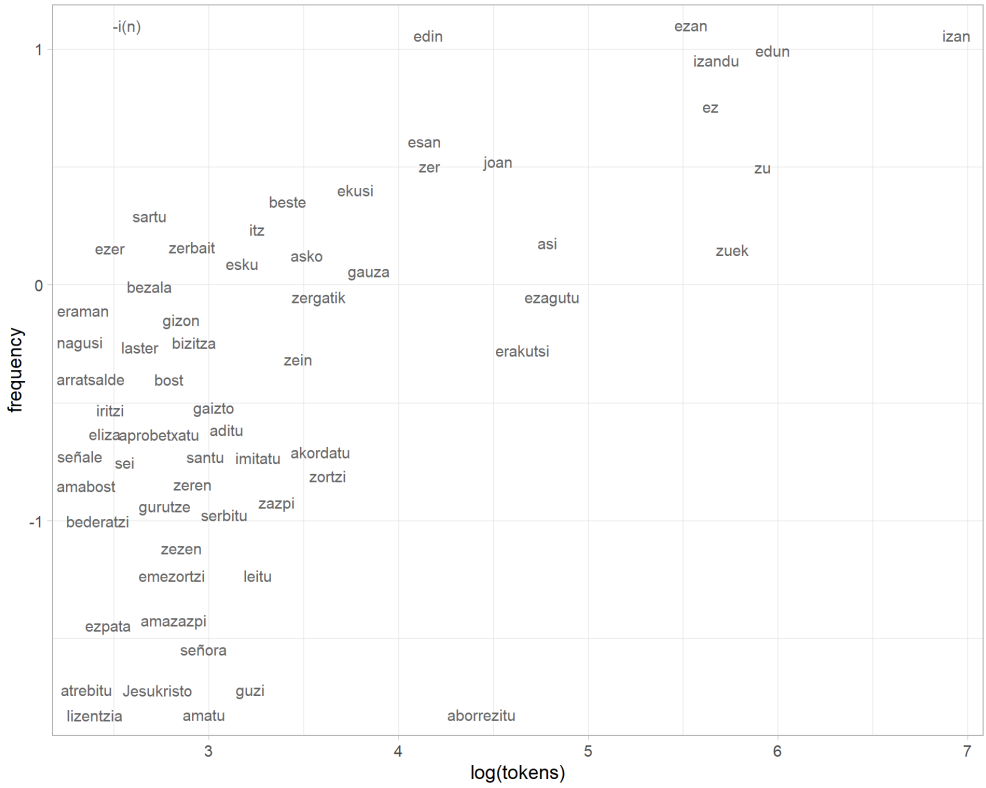


Figure 10: Log number of tokens and log frequency for lexemes with more than 10 examples in the text

As explained in 3.2, we grouped frequency values into three classes for statistical modelling. It is nevertheless important to explore the data using numeric values as well. As shown in Figure 11, forms of highest-frequency lexemes tend not to show InnSpell, but for those of slightly lower values InnSpell predominates. For mid- and low-frequency lexemes, the conservative spelling is more common.

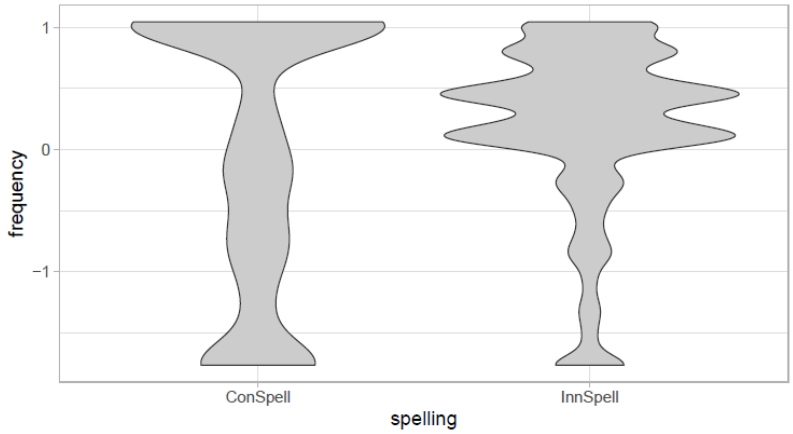


Figure 11: Violin plots for frequency and the variable SPELLING

Figure 12 shows how frequency interacts with other variables. The pattern just described (highest, mid and low values tend to show conservative spelling, and high-mid values show InnSpell) applies to laminal fricative sibilants appearing in native grammatical words. Items with the highest frequency values are auxiliary verbs (*izan* 'to be', **edin* 'to be (irrealis)', **edun* 'to have'...), which only show a significant proportion of InnSpell in forms where the sibilant appears before a consonant. In turn, lexemes in the lower part of the high-frequency class that are often spelled innovatively include forms of the content verb *joan* 'to go', and the grammatical items *zer* 'what' and *zu* 'you (sg.)'.

For apical sibilants, low-frequency lexemes (which tend to be borrowings) show InnSpell less frequently than mid- and high-frequency items. High-frequency lexemes with a high proportion of InnSpell are the following: *esan* 'say', *hasi* 'start', *ekusi* 'see', *oso* 'very' (all intervocalic). Examples of mid-frequency words with a high proportion of InnSpell include *nagusi* 'major', *arratsalde* 'afternoon', *eseri* 'sit', *asko* 'much', *osasun* 'health'.

For content words, InnSpell is slightly more common in the highest-frequency lexemes, but less common in the lowest-frequency ones. As regards manner of articulation, for affricate sounds InnSpell appears equally common for mid- and low-frequency lexemes, but less common for the high-frequency ones. Finally, loans have lower frequencies than inherited words, and their distribution has similar shapes for InnSpell and ConSpell.

Finally, Figure 13 plots the proportion of InnSpell for the three frequency classes in apical and laminal sibilants (see supplementary materials for plots showing interactions of the variable FREQUENCY_CLASS with other variables). It shows that the way in which we binned the data respects the original distribution fairly well: for laminal sibilants mid- and high-frequency items show high proportion of InnSpell, and for apical sibilants the proportion is highest in the high-frequency class, which, as discussed earlier, is rather small (few frequent lexemes have apical sibilants).

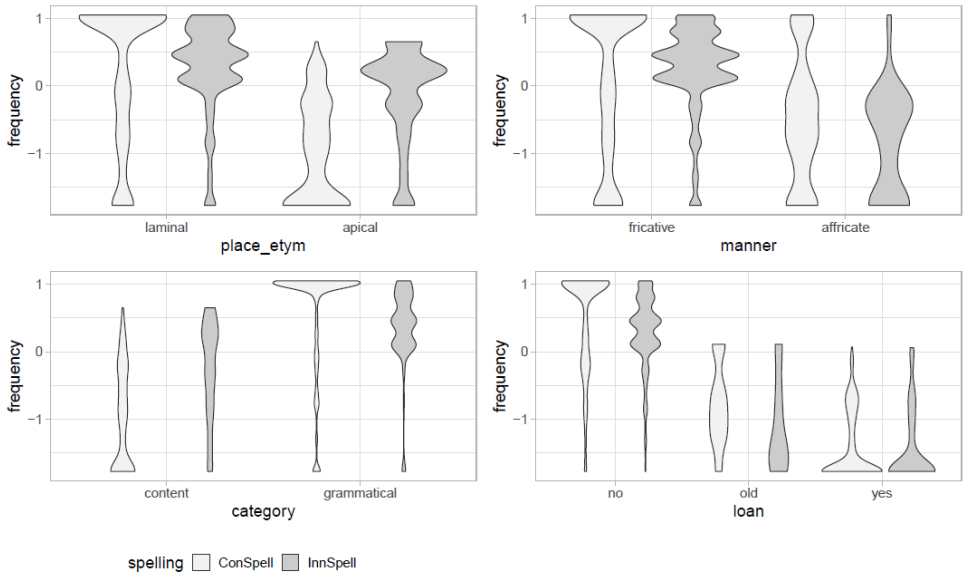


Figure 12: Violin plots for frequency and SPELLING for the variables PLACE_ETYM, MANNER, CATEGORY and LOAN

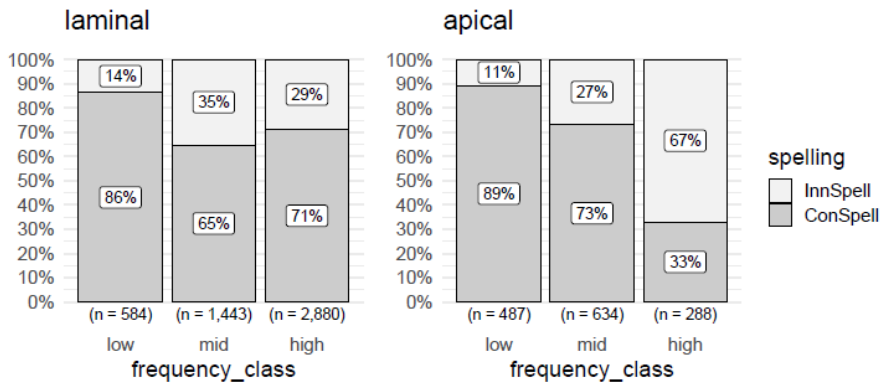


Figure 13: Proportions of InnSpell and lexeme's frequency class

4.7 Lexeme (random effect)

Figure 14 plots the lexeme frequency and the proportion of InnSpell for laminal and apical sibilants. The plot includes lexemes with 8 or more occurrences (not all labels are displayed to avoid overlap). Figure 15 shows histograms for the same data, i.e. the number of lexemes which have the given proportion of InnSpell.

Some differences in the distribution of InnSpell in laminal vs. apical sibilants seem to coincide with differences in lexical frequency between the two groups, as already discussed. Lower-frequency lexemes show ConSpell in apicals and a high proportion of InnSpell only occurs with higher-frequency items. Moreover, highest frequency lexemes with laminals (all grammatical) usually appear with ConSpell.

Generally speaking, for both laminal and apical sibilants, there are many lexemes which are spelled consistently in the conservative manner, but few lexemes show consistent InnSpell. Inconsistently spelled lexemes are more common for laminal sibilants. If we define consistent spelling as having 90% or more of examples spelled in the same manner, then for lexemes with 6 or more occurrences, for laminal sibilants 51% are spelled inconsistently compared to 36% for apical sibilants. Among the consistently written items (49% for laminals and 64% for apicals), ConSpell predominates: only a few lexemes are consistently spelled in the innovative way in each series of sibilants. They are *arrats* 'afternoon' (10/11 have InnSpell),

nagusi ‘major, senior’ (11/11), *ikasi* ‘learn’ (7/7), *zu* ‘you (sg.)’ (324/344) and *zuek* ‘you (pl.)’ (288/293). (More details on the most common lexemes in the corpus are given in the supplementary materials).

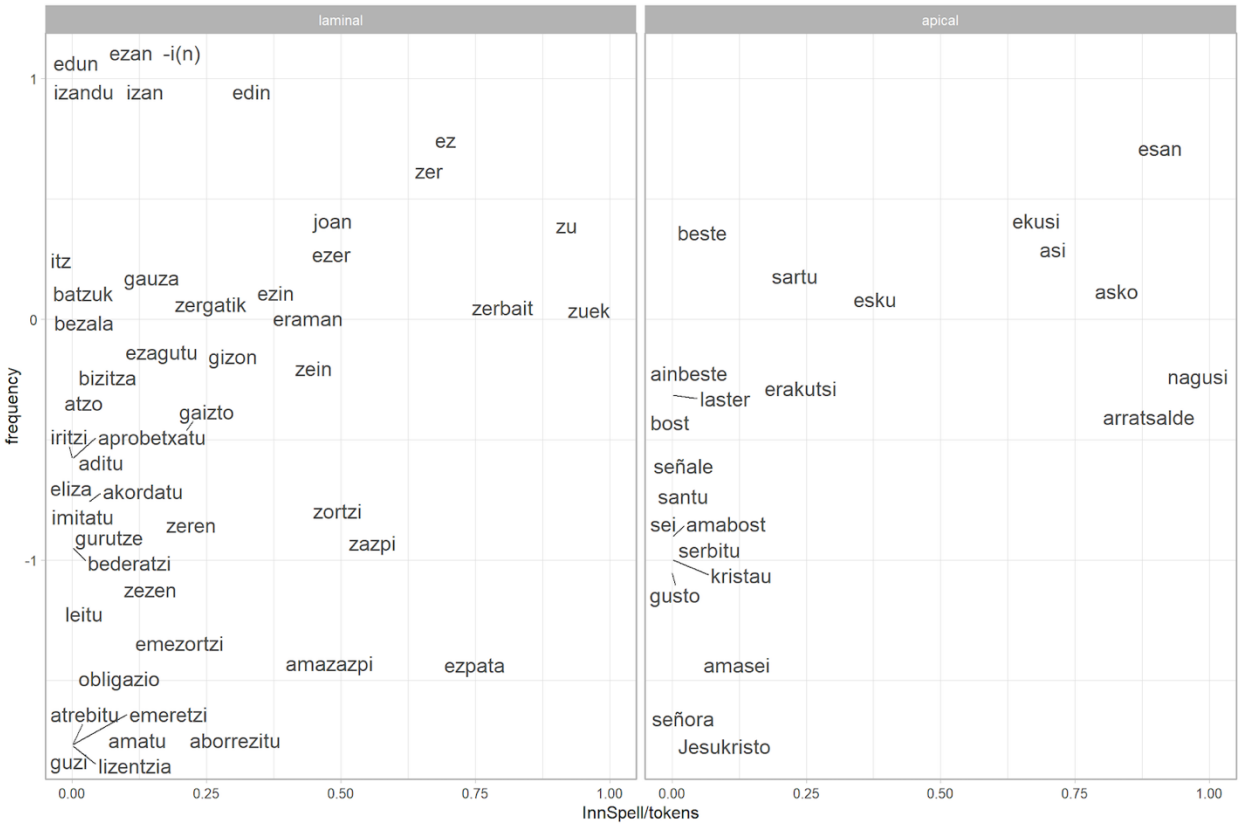


Figure 14: Lexemes’ frequency and the proportion of InnSpell for laminal and apical sibilants (for lexemes with 8 or more tokens)

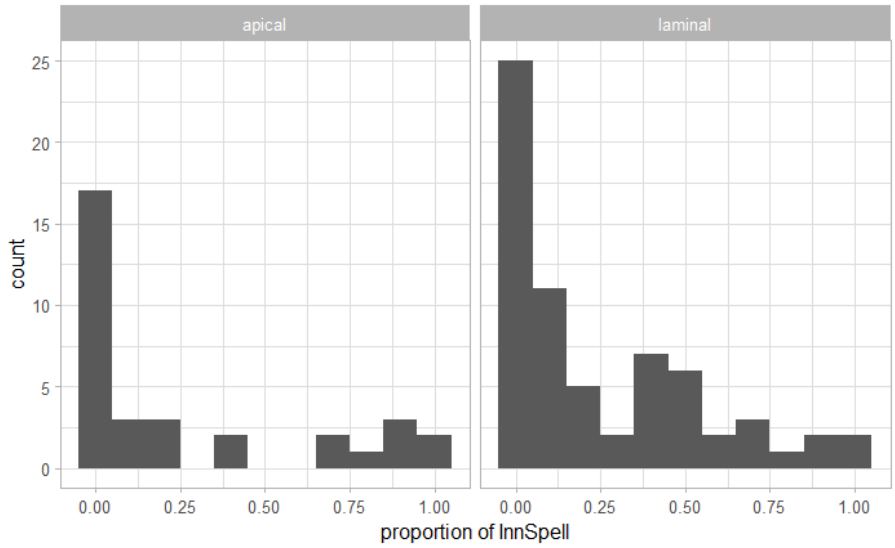


Figure 15: Histograms of the proportion of InnSpell for laminal and apical sibilants (for lexemes with 8 or more tokens)

4.8 Page (random effect)

The number or sibilants varies from page to page, the mean being 14 (SD = 7.8, min = 1 and max = 39). The proportion of InnSpell in sibilants is also different on different pages of the text: mean = 0.29, SD = 0.195, min = 0 and max = 1. Figure 16 shows the proportion of InnSpell for pages with more than 20 examples. As can be seen, the proportion of InnSpell does not change significantly throughout the text, but there are a few places where high values cluster (e.g. pages 101-104 and 213-214, which happen to contain the verbal paradigms with the participle *hasi*, which is frequently spelled innovatively).

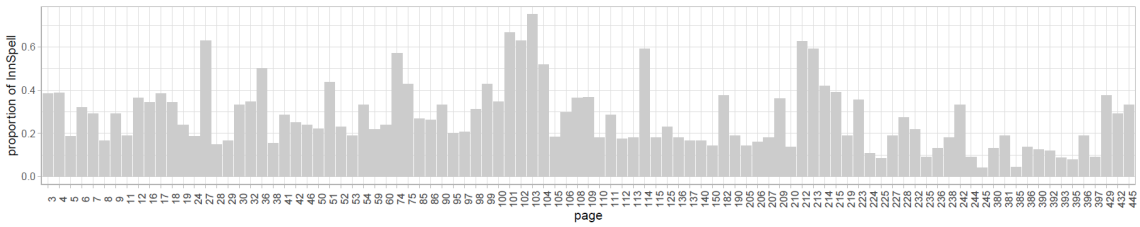


Figure 16: Proportions of InnSpell for pages in the text with more than 20 examples of sibilants

5 Modelling

5.1 Model building

The initial model consists of two random effects (LEXEME and PAGE), six independent variables (PLACE_ETYM, MANNER, CONTEXT, LOAN, CATEGORY, FREQUENCY_CLASS) and eight interactions (PLACE_ETYM – MANNER, FREQUENCY_CLASS – PLACE_ETYM, FREQUENCY_CLASS – MANNER, FREQUENCY_CLASS – CATEGORY, CONTEXT – PLACE_ETYM, LOAN – PLACE_ETYM, CATEGORY – PLACE_ETYM, CATEGORY – MANNER), as specified in Section 3.2. The model only includes random intercepts, as fitting random slopes proved unfeasible with our data.

As for the fixed effects, a likelihood ratio test indicates that excluding the MANNER – FREQUENCY_CLASS interaction results in a model which is not significantly worse ($\chi^2(2) = 4.68$, $p = 0.097$). Further deletions worsen the model's fit. Thus, the final model includes all the variables and interactions except for MANNER – FREQUENCY_CLASS.

Moving to the random effects, the inclusion of LEXEME improves the model. A likelihood ratio test shows that the model with the random effect is significantly better than the model without it ($\chi^2(1) = 1305.4$, $p < .0001$). The inclusion of the random effect PAGE similarly results in a better model ($\chi^2(1) = 133.3$, $p < .0001$).

The results of this model are given in Table 2 and Table 3 and will be discussed in the next section. Before that, we will comment on the model's goodness of fit.

We used the package DHARMa (Hartig 2022) to verify the model's assumptions. The checks of the model's residuals do not raise any concerns, with the exception of some issues caused by the random effect LEXEME, which shows strong intercept adjustments for some items and thus violates the assumption of uniformity for random effects. We decided to continue with this model despite this problem, because there are studies which suggest that this generalised linear mixed-effects models show robustness to misspecifications of random-effect distribution (Bell, Fairbrother & Jones 2019: 1051; Silk, Harrison & Hodgson 2020 and references therein). We include more details on the model's diagnostics in the supplementary materials (Section 5.3.1 in the document with the code used in the paper).

An assumption of regression models is the lack of collinearity. The model fitted here does not appear to have major collinearity issues: the condition number with the intercept included is 19.28, which is lower

than the level of 30, which indicates collinearity.[11] Nevertheless, a few variables and interactions have Variance Inflation Factor values higher than 10, which is often considered a level suggesting problems with collinearity. Those are MANNER (12.0), FREQUENCY_CLASS (11.6), PLACE_ETYM – FREQUENCY_CLASS (12.8), CATEGORY – FREQUENCY_CLASS (11.8), MANNER – CATEGORY (11.6). Taking into account the value corrected for the degrees of freedom, the highest value is that of MANNER (3.5) and MANNER – CATEGORY (3.4).

Table 2: Fixed effects in the final mixed-effects logistic regression model. Predicted odds are for InnSpell. The variables' reference levels are the following: PLACE_ETYM – laminal, MANNER – fricative, CONTEXT – _V, LOAN – no, CATEGORY – content, FREQUENCY_CLASS – mid

Predictors	Log-Odds	SE	*z* value	*p*	
(Intercept)	-3.33	0.48	-6.96	< 0.001	***
PLACE_ETYM [apical]	3.65	0.71	5.11	< 0.001	***
MANNER [affricate]	-3.82	0.84	-4.55	< 0.001	***
CONTEXT [#_]	1.86	0.22	8.34	< 0.001	***
CONTEXT [_#]	2.17	0.28	7.86	< 0.001	***
CONTEXT [_C]	3.82	0.30	12.92	< 0.001	***
LOAN [old]	-0.92	1.07	-0.87	0.387	
LOAN [yes]	0.40	0.58	0.68	0.494	
CATEGORY [grammatical]	-0.22	0.53	-0.41	0.678	
FREQUENCY_CLASS [high]	-3.51	0.72	-4.86	< 0.001	***
FREQUENCY_CLASS [low]	-0.55	1.31	-0.42	0.673	
PLACE_ETYM [apical] * MANNER [affricate]	6.07	1.19	5.09	< 0.001	***
PLACE_ETYM [apical] * FREQUENCY_CLASS [high]	4.13	0.89	4.65	< 0.001	***
PLACE_ETYM [apical] * FREQUENCY_CLASS [low]	3.29	1.50	2.19	0.028	
CATEGORY [grammatical] * FREQUENCY_CLASS [high]	5.57	0.76	7.31	< 0.001	***
CATEGORY [grammatical] * FREQUENCY_CLASS [low]	0.87	1.28	0.68	0.494	
PLACE_ETYM [apical] * CONTEXT [#_]	-4.90	0.88	-5.56	< 0.001	***
PLACE_ETYM [apical] * CONTEXT [_#]	-0.78	1.46	-0.54	0.593	
PLACE_ETYM [apical] * CONTEXT [_C]	-6.22	0.58	-10.70	< 0.001	***
PLACE_ETYM [apical] * LOAN [old]	-2.38	1.48	-1.61	0.108	
PLACE_ETYM [apical] * LOAN [yes]	-5.88	0.84	-6.99	<0.001	***
PLACE_ETYM [apical] * CATEGORY [grammatical]	-4.06	1.08	-3.75	<0.001	***
MANNER [affricate] * CATEGORY [grammatical]	2.38	0.90	2.65	0.008	**

Table 3: Random effects in the final mixed-effects logistic regression model

Random effect	*N* of groups	SD
LEXEME	871	3.43
PAGE	450	0.99

As for the model's goodness of fit, the model's index of concordance *C*, which measures how well the model predicts the data, is 0.97 (0.5 denotes random predictions and 1 a perfect correlation between data and model predictions). Another measure of goodness of fit we can use is pseudo-R^2 (Nakagawa &

[11] For the diagnostics we used the following tools: the index of concordance C and the condition number for collinearity were calculated with *JGmermod* library (Grafmiller 2019) and VIF values with *car* (Fox & Weisberg 2019). R^2 was obtained with the library MuMIn (Bartoń 2020).

Schielzeth 2013; Nakagawa, Johnson & Schielzeth 2017). The marginal R^2 measures the variance explained by the fixed effects, and the conditional R^2 expresses the variance explained by the entire model, i.e. including random effects. For the model fitted here, the marginal R^2 is 0.31 and the conditional R^2 is 0.86. The intra-class correlation coefficient (ICC) is 0.795, which means that as much as 80% of variance can be explained with the random effects and especially with the variable lexeme, which has the ICC of 0.733.

5.2 Effects of the predictors

The model's coefficients are listed in Table 3 and also represented visually in Figure 17, which orders the predictors by the value of the coefficients predicted by the model. Items to the right of the line at 0 show an increased probability of InnSpell, and those to the left show a decreased probability of InnSpell. Those with the confidence interval crossing the line at 0 have no significant effect. We can easily see, for example, that the strongest effects favouring InnSpell are related to the apical affricates and low-frequency grammatical lexemes. Another strong effect is CONTEXT: preconsonantal position is the context most strongly linked to InnSpell. At the other extreme, with the lowest coefficients (and thus the lowest probability for InnSpell) we find, for instance, preconsonantal apical sibilants and loanwords with an apical sibilant. The interpretation of coefficients is nevertheless not straightforward in a rather complex model as the one discussed here, and because of that we will use the plots in Figure 18. For each level of the predictors, they plot probabilities of InnSpell (expressed in percentages) computed while keeping all other factors in the model constant.[12] The lines represent 95% confidence intervals.

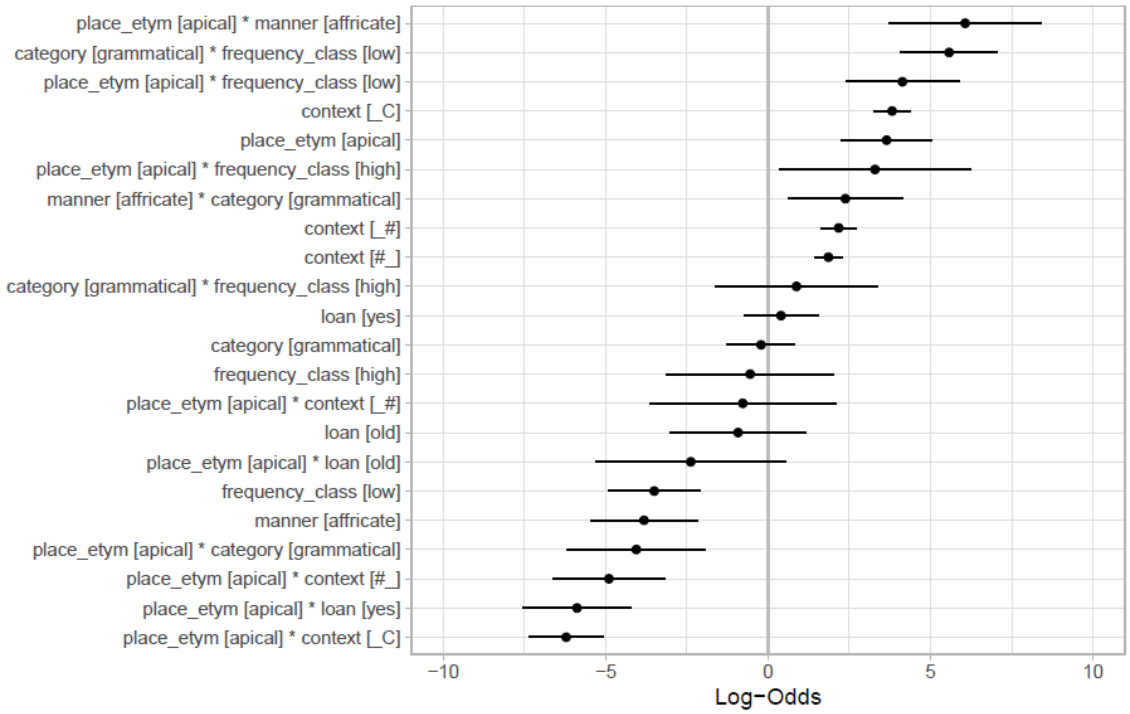


Figure 17: Log-odds predicted for InnSpell. Dots indicate the value predicted for each predictor, with horizontal lines showing 95% confidence intervals

[12] Calculated with the *ggeffects()* function in the *ggeffects* library (Lüdecke 2018).

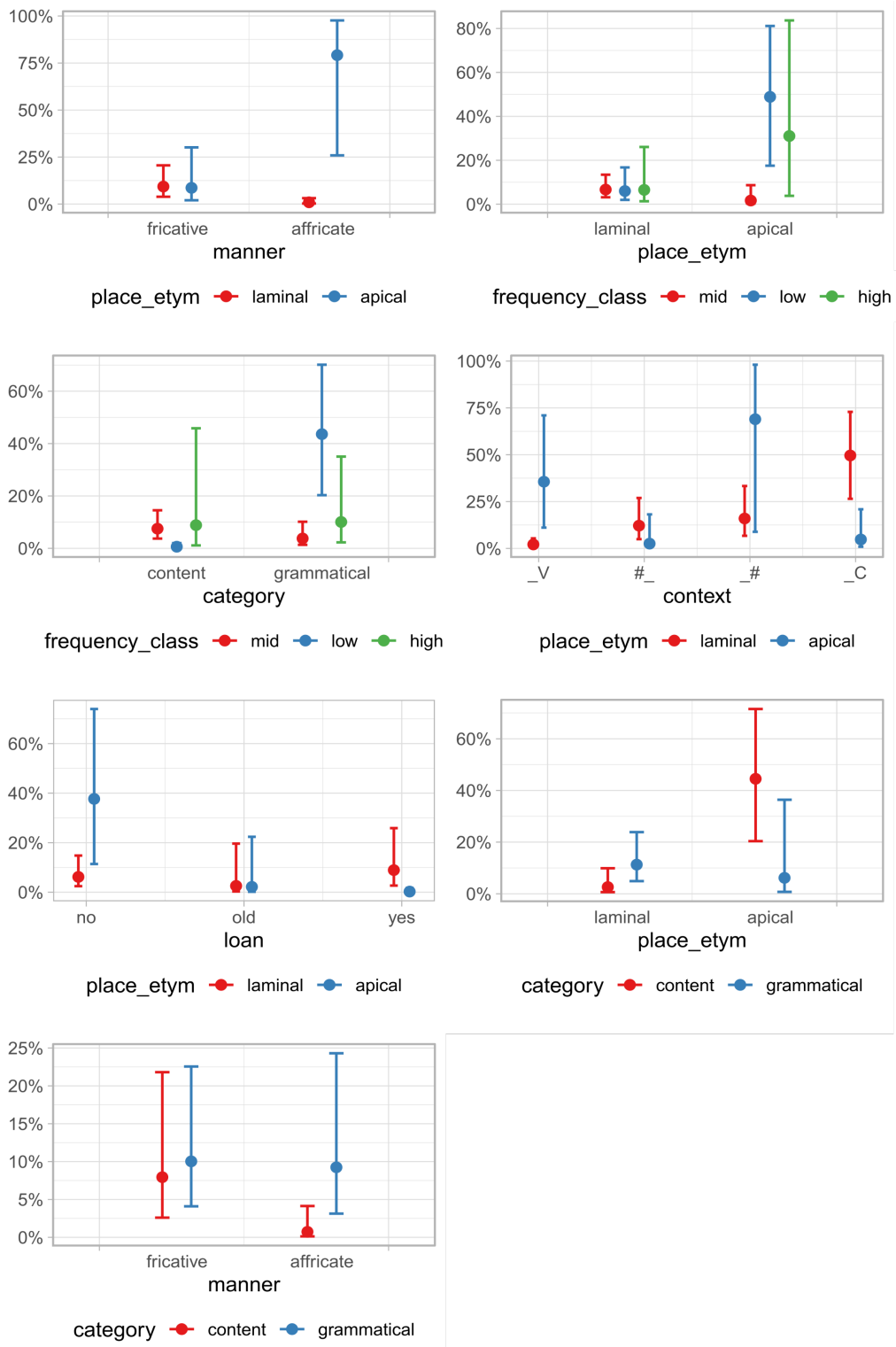


Figure 18: Interactions between predictors in the model (the *y* axis shows probabilities of InnSpell expressed in percentages)

Etymologically laminal and apical sibilants differ in various respects. Apical affricates show a very high probability of InnSpell (0.79) compared to their fricative counterpart (0.09). Among laminal

consonants, fricatives are slightly more likely to be written with InnSpell than affricates (0.09 vs 0.01). The variable CONTEXT shows clearly different patterns for apical and laminal sibilants. As regards laminal sibilants, the context where InnSpell is most likely to occur is the preconsonantal position (0.5), followed by word-final and word-initial contexts (0.16 and 0.12, respectively). The environment showing the least probability of InnSpell is the prevocalic position (0.02). For apical sibilants, word-final position shows the greatest probability of InnSpell (0.69), but the confidence interval is quite big (most probably due to the small number of examples). The prevocalic context comes next with 0.36. Word-initial and preconsonantal environments are much less likely to show InnSpell with apical consonants (0.03 and 0.05, respectively).

There are only non-significant differences between native words and loanwords for laminal sibilants (0.09 for recent loanwords, 0.06 for native words and 0.03 for older borrowings). Native words with an apical sibilant, however, show a high probability of InnSpell (0.38) as compared to older loanwords (0.02) and recent loanwords (0).

In the variable CATEGORY, grammatical words or affixes with a laminal sibilant show a slightly higher probability of InnSpell than content words (0.11 vs 0.03). For apical sibilants, there is a clearer contrast between categories, with content words showing a higher occurrence of InnSpell (0.45) than other items (0.06) (but it must be taken into account that the confidence intervals overlap). Category also interacts with manner, where content words with an affricate consonant show the lowest probability of InnSpell (0.01 as compared to around 0.1 for fricatives and grammatical items).

Finally, we observe a number of frequency-related effects. For laminal sibilants, the probability of InnSpell is low for words from all frequency levels (around 0.07). However, for apical sibilants, mid-frequency words have the lowest probability of InnSpell (0.02), as compared to high-frequency (0.31) and low-frequency words (0.49). Nevertheless, for both high- and low-frequency words confidence intervals are quite big and they have to be interpreted with caution. As regards the interaction between frequency and category, differences are small for content words: 0.09 for high, 0.07 for mid and 0.01 for low-frequency words. In grammatical items, the highest probability of InnSpell is associated with the low-frequency class (0.44) and the lowest with the high and mid-frequency groups (0.1 and 0.04, respectively).

In sum, the highest probability of InnSpell (>0.3) is predicted for: affricate apicals, apicals in high and low-frequency items, prevocalic apicals, word-final apicals, apicals in native words, apicals in content words, low frequency grammatical words and preconsonantal laminals. The lowest (<0.05) is predicted for: older loanwords, loanwords with an apical, word-initial apicals, prevocalic laminals, low-frequency content words, apicals in mid-frequency words and laminal affricates.

5.3 Random effects

Finally, let us have a look at the two random effects in the model, PAGE and LEXEME. As indicated earlier, the random effects' contributions to the final model are highly significant. Table 3 above shows that the standard deviation is 3.4 for LEXEME and 1 for PAGE, which means that there is much greater variability between different lexemes than there is between pages.

The lexemes with the highest and lowest intercept adjustments (among lexemes with more than 20 tokens in the text) are listed in Table 4.

As regards the effect for PAGE, it appears that negative adjustments are more common towards the end (Figure 19).

Table 4: Lexemes with highest and lowest intercept adjustments (for lexemes with more than 20 tokens)

Lexeme	Intercept adj.	Lexeme	Intercept adj.
asko 'much'	7.6	*hitz* 'word'	0
zuek 'you (pl.)'	5.8	*edun* 'have'	-1
zu 'you (sg.)'	4.3	*imitatu* 'imitate'	-1.4
zortzi 'eight'	4.2	*ekusi* 'see'	-1.5
aborrezitu 'abhor'	2.8	*asi* 'start'	-2
joan 'go'	2.3	*guzi* 'all'	-3.1
ez 'no'	0	*erakutsi* ' show'	-3.5
zer 'what'	1.9	*izandu* 'be'	-3.7

Figure 19: Intercept adjustments for the effect PAGE (for pages with more than 15 tokens)

6 Discussion

The statistical model explains only a part of the variation present in the data, but, at the same time, it shows that all the variables included in the model influence spelling. Of the variation explained by the model, the greatest part is explained by the random effect LEXEME. This is hardly surprising, given that we analyse spelling, which tends to be associated with lexemes. Nevertheless, other variables have also proven significant, which makes it possible to make inferences about the phonology of the language.[13] In this section, we thus propose answers to the questions posed in Section 2.4.

6.1 Neutralisation tendencies and the role of context and lexical effects

As regards the first question (Q1), our analysis suggests that the neutralising change that would result in a merger of alveolar sibilants was still in progress in the early 18th century Basque of San Sebastián, with the merger more advanced in the affricate than in the fricative class.

The neutralisation tendency of affricate sibilants is quite straightforward. Apical affricates are the most likely segments to show innovative spelling among the sibilants –most cases in our dataset being prevocalic (see Figure S11 in the supplementary materials), while laminal affricates remain largely unchanged with very few examples of innovative spelling. This is likely the consequence of no conflicting neutralisation patterns being found in the affricate class: the apical affricate becomes a laminal in both the Western and the Central pattern, and this is the only merger affecting affricates found in the area during the 19th and 20th centuries.

The case of the fricative sibilants, on the other hand, is more complex. Both patterns of neutralisation, the Western and the Central, show similar incidence in Lubieta's text, with roughly 1/3 of the sibilants

[13] It should be noted that the basis for our inferences, i.e. spelling, is not necessarily directly related to the pronunciation (cf. Section 2.1), but regularities and patterns found in spelling such as those we study in this paper are helpful to understand the sound system of earlier stages of a language.

showing innovative spelling: <z>/<c> are used instead of <s> in 34% of the cases, while <s> appears instead of <z>/<c> in 28% of the cases.

A closer look at the distribution of innovative spelling is necessary to understand the nature of each pattern. Generally speaking, our results show that the phonetic context is more relevant for explaining the changes in Basque sibilants than lexical or frequency effects. We will discuss both kinds of effects in turn.

Segmental context (Q2) has resulted highly significant in our model, thus suggesting that contextual neutralisation was a factor at this point in the evolution of the two mergers. Starting with the most pronounced trend, innovative spelling of laminal fricatives is most frequent in preconsonantal position, especially before /t/. In phonological terms, /s̻/ > /s̺/ /_t is almost systematic in our corpus (87%, n = 223). This is especially illustrative in combination with the extremely low rate of innovative spelling in etymologically apical sibilants (6%, n = 290) in this position. This is the only environment where we can observe a virtually complete neutralisation of the alveolar sibilants: with very few exceptions, all alveolar sibilants are produced as apicals before /t/ and, more generally, before voiceless stops. It is worth noting that this is one of the most common contexts for a sibilant to appear in Basque, and even more in running speech, where word-initial voiced stops show devoicing when preceded by a word-final sibilant in the previous word.

As for the word-final position, on the basis of previous descriptions of the Western merger, we could expect it to pattern with the preconsonantal context. Indeed, laminal sibilants are frequently spelled innovatively. However, unlike in preconsonantal position, apicals also frequently show innovative spelling word-finally. The word-final context should be interpreted cautiously, however, because of the low number of examples in the text.

Turning to word-initial and prevocalic contexts, these are usually considered the last stages in the spread of the Western neutralisation. The data presented here are in line with this. Innovative spelling appears quite frequently word-initially for laminals. Nevertheless, only a few lexemes with an etymological laminal fricative in this position show consistent innovative spelling (the personal pronouns *zu* 'you (sg.)' and *zuek* 'you (pl.)'). If we leave out these lexemes, the proportion of innovative spelling drops considerably in the word-initial position. Thus, the data suggest that in this context the change from laminal to apical was underway, but not very advanced. Apicals are mostly written conservatively in this position.

Finally, laminals are usually spelled conservatively before a vowel (as predicted from what we know about the evolution of the Western merger). Conversely, this is the context most linked to innovative spelling in apical sibilants.

These tendencies can be summarised as follows:

- _# – both (with a high proportion of innovative spelling in both series)
- _V – laminals maintained (half of the apicals show innovative spelling)
- _C – apicals only (most laminals show innovative spelling)
- #_ – both maintained (with exceptions with consistent innovative spelling)

Turning to lexical and frequency effects, as we have already said, the way in which a word is written is highly related to the word itself and less so to other factors (which is reflected in the importance of the random effect of lexeme). This is not surprising due to the indirect nature of the data we are dealing with, namely spelling. Nevertheless, differences between the two mergers may help us understand the processes behind each.

For etymologically laminal sibilants, grammatical words and affixes are slightly more likely to show innovative spelling than content words, even though in the statistical analysis the differences between frequency classes have not proven significant. However, as shown in 4.6, it appears that there are two classes among grammatical words: the most frequent ones are written conservatively, but those slightly less frequent (e.g. *zer* 'what', *zu* 'you (sg.)') are more innovative. Since the most frequent lexemes tend to be verbs (especially auxiliary and copular verbs), we might have an effect of word class too (cf. Phillips 2015: 366). Low-frequency words, in turn, tend to show conservative spelling. As noted by Phillips (2015: 367), the effects of word class and frequency might be independent. We observe that in our data: for content

words the proportion of innovative spelling decreases with frequency, but for grammatical words the most common items maintain conservative spelling.

The picture is even more complex in the apical series. Our prediction was that the apical to laminal neutralisation would start in lower frequency words. Generally speaking, the frequency values are lower for apicals (i.e. they do not appear in extremely frequent lexemes and appear in many very infrequent words). The most frequent words show innovative spelling often (contrary to what we predicted), and there are a few lexemes showing consistent innovative spelling. These words usually have an intervocalic sibilant, which makes it difficult to distinguish between contextual and lexical effects. On the other hand, in the low-frequency class, where we could expect high incidence of the innovation, most words are loans, and they are expected to follow the Spanish orthography. Native low-frequency lexemes show higher proportions of innovative spelling, but the results are hard to interpret because they usually only occur once in the text.

6.2 Accounting for the observed distributions of innovative spellings

In this section we focus on Q3: how to explain the emergence and spread of the two neutralisation patterns on the basis of the data from Lubieta's text.

We will start from the change from the lamino-alveolar /s̪/ to the apico-alveolar sibilant fricative /s̪/. We have shown that in Lubieta's variety the Western neutralisation had started but it was far from completed, and that we can order the contexts according to the strength of innovative spelling in the following way: preconsonantal > word-final > word-initial > prevocalic. Scholars agree that the order of extension of the Western neutralisation was from preconsonantal and word-final position to word-initial and prevocalic position. Data from our study allows us to better understand mechanisms behind this change.

Our account of the extension of the neutralising contexts in the Western merger can be summarised as follows (each Roman number represents a step in the extension process):

(1) Phases of the Western Basque merger
/s̪/ > /s̪/
i / _t
ii / _C
iii / _(C)$
iv / _(C)$ | #_
v / _(C)$ | $_V [general]

We propose that the process started as a phonetically conditioned sound change in preconsonantal position, and, more specifically, before /t/. In modern Basque, the canonical articulation of /t/ is usually described as apico-dental and transcribed as [t̪] (cf. Hualde 2003: 18).14 Although we cannot be certain about the articulation of any of these sounds in Lubieta's time, a classic description of Astarloa (1883: 179–180) states that in the pronunciation of [s̪] the tongue tip approximates the higher teeth while in that of [s̪] the tongue tip touches the lower incisors while folding towards the higher teeth (Michelena 1990: 279–280). Thus, unlike in the production of apico-dental [t̪], where the tongue tip goes upwards and touches the

[14] More research on the articulation of Basque dental stops is needed to determine to exact shape of the tongue. Due to the scarcity of accurate articulatory descriptions of Basque stops, they are often assumed to be similar to those of (Northern) Peninsular Spanish. In Peninsular Spanish, generally, the passive articulator is the back of the higher incisors, and the active articulator is the apex, with the lamina also contacting with the alveolar ridge (Martínez-Celdrán, Fernández-Planas & Carrera-Sabaté 2003). If /t/ was more laminal than apical at the time the Western merger was initiated, our explanation loses a bit of strength: as noted by one of the reviewers, a change from laminal to apical sibilant before a laminal stop is not so likely. Note, however, that the tongue tip would still point upwards (as in /s̪/) rather than downwards (as in /s̪/). In addition, a more general explanation would be still possible: the place retraction could have started more generally in preconsonantal position (fricatives are weakened in coda position, cf. e.g. Solé (2010)), though this would not directly account for the direction of the change.

higher incisors (and the alveolar ridge), in the lamino-(denti-)alveolar [s̪], the apex goes downwards and contacts with the lower incisors. Apico-alveolar [s̻] is like [t̪] in that the tongue tip stays upwards and approximates the base of the higher incisors and the alveolar ridge. A canonical modern Basque [t̪] would not have the exact same place of articulation of [s̻], but it would be close to it (much closer than it is to [s̪]), and the tongue shape would be way more similar.[15] Thus, coarticulation of a coda laminal sibilant to a following dental-alveolar stop may have resulted in the retraction of the place of articulation of this sibilant alongside the apicalisation of its lingual gesture. The change then extended first to all preconsonantal contexts and the perceptually ambiguous word-final position. In the next step the change reached word-initial position and, finally, any prevocalic position. Prevocalic positions were the last contexts of this change because, in word-initial position, coarticulation to a following vowel can enhance the apical/laminal distinction, and, in intervocalic position, both surrounding vowels likely helped the perception of various perceptually relevant phonetic cues.

Although occurring in a slightly different region of the mouth, this change observed in Basque would not be very different from sibilant retractions observed in many languages, which are especially common in preconsonantal position and syllable coda (cf. Kümmel 2007). For example, in English, /s/ (which is more posterior than Basque /s̪/) is retracted to /ʃ/ in /stɹ/ clusters (Shapiro 1995). Acoustically, the sibilant in /str/ clusters of Australian English has been shown to have a lower average M1 (first spectral moment) than these in /spr/ and /skr/, and these, in turn, lower than these in /sp, st, sk/; all clusters (also those without the rhotic) having lower average M1s than that in /sV/ (Stevens & Harrington 2016: 125–126, 133). In a subsequent perceptual experiment also involving speakers of Australian English, it was shown that sibilants in /str/ and /st/ clusters can be perceived as /ʃ/ by native speakers when extrapolated to prevocalic /_V contexts (Stevens & Harrington 2016: 133). In German, /s/-retraction occurs before any consonant in word-initial position, but it is extended to all coda clusters in some varieties (see, e.g., Alber, Kokkelmans & Rabanus 2021). Another example is Portuguese, where /s/ > /ʃ/ has extended to all codas (Zampaulo 2016). In all these cases, the sound change has been argued to begin in coarticulatory contexts where synchronic variation shows a bias in a given direction (Stevens, Bukmaier & Harrington 2015). In this situation, an extreme coarticulatory bias can be reinterpreted by a listener as an intended articulatory target (Baker, Archangeli & Mielke 2011). Once a conditioned merger is completed after the phonologisation of the variants in a coarticulatory context, a perceptual realignment favouring the innovative variants can extend to the rest of the instances of the affected phoneme (Harrington 2012).

Our proposal finds support within Basque, in both phonetic and historical data. The seed for this particular sound change can be observed in the modern varieties which did not develop the Western or the Central merger. One such case is that of Mixean Basque, in which a recent acoustical study has found evidence in line with the described preconsonantal sibilant retraction (Egurtzegi et al. 2022).[16] The second step, a conditional merger after the phonologisation of the neutralising change in a coarticulatory context, is found in historical records. In Kapanaga's text from 1656, written in the Western dialect, for instance, most cases of innovative spelling of <z, c> as <s> occur before a voiceless stop, with only a few examples elsewhere (Zuloaga 2020).

As regards the Central Basque merger, our study suggests that the order of extension of this neutralisation pattern was different:

[15] Nonetheless, note that a recent articulatory description (based on static MRI data) of the alveolar fricative sibilants of Basque highlights their high degree of variability in the modern language (Iribar Ibabe, Pagola Petrirena & Túrrez Aguirrezábal 2020).

[16] This variety has been proposed to show a third kind of merger that collapses the apico-alveolar and postalveolar sibilant categories for both fricatives and affricates (Egurtzegi & Carignan 2020).

(2) Phases of the Central Basque merger
/s̪/ > /s̻/
i / V_V
ii / $_V
iii / $_V | _#
iv / $_V | _(C)$ [general]

Innovative spelling appears mostly in prevocalic position, i.e. the context with the least incidence of innovative spelling in the Western pattern, and is uncommon in the preconsonantal position, the earliest and most prominent context of influence in the Western pattern. The Central neutralisation shows its biggest incidence in the intervocalic context, a position where we would expect a phonetically based neutralisation of place to have the least strength given that place-related acoustic cues are likely best perceived intervocalically.

The most straightforward way to account for an intervocalic apical to laminal change early in the merger process is through non-phonetic means, such as hypercorrection. Following Zuloaga (2020), we propose that the Central neutralisation resulted from hypercorrection, and we add that it started due to the ambiguous situation that was caused by the third step of the Western merger: after /s̻/ > /s̪/ was already a conditioned merger (as in Lubieta's Basque, where virtually only /s̪/ occurs before /t/), a perceptual realignment that slowly brought the alveolar exemplars towards the apical prototype gave rise to the first ambiguous realisations in non-coarticulatory contexts and speakers resolved this ambiguity by reverting the ongoing rule.

Thus, we propose that both mergers started due to the same coarticulatory bias, and we predict that the varieties that show the Central merger today initially developed the first stages of the Western merger, with /s̻/ > /s̪/ /_t, but then a hyper-corrective process due to the ambiguity resulting from a then non-robust opposition pushed the merger in the opposite direction. If this seemingly back-formed rule were the consequence of a certain degree of sociolinguistic awareness by the urban speakers of San Sebastián, then what we observe in the Central merger could be a case of what DeCamp (1972) called *rule symmetry* or the development of *mirror-image rules*:

> Whenever the application of a rule carries strong negative prestige, I suggest that the rule is not (within the same generation of adult speakers) dropped from the grammar or even skipped over in the derivation. Rather, a new rule may be added to the grammar, immediately following the offensive rule and symmetrical to it in form. If this new rule remains perfectly symmetrical, its effect is only to undo the effects of the offensive rule. It seldom remains perfectly symmetrical, however. Specifically, it is especially susceptible to rule generalization (Decamp 1972: 88)

We believe that Lubieta, who was familiar with other varieties of Basque, aimed to reflect the apical/laminal distinction that was gradually disappearing from his variety. The distinction was still preserved in those contexts to which the Western merger spread later. It was also present in many varieties spoken around San Sebastián and also reflected in most works printed in Basque. Thus, maintaining the distinction might be seen as more prestigious pronunciation. However, Lubieta lacked a standard variety or a reference grammar that would serve him as a guideline (the first Basque grammar was published just one year after his dictionary in 1729). This could have resulted in an over-representation of the presence of the laminal fricative sibilant in both Lubieta's written and spoken language. In the light of other sources of the period and current speech data, the application of rule-symmetry is not limited to Lubieta's manuscript: it was a more or less generalised phenomenon in the middle of the 18th century and, subsequently, it has been reinforced and extended in wide areas of the central dialect.

Thus, our account of changes in the sibilant system in the fricative class is summarised in Figure 20.

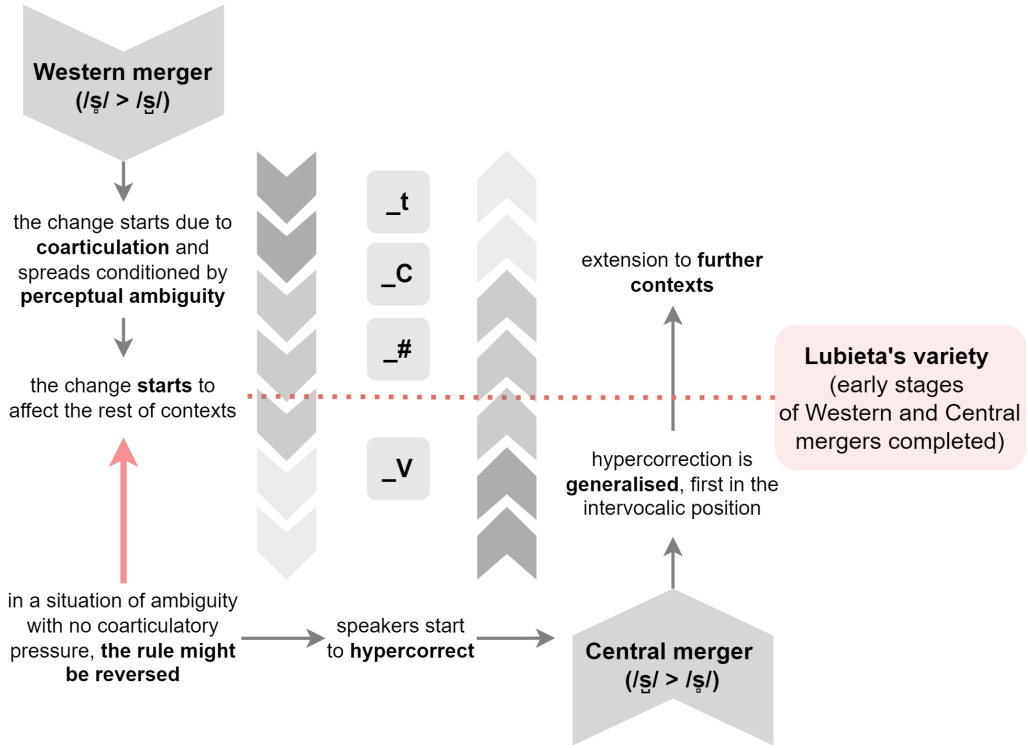


Figure 20: Changes in the sibilant system in Basque

As regards the nature of each merger, we can probably consider the Western pattern a merger by approximation, in which the phonetic targets of /s̪/ and /s̻/ gradually converge into /s̻/, while the Central pattern could be a merger by transfer, a unidirectional process in which one phoneme is replaced with another in a word-by-word manner (Labov 1994: 321). Under this account, we would expect the Western merger to be a sound change from below (cf. Labov 1994) that would gradually extend following a hierarchical model structure, but the hypercorrections that gave rise to the Central merger would be a case of a change from above, and thus its isogloss could show an uneven geographic spread, potentially with greater presence in urban areas than in rural areas (Zuloaga 2020).

As a final note, we want to emphasise that we were only able to understand our graphematic data by interpreting our results on the grounds of current theories of phonetically based sound change. Quantitative approaches to historical phonological datasets offer an attractive (although narrow) window into the intermediate steps of attested historical sound change. Nevertheless, we can only be able to understand what we observe by benefiting from an articulated theory of sound change, rooted in typological observations as well as in tested phonetic biases.

References

Alber, Birgit, Joachim Kokkelmans & Stefan Rabanus. 2021. Preconsonantal s-retraction in the Alps: Germanic, Romance, Slavic. *STUF - Language Typology and Universals* 74(1). 17–38. https://doi.org/10.1515/stuf-2021-1022.

Astarloa, Pablo Pedro. 1883. *Discursos filosóficos sobre la lengua primitiva o Gramática y análisis razonada del euskera o bascuence*. Bilbao: Pedro Velasco.

Baayen, Harald R. 2001. *Word frequency distributions*. Dordrecht: Kluwer.

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(3). 347–374. https://doi.org/10.1017/S0954394511000135.

Barteld, Fabian, Stefan Hartmann & Renata Szczepaniak. 2016. The usage and spread of sentence-internal capitalization in Early New High German: A multifactorial approach. *Folia Linguistica* 50(2). 385–412. https://doi.org/10.1515/flin-2016-0015.

Bartoń, Kamil. 2020. MuMIn: Multi-model inference. Software package. https://CRAN.R-project.org/package=MuMIn.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Bell, Andrew, Malcolm Fairbrother & Kelvyn Jones. 2019. Fixed and random effects models: Making an informed choice. *Quality & Quantity* 53(2). 1051–1074. https://doi.org/10.1007/s11135-018-0802-x.

Beristain, Ander. 2018. Azpeitiko euskararen hiztegi dialektal bat – I. *Fontes Linguae Vasconum* 125. 7–54. https://doi.org/10.35462/FLV125.1.

Beristain, Ander. 2019. Azpeitiko euskararen hiztegi dialektal bat – II. *Fontes Linguae Vasconum* 127. 7–58. https://doi.org/10.35462/FLV127.1.

Beristain, Ander. 2021. Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque. *Journal of the International Phonetic Association* 1–32. https://doi.org/10.1017/S0025100320000274.

Bilbao, Gidor. 2012. Lubieta, izen berri bat euskal testuen historian. In Joseba A. Lakarra, Joaquín Gorrochategui & Blanca Urgell (eds.), *2nd Conference of the Luis Michelena Chair - Koldo Mitxelena Katedraren II. Biltzarra - II. Congreso de la Cátedra Luis Michelena*, 621–649. Vitoria-Gasteiz: UPV/EHU.

Brown, Keith & Jim Miller. 2013. *The Cambridge dictionary of linguistics*. Cambridge: Cambridge University Press.

Bussmann, Hadumod, Kerstin Kazzazi & Gregory Trauth. 1998. *Routledge dictionary of language and linguistics*. London: Routledge.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2007. The phonology of the lexicon: evidence from lexical diffusion. In Joan Bybee (ed.), *Frequency of use and the organization of language*, 199–215. Oxford: Oxford University Press.

Bybee, Joan L. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production, and social factors*, 211–234. Amsterdam / Philadelphia: John Benjamins.

Camino, Iñaki. 2000. Abiaburu bat Donostiako euskera aztertzeko. In Koldo Zuazo (ed.), *Dialektologia gaiak*, 53–84. Vitoria-Gasteiz: Arabako Foru Aldundia.

Coupé, Christophe. 2018. Modeling linguistic variables with regression models: Addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in Psychology* 9(513). https://doi.org/10.3389/fpsyg.2018.00513.

De Smet, Isabeau & Freek Van de Velde. 2020. A corpus-based quantitative analysis of twelve centuries of preterite and past participle morphology in Dutch. *Language Variation and Change* 32(2). 1–25. https://doi.org/10.1017/S0954394520000101.

Decamp, David. 1972. Hypercorrection and rule generalization. *Language in Society* 1(1). 87–90.

Dücker, Lisa, Stefan Hartmann & Renata Szczepaniak. 2020. The emergence of sentence-internal capitalisation in Early New High German: Towards a multifactorial quantitative account. In Marco Condorelli (ed.), *Advances in historical orthography*, c. 1500–1800, 67–92. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108674171.004.

Echenique, Maria Teresa. 2013. La lengua vasca en la historia lingüística hispánica. In Rafael Cano (ed.), *Historia de la lengua española*, 59–82. Madrid: Ariel.

Egurtzegi, Ander. 2013. Phonetics and phonology. In Mikel Martínez-Areta (ed.), *Basque and proto-Basque. Language-internal and typological approaches to linguistic reconstruction*, 119–173. Frankfurt am Main: Peter Lang.

Egurtzegi, Ander & Christopher Carignan. 2020. An acoustic description of Mixean Basque. *The Journal of the Acoustical Society of America* 147(4). 2791–2802. https://doi.org/10.1121/10.0000996.

Egurtzegi, Ander, Dorota Krajewska, Iñigo Urrestarazu & Christopher Carignan. 2022. An acoustic description of sibilants in Mixean Basque. Manuscript.

Euskara Institutua. 2011. Goenkale Corpusa. Bilbao: UPV/EHU. https://www.ehu.eus/ehg/goenkale/.

Fox, John & Sanford Weisberg. 2019. *An R companion to applied regression*, 3rd edn. Thousand Oaks: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html.

Gordon, Matthew J. 2015. Exploring chain shifts, mergers, and near-mergers as changes in progress. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 173–190. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.009.

Grafmiller, Jason. 2019. JGmermod: custom functions for mixed-effects regression models. https://rdrr.io/github/jasongraf1/JGmermod/.

Gries, Stefan Th. 2010. Useful statistics for corpus linguistics. In Aquilinio Sánchez Pérez & Moisés Almela Sánchez (eds.), *A mosaic of corpus linguistics: Selected approaches*, 269–291. Frankfurt am Main: Peter Lang.

Gries, Stefan Th. 2013. *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter.

Gries, Stefan Th. 2015a. Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics* 16(1). 93–117. https://doi.org/10.1177/1606822X14556606.

Gries, Stefan Th. 2015b. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125. https://doi.org/10.3366/cor.2015.0068.

Gries, Stefan Th. & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: a multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics* 14(3). 293–320. https://doi.org/10.1017/S1360674310000092.

Harrington, Jonathan. 2012. The coarticulatory basis of diachronic high back vowel fronting. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production and social factors*, 103–122. Amsterdam / Philadelphia: John Benjamins.

Hartig, Florian. 2022. DHARMa: Residual diagnostics for HierARchical regression models. Software package. http://florianhartig.github.io/DHARMa/.

Herold, Ruth. 1990. *Mechanisms of merger: The implementation and distribution of the low back merger in eastern Pennsylvania*. Philadelphia: University of Pennsylvania dissertation. https://repository.upenn.edu/dissertations/AAI9026574.

Hualde, José Ignacio. 2003. Segmental phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 15–64. Berlin: De Gruyter Mouton.

Hualde, José Ignacio. 2010. Neutralización de sibilantes vascas y seseo en castellano. *Oihenart: Cuadernos de Lengua y Literatura* 25. 89–116.

Igartua, Iván. & Xabier Zabaltza. 2012. *Euskararen historia laburra* [A brief history of the Basque language]. Donostia: Etxepare Euskal Institutua.

Iribar-Ibabe, Alexander, Rosa Miren Pagola Petrirena & Itziar Túrrez Aguirrezábal. 2020. Caracterización articulatoria de las fricativas sibilantes alveolares del euskera: Primera parte. *Loquens* 7(2). e073. https://doi.org/10.3989/loquens.2020.073.

Kang, Yoonjung, Tae-Jin Yoon & Sungwoo Han. 2015. Frequency effects on the vowel length contrast merger in Seoul Korean. *Laboratory Phonology* 6. 469–503. https://doi.org/10.1515/lp-2015-0014.

Kümmel, Martin. 2007. *Konsonantenwandel bausteine zu einer typologie des lautwandels und ihre konsequenzen für die vergleichende rekonstruktion*. Wiesbaden: Reichert Verlag.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell Scientific Publications.

Lakarra, Joseba A. 1995. Reconstructing the Pre-proto-Basque root. In José Ignacio Hualde, Joseba A. Lakarra & Robert L. Trask (eds.), *Towards a history of the Basque language*, 189–206. Amsterdam: John Benjamins.

Lakarra, Joseba A. 2013. Root structure and the reconstruction of Proto-Basque. In Mikel Martínez-Areta (ed.), *Basque and proto-Basque. Language-internal and typological approaches to linguistic reconstruction*, 173–221. Frankfurt am Main: Peter Lang.

Lakarra, Joseba A., Julen Manterola & Iñaki Segurola. 2019. *Euskal hiztegi historiko-etimologikoa* (EHHE-200). Bilbao: Euskaltzaindia.

Lüdecke, Daniel. 2018. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 3(26). 772. https://doi.org/10.21105/joss.00772.

Lüdecke, Daniel. 2020. sjPlot: Data visualization for statistics in social science. Software package. https://CRAN.R-project.org/package=sjPlot.

Martínez Alcaide, María José. 2010. *La fijación ortográfica del español: Norma y argumento historiográfico*. Bern: Peter Lang.

Martínez-Celdrán, Eugenio, Ana María Fernández-Planas & Josefina Carrera-Sabaté. 2003. Castilian Spanish. *Journal of the International Phonetic Association* 33(2). 255–259. https://doi.org/10.1017/S0025100303001373.

Michelena, Luis. 1990. *Fonética histórica vasca*, 2nd edn. San Sebastián: Diputación Foral de Gipuzkoa.

Minkova, Donka. 2015. Establishing phonemic contrast in written sources. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 72–85. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.024.

Mounole, Céline & Joseba A. Lakarra. 2018. Euskara Arkaikoa. In Joakin Gorrotxategi, Iván Igartua & Joseba A. Lakarra (eds.), *Euskararen historia*, 345–468. Vitoria-Gasteiz: Eusko Jaurlaritza.

Mujika, José Antonio. 2002. *Euskal ortografiaren hastapenak iparraldeko literaturan*. Bilbao: UPV/EHU.

Muxika-Loitzate, Oihane. 2017. Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano. *Languages* 2(4). 25. https://doi.org/10.3390/languages2040025.

Nakagawa, Shinichi, Paul C. D. Johnson & Holger Schielzeth. 2017. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface* 14(134). https://doi.org/10.1098/rsif.2017.0213.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2). 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Oyosa, Anabel Eugenia. 2015. Consideraciones sobre grafemática en el ámbito hispánico: en torno a los usos gráficos en la documentación medieval de las lenguas romances peninsulares. *Anuario de Letras. Lingüística y Filología* 1(1). 273–347.

Patil, Indrajeet. 2018. ggstatsplot: "ggplot2" based plots with statistical details. Software package. https://doi.org/10.5281/zenodo.2074621.

Phillips, Betty S. 2015. Lexical diffusion in historical phonology. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 359–373. Oxford: Oxford University Press.

R Core Team. 2021. R: A language and environment for statistical computing. Software package. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rutkowska, Hanna & Paul Rössler. 2012. Orthographic variables. In Juan Manuel Hernández Campoy & Juan Camilo Conde Silvestre (eds.), *The handbook of historical sociolinguistics*, 213–236. Oxford: Blackwell.

Sagarzazu, Txomin. 2005. *Hondarribiko eta Irungo euskara*. Irun: Alberdania.

Shapiro, Michael. 1995. A case of distant assimilation: /str/ → /ʃtr/. *American Speech* 70(1). 101–107. https://doi.org/10.2307/455876.

Silk, Matthew J., Xavier A. Harrison & David J. Hodgson. 2020. Perils and pitfalls of mixed-effects regression models in biology. *PeerJ* 8. e9522. https://doi.org/10.7717/peerj.9522.

Solé, Maria-Josep. 2010. Effects of syllable position on sound change: An aerodynamic study of final fricative weakening. *Journal of Phonetics* 38(2). 289–305. https://doi.org/10.1016/j.wocn.2010.02.001.

Stenroos, Merja. 2002. Free variation and other myths: Interpreting historical English spelling. *Studia Anglica Posnaniensia: International Review of English Studies* 38. 445–469.

Stevens, Mary, Véronique Bukmaier & Jonathan Harrington. 2015. Pre-consonantal /s/ retraction. Presented at the International Congress of Phonetic Sciences, Glasgow, University of Glasgow. https://www.phonetik.uni-muenchen.de/~jmh/papers/sretractionicphs15.pdf.

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–134. https://doi.org/10.1016/j.wocn.2016.08.003.

Trask, Robert Lawrence. 1997. *The history of Basque*. London: Routledge.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag New York. https://ggplot2.tidyverse.org.

Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York: Routledge.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419. https://doi.org/10.1075/dia.30.3.04wol.

Zampaulo, Andre. 2016. Sibilant sound change in the history of Portuguese: An information-theoretic approach. *Diachronica* 33(4). 503–529. https://doi.org/10.1075/dia.33.4.03zam.

Zuloaga, Eneko. 2020. *Mendebaleko euskararen azterketa dialektologiko-diakronikorantz*. Bilbao: UPV/EHU dissertation. https://addi.ehu.es/handle/10810/49846.

Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer New York.

Dorota Krajewska
Department of Linguistics and Basque Studies
University of the Basque Country UPV/EHU
Unibertsitateko Ibilbidea, 5
01006 Vitoria-Gasteiz, Spain
dorota.krajewska@ehu.eus

Eneko Zuloaga
Department of Linguistics and Basque Studies
University of the Basque Country UPV/EHU
Unibertsitateko Ibilbidea, 5
01006 Vitoria-Gasteiz, Spain
eneko.zuloaga@ehu.eus

Ander Egurtzegi
CNRS IKER (UMR5478)
Campus de la Nive, Château-Neuf
15, Place Paul Bert
64100 Bayonne, France
ander.egurtzegi@iker.cnrs.fr