



IKASTORRATZA, e-Revista de Didáctica, es una revista en formato digital que publica artículos relacionados con los procesos de enseñanza y aprendizaje, a través de Internet y bajo la licencia Creative Commons.

IKASTORRATZA, e-Revista de Didáctica, es una publicación seriada, gratuita y libre de ser impresa que cada seis meses divulga artículos científicos, propuestas didácticas y artículos de opinión sobre cuestiones relativas al mundo de la didáctica.

IKASTORRATZA, e-Revista de Didáctica, asume como objetivo principal la difusión del conocimiento pedagógico y de metodologías didácticas que favorezca la expansión de prácticas de educativas efectivas.

IKASTORRATZA, e-Revista de Didáctica, es una revista bilingüe, abierta a propuestas de autores y autoras que deseen publicar trabajos inéditos tanto en euskara como en castellano.

# IKASTORRATZA. Didaktikarako e-aldizkaria

*IKASTORRATZA. e-journal on Didactics*

## IKASTORRATZA. e-Revista de Didáctica

ISSN: 1988-5911 (Online) Journal homepage: <http://www.ehu.es/ikastorratza/>

---

## Populazioaren sailkapena lexikoaren arabera

Iñaki Gaminde<sup>1</sup>, Ander Olalde<sup>2</sup>, Ursua Gaminde<sup>3</sup> y Aintzane Etxebarria<sup>4</sup>

<sup>1</sup>[inaki.gaminde@gmail.com](mailto:inaki.gaminde@gmail.com)

<sup>2</sup>[ander.olalde@ehu.es](mailto:ander.olalde@ehu.es)

<sup>3</sup>[ursua.gaminde@gmail.com](mailto:ursua.gaminde@gmail.com)

<sup>4</sup>[aintzane.etxebarria@ehu.es](mailto:aintzane.etxebarria@ehu.es)

To cite this article:

Gaminde, I., Olalde, A., Gaminde, U. & Etxebarria, A. (2016). Populazioaren sailkapena lexikoaren arabera. *IKASTORRATZA. e-Revista de Didáctica*, 16, 68-84. Retrieved from [http://www.ehu.es/ikastorratza/16\\_alea/4.pdf](http://www.ehu.es/ikastorratza/16_alea/4.pdf)

To link to this article:

[http://www.ehu.es/ikastorratza/16\\_alea/4.pdf](http://www.ehu.es/ikastorratza/16_alea/4.pdf)

Published online: 20 Jun 2016.

# Populazioaren sailkapena lexikoaren arabera

Iñaki Gaminde<sup>1</sup>, Ander Olalde<sup>2</sup>, Ursua Gaminde<sup>3</sup> y Aintzane Etxebarria<sup>4</sup>

<sup>1</sup>Iñaki Gaminde

inaki.gaminde@gmail.com

<sup>2</sup>Ander Olalde

ander.olalde@ehu.es

<sup>3</sup>Ursua Gaminde

ursua.gaminde@gmail.com

<sup>4</sup>Aintzane Etxebarria

aintzane.etxebarria@ehu.es

## Laburpena

Hizkuntza sistema konplexua, dinamikoa eta moldagarria da, hau da, osagaiak eta euren arteko erlazioak denboran zehar aldatu ez ezik, interakzioen ondorioz moldatu ere egiten dira, gainera ez dute pisu bera izaten, batzuk beste batzuk baino gehiago erabiltzen baitira. Teoria horretan oinarrituta, lan honetan Larrabetzuko hamabi gazte euskaldunen lexikoaren erabilgarritasun maila, kontzeptu bera adierazteko erabiltzen duten lemen pisua eta lema bakoitzaren aldaki fonetikoaren pisua kontuan hartu dituen sistema aurkeztu da. Hiru tasun multzo horiek doitzeko modu bat ere proposatu da lan honetan. Horretarako, Leveinshtein distantziaren bidez lemak elkarrengandik zenbatean aldentzen diren kalkulatu da, eta ondoren, aldaki bakoitzaren pisua kalkulatu da. Azkenik, mutilen lexikoaren bilakaera neskena baino arinago gertatzen dela ondorioztatu da.

## Resumen

La lengua es un sistema adaptativo, complejo, y dinámico, es decir, los componentes y sus relaciones cambian en el tiempo y se adaptan por las distintas interacciones, además estos componentes no tienen el mismo peso dentro del sistema lingüístico, ya que algunos se usan más que otros. Basándonos en esta teoría, presentamos un sistema en el que se tienen en cuenta tres aspectos de la lengua que son el nivel de disponibilidad léxica, el peso de los lemas que hacen referencia a un mismo concepto y el peso de las variantes fonéticas de cada lema. Para adecuar los tres grupos mencionados se han

llevado a cabo dos cálculos, primero, la distancia de Levenshtein para saber en qué medida se alejan los lemas, y posteriormente, el peso de cada variante. Por último se concluye que la evolución léxica es más rápida en el grupo del género masculino que en el del femenino.

### **Abstract**

Language is an adaptive, complex, and dynamic system, that is, the components and their relationships change over time and they adapt to different interactions, in addition, all of these components do not hold the same importance in the language system inasmuch as some of these are used more than others. Based on complexity adaptive systems theory we present a system and it takes into account three aspects of language: lexical availability, the weight of lemmas that refer to a concept and the weight of phonetic variants of a lemma. To adjust three mentioned data sets we made two calculations: Levenshtein distance between lemmas and the weight of each variant. Finally, this study concludes that lexical development in the boy group is faster than in the girl group.

**Hitz gakoak:** Sistema konplexua, lexikoa, lema, aldaki fonetikoak, doitu.

**Palabras clave:** Sistema complejo, léxico, lema, variante fonética, adecuar.

**Key words:** Complex system, lexicon, lemma, phonetic variant, to adjust.

## 1. Sarrera

Hiztun komunitate baten lexikoa aztertzen den kasu askotan irudi luke jasotzen den lexikoa homogeneous dela komunitate horretan eta hiztun guztiek berdin erabiltzen dituztela jasotzen diren molde guztiak.

Erazko lexikoaren gainean egin diren azterketa askotan erakutsi den bezala (Etxebarria, 1996; Gomez, 2004; López Morales, 1996; Paredes, 2012; Rodríguez eta Muñoz, 2009; Serrano, 2004; besteak beste), hitz berberak jakin arren, hiztun guztiok denak ez ditugu berdin erabilgarri. Bestalde, ezin ahatz daiteke lexikoa kanpo faktore askoren menpe dagoena, hala nola faktore kulturalak, bizimoduaren aldaketa arina, euskara batuaren eragina edo inguruko erdaren presioa gure kasuan, e.a.

Kontzeptu edo ideia bera adierazteko hitz bat baino gehiago ematen direnean ere, pentsa daiteke sinonimotzat edo jotzen diren horiek maila berean daudela. Argi dago hori ez dela horrela gertatzen, baina nekez aurki daitezke hitzen doiketak egiten diren lanak.

Azkenik, hitzen aldaki fonetikoekin antzera gertatzen da; hitz baten aldaki bat baino gehiago erabiltzen direnean, ez dago argi zein den bakoitzaren pisua erabilera errealean, edo hiztun guztiek aldaki guztiak darabiltzaten edo bakarren bat baino ez.

Lexikoari eragiten dioten kanpo faktoreak gure lan honetatik kanpo utzita, hemen aurkeztu nahi dugu sistema bat aipatu berri ditugun beste hiru barne faktoreak kontuan hartzen dituenena. Hau da, lexikoaren erabilgarritasun maila, kontzeptu berbera adierazteko erabiltzen diren lehen pisua eta lema bakoitzaren aldaki fonetikoaren pisua kontuan hartzen dituen sistema aurkeztuko dugu eta Larrabetzuko gazte euskaldunen lexikoaren atal baten sarrerei aplikatuko diegu.

Bestalde, aipatu berri ditugun hiru tasun multzo hauek, erabilgarritasuna, lema eta aldaki fonetikoak, ezin eduki dezakete pisu bera populazioaren egituraketa aztertzean, horregatik doitzeko modu bat ere proposatzen dugu.

Kontuan hartzen badugu ez dagoela hizkuntzarik hiztun barik, oso garrantzitsua da hizkuntza aztertzean hiztunen erabilera erreala. Sistema konplexuen teoriaren arabera, hizkuntza sistema konplexua da, hau da, era ezberdinetako elementuek osatzen dute, eta modu ezberdinetara elkarri eragiten diete. Gainera sistema konplexu hori dinamikoa ere bada, alegia, hizkuntza osatzen duten osagaiak denboran zehar aldatzen dira, baina baita beraien arteko erlazioak ere, moldatu egiten baitira elkarreraginetan. Hizkuntza formak interakzioaren epifenomenoak dira, eta azaleratzen diren estabilitateak dira sistema

dinamiko baten barnean. Azaleratzen diren forma horiek hizkuntza komunitateko kideen adaptazioak dira hizkuntza elkarreraginak daudenean, eta batzuk beste batzuk baino pribilegiatuagoak izaten dira, gehiago erabiltzen direlako, eremu jakin bateko dialektoan erabiltzen direlako edota funtzio espezializatua dutelako. (Cameron eta Larsen Freeman, 2007)

Adierazitako sistema konplexuen teoriaren ildotik, hizkuntzaren irakaskuntzarako irizpideak zehazteaz gain, hizkuntza sistema konplexu gisa ere azter daiteke (Moreno Cabrera, 2008; Léonard, 2015) hizkuntza formen adaptazio-arrazoi ezberdinak (geografia, adina, jabekuntzaren garapen-maila, besteak beste) kontuan izanik. Horretarako, datu multzotik abiatuta euron modelizazioa egiten da eta algoritmo batzuen bidez azpiko egiturak azaleratzen dira (Leonard, 2015).

Horrenbestez, gure lan honen helburu nagusia erraz labur daiteke: hiztun errealen lexikoaren gaineko datu multzo batetik abiatuta, datuen modelizazio doitua egitea proposatzen dugu, ondorengo tratamendu matematikoaren bidez, populazioaren sailkapenean egon daitezkeen egiturak eta joerak azaleratzeko.

Lana lau atal nagusitan banatuta aurkezten dugu, sarrera honen osteko bigarren atalean datuak jasotzeko eta prestatzeko erabili den metodologiaz jardungo dugu, hirugarren atalean datuen azterketa egingo da eta, azkenik, laugarren atalean, ondorioak eta hurrengo lanetarako iradokizun batzuk aurkeztuko dira.

## **2. Metodologia**

Metodologiari dagokion atal hau azpi-atal bitan banatuta aurkezten dugu. Lehen azpi-atalean datu jasotzeaz jardungo dugu eta bigarreanean datu prestatetaz.

### ***2.1. Datu jasotzea***

Lexikoaren gaineko datuak jasotzeko, gure aurreko lan baten ildotik (Gaminde eta lank. 2012), giza gorputzaren arlo semantikoa aukeratu dugu. Gai hau aukeratu dugu teknikoa ez delako eta nahikoa egonkorra izan daitekeelako belaunaldien artean. Beste gai batzuk bizimoduarekin lotuago egon daitezke eta bizimoduaren aldaketak gehiago eragin diezaieke. Denetara 40 sarrera aukeratu dira (1. taula); gehienak oso arruntak eta bakar batzuk ez hain arruntak (“begitxindor”, “eskumutur”, “azkordin”, “zotin” eta “zil”, esaterako).

**Taula 1.** Sistema aplikatzeko aukeratu diren 40 hitz.

buru	belarri	sudur	kokots	bekoki
hagin	azkordin	txistu	sorbalda	lepo
sama	beso	atzamar	eskutur	hanka
orpo	zain	giltzurrun	azal	hazkura
begitxindor	doministiku	aurpegi	begi	aho
ile	ezpain	mihi	zotin	zil
gerri	ukondo	ukabila	atzazal	belaun
orkatila	gibel	hezur	gari	aharrausi

Datuak lortzeko 12 informatzaile izan ditugu, 6 neska eta 6 mutil. Denak Larrabetzuko euskaldunak dira, euron ama hizkuntza bertako barietatea izan da eta D ereduan egin dituzte euren oinarrizko ikasketak. Adinari dagokionez 1977-1994 urte artean jaiokak dira eta adin talde bi egin daitezke: (1) 1977-1985 urte bitartean jaiokak eta (2) 1986-1994 urte bitartekoak. Neskei “n” ezarriko diegu eta mutilei “m”<sup>1</sup>; 2. taulan eduki ditugun informatzaileak ematen ditugu adin tarte eta generoen arabera sailkatuta<sup>1</sup>.

**Taula 2.** Aukeratu diren informatzaileak euren adin tarte eta generoen arabera.

	Adin tarte (1)		Adin tarte (2)	
	Lekukoa	Urtea	Lekukoa	Urtea
Neskak	N1	1977	N4	1988
	N2	1978	N5	1991
	N3	1985	N6	1994
Mutilak	M1	1977	M4	1988
	M2	1979	M5	1992
	M3	1985	M6	1994

Inkesta egin zenean informatzaileei eskatzen zitzairen itzultzeko berbak gaztelaniatik; kasu batzuetan azaldu egiten zitzairen euskaraz eskatzen zen berba zeri zegokion (“begitxindor” edo “azkordin” berben kasuetan, adibidez). Inkesta osoa grabatuta jaso

<sup>1</sup> Informatzailearen identifikatzailea ipini dugu “Lek” eta jaiotze urtea “urtea”.

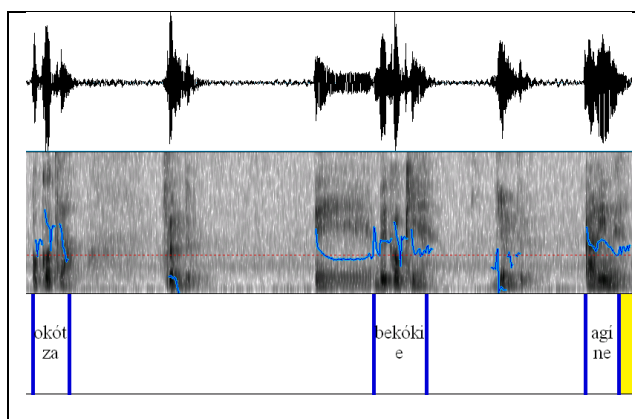
zen “MarantzPMD610” izeneko grabagailu baten bitartez. Informatzaileak dudan egon diren kasuetan lehendik jasota eduki ditugun berba batzuk proposatu zaizkie eta horien inguruko informazioa ere jaso da.

## 2.2. *Datu prestaketa*

Aztertu nahi izan den tasun multzo bakoitzeko datu prestaketa berezitua egin behar izan denez gero, aldiak banan-banan aztertuko ditugu.

Grabaketak egin bezain laster “praat” (Boersma eta Weenink, 2015) izeneko programaren bidez transkripzioak eta etiketazioak egin ziren 1. irudian erakusten den bezala, datu basea osatu ahal izateko.

**Irudia 1.** “Praat” programaren bidez egindako etiketazio adibidea.



Hitzaren transkripzioarekin batera erabileraren datuak ere jaso dira. Informatzaileek beti ez dute itzulpena zuzenean ematen, zalantzan geratzen dira sarritan; honelako kasuetan herrian bertan jasotako erantzunen bat proposatu zaie, proposatutako hitzen gainean jasotako erantzunak era bitara sailkatu dira; zuzenean onartu dutenean bi puntu eman zaizkio eta ezaguntzat edo entzundakotzat jo dutenean puntu bat eman zaio. Kasu batzuetan, lagunduta ere, ez dute erantzuten jakin eta orduan sarrera horri 0 eman zaio. Era honetara, hitzen erabilera Otik, hau da ezagutza eza, 3ra bitarteko eskala baten kokatzen da. Denetara 480 hitz jaso dira.

Transkripzioak lortu ondoren hitzen lematizazioa egin da horretarako hitz bakoitzari lema bat esleitu zaio dauzkan aldaki fonetikoak kontuan hartu barik. “giltzurrun” izan da erabili den lema “giltzurrun”, “gintzurrun”, “gultzurrun”, “guntzurrun” eta “kuntzurrun” aldakiak ordezkatzeko. Beste adibide bat “belaun” izan daiteke, lema honen bidez “belaun”, “beleun” eta “belun” aldakiak ordezkatu dira. Denetara 47 lema

batu ditugu, izan ere, galdera batzuek lema bat baino gehiago izan dute; adibidez “lepo” eta “bizkar” sinonimotzat eman dira, edo “begitxindor” eta “txindor”. Behin lematizazio prozesua eginez gero, lema bakoitzaren pisua kalkulatu da.

Lemen pisua kalkulatzeko lema bakoitza zenbat informatzailek erabili duten zenbatu da eta kopuru hori zati informatzaile kopurua egin da. Lema bakoitzari aldaki fonetikoei baino pisu gehiago emateko, emaitza bider 2 egin da. Beraz, aplikatu den kalkulua lemaren agerpen kopurua (LAK) zati 12 bider 2 izan da:

$$(LAK/12) \times 2$$

Aldaki fonetikoaren pisua kalkulatzeko prozesu bi erabili dira. Hasieran Leveinshtein distantziaren bidez (Vals eta lank., 2012) lema bakoitza besteetatik zenbatean aldentzen den kalkulatu da. Distantzia honekin lema bien artean dauden desberdintasun kopuruak zenbatzen dira; adibidez, “ago” eta “ao” aldakien artean batetik bestera pasatzeko aldaketa bat egin behar da:

1. Barietatea:	a	g	o
2. Barietatea	a		o
Orotara	1		

Lema batek aldaki bi baino gehiago daukanean, distantziak banan-banan kalkulatu dira (3. taula)

**3. Taula.** Sudur lemaren aldaki fonetikoaren arteko distantziak.

	Sudur	Suur	Sur	Orotara
Sudur	0	1	2	3
Suur	1	0	1	2
Sur	2	1	0	3



Aldaki bakoitzaren agerpen kopuruen arabera, distantzia guztien batuketa egiten da. Kasu honetan “sudur”, “suur” eta “sur” launa aldiz agertzen direnez gero, “sudur” eta “sur” aldakien pisua 12koa izango da eta “suur” aldakiarena 8koa. Distantzia hauen guztien batura 128koa da. Hurrengo pausua aldaki bakoitzaren pisua zati 128 egitea da eta lortzen ditugu 4. taulako emaitzak.

**4. Taula.** Aldaki fonetiko bakoitzaren pisua.

<b>Aldakia</b>	<b>Indizea</b>
Surre	0,09375
Surre	0,09375
Sudurre	0,09375
Suurre	0,0625
Sudurre	0,09375
Suurre	0,0625
Suurre	0,09375
Sudurre	0,09375
Surre	0,09375
Sudurre	0,09375
Suurre	0,0625
Suurre	0,0625

Aldaki fonetikoei lemei eta erabilgarritasunari garrantzi apalagoa ematekotan, doitzean puntu bakarra emango zaio aldaki bera informatzaile guztiek erabiltzen dutenean. Orduan, 12 informatzaileen datuekin lortu ditugun indizeak proportzionatzeko puntu batekin hiruko erregela bat erabili da. Azkenik, lortu ditugun 69 aldaki fonetikoak 0 eta 1 bitarteko indizeekin agertzen dira. Informatzaile guztiek aldaki fonetiko bakarra darabiltenean orduan kasu guztietan 1 izango den moduan, informatzailearen batek aldakirik ez badu erabili 0 izango da bere puntuazioa.

Lexikoaren erabileraz hartu ditugun hiru mailatako informazioak banaka kalkulatzeko dira eta gero denen arteko indizea kalkulatzeko lortu diren emaitza doituena batuketa egiten da. Datu basearen sarrera bat lehen kolpean itzuli bada (3 puntu), lema bakarra badauka (puntu bi) multzo osoan eta aldaki fonetiko bakarrarekin erabili bada (puntu bat), orduan berorren pisua 6koa izango da.

### **3. Datuen azterketa**

Atal honetan aukeratu ditugun hitzen tasun multzoen azterketarekin batera, populazioaren sailkapena egingo dugu tasun multzo horien arabera eta informatzaileen adina eta generoa kontuan hartuta (aldagai bi hauek orain arte egin diren lan askotan gehien erabili izan direnak dira; Ariztimuño, 2010; Aurrekoetxea, 2008, Ezenarro, 2008, Gaminde eta Romero, 2011, Gaminde eta lank., 2012, Ormaetxea, 2008, Unamuno, 2010, Unamuno eta lank., 2012; Zubillaga eta Gaminde, 2010).. Azterketa argiago azaltzeko atala lau azpi-ataletan banatuta aurkeztzen dugu; lehen azpi-atalean hitzen erabilgarritasunaren datuen azterketa emango dugu, bigarrenean hitzen lemei dagokien azterketa agertuko da, hirugarrenean aldaki fonetikoaren azterketa ikusiko da eta laugarrenean, azkenik, tasun guztien datu doituen azterketa egingo da.

#### ***3.1. Erabilgarritasunaren datuen azterketa***

Sarreren azterketa egiten badugu, ikusten dugu 40 sarreretatik 21 (%52,5) puntuazio maximoarekin (36 puntu) erabili dituztela informatzaile guztiek: “aho”, “beso”, “buru”, “aurpegi”, “ukondo”, “atzamar”, “hagin”, “lepo”, “doministiku”, “hezur”, “ezpain”, “mihi”, “eskutur”, “sudur”, “begi”, “belarri”, “ile”, “hanka”, “azala”, “belaun” eta “atzazal”.

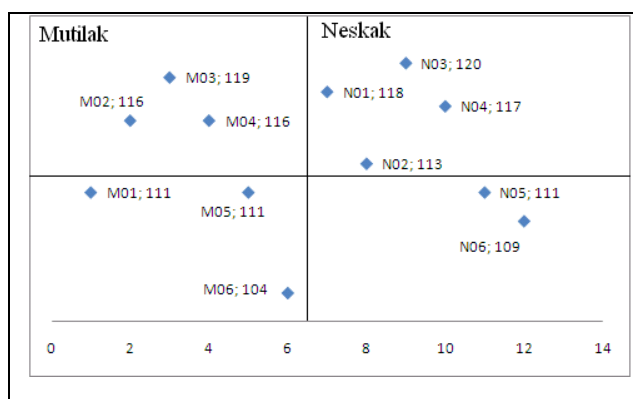
Puntuaziorik baxuena eduki duen sarrera “azkordin” izan da (13 puntu), gainerako sarrerak 20 eta 35 artean kokatzen dira: (35) “kokots”, “aharrausi”, “gerri”, “sama”, “bekoki”, “gibel”, “zotin”, “hazkura”, “giltzurrun” eta “txistu”; (34) “sorbalda”; (33) “zil”; (32) “orkatila”; (31) “ukabil” eta “zain”; (30) “orpo”; (29) “begitxindor”; (26) “gari”.

Informatzaileek lortu dituzten puntuazioak aztertzen baditugu, ikus dezakegu gehien eduki duenak 120 puntu eduki dituela (N3), hau da maximoa eta gutxien eduki duenak 104 (M6). Informatzaile guztien puntuazioen batezbestekoa 113,75koa da (des. 4,731). Puntuazioak informatzaileen adin taldeen arabera aztertzen baditugu, (1) adin taldeak 116,17ko batezbestekoa dauka (des: 3,545) eta (2) adin taldeak 111,33 (des: 4,761). Alde hauek joera bat adierazten dute, izan ere, ez dira estatistikoki esanguratsuak.

Emaitzak informatzaileen generoaren arabera aztertuta, ikusten dugu mutilen batezbestekoa 112,83koa dela (des: 5,344) eta neskena 114,67koa (des: 4,320). Aldea ez da estatistikoki esanguratsua.

Adinaren eta generoaren arabera alde txiki hauek 2. irudiko dispersio grafikoan islatzen dira.

**2. Irudia.** Hitz erabilgarritasun indizeen arabera dispersioa.



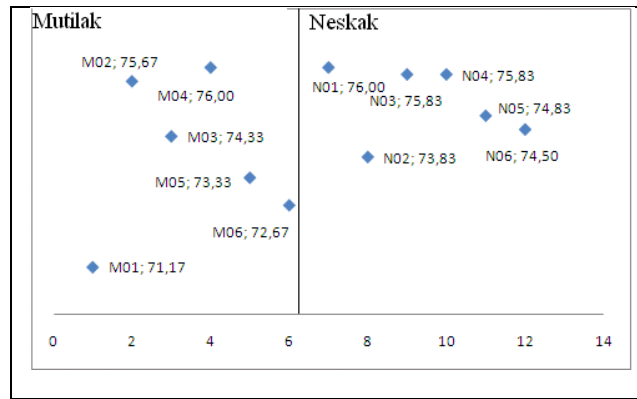
### 3.2. Lemen azterketa

Aukeratu ditugun 40 sarrera gauzatzeko 47 lema erabili dira. 35 sarrerak lema bakarra daukate (%87,5); beste bostek (%12,5) 12 lema biltzen dituzte; sarrera horiek honako hauek dira: lepo: “lepo” eta bizkar”; zotin: “zotin” eta “txokin”; eskutur: “eskubitor” eta “eskumutur”; zil: “zil”, “tripako zil”, “tirrin” eta “zilbot” eta begitxindor: “begitxindor” eta “txindor”.

Informatzaileek lortu dituzten puntuazioak aztertzen baditugu, ikus dezakegu gehien eduki dutenek 76 puntu eduki dituztela (M4 eta N1) eta gutxien eduki duenak 71,17 (M1). Informatzaile guztien puntuazioen batezbestekoa 74,5koa da (des. 1,532). Puntuazioak informatzaileen adin taldeen arabera aztertzen baditugu, (1) adin taldeak 74,47ko batezbestekoa dauka (des: 1,842) eta (2) adin taldeak 74,53 (des: 1,331). Alde hau oso txikia da eta ez da estatistikoki esanguratsua.

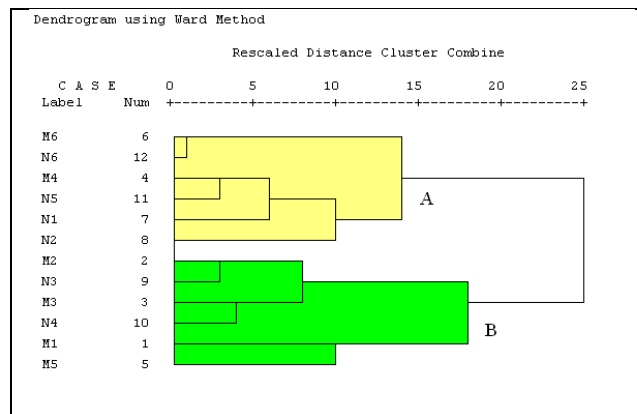
Emaitzak informatzaileen generoaren arabera aztertuta, ikusten dugu mutilen batezbestekoa 73,86koa dela (des: 1,845) eta neskena 75,14koa (des: 0,885). Aldea ez da estatistikoki esanguratsua. 3. irudiko dispersio grafikoan ikusten den bezala, nesken puntuazioak batuago agertzen dira mutilenak baino.

**3. Irudia.** Lemen indizeen arabera dispersioa.



Informatzaileen lemen gaineko emaitzen bidez multzokatze azterketa egiten badugu 4. irudiko dendograma lortzen dugu, bertan ikusten den moduan, A taldean mutil bi batzen dira (M4 eta M6) (2) adin taldekoak eta lau neska (N1, N2, N5 eta N6), neskak (1) eta (2) adin taldekoak dira; B taldean lau mutil batzen dira (M1, M2, M3 eta M5), lehen hirurak (1) adin taldekoak eta azkena (2) adin taldekoa, eta neska bi (N3 eta N4) bakoitza adin talde batekoa.

#### 4. Irudia. Informatzaileen multzokatzea lemen arabera.



### 3.3. Aldaketa fonetikoaren azterketa

Aukeratu ditugun 47 lema gauzatzeko agertu zaizkigun aldaki fonetikoak 69 izan dira. 34 lemak aldaki fonetiko bakarra daukate (%72,34). Gainerako hamahiruek (%27,66) 35 aldaki batzen dituzte. Lema horien aldaki fonetikoak 5. taulan agertzen direnak dira:

**5. Taula.** Corpusean batu diren lemen aldaki fonetikoak.

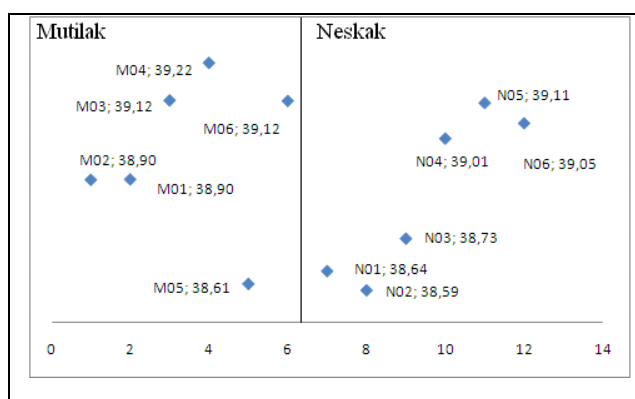
<b>Lema</b>	<b>Aldaki fonetikoak</b>
судар	“судар”, “suur” eta “sur”
kokots	“kokots” eta “okots”
giltzurrun	“giltzurrun”, “gintzurrun”, “gultzurrun”, “guntzurrun” eta “kuntzurrun”
doministiku	“doministiku” eta “doministikun”
aurpegi	“aurpegi”, “arpegi”, “arpei” eta “arpi”
aho	“ago” eta “ao”
zotin	“zotin” eta “zoten”
gerri	“gerri” eta “garri”
ukabila	“ukabiye”, “ukabile”, “ukubiye” eta “ukubile”
atzazala	“atzazala” eta “azatzala”
belaun	“belaun”, “beleun” eta “belun”
orkatila	“orkatille” eta “orkatiye”
aharrausi	“arrausi” eta “arrosi”

Informatzaileek lortu dituzten puntuazioak aztertzen baditugu, ikus dezakegu gehien eduki duenak 39,22 puntu eduki dituela (M4) eta gutxien eduki duenak 38,59 (N2). Informatzaile guztien puntuazioen batezbestekoa 38,92koa da (des. 0,224). Puntuazioak informatzaileen adin taldeen arabera aztertzen baditugu, (1) adin taldeak 38,81ko batezbestekoa dauka (des: 0,196) eta (2) adin taldeak 39,02 (des: 0,214). Alde hau oso txikia da eta ez da estatistikoki esanguratsua.

Emaitzak informatzaileen generoaren arabera aztertuta, ikusten dugu mutilen batezbestekoa 38,98koa dela (des: 0,223) eta neskena 38,86koa (des: 0,228). Aldea ez da estatistikoki esanguratsua.

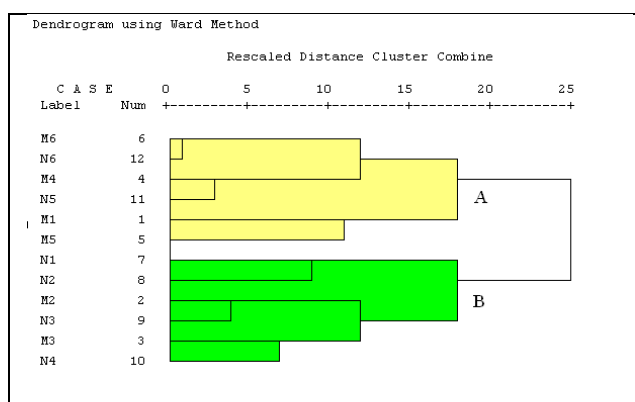
5. irudiko dispertsio grafikoan ikusten den bezala, nesken puntuazioetan eremu argi bi agertzen dira adin taldeen arabera, mutilen kasuan, joeratzat baino ezin har daiteke hau.

**5. Irudia.** Lemen aldaki fonetikoaren indizeen arabera dispersioa.



Informatzaileen aldaki fonetikoaren gaineko emaitzen bidez multzokatze azterketa egiten badugu 6. irudiko dendograma lortzen dugu, bertan ikusten den moduan, A taldean lau mutil batzen dira (M1, M4, M5 eta M6), hiru (2) adin taldekoak eta bat (1) adin taldekoa, eta neska bi (N5 eta N6), biak (2) adin taldekoak. B taldean mutil bi batzen dira (M2 eta M3), biak (1) adin taldekoak, eta lau neska (N1, N2, N3 eta N4), lehen hirurak (1) adin taldekoa eta bestea (2) adin taldekoa.

**6. Irudia.** Informatzaileen multzokatzea lemen aldaki fonetikoaren arabera.



**3.4. Atal guztien azterketa**

Azken azpi-atal honetan informatzaileak sailkatuko ditugu aurreko azpi-ataletan aztertu ditugun puntuazio doituaren arabera.

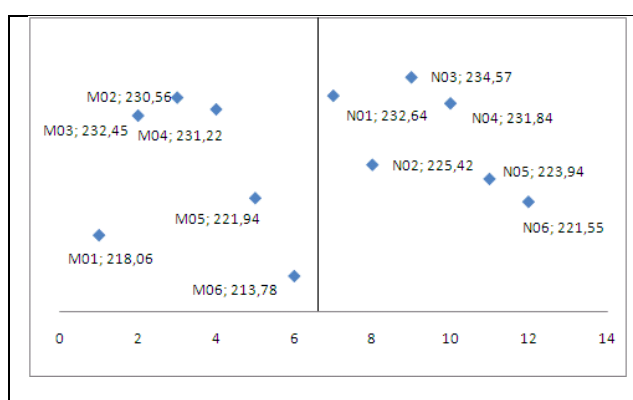
Informatzaileek lortu dituzten puntuazio orokorrak aztertuta, ikus dezakegu gehien eduki duenak 234,57 puntu eduki dituela (N3) eta gutxien eduki duenak 213,78 (M6). Informatzaile guztien puntuazioen batezbestekoa 226,5koa da (des. 6,685). Puntuazioak informatzaileen adin taldeen arabera aztertzen baditugu, (1) adin taldeak 228,95ko

batezbestekoa dauka (des: 6,182) eta (2) adin taldeak 224,05 (des: 6,758). Alde hau txikia da eta ez da estatistikoki esanguratsua.

Emaitzak informatzaileen generoaren arabera aztertuta, ikusten dugu mutilen batezbestekoa 224,67koa dela (des: 7,848) eta neskena 228,33koa (des: 5,358). Aldea ez da estatistikoki esanguratsua.

7. irudiko dispersio grafikoan ikusten den bezala, nesken puntuazioak batuago agertzen dira mutilenak baino; bestalde adina faktore garrantzitsua bada ere, (1) adin taldeko mutil bat (2) adin taldekoekin agertzen da eta nesken kasuan alderantziz.

**7. Irudia.** Indize orokorren araberako dispersioa.



#### 4. Ondorioak eta hurrengo lanak

Lan honetan zehar erakutsi dugun moduan, lexikoaren arabera populazioa sailkatzeko aurkeztu dugun sistemak onura batzuk ekartzen dituela uste dugu. Sarreran esaneko moduan, zaila izan daiteke aukeratzen diren aldagaien arabera populazioaren era bateko sailkapena egitea, baina hemen aurkezten dugun sistemaren bidez populazioaren egituraketan agertzen diren joera nagusiak agirian utzi daitezke.

Hemen erabili dugun Larrabetzuko informatzaile gazteen multzoan adinaren arabeko aldea oso handia ez den arren, aukeratutako adin tartea txikia delako besteak beste, argi agertzen da berau faktore garrantzitsua dena lexikoaren bilakaeran. Hemen aukeratu den 18 urteko adin tartea handiagotuz gero, baliteke emaitzak bestelakoak izatea. Edozelan ere, lexikoaren bilakaera ez dela guztiz lineala ere argi agertu da; badira adin talde bateko informatzaile batzuk ez dagokien taldean agertzen direnak euren bilakaera arinagoa edo geldoagoa izan delako.

Era berean, agirian utzi den bezala, nesken puntuazioen desbiderapenak txikiagoak dira mutilenak baino, honek argi adierazten du talde tinkoagoa osatzen dutela; mutilen lexikoaren bilakaera arinago gertatzen den artean.

Kasu honetan erabili ditugun aldagaiak, adina eta generoa, sozialak izan dira. Beste lan batzuetarako beste aldagai sozial batzuk ere proposa daitezke, hala nola ikasketa maila, informatzaileen ama hizkuntza, maila ekonomikoa, e.a. Aldagai geografikoak ere erabil daitezke, jatorri batzuetako informatzaileak erkatu nahi izanez gero. Era berean, informatzaile berberen erabilerak azter daitezke estiloen arabera. Azkenik, lan honen oinarria den sistema konplexuen teoriaren arabera (Cameron eta Larsen-Freeman, 2007) hizkuntza erabiltzen ikasten da, bariazioak garrantzia du eta hiztunek hautatzen dituzte hizkuntza egiturak. Izan ere, xede hizkuntza konplexua eta dinamikoa izanik, hizkuntza ikasleei lagundu egin behar zaie era ezberdinetara; lehen pausua, helburuaren deskribapena da, eta ekintza hori sistema konplexuen ikuspuntutik egin beharrekoa da. Ikasleen jabekuntza-maila deskribatzeak ere testuinguru ezberdinetara moldatzerako orduan nola egin jakiteko lagun dezake:

“A more practical lesson from the notion of language as a complex adaptive system is my proposal that we should not only be teaching students language, but we should also be teaching them to adapt (Larsen-Freeman, 2013a), i.e., teaching them to take their current language resources and mold them to new situations.”  
([languagemagazine.com](http://languagemagazine.com)).

Ikus daitekeenez, deskribapena ez da nahikoa, eta ikasleek erabili beharreko hizkuntza formak aurkeztu behar zaizkie eta bariazio linguistikoa ulertu egin behar dute, beti ere ikasleen garapen mailaren arabera. Azkenik, ikasleek hizkuntzarekin lan egin behar dute, eta hizkuntzaren adaptazioa landu behar da interakzioaren bidez, jardueren edota atazen baldintzak aldatuta ([languagemagazine.com](http://languagemagazine.com)). Azken ideia hau etorkizunean egin beharreko atazatzat hartzen da, lan honetan lexikoari dagokionez lehen pausua eman baita.

## **Bibliografia**

- Ariztimuño, B. (2010). Tolosako eta Ataungo hizkerak: hizkuntz bariazioa eta konbergentzia-joerak. *Uztaro*, 72, 79-96.
- Aurrekoetxea, G. (2008). Bariazio soziolinguistikoa Dimako euskaran. *Euskalingua*, 12, 17-26.



- Boersma, P. eta D. Weenink, D. (2015). Praat: doing phonetics by computer. <http://www.praat.org/> [15/09/2015]
- Cameron, L. eta Larsen-Freeman, D. (2007). Complex systems and applied linguistics. *International Journal of Applied Linguistics*, 17(2), 226–239.
- Etxebarria, M. (1996). Disponibilidad léxica en escolares del País Vasco. Variación sociolingüística y modelos de enseñanza bilingüe. *Revista Española de Lingüística*, 26, 301-325.
- Ezenarro, A. (2008). Etxebarri eta Bolibarko bariazio linguistikoa. *Uztaro*, 67, 59-84.
- Gaminde, I. eta Romero, A. (2011). Genero eta adina Bermeoko berbaldiaren fonemen eta hotsen frekuentzian. *FLV*, 113, 115-138.
- Gaminde, I.; Romero, A. eta Legarra, H. (2012). *Gramatika eta Hizkuntz Bariazioa Bermeon*. Bermeo: Bermeoko Udala eta Campos Hegaluzea.
- Gaminde, I; Etxebarria, A.; Garay, U. eta Romero, A. (2012) Lehen hizkuntzaren eragina Bizkaiko gazteen lexikoaren erabileran. *Ikastorratza*, 9, 1-16.
- Gómez, M. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Doktorego tesia, Valentziako Unibertsitatea.
- Language Magazine. (July, 2015). [http://languagemagazine.com/?page\\_id=35518](http://languagemagazine.com/?page_id=35518).
- Léonard, J. L. (2015). Modeling Regional Variation from EAS: Complexity and communal aggregates. EUDIAREN 4. Jardunaldietan emandako hitzaldia (agertzeko).
- López Morales, H. (1996). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología de la Universidad de Chile*, 35, 245-259.
- Moreno Cabrera, J. C. (2008). *El Nacionalismo Lingüístico: Una ideología destructiva*. Bartzelona: Ediciones Península.
- Ormaetxea, J.L. (2008). Otxandioko hizkera: adinaren araberrako bariazioa, *FLV*, 108, 249-262.
- Paredes, F. (2012). Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica. *Revista Nebrija de Lingüística Aplicada*, 11(6), 78-100.
- Rodríguez, F. J. eta Muñoz, I. O. (2009). De la disponibilidad a la didáctica léxica. *Tejuelo*, 4, 8-18.

- Serrano, M. (2004). Aspectos sociolingüísticos del léxico disponible castellano de los preuniversitarios leridanos. *Pragmalingüística*, 12, 147-165.
- Unamuno, L. (2010). Adinaren araberako bariazioa Gizaburuagako hizkeran. *Euskalingua*, 16, 41-48
- Unamuno, L., Ensunza, A., Ormaetxea, J. L., Iglesias, A eta Aurrekoetxea, G. (2012) Euskararen bariazio sintaktikoaz lehen datuak. *Euskalingua*, 12, 6-10
- Vals, E.; Nerbonne, J.; Prokic, J.; Wieling, M.; Clua, E. eta Lloret, M. R. (2012). Applying the Levenshtein Distance to Catalan dialects: A brief comparison of two dialectometric approaches. *Verba*, 39, 35-61.
- Zubillaga, H. eta Gaminde, I. (2010). /t/ren palatalizazioa Lekeitioko euskaran. *Ikastorratza*, 3, 1-9.