

Hizkuntzaren Tratamendurako tresnak.

Mikel Iruskieta
IXA taldea (UPV/EHU)

2019.eko uztailaren 4

Eskema

- 1 Hizkuntzaren Prozesamendua (HP) eta aplikazioak

Eten digitala eta tekno-pedagogikoa

Zer eskaintzen dio HPak irakasleari?

- Haur Hezkuntzatik hasi eta maila guztietarako baliabide digitalak daude
- Orain arte egin ezin ziren atazak (edo kostatakoak) egin daitezke, osagarriak dira (oraingoz)
- Baliabide digitaletan etena handia bada, handiago da corpusen erabileran eta hizkuntza prozesatzeko tresnen erabileran
- Berrikuntza honen kontrako iritziak ere badaude (baita teknologia azpigaratutako eremuetan ere)

Berritasunak

- Ebidentzian oinarrituriko azterketak sustatu ([PDF](#))
- Testu kopuru handiak aztertu ahal izateko modua
 - Materialgintza aztertu >> hobetu
- Zenbait fenomenoren ebaluazio zehatzagoa
- Konparazio bidezko ikasketa/ebaluazioa

Gaitasunak

- HIZKUNTZEN IKASKUNTZA. KULTURA ANITZEKO TESTUINGURUAK (G03)
 - Hezkuntza-proiektu eta ekimenetan hizkuntza eta kultura aniztasunarekiko sentsibilitatez jokatzeari buruzko jabetza
- PRAKTIKARI BURUZKO HAUSNARKETA (G10)
 - IKTek hizkuntzaren ikaskuntzarako duten balioaz jabetza
- DOKUMENTAZIOA ETA INFORMAZIOA AUKERATZEA (IKT) (G11)
 - Hizkuntzaren (eta literaturaren) inguruko dokumentazio-guneak ezagutzeari buruzko jabetza
 - IKTak baliabide-iturri gisa erabiltzea eta kritikoki baloratzea
- KOMUNIKAZIO-KONPETENTZIAK (G13)
 - Komunikatzeko erabil daitezkeen eta erabiltzen diren baliabide, estrategia, teknologia eta askotariko moduak ulertzea
 - Komunikazio-kompetentziak eta teknologiak erlazionatzea

Outline

- 1 Hizkuntzaren Prozesamendua (HP) eta aplikazioak
 - Hizkuntza-mailak: fonetikatik testura
 - Corpusak
 - Bibliografia

Irakurri vs Entzun: Irakurle digitala behar berezidunentzat

Probatu ondorengo **irakurle digital** helbide honetan:

- <http://tts.elhuyar.eus/>: gorde ipuinaren soinua.
Hizkuntza ezezagun batean probatu eta sortu ipuin elebiduna.

Autismoa dutenentzat euskaraz komunikatzeko programa:

E-mintza.

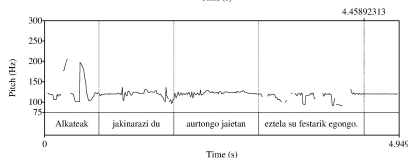
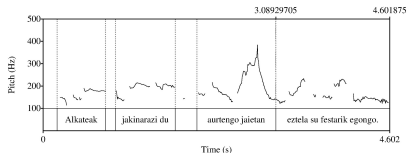
- Aipatu zer motatako ekintzak komunikatu daitezkeen
E-mintzarekin?

Hitz egin ta aztertu: Fonetika ikertzeko

Fonetika ikertzeko edota irakurketa ozena aztertzeko tresna:

PRATT

- Nabarmendu intonazioarekin NOIZ eta NORK, esaldi honetan eta aztertu PRAATEkin



Fonetika denbora errealean aztertzeko tresnak

Esaldi batetako hitzak zein kategoriakoak diren jakiteko tresna

Erabili **Eustagger** hitzen kategoriak zein diren jakiteko:

- Zapata berriak erosi nituen, Zaragozan. Polit askoak, gainera!
- Non dago zuk aipatu duzun zentro komertziala?
- Buruko mina al daukazu? Har ezazu aspirina eta oheratu zaitez.
- Lehen, Zaramagan bizi izan nintzen, eta, orain, Adurtzan.
- Nire lagun batek biolina jotzen du, oso belarri fina baitu.
- Mila esker! Bihotz handia duzu, eta, gainera, atsegina zara!
ADIBIDEEN ITURRIA: <https://vimeo.com/11489099>

Kategoriaren esanahia ikusteko clickatu [hemen](#).

<u>Mila</u>	<u>esker</u>	<u>!</u>	<u>Bihotz</u>	<u>handia</u>	<u>duzu</u>	<u>,</u>	<u>eta</u>	<u>,</u>	<u>gainera</u>	<u>,</u>	<u>atsegina</u>	<u>zara</u>	<u>!</u>
<i>mila</i>	<i>esker</i>	<i>!</i>	<i>bihotz</i>	<i>handi</i>	<i>ukan</i>	<i>,</i>	<i>eta</i>	<i>,</i>	<i>gain</i>	<i>,</i>	<i>atsegin</i>	<i>izan</i>	<i>!</i>
DET	IZEARR	PUNT_ESKL	IZEARR	ADJIZO	ADT	PUNT_KOMA	LOTJNT	PUNT_KOMA	IZEARR	PUNT_KOMA	IZEARR	ADT	PUNT

Zuzentzaile ortografikoak: XUXEN eta HOBELEX

Zuzendu ondorengo testua **XUXEN** zuzentzaile ortografiko eta gramatikalarekin

- Erregen aurretxean zan. Umia arrantza batian negarrez euan ta amak ixiltze alde: Ixildu zaitetz. Ume negartieri erregiak eztetse kopiñik pe ekartzen, eta negarrez jarraitzen badozu estira pasauko erregiak. Eztetsa ardura, maistriak esan zestan, ez juateko eskolara erregiak pasau arte, ta pasatzen ez badira, eskolara juan biharrik ez.

Zer ez du zuzentzen? Zergatik?

ADIBIDEAREN ITURRIA: IXA taldea

HOBELEX zuzentzaile ortografiko eta lexikoa ere proba dezakezu.

Lexikoa eta morfologia: Euskal Datu Base Digitala (EDBL)

Hizkuntza-forma batekin zalantzarik baduzu, bilatu **EDBL**n:

- Zer kategoria eskatzen du *–tasun* atzizkiak eta zer kategoria ematen du?
- *Gorritasun*, *edertasun*, **etxetasun*, **egontasun*, **gaizkitasun*
- Eta *–garr*i atzizkiak?
- Eta *–tzar* atzizkiak?
- Zer erlazio adierazten du *–lako* erlazio morfemak?
- Eta *–bait* erlazio morfemak?
- Eta *–la* erlazio morfema anbiguoak?

Hiztegiak

Elhuyar Hiztegia erabiliz, **eman euskaraz** ondorengo hitzak eta **entzun** euskarazkoak nola esaten dituen gizakiak eta makinak:

- Darse por vencido, excelente, rima, manjar
- Inapetencia, rechazar...

Proposamen interesgarririk?

Euskalbar euskarazko hiztegiak nabigatzailean izateko azalpena irakurri eta instalatu.

Bestelako hiztegiak:

- Gaika antolatuta: haurrekin hitz egiteko hiztegia, esaterako:
<http://www.hiztegia.net/>
- Ontologia eleanitza: **WordNet**

EUSKALTERM

Zuzendariak klaustroan txosten edo aurkezpen bat egiteko eskatu dizu, horretarako irakurri duzun artikuluko kontzeptu hauek **EUSKALTERM**ekin nola itzuliko zenituzke?

- Didáctica y Organización Escolar
- Padres sustitutos
- Maestro puericultor
- Relation parent école
- Puériculture
- Apprendre par coeur
- Parental attitude
- Parent training
- Mother child interaction
- Acquired behavior

Sintaxia automatikoki aztertzeko tresnak

Aztertu honako esaldiko izen sintagma eta aditz sintagma, azentuak eta intonazioa markatzeko: **lxati**

- Nire lagun batek biolina jotzen du, oso belarri fina baitu.

Izen Sintagma Aditz Sintagma

Emaitza:

Nire lagun batek biolina jotzen du, oso belarri fina baitu .

Nire lagun batek	BIOLINA	jotzen du,	oso belarri fina	baitu.
oO-oo-oO	oOoO	ooO,	oO-ooo-oO	oO.
+2-1	+2-1	-1,	+2-1	-1.

Sintaxia automatikoki aztertzekeo tresnak

Aztertu honako esaldiaren zuhaitz sintaktikoa [Maltixarekin](#)

- Nire lagun batek biolina jotzen du, oso belarri fina baitu.

Syntax analysis of: Nire lagun batek biolina jotzen du, oso belarri fina baitu. is:

Index	Word	Head	Relation	Lemma	Features	Category	Subcategory
1	Nire	2	ncmod	ni	KAS:GEN—NUM:S	IOR	IOR_PERARR
2	lagun	5	ncsubj	lagun	-	IZE	IZE_ARR
3	batek	2	detmod	bat	KAS:ERG	DET	DET_DZH
4	biolina	5	ncmod	biolin	KAS:ABS—NUM:S	IZE	IZE_ARR
5	jotzen	0	ROOT	jo	ASP:EZBU	ADI	ADI_SIN
6	du	5	auxmod	*edun	-	ADL	ADL
7	0	6	PUNC	0	-	PUNT	PUNT_KOMA
8	oso	10	gradmod	oso	-	ADB	ADB_ARR
9	belarri	11	ncobj	belarri	-	IZE	IZE_ARR
10	fina	9	ncmod	fin	KAS:ABS—NUM:S	ADJ	ADJ_ARR
11	baitu	5	cmmod	ukan	ERL:KAUS—ASP:PNT	ADT	ADT
12	.	11	PUNC	.	-	PUNT	PUNT_PUNT

XX. mendeko Corpusa

a) Aztertu **XX. mendeko Corpusean** nola erabiltzen diren ondorengo hitzak:

- *Eraso* hitzaren erabileraren berri eman.
 - Hitz berria da? Noiz erabili da gehien?
 - Zein euskalkitan erabiltzen da hitza?
 - Zein testu motatan agertzen da?
 - Administrazioko testuetan agertzen da? Zergatik?
.....
- *Azterketa* hitzaren erabileraren berri eman
 - Hitz zaharra al da? Zergatik?
- Eta *eldarnio* hitzari zer gertatzen zaio?
- Nola esaten zen lehen *sexu*?
- Zure euskalkiko zein hitz aztertu zenuke?

b) **Lexikoaren Behatokia**: bilaketak kategoriaka eta bilaketaren testuingurua molda daiteke (aurreko edota ondorengo hitzak)

Outline

- 1 Hizkuntzaren Prozesamendua (HP) eta aplikazioak
 - Hizkuntza-mailak: fonetikatik testura
 - Corpusak
 - Bibliografia

Corpusa da

- Antzeko ezaugarriak dituen testu-multzo prozesagarria
 - Ezaugarri horiei esker hizkuntza-fenomenoak azter daitezke

OHARRA: Atazaren, tamainuaren eta kalitatearen araberakoa izango da ikerketaren emaitza

Corpus Baliagarriak

HHrako baliagarriak diren euskal corpusak

- **LingoLang**: corpus eleanitza, hizkuntza ukipenean dauden hizkuntzen hiztegiak eta esapide oinarritzkoak, herriz herri
- Haur eta gazteentzat corpusa **ZerNola** testu zientifikoaren corpusa
- **Ahotsak** euskalkiak lantzeko ikus-entzunezkoak
- **Euskalkiak**, gazteen euskalkiak lantzeko
- **Childes** *Child Language Data Exchange System* Ahozko elkarrizketak aztertzeko
- **Eskola datu basea** (ENG)

Corpusetan oinarrituriko hizkuntza irakaskuntza

- Corpusean oinarrituriko hizkuntzaren irakaskuntzan aldaketak ekarri ditu:
 - Materialgintzan: corpusetan oinarrituriko hiztegiak, testu errealetan oinarrituriko ariketak
 - Metodologian ikuspegi orokorra eta konparazio-lan asko
 - Ez-aditu VS adituen corpusen alderaketa
 - Ikasleek corpusak erabiltzea eta bestelako ezaugarriak ikastea (hipotesiak eraikitzea)
- Irakurgai praktikoa:
<http://www.tesl-ej.org/ej32/a1.html>

ANALHITZAko etiketak

Labur.	Kategoria
A	Adberbioa
C	Diskurtso-markatzailea
D	Determinatzailea
G	Adjektiboa
N	Izen arrunta
O	Puntuazioa
R	Izen berezia
V	Aditza

Erauzi hitzak abestietatik

Erabili ANALHITZA:

<http://ixa2.si.ehu.es/clarink/analhitza.php?lang=eu>

abestien letrak aztertzeko:

- Aldapan gora (HUNTZA)

<https://txantxangorria.eus/?p=6546>

- Zirkorrika (AEK) <https://txantxangorria.eus/?p=6667>

Zein da hitzik (leमारिक) errepikatuena?

Nire lehen corpusa

- Bilatu testu bat PDF formatukoa (edo txt)
- Pasatu PDFtik TXTra (UTF8ra)
- Beharbada eskuz egin behar da, garbiketaren bat?
- Testutik corpusera (Windowsen)
 - `for %f in (*.txt) do type "%f" >> corpus.txt`

Bibliografia

- Aldabe, I., Arrieta, B., de Ilarraza, A. D., Maritxalar, M., & Niebla, I. (2008) Hizkuntzaren Tratamendu Automatikoa Euskararen Irakaskuntzan. BAT aldizkaria.[URL](#)
- Aldezabal, I., Arriola, J. M., Diaz de Ilarraza, A., & Sarasola, K. (2005). Hizkuntzalaritza konputazionala. UEU. [URL](#)
- Alegria, I., Artola, X., Diaz de Ilarraza, A., Sarasola, K., & Aduriz, I. (2011). Teknologia garatzeko estrategiak baliabide urriko hizkuntzetarako: euskararen eta Ixa taldearen adibidea. *Linguamática*, 3(1), 13-31. [URL](#)
- Ezeiza, N. (2003). Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketatzaile sintaktiko sendo eta malgua (Doctoral dissertation, PhD thesis, University of the Basque Country).[URL](#)