

Euskarazko testuen komunikagaitasun-maila automatikoki sailkatzeko lehendabiziko urratsak

Artikulu honetan, HABEren eta IXA ikerketa-taldearen arteko egitasmo baten lehen urratsak azaltzen dira. Proiektuaren helburua komunikagaitasun-mailen arabera testu-sailkatzailer automatiko bat sortzea da, ikasleen ulermen-mailarako egokiak diren idatzizko testuak objektiboki hautatzeko aukera emango duena. Horretarako, HEOCeko mailaz mailako adierazpide linguistikoak literaturan erabili izan diren bigarren hizkuntzen jabeakuntzako (SLA, *Second Language Acquisition*) parametroen gisakotzat hartu dira, eta horiekin informatika-aplikazioetarako linguagrama bat sortu.

Gako-hitzak

informatika-aplikazioetarako linguagrama, irakurgarritasun-mailak komunikagaitasun-mailak, sailkatzailer automatikoa, SLA parametroak,

1. SARRERA

2020ko irailaren 15ean, Eusko Jaurlaritzako Kultura eta Hizkuntza Politika Sailak (HABEren bidez) eta Euskal Herriko Unibertsitateak (IXA ikerketa-taldearen bidez) lankidetzeta-hitzarmena¹ sinatu zuten, HABEk azken urteotan sortu eta batu duen informazioa egituratzeko eta, hala, euskara ikasteko baliagarriak izan daitezkeen hizkuntza-teknologiak garatzeko. Hitzarmen horren baitan sortutako betebeharrei erantzuteko sortu da "HABEko testu-bildumen azterketa eta egituraketa-proposamena, hizkuntza-teknologiak garatzeko" proiektua HABEn, artikulu honetan azalduko duguna.

Hitzarmen horretara iristeko arrazoia da bi alderdiek zenbait interesetan bat egiten dutela. Izan ere, IXA taldeak UPV/EHUko Bilboko Hezkuntza Fakultateko ikasleen euskarazko komunikagaitasun-maila ebaluatzeko beharra atzeman du, eta, behar horri

¹ https://www.habe.euskadi.eus/contenidos/informacion/lankidetzeta_hitzarmenak/eu_def/adjuntos/HABEk-eta-EHUK-onartua-GARBIA-Hitzarmena_eus.pdf

erantzuteko, ebaluazio hori automatikoki egingo duen tresnen garapenean jarri du begirada. HABErentzat, berriz, interes handiko ideia da hizkuntza-teknologiako tresnen bidez testuak mailakatzeko betebeharra modu objektiboago batean aurrera eramatea. Teknologia horren bidez, mailakatzearen objektibitatea hobetzeaz gain, egiteko horiek gauzatzeko beharrezko denbora ere nabarmen murriztuko litzateke.

Hitzarmen horretan, 4 urteko lankidetzaz adostu zuten HABEk eta IXA taldeak, eta, horrez gain, alde bakoitzak lankidetzaz horri dagokion lana egiteko konpromisoa adierazi zuten. HABEk bere fitxategiak aztertze eta egituratzeko betebeharra hartu zuen bere gain, UPV/EHUko ikerlariekin beharrezko elkarlana mantenduz. IXA taldeak, berriz, bere eskarmentua, baliabide informatikoak eta bere jakintza guztia sailkatzaile automatikoen programazioan jartzeko konpromisoa hartu zuen, baita sailkatzaile automatikoa garatzeko bidean sortzen dituen hizkuntza-teknologiako tresnak HABEren esku jartzekoa ere.

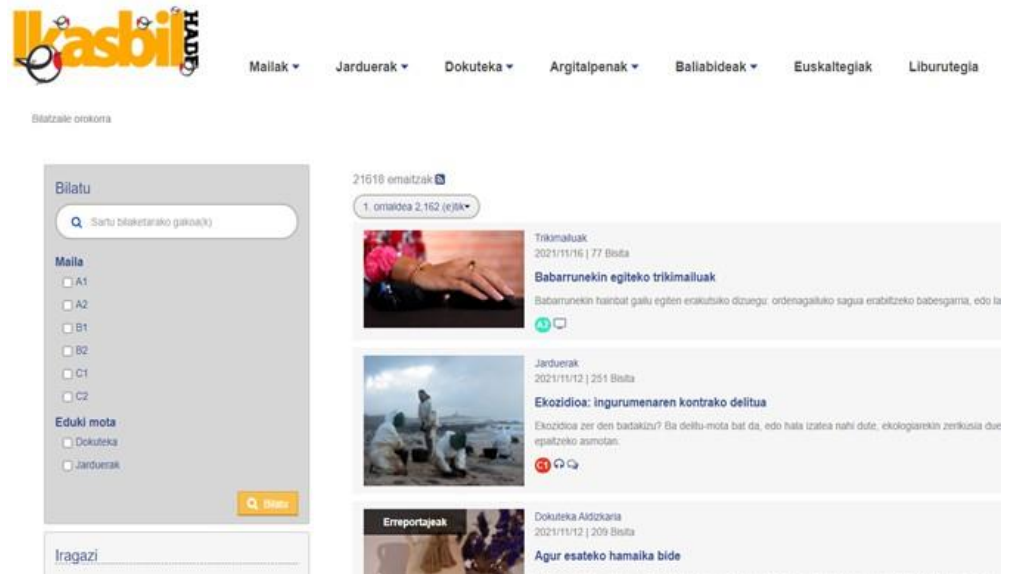
Lehen esan bezala, hemen azalduko den proiektua hitzarmen horretatik jaio da, eta proiektuaren betebeharrei erantzunez, bi lan-ildo izan dituen proiektu bat garatu da:

- Hizkuntza-teknologietan oinarritutako tresnen garapena, Ikasbil atarirako (IXAekin elkarlanean). Atal horretan, sailkatzaile automatikoaren bidean sor daitezkeen tresnen garapena, bistaratze-teknikak eta irizpide-proposamenak landu ditugu.
- HABEko testu-bildumen azterketa eta egituraketa-proposamena. Atal horretan, HABEren egiaztatze-gintza-prozesuko testu-bildumak aztertu, egituratu eta fitxategi horien transkripzioak lortzeko beharrezko lana egin da, gero testu-bilduma hori hizkuntza-teknologiaren garapenean erabilgarria izan dadin.

Artikulu honetan azaltzen den atala proiektuaren lehen lan-ildoari dagokio; hizkuntza-teknologietan oinarritutako tresnen garapenari, hain zuzen. Ildo horretan, besteak beste, sailkatzaile automatikoen irizpideak HEOCarekin (Gonzalez de Txabarri et al., 2015) uztartzeko aukera aztertu da. Behin sailkatzaile automatikoaren programazioa burututa, sailkatzaile hori Ikasbilen jarriko da interesa duen ikasle edo irakasleek erabil dezaten. Ikasbil euskararen ikas-baliabide ugari biltzen dituen web-ataria da, eta erreferentea da euskarazko irakaskuntzarako erabilgarriak diren ariketak bilatzen dituzten irakasleentzat, baita euskarazko autoikaskuntza egiten ari diren ikasleentzat ere.

Web-atari horrek eskaintzen dituen baliabideen artean, idatziak, *online* egiteko moduko ariketak, entzungaiak, argitalpenak, hiztegiak, denbora-pasak, agenda eta abar aurki ditzakegu. Gainera, baliabide horiek euskararen ikaskuntza-prozesuko maila ezberdinetan dauden ikasleentzat daude bideratuta: bai euskaraz lehen urratsak ematen ari direnentzat, baita beren euskara maila altuagora eramanez nahi dutenentzat ere (ikus 1. irudia).

1. irudia. Ikasbil atariko orrialde bat



2. SAILKATZAILE AUTOMATIKOEN ERABILERA: ABIAPUNTUA ETA ONDORENGOAK

Urte askoko ibilbidean zehar, hizkuntzalaritzak frogatu du ikasleek beren mailatik gertuko testuak behar dituztela hizkuntzaren ikasketarako. Hainbat lanetan frogatu izan da ikasleek motibazio eta jarrera hobea izaten dutela beren mailako testuak lantzerakoan (Jorgenson, 1977).

Testu batek ikaslearekiko eduki dezakeen egokitasun-mailaren neurketa, era subjektiboan egin ohi da. Irakasle edo teknikari batek testu bat aztertu eta irizpide batzuen arabera aukeratzen du testu hori zein mailatakoa den. Hala ere, lan horretan subjektibitateak leku handia izaten du, eta mailakatze-prozesu hori sailkatzaile automatikoez bideratzeak objektibitatea gehi liezaioke (Santucci et al., 2020).

Hizkuntza handietan, urte asko daramate sailkatzaile automatikoekin lanean (ingelesez, adibidez, 80 urtetik gora). Alabaina, hizkuntza-teknologiek azken 10 urtean izan duten gorakadak tresna berriak eskaini dizkio alor horri, sailkatzaile automatikoak hizkuntza txikiagoentzat eskuragarri jartzeaz gain.

Tradizioz, irakurgarritasun-mailaren sailkapen automatikorako parametroak testuaren azaleko analisi batean oinarritu izan dira; adibidez, Rudolf Flesch eta J. Peter Kincaid ikerlariak proposatu zuten hitz bakoitzeko silaba kopurua eta esaldi bakoitzeko hitz kopurua (Kincaid et al., 1975). Parametro horiek, nahiz eta oraindik erabiltzen diren, ez dituzte kontuan hartzen hizkuntzaren konplexutasunean eragiten duten hainbat aspektu, eta, horregatik, ikerlariak arazo hori konpontzen saiatu izan dira urteetan zehar.

Sailkatzaileak hizkuntza-didaktika kontuan hartuko duten irizpideen arabera testuak mailakatzeko nahiari erantzunez, hainbat ikerlarik parametro berriak proposatu izan dituzte. Adibidez, testu baten zailtasun lexikoa neurtzeko, hiztegian oinarritutako Dale-Chall formula aurkeztu zuten Jeanne S. Chall eta Edgar Dale ikerlariak (Chall eta Dale, 1995). Berrikiago, ezaugarri sintaktikoen neurgarritasuna aztertu duten lanak ere publikatu izan dira (Heilman et al. 2007; Petersen eta Ostendorf, 2009; Scharm eta Ostendorf, 2005). Bide horri jarraituz publikatu da bigarren hizkuntzaren irakaskuntza eta sailkatzaile automatikoak lotzea helburu duen *On Improving the Accuracy of*

Readability Classification using Insights from Second Language Acquisition lana (Vajjala eta Meurers, 2012).

3. SLA PARAMETROAK

Arestian aipatutako lanean (Vajjala eta Meurers, 2012), bi motatako parametroak erabiltzen dituzte sailkatzaile automatikoa entrenatzeko: tradizionalak, batetik, eta bigarren hizkuntzaren jabekuntzan (SLA, *Second Language Acquisition*) oinarritutako parametroak, bestetik. SLA parametroak bigarren hizkuntzaren irakaskuntza oinarritzat hartuta zehazturikoak dira; horren erakusgarri dira, adibidez:

- *Adjective variation* (adjektibo-bariazioa): testu batek zenbat adjektibo mota dituen.
- *Verb variation* (aditz-bariazioa): testu batek zenbat aditz mota erabiltzen dituen.
- *Lexical density* (lexiko-dentsitatea): testuan ageri diren hitz lexiko moten kopurua, testuko hitz kopuruarekin alderatuta.
- *Number of Clauses per Sentence* (perpaus kopurua esaldiko): zenbat esaldi konplexu eta haien konplexutasun maila.

Aipatutako lanaren asmoa sailkatzailearen zehaztasuna hobetzea da, eta, horretarako, proposatzen dute SLA parametroak gehitzea. Proposamen hori baliozkotzeko, proba batzuen berri ematen dute, emaitzarik onenak zer parametro multzorekin lortu diren erakutsiz. Proben emaitzak nahiko esanguratsuak dira: parametro tradizionalekin baino emaitza hobek lortzen dira SLA parametroekin; denak batera erabilia, berriz, emaitza nabarmen hobek lortzen dira (ikus 1. taula).

1. taula. Parametro motak eta haiekin lortutako emaitzak (Vajjala & Meurers, 2012, 169. or.; moldatua)

Parametro mota	Zehaztasuna
Parametro tradizionalak	% 72,3
SLA parametroak	% 82,3
Parametro guztiak	% 93,3

Vajjala eta Meurers (2012) lanean proposatzen diren SLA parametroak 2. taulan bildu ditugu. Lan horretan ikerlariak egiten duten proposamenak oso bateragarria dirudi HABEk proiektu honen inguruan duen asmo nagusietako batekin: HEOCa sailkatzaile automatikoaren irizpideen artean islatzearekin, hain zuzen ere. Izan ere, SLA parametroen zerrendak, forma aldetik, antzekotasun zenbait ditu HEOCean egiten den mailaz mailako adierazpide linguistiko eta testualen zerrendekin.

2. taula. SLA parametro lexikoak eta sintaktikoak (Vajjala eta Meurers, 2012, 167-168. or.)

SLA parametro lexikoak	SLA parametro sintaktikoak
-Lexical Density (LD)	-Mean length of clause (MLC)
-Type-token ratio (TTR)	-Mean length of a sentence (MLS)
-Corrected TTR (CTTR)	-Mean length of T-unit (MLT)
-Root TTR (RTTR)	-Num. of Clauses per Sentence (C/S)
-Bilogarithmic TTR (LogTTR)	-Num. of T-units per sentence (T/S)
-Uber Index (Uber)	-Num. of clauses per T-unit (C/T)
-Lexical Word Variation (LV)	-Num. of Complex T-units per T-unit (CT/T)
-Verb Variation-1 (VVI)	-Dependent Clause to Clause Ratio (DC/C)
-Squared VVI (CVVI)	-Dependent Clause to T-unit Ratio (DC/T)
-Corrected VVI (CVVI)	-Co-ordinate Phrases per T-unit (CP/T)
-Verb variation 2 (VV2)	-Complex Nominals per Clause (CN/C)
-Noun Variation (NV)	-Complex Nominals per T-unit (CN/T)
-Adjective Variation (AdjV)	-Verb phrases per T-unit (VP/T)
-Adverb variation (AdvV)	
-Modifier Variation (ModV)	
-Mean Textual Lexical Density (MTLD)	

4. HEOC-A SAILKATZAILEAN NOLA ISLATU

HEOCean, komunikagaitasun-mailen deskribapenak egitean, bakoitzari dagozkion adierazpide linguistiko eta testualak zerrendatzen dira. Gainbegiratu bat emanda, HEOCaren eta SLA parametroen artean antzekotasun zenbait daudela ikus daiteke.

Hala ere, HEOCari begiratzen badiogu, atal batean baino gehiagotan aipatzen da gaitasun linguistikoko ezaugarriak ezin direla beren testuinguru eta funtzio komunikatiboetatik at landu:

Komunikazioaren eta ekintza-ikuspegiaren baitan aurkezten eta zehazten dira mailaz mailako adierazpide linguistikoak. Baliabide horiek, nahiz eta, laburtze aldera, guztiak jarraian eskema baten arabera aurkeztu, ez dira aipatutako eskema horren ildotik programatu behar, ezta landu behar ere, baizik eta adierazten dituzten funtzio komunikatiboan arabera. **Izan ere, adierazpide linguistikoak ez dira curriculum, ezta euskalduntzearen berariazko helburu ere; helburu den gaitasun komunikatiboranzko bitarteko baizik.** (Gonzalez de Txabarri et al., 2015, 63. or) [letrakera lodia gurea da].

Kontuan hartu behar da, gainera, testuinguruak HEOCean duen garrantzia:

Helduen Euskalduntzearen Oinarrizko Curriculum, beraz, sei mailatan egituratuta aurkezten da; **elkarreragin soziodiskurtsiboan ardaztua, hizkuntzaren ikuspegi komunikatiboan oinarritua, ekintza komunikatiboan definitua, eta euskararen ezaugarri eta egoera soziolinguistikora egokitua.** (Gonzalez de Txabarri et al., 2015, 29. or.) [letrakera lodia gurea da].

Testuinguruaren arabera jakintza linguistikoa garatzearen kontzeptua Europako Erreferentzia Marko Bateratuan ere ageri da:

Hizkuntzaren erabileraz hitz egitean, pertsonen ekintzaz ari gara. Horiek, gizabanako eta gizarte-eragile izaki, gaitasun orokor batzuk dituzte eta, bereziki, komunikazio-gaitasuna. Hizkuntza-ekintzak gauzatzeko, gaitasun horiek erabiltzen dituzte, testuinguru, baldintza eta muga desberdinen arabera. Hala lantzen (ulertzen eta sortzen) dituzte testuak, **arloan espezifikoko gai batzuk garatuz, egokien irizten dieten estrategiak erabiliz eta atazak gauzatuz**. Parte-hartzaileek, jarduera horiek egin ahala, gaitasunak sendotu edo hobetuko dituzte. (Gonzalez de Txabarri et al., 2015, 22. or.) [letrakera lodia gurea da].

Gaur-gaurkoz, ez dago teknologiarik HEOCeko ezagutza osoa prozesatu eta horrekin sailkatzaile automatiko bat garatzeko gauza denik; ezinezkoa da, oraingoz, jakintza hori programa informatiko bati atxikitzea. Hortaz, bertsio laburtu eta sinplifikatuago bat osatuko da, *informatika-aplikazioetarako linguagrama*, alegia, betiere garatuko dugun sistemaren mugak kontuan izanda.

Linguagrama bat, funtsean, curriculum linguistikoa bat da (Aldezabal et al. 2019; Arriola et al. 2020), baina guk garatuko dugunaren bereizgarria da berariaz informatikan erabiltzeko diseinatua izango dela. Linguagramaren proposamena bat dator hizkuntza-didaktikako adierazleak sailkatzaile automatikoen parametroen artean erabiltzearen ideiarekin (Vajjala eta Meurers, 2012), baina beste ikuspuntu batetik lantzen du. Azken finean, euskararen ezaugarri tipologiko bereizgarriak kontuan izanik (ergatiboa, hitz ordena eta abar), pentsatzekoa da beste hizkuntzetako sailkatzaileetan erabiltzen diren konplexutasun-ezaugarriak (SLA parametroak) ez direla gure kasuan baliagarriak izango.

Bestalde, kontuan izan behar dugu guk helburutzat dugun aplikazioa ez dela azterketa-zuzentzaile bat, ikasle eta irakasleei lagungarri gerta dakiekeen tresna bat baizik: testu baten gutxi gorabeherako komunikagaitasun-maila aurreikusteko, testu baten mailatik gertu dauden beste testu batzuk bilatu eta identifikatzeko, interesatzen zaizkigun zenbait ezaugarri dituen testu bat bilatzeko, eta abarretarako. Izan ere, HEOCaren adierazpide linguistikoa eta testualak baino ez dira kontuan hartuko, teknologiak besterik egiteko aukerarik ematen ez digun bitartean.

Aplikazio informatikoetarako linguagrama bat sortuz, bi helburu garrantzitsu betetzen dira:

- Mailakatzeko eginko duen aplikazioaren zehaztasuna hobetzea.
- HEOCaren presentzia bermatzea aplikazioaren mailakatzeko-parametroetan.

Horiek horrela, HEOCean zehazten diren mailaz mailako gaitasun linguistikoa eta testualak SLA parametroekin uztartzeko aukerarik dagoen aztertu da, eta, IXA taldekoekin elkarlanean, *informatika-aplikazioetarako linguagrama* bat sortu.

Horretarako, HABEK aurretik egindako lana baliatu da. Izan ere, komunikagaitasun-mailaren arabera banatutako adierazpide-zerrenda bat eskaintzen du HEOCak, eta hemen proposaturiko linguagramak informazio hori sailkatzaile batean erabilgarri izateko eran aurkeztuko du.

Hori guztia kontuan hartuta, zerrenda bat egin da mailaz maila banatutako adierazpideekin. Zerrenda horren adibide gisa, A1 mailako linguagrama 1. eranskinetan dago jasota.

5. ONDORIOAK

Hasteko, sailkatzaile automatikoen parametroetan ikusitako eboluzioa aztertuta, ikus daiteke hizkuntza-teknologiaren alorrean gero eta garrantzi handiagoa ematen ari zaiola informatika eta hizkuntzalaritza uztartzeari. Nahiz eta testuinguru hain desberdina duten bi alor horiek uztartzeak hasiera batean arazotsua dirudien, ikerketa ugariaren ondorioek azpimarratzen dute elkarlana onuragarria izango dela batarentzat zein bestearentzat. Adibidez, artikuluko honetako 3. atalean azaltzen den bezala, sailkatzaile automatikoetan hizkuntzalaritzan oinarritutako irizpideak gehituz, sailkatzaile horien zehaztasuna nabarmen handitzen da.

Horrez gain, linguagramaren bidez sailkatzaileen irizpideak HEOCarekin uztartzeko asmoaren egingarritasuna ikusirik, hizkuntza-teknologiek euskaraz etorkizunean izan dezaketen garapenaz jabetu beharra dago. Garapen horrek berebiziko garrantzia du euskararena bezalako hizkuntza-komunitate txikientzat. Izan ere, garrantzi hori ez dagokie bakarrik teknologia horien garapenaz lortuko ditugun onura argiei (mailakatzean denbora eta objektibitatea irabazteari); horrez gain, euskarak hizkuntza-teknologietan aurreratzeak hizkuntzen arteko egoera diglosikoa eragiten duten arrazoi soziolinguistiko zenbait ekiditen ere lagundu diezaguke. Hau da, hizkuntza-teknologiaren garapenak euskara beste hizkuntza handiagoen mailan jarriko luke alor horretan, eta oso garrantzitsua da, ahal den heinean behintzat, eremu horretan ere atzean ez geratzea.

Bibliografia

- Aldezabal, I., Arriola, J. eta Estarrona, A. (2019). A modular grammar-helping tool for Basque: work in progress. Turku, Finland: *Proceedings of the NoDaLiDa 2019 Workshop on Constraint Grammar - Methods, Tools and Applications*.
<https://www.ixaeu.com/sites/default/files/dokumentuak/13067/ecp19168001.pdf>
- Arriola, J. M., Irukieta, M. eta Goikoetxea, J. (2020). Hizkuntza teknologiak hizkuntzen ikas-irakaskuntzan: zenbat aukera hainbat erronka. *eHizpide*, 95.
<http://www.ikasten.ikasbil.eus/mod/habecms/view.php/irakasbil/argitalpenak/hizkuntza-teknologiak-hizkuntzen-ikas-irakaskuntzan>
- Chall, J. S. eta Dale, E. (1995). *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline Books.
- Gonzalez de Txabarri, J., Elozegi, L., Agirre, H., Larramendi, A., Maiz, A. eta Gonzalez de Txabarri, L. M. (2015). *Helduen Euskalduntzearen Oinarrizko Curriculum (HEOC)*. HABE, Helduen Alfabetatze eta Berreuskalduntzerako Erakundea.
<http://www.ikasten.ikasbil.eus/mod/habecms/view.php/irakasbil/didakteka/heoc-liburua-pdf-bertsioa>
- Heilman, M. J., Collins, K. eta Callan, J. (2007). Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts. *Proceedings of the Human Language Technology Conference*, 460-467. <https://aclanthology.org/N07-1058.pdf>
- Jorgenson, G. W. (1977). Relationship of Classroom Behaviour to the Accuracy of the Match Between Material Difficulty and Student Ability. *Journal of Educational Psychology*, 69(1), 24-32.
<https://doi.org/10.1037/0022-0663.69.1.24>
- Kincaid, J., Fishburne, R. J., Rogers, R. L. eta Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease formula) for navy enlisted personnel*. Millington, TN: Research Branch Report 8-75, Naval Technical Training Command.
<https://stars.library.ucf.edu/istlibrary/56/>
- Petersen, S. E. eta Ostendorf, M. (2009). A machine learning approach to reading level assessment. *Computer Speech and Language*, 23, 89-106.
<https://dada.cs.washington.edu/research/tr/2006/06/UW-CSE-06-06-06.pdf>
- Santucci, V., Santarelli, F., Forti, L. eta Spina, S. (2020). Automatic Classification of Text Complexity. *Applied Sciences* 10(20), 72-85.
<https://doi.org/10.3390/app10207285>
- Scharm, S. E. eta Ostendorf, M. (2005). Reading level assesment using support vector machines and statistical language models. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (523-530. or.). <https://doi.org/10.3115/1219840.1219905>
- Vajjala, S. eta Meurers, D. (2012). On Improving the Accuracy of Readability Classification using Insights from Second Language Acquisition. In *NAACL HLT '12: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP* (163-173. or.). Association for Computational Linguistics .

Gehiago jakiteko

- Allington, R. L., McCuisson, K. eta Billen, M. (2015). What research says about text complexity and learning to read. *The Reading Teacher*, 68(7), 491-501.
<https://horizonsd.ca/Services/Literacy/Documents/What%20Research%20Says.pdf>
- Chen, X. eta Meurers, D. (2016). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity at COLING*, 113-119.
<https://doi.org/10.17863/CAM.39630>
- Coleman, M. eta Liau, T. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2), 283-284.
<https://doi.org/10.1037/h0076540>
- Dell'Orleta, F., Montemagni, S. eta Venturi, G. (2011). READ-IT: Assesing Readability of Italian Texts with a View to Text Simplification. *Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies*, 73-83. <https://aclanthology.org/W11-2308.pdf>
- Nikleva, D. G. (2008). La oposición oral/escrito: consideraciones terminológicas, históricas y pedagógicas. *Didáctica. Lengua y Literatura*, 20, 211-227.
<https://revistas.ucm.es/index.php/DIDA/article/view/DIDA0808110211A>
- Sinclair, J. (2004). Corpus and Text - Basic principles. In M. Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice* (1-16. or.). Oxbow Books.

1. eranskina. HEOCeko A1 mailan ageri diren adierazpideak**A1****Gaitasun linguistikoaren garapena**

Adierazpide linguistikoak

ALDERDI ORTOGRAFIKOA**Ortografia:**

- -r-ren aurrean bokala mailegu zaharretan: *arropa, errege, erloju...*
- *nb* eta *np* grafiak
- Euskararen fonema txistukariak: /tz/, /tx/, /ts/, /z/, /s/, /x/
- Maileguen egokitzapena: *funtzionario, erresistentzia...*

ALDERDI MORFOSINTAKTIKOA**Izen-sintagma****Deklinabidea**

Mugatu singularra, mugatu plurala eta mugagabea

- | | |
|-----------------------------|-----------|
| • Absolutua | NOR |
| • Ergatiboa | NORK |
| • Partitiboa | ZERIK |
| • Datiboa | NORI |
| • Inesiboa | NON |
| • Genitibo edutezkoa | NOREN |
| • Adlatiboa | NORA |
| • Ablatiboa | NONDIK |
| • Soziatiboa | NOREKIN |
| • Genitibo leku-denborazkoa | NONGO |
| • Adlatibo bukatuzkoa | NORAINO |
| • Destinatiboa | NORENTZAT |
| • Instrumentala | ZERTAZ |

Galdetzaileak

Nor (nork, nori, noren, norekin)
Zer (zerk, zer ordutan, zer ordutatik zer ordutara)
Non (nongoa, nora, nondik)
Noiz
Nola
Zenbat (zenbatek)

Izenordainak

Pertsona-izenordainak: ni, zu, gu...
Erakusleak: hau, hori, hura, hauek, horiek, haiek
Izenordain mugagabeak:

- Zerbait, norbait, zerbaitek, norbaitek, zerbaiti, norbaiti
- Inor ez, ezer ez, inork ez

Izenlagunak:

Noren eta Nongo

Determinatzaileak

Mugatzaileak

- Artikuluak:
 - Zehaztuak (-a/ -ak)
 - Zehaztugabeak (bat/ batzuk)
- Erakusleak: hau, hori, hura, hauek, horiek, haiek
 - Zenbatzaileak
- Zehaztuak
 - Zenbakiak: (0-100)
 - Ordinalak: (lehen, bigarrena...)
 - Kopuru-adierazpenak: kilo bat, lau ordu...
 - Orokorrak: oso, guztia, dena
- Zehaztugabeak: asko, gutxi, batzuk, nahiko(a), pixka bat

Adjektiboa

Motak:

- Bakunak: eder, polit, itsusi...
- Graduatzaleak: oso, nahiko...
- Errepikapenaren bidezkoa: txiki-txikia...

Adberbioak

Leku-adberbioak: hemen, hor, han...

Denbora-adberbioak: gaur, atzo, bihar, etzi, orain, gero, lehen...

Modu-adberbioak: gaizki, txarto, ondo, poliki...

Graduatzaleak: batere, ia-ia, guztiz, izugarri, ikaragarri

Postposizioen bidez osatuak:

- Nor edo partitiboa eskatzen dutenak: -rik gabe
- Noren eskatzen dutenak: -(r)en gainean, azpian, aurrean, atzean, barruan, kanpoan...

Hitz-sortzaileak

Elkarketa

- Seme-alabak, aitona-amonak, patata-tortilla...

Aditza**Ezaugarriak**

Aditz iragangaitzak eta iragankorrak

- Jon sartu da
- Jonek sartu du autoa

Jokatuak eta jokatugabeak

- Etorri naiz, egin dut...
- Etorzea, egiteko

Aditz perifrastikoak

- Joango naiz, etorri da, eramango du

Aditz trinkoak

- Dago, dauka

Aspektua

Orainaldi burutua: joan naiz

Orainaldi ez burutua: joaten naiz

Gertakizuna: joango naiz

Puntuakaria: ikasten ari naiz, dago, dauka

Aditz-lokuzioen aspektua: bizi izan, ari izan, behar izan, balio izan, nahi izan, ahal izan eta ezin izan

Modua

Indikatiboko orainaldia

- NOR izan, egon
- NOR-NORK: ukan, eduki
- NOR-NORI: zait, zaizkit... (3. pertsona)
- ZER-NORI-NORK: diot, didazu...

Indikatiboko lehenaldia

- NOR: izan
- NOR-NORK: ukan, eduki

Ahalera

- Ahal izan
- Ezin izan

Agintera

- Partizipio burutua: Etorri! Segi!
- NOR: izan (zaitez, zaitetze)
- NOR-NORK: ukan (ezazu, itzazu)

Bestelako erabilerak:

- Eguraldia: euria / elurra... ari du. Haizea dabil...

Sintaxia**Perpaus bakuna**

Galderazko perpausa. Galdegaia galderazko perpausetan

Adierazpen-perpausak

- Baiezkoak. Galdegaia baiezko perpausetan
- Ezezkoak. Galdegaia ezezko perpausetan

Harridura-perpausak

- Erakusleen bidez osatuak: Hau hotza!
- Galdetzaileen bidez:
 - ZER / ZEIN itsusia (den)!
 - NOLAKO aurpegia (duen)!
 - ZENBAT dituen!

Aginte-perpausak

- Geldi hor!
- Ez egin hori!

Perpaus elkartuak

Juntadura

- Emendiozkoak: eta, ere
- Hautakariak: ala, edo
- Aurkaritzakoak: baina

Menderakuntza

- Izen-funtziozko menderakuntza
 - -tzea
 - Agindua edo debekua: -tzeko
- Baldintzazko perpausak (errealak):
 - -ba, baldin ba- (ostiralean bada, joango naiz...)