



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

V. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2023ko maiatzaren 17, 18 eta 19a
Donostia, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)



Aitortu-PartekatuBerdin 3.0

GIZA ZIENTZIAK ETA ARTEA

**Easy-to-Read Language: baliabide
linguistikoen eta testuen
egokitzapena eta tresna
automatikoen garapena**

*Margot Madina, Itziar Gonzalez-Dios
eta Melanie Siegel*

35-42 or.

<https://dx.doi.org/10.26876/ikergazte.v.01.04>

ANTOLATZAILEA:



BABESLEAK:



LAGUNTZAILEAK:



Easy-to-Read Language: baliabide linguistikoaren eta testuen egokitzapena eta tresna automatikoen garapena

Margot Madina¹, Itziar Gonzalez-Dios², Melanie Siegel¹

¹ Darmstadt university of Applied Sciences (Hochschule Darmstadt)

² HiTZ Zentroa-Ixa, Euskal Herriko Unibertsitatea (UPV/EHU)

margot.madina-gonzalez@h-da.de, itziar.gonzalezd@ehu.eus,
melanie.siegel@h-da.de

Laburpena

Informazioa, ezagutza eta kultura eskuratzea herritar guztien eskubidea da. Hala ere, idatzizko testua ez da beti bat egiten irakurtzen duguna ulertzeko dugun gaitasunarekin. *Easy-to-Read Language* (E2R) ulermen oztopo horiek gainditzea du helburu, beraz, aukera berdintasuna sustatuz. E2R hizkuntzaren arabera alda daitezkeen arau-multzo bat dauka, eta sorburu-testuan ez dauden adibideak, ikusmen eta/edo entzumen-laguntzak ere izan ditzake. Zoritxarrez, ez dago hizkuntza-baliabide askorik, ezta egokitzapen-tresnarik ere erabilgarri E2R-erako. Tesi honek E2R egokitzapen automatikorako dauden baliabideak ebaluatu eta baliabide eta egokitzapen tresna berriak sortzea du helburu. Europako hiru hizkuntzatan zentratuko da: alemana, gaztelania eta euskara.

Hitz gakoak: Easy-to-Read, Leichte Sprache, Lectura Fácil, Irakurketa Erraza, Testuen Sinplifikazio Automatikoa, Testuen Konplexutasunaren Ebaluazioa

Abstract

Access to information, knowledge and culture is a right of all citizens. However, written text does not always match our ability to understand what we read. Easy to Read Language (E2R) aims to overcome these comprehension barriers, thus promoting equal opportunities. E2R contains a set of rules that may vary by language, and may also include examples, visual and/or auditory aids not found in the source text. Unfortunately, there are not many language resources or adaptation tools available for E2R. This thesis aims to evaluate existing resources for automatic E2R adaptation and to create new resources and adaptation tools. It will focus on three European languages: German, Spanish and Basque.

Keywords: *Easy-to-Read, Leichte Sprache, Lectura Fácil, Irakurketa Erraza, Automatic Text Simplification, Readability*

1 Sarrera eta motibazioa

Testu idatzia konplexuegia izan daiteke komunikazio-urritasuna duten pertsonen ulertzeko, eta horrek gizartetik baztertua eragiten du. *Easy-to-Read Language* (E2R) komunikazio-urritasuna duten pertsonen inklusioaren funtsezko zati bat da (Hansen-Schirra et al., 2020; Maaß eta Hansen-Schirra, 2022). E2R hizkuntza estandar baten hizkuntza-aldaera da, konplexutasun gutxikoa, eta testuen ulergarritasuna hobetzea du helburu (Nitzke et al., 2022). Bere eginkizunetako bat edukiak eskuragarri jartzea da, eta komunikazio-urritasuna duten pertsonen parte-hartzea bermatzea (Hansen-Schirra eta Maaß, 2020). E2R-ren hiztegia eta gramatika egiturak hizkuntza natural jakin baten oinarritzko hiztegi eta gramatikara mugatzen dira (Hansen-Schirra et al., 2020; Nitzke et al., 2022). Artikulu honetan, E2R terminoa erabiliko dugu E2R hizkuntzaren edozein aldaera izendatzeko, gehien erabiltzen den terminoa baita. Hizkuntza zehatz bati buruz hitz egiterakoan, horri dagokion terminoa erabiliko dugu; *Leichte Sprache* alemanaren kasuan, *Lectura Fácil* gaztelaniaren kasuan, eta *Irakurketa Erraza* euskaraz.

E2R ez da *Plain Language* (hizkera sinplea) (PL) deritzonarekin nahastu behar, haien artean desberdintasun garrantzitsu batzuk baitaude: (1) PL-ak testuari erreparatzen dio, eta E2R-ak, berriz, testuari, ilustrazioei eta maketazioari; (2) E2R normalean desgaitasun intelektuala duten pertsonen zuzenduta dago, PL-ak erronka handiegia

ekar liezaieke-eta; (3) PL, hasiera batean, testu juridikoetan eta gobernu-testuetan oinarritzen zen, esanahi in-trintsekoa dutelako; E2R, berriz, mota guztietako testuetan aplikatzen da (García Muñoz, O., 2013; Maaß, 2020; Balmford, 2002).

E2R eta PL-aren arteko desberdintasunak gorabehera, bereizketa hori oraindik ez da hizkuntza guztietan egi-ten. E2R eta PL ingelesez aurki ditzakegu, baita *Leichte Sprache* eta *einfache Sprache* alemanez ere. Gaztelaniaz, *Lectura Fácil* (E2R-ren baliokidea), *Lenguaje Claro* (PL-ren baliokidea), *Comunicación Clara* (komunikazio ar-gia) eta *Lenguaje Ciudadano* (herritar hizkuntza) daude. Azkenik, euskaraz, *Irakurketa Erraza* baino ez dago. E2R urritasun kognitibo edo intelektuala duten pertsonen zuzenduta dago batez ere, baina beste xede-talde ba-tzuentzat ere onuragarria izan daitezke: adimen-, kognizio- edo garapen-desgaitasuna duten pertsonak, entzumen-desgaitasuna duten pertsonak, alfabetatze txikia duten pertsonak, etorkinak edo irakurketa-errefortzua behar duten haurrak (Bredel eta Maaß, 2016; Hansen-Schirra eta Maaß, 2020; Maaß eta Garrido, 2020; Maaß, 2019). Hizkun-tzaren arabera, E2R-ren xede-irakurle edo erabiltzaileak ere alda daitezke. Esaterako, frogatu da *Leichte Sprache* ez dela egokia alemanaren ikasleei bigarren hizkuntza gisa irakasteko (Ahrens, 2022; Heine, 2017; Kilian, 2017). Hala ere, Irakurketa Erraza euskara ikasten dutenek bigarren hizkuntza gisa edo euskara ulertzeko zailtasunak dituztenek erabil dezakete ¹.

Egokitutako testu kopuru oso mugatua dago, eta daudenak eskuz sortu dira; hala ere, metodo hau oso garestia da, bai denboran, bai diruan ere (Drndarević et al., 2013). Denbora-tarte mugatu baterako soilik garrantzitsuak di-ren kasuetan, albiste testuen kasuan adibidez, oso zaila da datozen edukien erritmoa mantentzea (Bott et al., 2012b; Drndarević et al., 2013). Jadanik egon dira testu estandarrek E2R-era automatikoki egokitzeko saiakera batzuk, HP (Hizkuntzaren Prozesamendua, ingelesez *Natural Language Processing* bezela ezagutzen dena) teknikak erabi-liz. ATS (testuen sinplifikazio automatikoa, *Automatic Text Simplification*) xede-populazio edo gizabanakoentzako testu sinplifikatuak sortzeko HP teknologia da; teknologia honek, testuen edukia egokitzen du ulergarriagoa izan dadin eta hauek sortzeko behar diren eskuzko esfortzuak murrizten ditu (Bott et al., 2012b; Saggion et al., 2015). Dakigunez, gaur egun ez dago hizkuntza estandar batetik E2R bertsiora zehatz-mehatz egokitzea ahalbidetzen duen ATS edo HP teknologiarik. Gainera, E2R egokitzeak ez du esan nahi esaldi edo hitz konplikatuek saihestea soilik. Batzuetan, xede-testuan adibideak sartu behar dira, edo esamoldeak beste batzuegatik aldatu behar dira. E2R egokitzapenaren erronka handienetako bat erabiltzaileen aurretiko ezagutzak aurreikustea eta ezagutza hori ez dagoenean garatzea da (Maaß eta Rink, 2020). Aipatu dena kontuan hartuta, E2R-erako egokitzapen automatikoko tresnen behar handia dagoela ondoriozta dezakegu. E2R-era egokitzea ez da itzulpen-arazo tradizional gisa ulertu behar, beste erronka batzuk izatea ere suposatzen baitu, hala nola informazio berria sartzea. Egokitzapen prozesu hau automatizatzeak, informazioa guztientzako egokituta egotea suposatuko luke, beraz, irisgarritasuna hobetzea.

Tesi honen ekarpen zientifiko nagusiak hurrengoak dira:

- Lehenik eta behin, E2R-aren egokitzapen automatikorako oinarri teorikoa ezartzen dugu, ATS teknikak eta emaitzan eragina duten hizkuntza-ezaugarriak kontuan hartuta.
- Bigarrenik, dauden datu-base paraleloak bildu eta berriak sortuko ditugu, beharrezkoa denean.
- Hirugarrenik, ATS teknikak proposatzen ditugu hizkuntza estandarra E2R-era egokitzeko. Teknika horiek hizkuntza desberdinetara egokitzeko diseinatuko dira, eta, horrela, ingurune eleanitzetan eta hizkuntza ani-tzetan erabili ahal izateko.

2 Arloko egoera eta ikerketaren helburuak

Bigarren atal honetan ATS eta E2R egokitzapenerako egin diren planteamendu desberdinak aztertuko dira. Proiektu honek aztertuko dituen hiru hizkuntzatan (alemana, gaztelania eta euskara) zentratuko da eta hizkuntza horietan garatutako baliabide edota tresna nagusiak deskribatuko ditu. Baliabide hauek hainbat motatakoak izan daitez-ke: corpusak (testu kolekzioak dira, hizkuntzaren konplexutasun maila desberdinetan idatzita edon daitezkeenak), testuen konplexutasunaren ebaluazioak (testuak konplexutasunaren mailaren arabera sailkatzeko erabiltzen dira), hiztegiak (E2R hiztegiek hizkuntza estandarreko termino baten E2R baliokideak eskaintzen dituzte), hizkuntza zuzentzaileak (E2R arauren bat hautsi den agiaztatzen lagun dezakete) eta E2R testuak egokitzeko tresnak (testu bat E2R-era egokitzeko jada diseinatuta dauden sistema edo tresnei deritzo).

¹<https://lecturafacileuskadi.net/blog/irakurketa-erraza-ezinbesteko/>

2.1 Alemana

Leichte Sprache-rako lehen arauak *Inclusion Europe*-ek garatu zituen 1998an (Pottmann, 2019). Hala ere, nahiz eta lehen arau-multzo hau alemanez idatzita egon, E2R-rako arau generikoak ziren, edozein hizkuntza estandarrentzat aplikagarriak izan zitekeenak. 2002-an bi gertakari garrantzitsuen ezarpenek eraman zuten Alemaniako gobernuaren denentzako informazio eskuragarria eskaintzera: *Gesetz zur Gleichstellung von Menschen mit Behinderungen* (elbarrientzako Alemaniako berdintasun-legea) eta *Barrierefreie-Informationstechnik-Verordnung* (teknologia irisgarriaren promulgazioa). 2006-an, *Netzwerk Leichte Sprache* (hizkuntza arrunteko elkarteak) sortu zen, urritasun kognitiboak dituzten pertsonen eta haien arreta-zerbitzuen hornitzaileen ahalduzko mugimenduak eusten dituen elkarteak (Lindholm eta Vanhatalo, 2021); *Leichte Sprache*-ren arauak *Netzwerk Leichte Sprache*-k garatu zituen 2013-an (Pottmann, 2019). Gaur egun *Leichte Sprache* da Alemaniaren gehien erabiltzen den komunikazio erabilera, *einfache Sprache* baino, ziurrenik publiko zabalago batengana hel daitekeelako (Maaß, 2020).

Hainbat corpus daude *Leichte Sprache*-aren inguruan. Klaper et al. (2013) lehen corpus paraleloa garatu zuten, alemana eta *Leichte Sprache* esaldi lerrokatuekin. Harrezkero, beste corpus asko garatu dira: *LeiSa* (2018), Battisti et al. (2019), *TextComplexityDE* (2019), *APA* (2020), *LeiKo* (2020), *KED* (2020), *20m* (2021), *Geasy* (2021), Toborek et al. (2022), *SNIML* (2022), eta *Klexicon* (2022).

Alemanak *Hurraki*² hiztegia ere badauka, *Leichte Sprache*-n idatzitako alemanezko hitzen azalpenekin. E2R-eko arauak hizkuntza zuzentzaile batzuetan ere ezarri dira. Siegel eta Lieske (2014; 2015) aurki ditzakegu, E2R-en arau batzuk ezarri zituztenak *Acrolinx*-en³ eta *LanguageTool*-en⁴. E2R testuak egokitzeko tresnei dagokienez, *EasyTalk* (Steinmetz eta Harbusch, 2020, 2022) aurki dezakegu, *Leichte Sprache*-n idazketan laguntzeko sistema bat. SUMM⁵ ere badugu, edozein testu automatikoki E2R-era bihurtzen laguntzen duen adimen artifizialeko lehen tresna.

Alemanaren irakurgarritasuna ebaluatzeko ikerketa asko egin dira, berriena Mohtaj et al. (2022) izanda. Ikerketa honetan GermEval 2022 alemaneko testuaren konplexutasunaren ebaluazioari buruzko arazo partekatuari bidalitako lan gutzien ikuspegi orokorra eskaintzen dute.⁶

2.2 Gaztelera

Mayol eta Salvador-ek (1999) E2R-aren egoera Katalunian eta Europan zein zen aztertu zuten, eta Katalunian nola ezartzeko gomendio batzuk eman zituzten. 2001-ean, *Easy Language* batzorde bat sortu zen (Lindholm eta Vanhatalo, 2021). *Asociació de Lectura Fácil* (Irakurketa Erraza Elkarteak) 2003-an sortu zen legez; harrezkero, literatur lanen sorkuntzan, E2R klubak sustapenean eta informazio irisgarria sortzeko prestakuntza ematen aritu da. Nomura et al. (2012) *Directrices para materiales de lectura fácil*-en (Irakurketa Erraza materialerako gidalerroak) bigarren edizioa itzuli zuen. 2016-an, eskualdeko hainbat elkarteek sortu zuten *Red de Lectura Fácil* (Irakurketa Errazeko Sarea).

Espainiako ATS-en proiektu nagusietako bat *Simplext* (Saggion et al., 2015) da. Proiektu hau 2010-etik 2013-ra arte egon da martxan, eta gaztelaniazko testuetan datu base paralelo bat sortzea barne hartu du (Bott et al., 2012b), baita ATS-erako aurretiazko tresnak garatu ere. *Lectura Fácil*-erako beste corpus batzuk hauek dira: *Mitkov eta Štajner* (2014; 2019), *Newsela* (Xu et al., 2015), *EASIER* (Alarcón García, 2022), *ALEXIS* (Ferrés eta Saggion, 2022), *Simple TICO-19* (Shardlow eta Alva-Manchego, 2022), *IrekiaLF.es* (2022) eta *CLARA-MeD* 2022.

Testuen konplexutasunaren ebaluazioari dagokionez, *Coh-Meetrix-Esp* (2016) dugu, 45 irakurgarritasun-indize kalkulatu dituenak. *MultiAzterTest* (2021) ingelesez, gaztelaniaz eta euskaraz lan egiten duen analizatzaile eleantun bat da, eta kohesio, hizkuntza eta irakurgarritasun 125 neurri baino gehiago aztertzen ditu. *Ventosa Pérez*-en proiektuan (2022), irakurgarritasuna ebaluatzeko web aplikazio bat ezarri zuten. *BERTIN for RA* (2022) hiru testu-maila sailkatzeko gai da. Azkenik, *Clara*⁷ (2021) dago, testuen argitasuna egiaztatzeko aukera ematen duen tresna. Dagoeneko aipatu dugu *Simplex* (2012b) E2R egokitzeko tresna, baina gehiago daude. *LexSIS* (Bott et al., 2012a) izan zen gaztelaniazko sinplifikazio lexikoaren lehen hurbilketa. *CASSA* (2015) hitz baten sinonimo sinpleagoak sortzen dituen algoritmoa da. Azkenik, *Cumbicus-Pineda et al.* (2021) informazio sintaktikoa gehitu du edizioan

²<https://hurraki.de/wiki/Hauptseite> (azken kontsulta: 2023-01-25)

³Acrolinx dokumentazio teknikoaren egileei laguntzeko software paketea da

⁴LanguageTool 2003-az geroztik garatutako kode irekiko testuak egiaztatzeko software bat da.

⁵<https://summ-ai.com/en/> (azken kontsulta: 2023-02-01)

⁶*Leichte Sprache*-ko korronteen ikuspegi zehatzagoa eta corpus eta tresnei buruzko xehetasun gehiago lortzeko, kontsultatu Madina et al. (2023)

⁷<https://comunicacionclara.com> (azken kontsulta: 2023-01-24)

oinarritutako *EditNTS* (Dong et al., 2019) sisteman.

2.3 Euskara

Euskal Herriko E2R Elkarteak ⁸ literatura, kultura eta informazioa eskura jartzen ditu. E2R proiektua Kataluniatik ekarri zuten 2012-an. E2R-en formazioak, liburuak eta berriak eskaintzen dituzte. *Irakurketa Erraza* lehen hizkuntza ezberdina dutenei euskara irakasteko ere erabiltzen da batzuetan. *Irakurketa Erraza*-rako dauden ATS tresna gehienak IXA taldeak garatu ditu ⁹.

Bi *Irakurketa Erraza* corpus daude. *CBST* (2018) (*the Corpus of Basque Simplified Texts (CBST)*) edo Euskarazko Testu Sinplifikatuen Corpora (ETSC) euskaraz 227 esaldi original eta esaldi bakoitzaren bi bertsio sinplifikatu biltzen ditu. Esaldiak gizarte zientzien, medikuntzaren eta teknologiaren gaietarako dagozkie. Perpausen bertsio sinplifikatuak bi ikuspegi ezberdin jarraituz sortu ziren, egiturazkoa eta intuitiboa. Bigarren corpusa *LBSPC* (2014) da (*The Leveled Basque Science Popularisation Corpus (LBSPC)*) 2 mailatako 400 testuz osatuta dago. Testuak *ZerNola* ¹⁰ webgunetik atera ziren, 12 urte arteko haurren artean zientzia ezagutarazteko webgunetik; hau testu bilduma “erraza” zen. Gainerako 200 testuak *Elhuyar Aldizkaria* ¹¹ euskarazko zientzia eta teknologiari buruzko aldizkarietik atera ziren; hau testu bilduma “konplexu” moduan. Corpus hau bai *ErreXail* (2014) bai *MultiAzter-Test* (2021) ikasketetan erabili zen, hauek baitira euskaraz idatzitako testuen konplexutasuna ebaluatzeko azterketa tresna bakarrak.

E2R egokitzapen tresnei dagokienez, *Aranzabe et al.* (2012) bi corpusetatik (EPEC eta Consumer) hartutako esaldi luzei buruzko azterketa linguistiko bat egin zuten, eta hortik abiatuta, sintaxi-zuhaitz bati aplikatutako eskuz idatzitako arauak erabiliz, sinplifikazio sintaktikoa proposatu zuten. Lan honen ostean, *Gonzalez-Dios*-ek (2016) sinplifikazio sintaktikorako arauetan oinarritutako sistema bat aurkeztu zuen, testuak maila ezberdinetan sinplifikatzen zituen.

2.4 Helburua

E2R-eko erabiltzaileen egokitzapen beharrak identifikatu dira dagoeneko neurri batean. Hala ere, oraindik ez da konponbiderik lortu egokitzapen automatikorako tresnak garatzeko. Lehen esan bezala, itzulpen automatiko tradizionalak eta ATS teknikak ez dira egokiak E2R egokitzapena egiteko. Proiektu honen helburu nagusia HP tresna eta baliabideak garatzea da, hizkuntza estandarren eta E2R-en arteko egokitzapen prozesua automatizatzeko.

3 Ikerketaren muina

Aurreko atalean adierazi bezala, hizkuntza bakoitzerako dauden baliabideak desberdinak dira. Proiektu honetan, *Leichte Sprache*-aren artearen egoera abiapuntu eta inspirazio gisa har daiteke E2R aldaerak hainbeste garatu ez dituzten, edota erakunde publiko eta ofizialetan alemanak bezainbeste integratu ez dituzten beste hizkuntzentzat (gaztelania eta euskara, esaterako). Gaztelaren kasua dugu, alemana bezalako hizkuntza flexio/fusioa dena, eta euskararen kasua, aglutinatzailea. Interesgarria da ikustea ea alderik dagoen dauden tresnak eleaniztunak bihurtzeko orduan, bi hizkuntza mota desberdinetara egokituko baikenituzke. Gainera, euskara inguruko hizkuntzetatik ezberdina da, eta ikasten ari den jende askok etekina atera diezaioke *Irakurketa Erraza*-n idatzitako testuei. Azpimarratu behar da, halaber, euskara eta gaztelania eskualde berean hitz egiten direla, non hizkuntza biak ofizialak diren eta lege-testuetan, albisteetan eta bestelakoetan agertzen diren ¹². Interesgarria litzateke kasu honetaz eztabaidatzea (baita Kanadaren edo hizkuntza ofizial bat baino gehiago duten beste herrialde/eskualdeen kasuan ere hitz egin litekeen bezala).

Erakunde ofizialen testu administratiboetan jarriko dugu arreta, hala nola, estatuko webguneak edo herritarren informaziorako webguneak. Izan ere, webgune hauek biztanleria osoarentzat garrantzitsua den informazioa daukate, eta beti izan behar dute ahalik eta eskuragarrien alderdi guztietan, bai konplexutasun sintaktiko edo semantikoan, bai diseinuan. Gainera, testu hauek ez dute eleberriek edo beste testu mota batzuek izan dezaketen konplexutasun semantiko edo metaforikoa adierazten. Gogoratu behar dugu E2R-erako egokitzapen automatikorako tresna baterako lehen urratsa dela, eta denbora mugak ere badaudela, beraz, ezin ditugu mota guztietako

⁸<https://lecturafacileuskadi.net/blog/irakurketa-erraza-ezinbesteko/> (azken kontsulta: 2023-01-24)

⁹<http://www.ix.a.es/> (azken kontsulta: 2023-01-24)

¹⁰<http://www.zernola.net/>

¹¹<https://aldizkaria.elhuyar.eus>

¹²Frantsesa ere hizkuntza ofiziala da lurralde batzuetan euskararekin batera, hala ere, bere analisia artikulu honen irismenetik kanpo dago.

testuak landu. Ikerketa gehiago bideratu liteke tresna horiek beste testu mota batzuekin ebaluatzerako. Hizkuntzak funtsezko zeregina du gaur egungo gizartean; E2R-aren hedapenak eta E2Ran oinarritutako hizkuntza baliabideak eta egokitzapen tresnak sortzeak, hainbat arrazoiengatik hizkuntza estandarra ulertzeko zailtasunak dituztenak barneratzen lagunduko dute.

4 Ondorioak

Baliabide mota eta kopuru desberdinak daude E2R oinarritzen den hizkuntzaren arabera. Baliabideak ez dira berdinak hizkuntza guztietan, batzuek besteek baino gehiago dituzte; hala ere, ez dago argi zein den horren arrazoia. E2R-ren beharraz jabetu ez izanagatik, inbertsio faltagatik edo beste baliabide mota batzuen faltagatik izan daiteke. Bestalde, ATS sistemetarako ahalegin asko egin dira; hala ere, sistema hauek ez dituzte E2R arauak betetzen eta ez dute laborpenik egiten. Hori dela eta, nahiz eta testuak E2R-era egokitzeko lagungarri izan, oraindik ez dira nahikoak. Orokorrean, E2R baliabide batzuk existitzen badira ere, hainbat izaeratakoak dira.

Hizkuntza-baliokideen barruan ez dago koherentziarik; paraleloak, konparagarriak, edota lerrokatuak izan daitezke eta konplexutasun maila ezberdineko testuak izan ditzakete. Arrazoi ezberdinengatik izan daiteke hori: (1) helburu ezberdinetarako diseinatu dira, edo (2) ez dago corpusak sortzeko datu nahikorik. Web-orri batzuek informazioa hizkuntza estandarrean eta E2R-an eskaintzen dute, baina beste batzuek ez. Horrek zaildu egiten du esaldiekin lerrokaturiko corpus paraleloak sortzea. Bestalde, webguneetan aurki daitezkeen E2R deritzon baliabideak askotan ez daude dagozkien aleman dokumentu bakar bati lotuta, alemaniar dokumentu anitzen goi-mailako laborpenak dira baizik (Klaper et al., 2013). Gaztelaniaren kasuan, E2R testuak sortzeko laguntzetarako ekarpen berri batzuk daudela ikus dezakegu. Hala ere, Bernal eta Grossi-k (2022) nabarmendu duenez, E2R sustatzeko proposatutako ekimenak ez dira eraginkortasunez aplikatu. Komunitate batzuek testu estandarrek E2R-era egokitze prozesua hobetzeko ahaleginak egin arren, oraindik lan asko dago egiteko. Euskararen kasuan, adibidez, testuen sinplifikazioaren inguruan ikerketa asko egiten ari dela ikus dezakegu. Hala ere, ez daude corpus nahikoak eskuragarri ikerketa horiek egiteko.

Esandako guztia kontuan hartuta, esan genezake E2R eta ATS ikasketa-esparru errepikakorrak izan direla azken urte hauetan eta badirudi horrela jarraituko dutela. Hala ere, oraindik asko dago egiteko, bereziki sortzen ari diren tresna eta baliabide guztien ezarpenari dagokionez. Horiek ezartzeak komunikaziorako ezintasunak dituzten pertsona guztientzako informazioa eskura izango luke.

5 Etorkizunerako planteatzen den norabidea

Artikulu honetan, E2R-aren artearen egoera aurkeztu dugu alemanez, gaztelaniaz eta euskaraz. Lehen urratsa corpus handi bat sortzea izango litzateke, lehendik dauden corpusak batuz. Beste aukera bat webean arakaturik eta E2R itzultzaile profesionalen laguntzarekin corpus berriak sortzea izango litzateke. Baliabide gehien dituen hizkuntza alemana denez, E2R egokitzapen tresnak hizkuntza horretan oinarritu beharko lirateke lehenik. Aleman estandarretik testuak *Leichte Sprache*-ra egokitzeko tresna bat sortu ondoren, eleaniztun bilakatzeari ekin genioke.

Ebaluazio prozesua hiru zatiz osatuko da: lehenik eta behin, xede-testuen konplexutasunaren ebaluazioa egingo dugu haren zailtasuna neurtzeko, eta tresna automatikoak erabiliko ditugu baldintzazko perpausak, esamolde pasiboak eta E2R testu batean egon behar ez diren beste ezaugarri linguistiko batzuk detektatzeko. Ondoren, eskuzko ebaluazioa egingo dugu sorburu-testuaren informazio guztia xede-testuan sartuta dagoela ziurtatzeko, eta esanahia gordetzen dela ziurtatzeko. Edizio osteko fasea izango da, non beharrezko aldaketa guztiak egingo diren. Azkenik, hartzailearen ebaluazioa egingo da, hau da, ebaluazio bat non xede-erabiltzaileek xede-testua ontzat emango duten edo ez. Ebaluazio honek aukera emango digu ez bakarrik sistemak ondo edo gaizki funtzionatzen duen ikusteko, baita E2R-eko etorkizuneko erabiltzaileek emaitza nola hautematen duten jakiteko ere.

Erreferentziak

- Ahrens, S. (2022). Easy Language Translations for Second Language Learners–Worthwhile Concept or Didactic Mistake? *Translation, Mediation and Accessibility for Linguistic Minorities*, 128:175.
- Alarcón García, R. (2022). Lexical Simplification for the Systematic Support of Cognitive Accessibility Guidelines. <https://doi.org/10.1145/3471391.3471400>.
- Aranzabe, M. J., De Ilarraz, A. D., & Gonzalez-Dios, I. (2012). First Approach to Automatic Text Simplification in Basque. In *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) workshop (LREC 2012)*, 1–8.

- Aumiller, D. & Gertz, M. (2022). Klexikon: A German Dataset for Joint Summarization and Simplification. *arXiv preprint arXiv:2201.07198*.
- Baeza-Yates, R., Rello, L., & Dembowski, J. (2015). Cassa: A context-aware synonym simplification algorithm. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1380–1385.
- Balmford, C. (2002). Plain Language: Beyond a 'Movement'. Repositioning Clear Communication in the Minds of Decision-Makers. In *Fourth Biennial Conference of the PLAIN Language Association International, Toronto, Canada*.
- Battisti, A. & Ebling, S. (2019). A corpus for Automatic Readability Assessment and Text Simplification of German. *arXiv preprint arXiv:1909.09067*.
- Bengoetxea, K. & Gonzalez-Dios, I. (2021). MultiAzterTest: a Multilingual Analyzer on Multiple Levels of Language for Readability Assessment. *arXiv preprint arXiv:2109.04870*.
- Bernal, M. P. C. & Grossi, M. E. (2022). *Translation, Mediation and Accessibility for Linguistic Minorities*, volume 128. Frank & Timme GmbH.
- Bott, S., Rello, L., Drndarevic, B., & Saggion, H. (2012a). Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *COLING*.
- Bott, S., Saggion, H., & Mille, S. (2012b). Text simplification tools for spanish. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 1665–1671.
- Bredel, U. & Maaß, C. (2016). *Leichte Sprache: Theoretische Grundlagen? Orientierung für die Praxis*. Bibliographisches Institut GmbH.
- Campillos Llanos, L., Terroba Reinares, A. R., Zakhir Puig, S., Valverde, A., & Capllonch-Carrión, A. (2022). Building a comparable corpus and a benchmark for spanish medical text simplification.
- Cumbicus-Pineda, O. M., Gonzalez-Dios, I., & Soroa, A. (2021). A Syntax-Aware Edit-based System for Text Simplification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, 324–334.
- Das Netzwerk Leichte Sprache (2013). Die Regeln für Leichte Sprache. https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_\Leichte_\Sprache.pdf (last accessed: 2023-02-14),.
- de la Rosa, J., Ponferrada, E. G., Villegas, P., de Prado Salas, P. G., Romero, M., & Grandury, M. (2022). BERTIN: Efficient Pre-Training of a Spanish Language Model using Perplexity Sampling. *Procesamiento del Lenguaje Natural*, 68:13–23.
- Dong, Y., Li, Z., Rezagholizadeh, M., & Cheung, J. C. K. (2019). EditNTS: A Neural Programmer-Interpreter Model for Sentence Simplification through Explicit Editing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3393–3402, Florence, Italy. Association for Computational Linguistics.
- Drndarević, B., Štajner, S., Bott, S., Bautista, S., & Saggion, H. (2013). Automatic text simplification in spanish: A comparative evaluation of complementing modules. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 488–500. Springer.
- Ferrés, D. & Saggion, H. (2022). ALEXSIS: A Dataset for Lexical Simplification in Spanish. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, 3582–3594.
- García Muñoz, O. (2013). *Lectura fácil - Métodos de Redacción y Evaluación*. Ministerio de Educación.
- Gonzalez-Dios, I. (2016). *Euskarazko Egitura Sintaktiko Konplexuen Analisisirako eta Testuen Sinplifikazio Automatikorako Proposamena. Readability Assessment and Automatic Text Simplification. The Analysis of Basque Complex Structures*. PhD thesis, University of the Basque Country (UPV/EHU).
- Gonzalez-Dios, I., Aranzabe, M. J., de Ilarraza, A. D., & Salaberri, H. (2014). Simple or Complex? Assessing the Readability of Basque Texts. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*, 334–344.
- Gonzalez-Dios, I., Aranzabe, M. J., & Díaz de Ilarraza, A. (2018). The corpus of Basque simplified texts (CBST). *Language Resources and Evaluation*, 52(1):217–247.

- Gonzalez-Dios, I., Gutiérrez-Fandiño, I., Cumbicus-Pineda, O. M., & Soroa, A. (2022). IrekiaLF.es: a New Open Benchmark and Baseline Systems for Spanish Automatic Text Simplification. In *Proceedings of the TSAR 2022 Workshop*.
- Hansen-Schirra, S., Bisang, W., Nagels, A., Gutermuth, S., Fuchs, J., Borghardt, L., DEILEN, S., Gros, A.-K., Schiff, L., & Sommer, J. (2020). Intralingual Translation into Easy Language—or How to Reduce Cognitive Processing Costs. *Easy Language Research: Text and User Perspectives*. Berlin: Frank & Timme, 197–225.
- Hansen-Schirra, S. & Maaß, C. (2020). Easy Language, Plain Language, Easy Language Plus: perspectives on comprehensibility and stigmatisation. *Easy Language Research: Text and User Perspectives*, 2:17.
- Hansen-Schirra, S., Nitzke, J., & Gutermuth, S. (2021). An Intralingual Parallel Corpus of Translations into German Easy Language (Geasy Corpus): What Sentence Alignments Can Tell Us About Translation Strategies in Intralingual Translation. In *New Perspectives on Corpus Translation Studies*, 281–298. Springer.
- Hauser, R., Vamvas, J., Ebling, S., & Volk, M. (2022). A Multilingual Simplified Language News Corpus. In *2nd Workshop on Tools and Resources for READING Difficulties (READI)*, 25.
- Heine, A. (2017). Deutsch als Fremd- und Zweitsprache—eine besondere Form Leichter Sprache? Überlegungen aus der Perspektive des Faches DaF/DaZ. *Leichte Sprache “im Spiegel theoretischer und angewandter Forschung*. Berlin: Frank & Timme, 401–414.
- Jablotschkin, S. & Zinsmeister, H. (2020). LeiKo: A Corpus of Easy-to-Read German.
- Jach, D. (2020). Korpus Einfaches Deutsch (KED). <https://daniel-jach.github.io/simple-german/simple-german.html> (last accessed: 2022-09-08).
- Kilian, J. (2017). „Leichte Sprache“, Bildungssprache und Wortschatz—Zur sprach- und fachdidaktischen Wertigkeit der Regelkonzepte für „leichte Wörter“. Eds. BETTINA BOCK, ULLA FIX, and DAISY LANGE. *“Leichte Sprache” im Spiegel theoretischer und angewandter Forschung*, 189–209.
- Klaper, D., Ebling, S., & Volk, M. (2013). Building a German/Simple German Parallel Corpus for Automatic Text Simplification.
- Lange, D. (2018). Comparing ‘Leichte Sprache’, ‘einfache Sprache’ and ‘Leicht Lesen’: A Corpus-Based Descriptive Approach. In Eds. SUSANNE J. JEKAT, and GARY MASSEY. *Barrier-free communication: methods and products: proceedings of the 1st Swiss conference on barrier-free communication*. Winterthur: ZHAW Digital Collection, 75–91.
- Lieske, C. & Siegel, M. (2014). Verstehen Leicht Gemacht. *technische kommunikation*, 1:44–49.
- Lindholm, C. & Vanhatalo, U. (2021). *Handbook of Easy Languages in Europe*. Frank & Timme.
- Maaß, C. (2019). Easy Language and Beyond: How to Maximize the Accessibility of Communication. Invited Plenary Speech at the Klaara 2019 Conference on Easy-to-Read Language Research (Helsinki, Finland. 19-20 September 2019).
- Maaß, C. (2020). *Easy Language—Plain Language—Easy Language Plus: Balancing comprehensibility and acceptability*. Frank & Timme.
- Maaß, C. & Garrido, S. H. (2020). Easy and Plain Language in Audiovisual Translation. *Easy language research: Text and user perspectives*, 2:131.
- Maaß, C. & Hansen-Schirra, S. (2022). Removing Barriers: Accessibility as the Primary Purpose and Main Goal of Translation. *Translation, Mediation and Accessibility for Linguistic Minorities*, 128:33.
- Maaß, C. & Rink, I. (2020). Scenarios for easy language translation: How to produce accessible content for users with diverse needs. *Easy Language Research: Text and User Perspectives*. Berlin: Frank & Timme, 41–56.
- Madina, M., Gonzalez-Dios, I., & Siegel, M. (2023). Easy-to-Read in Germany: a Survey on its Current State and Available Resources. In *To appear in the proceedings of the LCT’23 conference*.
- Mayol, M. C. & Salvador, E. (1999). *Materials de lectura-fàcil: Anàlisi, directrius internacionals i proposta per a elaborar aquests materials a Catalunya*. FUS, Grup de Fundacions.
- Mitkov, R. & Štajner, S. (2014). The fewer, the better? A Contrastive Study About ways to Simplify. In *Proceedings of the Workshop on Automatic Text Simplification-Methods and Applications in the Multilingual Society (ATS-MA 2014)*, 30–40.

- Mohtaj, S., Naderi, B., & Möller, S. (2022). Overview of the GermEval 2022 Shared Task on Text Complexity Assessment of German Text. In *Proceedings of the GermEval 2022 Workshop on Text Complexity Assessment of German Text*, 1–9.
- Naderi, B., Mohtaj, S., Ensikat, K., & Möller, S. (2019). Subjective Assessment of Text Complexity: A Dataset for German Ganguage. *arXiv preprint arXiv:1904.07733*.
- Nitzke, J., Hansen-Schirra, S., Habig, A.-K., & Gutermuth, S. (2022). Translating Subtitles into Easy Language: First Considerations and Empirical Investigations. *Translation, Mediation and Accessibility for Linguistic Minorities*, 128:127.
- Nomura, M., Skat, Nielsen, G., & Tronbacke, B. (2012). Directrices materiales de lectura fácil.
- Pottmann, D. M. (2019). „Leichte Sprache and Einfache Sprache”–German Plain Language and teaching DaF German as a Foreign Language. *Studia Linguistica*, 38:81–94.
- Quispesaravia, A., Perez, W., Cabezudo, M. S., & Alva-Manchego, F. (2016). Coh-Metrix-Esp: A complexity analysis tool for documents written in Spanish. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4694–4698.
- Rios, A., Spring, N., Kew, T., Kostrzewa, M., Säuberli, A., Müller, M., & Ebling, S. (2021). A New Dataset and Efficient Baselines for Document-level Text Simplification in German. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, 152–161, Online and in Dominican Republic. Association for Computational Linguistics.
- Saggion, H., Štajner, S., Bott, S., Mille, S., Rello, L., & Drndarevic, B. (2015). Making it simplext: Implementation and evaluation of a text simplification system for spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):1–36.
- Säuberli, A., Ebling, S., & Volk, M. (2020). Benchmarking Data-driven Automatic Text Simplification for German. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding Difficulties (READI)*, 41–48, Marseille, France. European Language Resources Association.
- Shardlow, M. & Alva-Manchego, F. (2022). Simple TICO-19: A Dataset for Joint Translation and Simplification of Covid-19 Texts. In *Proceedings of the 13th Language Resources and Evaluation Conference, Marseille, France, June. European Language Resources Association*.
- Siegel, M. & Lieske, C. (2015). Beitrag der Sprachtechnologie zur Barrierefreiheit: Unterstützung für Leichte Sprache. *Zeitschrift für Translationswissenschaft und Fachkommunikation*, 8(1):40–78.
- Štajner, S., Saggion, H., & Ponzetto, S. P. (2019). Improving Lexical Coverage of Text Simplification Systems for Spanish. *Expert Systems with Applications*, 118:80–91.
- Steinmetz, I. & Harbusch, K. (2020). Enabling Fast and Correct Typing in ‘Leichte Sprache’ (Easy Language). In *Proceedings of the The Fourth Widening Natural Language Processing Workshop*, 64–67.
- Steinmetz, I. & Harbusch, K. (2022). A text-writing system for easy-to-read german evaluated with low-literate users with cognitive impairment. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, 27–38.
- Toborek, V., Busch, M., Boßert, M., Welke, P., & Bauckhage, C. (2022). A New Aligned Simple German Corpus. *arXiv preprint arXiv:2209.01106*.
- Torrijos, C. & Oquendo, S. (2021). ¡Hola! Soy Clara y mido la claridad de tu texto. *Archiletras científica: revista de investigación de lengua y letras*, (6):119–134.
- Ventosa Pérez, A. (2022). Servicio Web de Lectura Fácil: Calculador de Índices de Lecturabilidad.
- Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in Current Text Simplification Research: New Data can Help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

6 Eskerrak eta oharrak

Lan hau Melanie Siegel (Darmstadt University of Applied Sciences), eta Itziar Gonzalez Dios-en (Euskal Herriko Unibertsitatea) zuzendaritzapean egiten ari naizen doktorego tesi batean oinarritzen da.