

Testu-ahots bihurketa (tab) euskaraz gaur egun

Inmaculada Hernáez

Elektronika eta Telekomunikazio Saila. UPV/EHU.

Iñaki Gaminde

Hizkuntza eta Literaturaren Saila. UPV/EHU.

Txillardegi

Pilartzo Etxeberria

Rosa Gandarias

Euskal Filologia Saila. UPV/EHU.

Artikulu honetan euskaraz gaur egun testu-ahots bihurketa zertan den azaltzen da, arazo teknikoak eta linguistikoak kontutan hartzen direlarik. Prozesaketa linguistikoaren atal nagusiak, aurreprozesaketa linguistikoa, transkripzio fonetikoak, azentu azterketa eta azterketa prosodikoa deskribatu ondoren, teknologia honek informazio sistemetan, hizkuntz irakaskuntzan, pertsona gutxituen laguntzan, ahots ikerketa oinarritzekoan, ikerketa linguistikoan eta multimedietan euki ditzakeen aplikazioak aurkezten dira. Azkenik, euskara alorrean topatu ditugun arazo linguistiko nagusiak eta hartu behar izan ditugun behin behineko erabakiak agertu ostean, hurrengo urratsak iradokitzen dira.

Giltza hitzak: *Sintesia, fonetika, hizkuntz-normalizazioa.*

En este artículo se describe la situación de la conversión de texto a voz en euskara, teniendo en cuenta tanto los problemas lingüísticos como técnicos. Después de describir las diferentes etapas del procesado lingüístico, preprocesado, transcripción fonética, análisis acentual y prosódico, se contemplan las aplicaciones que puede tener esta tecnología en sistemas de información, enseñanza de idiomas, ayudas a personas discapacitadas, investigación básica de voz, investigación lingüística y en multimedia. Finalmente tras mostrar los principales problemas lingüísticos que hemos encontrado con el euskara así como las decisiones tomadas durante el desarrollo del sistema, se sugieren los próximos pasos.

Palabras clave: *Sintaxis, fonética, normalización lingüística.*

In this article text to speech conversion for the basque language -euskara- is described, taking in to account linguistic and technical difficulties. After a description of the different phases of the linguistic processing module, i.e. pre-processing, phonetic transcription, accentual and prosodic analysis, some applications of this technology are presented, such as information systems, languages learning, handicapped aids, basic speech and linguistic research and multimedia applications. Finally we show some of the linguistic problems taht we have found, and the decisions taken in our development, and we suggest the next steps to be done.

Key words: *Sintaxis, fonetica, linguistic normalization.*

SARRERA

Elektronika eta konputagailuen teknologiaren garapenak makinaren erabilpen masiboa ekarri du informazio-prozesuan. Kasu gehienetan informazio honen iturria gizakia izaten da eta berau azken erabiltzailea, informazio horren hartzailea, hain zuzen. Halandaze, gizaki eta makinaren artean, norabide bietan, informazioa trukatzeko sistimak garatzeko beharizan handia dago. Komunikazio honetan bitartekorik egokiena ahotsa bera da, zeren berau baita komunikazioetan erabiltzen den bitartekorik garrantzitsuenena. Ahots-sintesiak gizaki eta makinaren arteko komunikazioa ahalbidetzen du makinatik gizakiaganako norabidean; aitzitik, ahots-errekonozimenduak komunikazio horren kontrako norabidean iharduten du.

Jakina denez, sarri askotan makina batekin komunikatzeko bitartekorik onena ez da izaten ahotsa. Testu luzeak irakurri egin behar badira, adibidez, askozaz errazagoa da oradenagailuaren pantailan bertan egitea. Hala ere, gizaki eta makinaren arteko elkarrizketa interaktiboa gertatzen den aplikazioetan ahotsaren bidezko komunikazioak abantaila handiak ematen ditu (Witten,1982; Mariño, Nadeu eta Llisterri,1987), arrazoi nagusi bigatik: batetik, gizakiak eskuz edo ikusmenaz beste zerbait egiten duen artean daukan mezu bat entzuteko ahalmenagatik, eta bestetik, telefono sarearen erabilera hedatu unibertsalak ahotsaren bidez ordenagailu batekin konektatzea ahalbidetzen du, gainerako tresneria berezirik gabe (datu garraiorako esate baterako behar izaten dena barik). Ez da ahantzi behar, bestalde, ikusmen, entzumen edo mintzamen gabeziak dauzkaten pertsonen komunikazioan sintesi sistimek euki dezaketen garrantzia, berau arrazoi nahikoa izango litzateke sistema hauen erabilera suspertzeko.

Sintesi sistimek jadetsi duten ulergarritasun maila nahikoa altua bada ere, sistema guztiek akats multzo bat daukate, akatsok naturaltasun eza izenaren bidez labur daitezke. Sintesi sistimek dauzkaten prosodia ereduak (iraupena, energia eta oinarritzko maiztasuna) naturaltasun eza horren errudunak izaten dira (Kaplan eta Lerner,1987).

Lan honetan, laburki bada ere, testu-ahots bihurteta (TAB) euskeraz gaur egun zertan den azaldu nahi dugu, nola baita zeintzuk diren alor honetan dauzkagun arazo nagusiak, teknikoak zein linguistikoak, eta aurrerantzean eginkizun dauzkagunak ere. Artikuluaren lehen sekzioan TAB zer den azalduko dugu, jarraian prozesaketa linguistikoaren prozesua azalduko dugu euskarari dagokion esparruan. Hirugarren sekzioan TAB-en aplikazioak aurkeztuko ditugu. Jarraian, euskaraz teknologia honen garapenean dauzkagun arazo linguistikoak aurkeztuko ditugu, hurrengo sekzioan hartu behar izan ditugun behin behineko erabakiak gain-begiratzeko. Azkenik, seigarren sekzioan TAB-en garapenerako euskaraz ikusten ditugun eman beharreko urratsak proposatuko ditugu.

1. TAB ZER DEN.

Testu-ahots bihurketa egiten duen sistema **edozein** testu irakurtzeko gai izan behar da eta gainera, irakurketa era ulergarri eta naturalez egin behar da. Bestalde, ez dira nahastu behar testu-ahots bihurgailuak eta hitz egiten duten makinak, izan ere, beti ez dira zertan berdinak izan; gure interesa esaldi berriak automatikoki jaulkitzeko gai diren sistemetan kokatzen da. Hitz isolatuak edo esaldi zatiak kateatzeko gai diren sistimak, ahots-erantzun sistema deritzenak, oso lexiko murrizta behar izaten denean (ehun bat hitz gutxi gora behera) baino ezin aplika daitezke, hala nola jaulki behar diren esaldiek oso egitura mugatua daukatenean ere; honelako sistimak tren geltokietan sarrerak eta irteerak adierazteko erabiltzen dira adibidez. TAB-en esparruan ezinezkoa izango litzateke hizkuntza naturalean egin daitezkeen hitz eta esaldi guztiak grabatzea eta gordetzea; guzti horregatik, egokiagoa litzateke TAB ahotsa automatikoki jaulkitzeko eratzat jotzea, ahoskatzeko esaldien grafema fonema transkripzio baten bidez egina.

Seinalaren prozesu digitalean eta adimen artifizialean egin diren aurrerapausoengatik ere, egun giza-ahotsen antzera agin dezakeen ordenagailu sistimatik nahikoa urrun gaudela esan beharrean gaude. Azken batez, eginahalak egiten ari gara giza-irakurketa eta mintzamina simulatzeko, testu baten irakurketa eta enuntziatio prozesua oso konplexua da eta faktore askoren menpe dago.

Labur-bilduz, prozesu hau honako honetan datzala esan dezakegu, hots, lehenbizi irudia begien sentsoreek harrapatuta, burmuinera eramaten da inpultso elektrikoen bidez, bertan prozesatu egiten da eta aginduak bialtzen dira biriki, ahokorda eta artikulazio organoetara soinuak jaulkitzeko. Soinu hau norberaren belarriaren bidez atzera hartzen da eta burmuinak denbora errealean monitorizatu izaten da aho-barrunbearen osaketa, ebakitako seinalearen energia eta intonazioa egokitzen.

1. Ahozko irakurketaren prozesuaren irudia.

Seinalaren prozesu digitalean eta adimen artifizialean egin diren aurrerapau-soengatik ere, egun giza-ahotsen antzera agin dezakeen ordenagailu sistimatik nahikoa urrun gaudela esan beharrean gaude. Azken batez, eginahalak egiten ari gara giza-irakurketa eta mintzamina simulatzeko, testu baten irakurketa eta enuntziatio prozesua oso konplexua da eta faktore askoren menpe dago.

Labur-bilduz, prozesu hau honako honetan datzala esan dezakegu, hots, lehendabizi irudia begien sentsoreek harrapatuta, burmuinera eramaten da inpultso elektrikoan bidez, bertan prozesatu egiten da eta aginduak bialtzen dira biriki, ahokorda eta artikulazio organoetara soinuak jaulkitzeko. Soinu hau norberaren belarriaren bidez atzera hartzen da eta burmuinak denbora errealean monitorizatu izaten da aho-barrunbearen osaketa, ebakitako seinalearen energia eta intonazioa egokitze-ko.

2. PROZESAKETA LINGUISTIKOA EUSKARAZ

Sekzio honetan prozesaketa linguistikoari dagokion prozesu osoaren lehen atalaren aurkezpena egingo dugu; prozesu hau egiten den hizkuntzaren menpe dagoenez gero, hemen aurkezten duguna geuk, euskararen arabera, egin duguna azalduko dugu. Gauza jakina da karakterren bidezko idazketa hizkuntza baten hotsak kodifikatzeko bitartekoa dena. Orohar, ez da egoten mezua kodifikatzeko erabiltzen diren karakter eta aipatutako mezu horren ahozko adierazpenaren hotsen arteko korrespondentziarik. Testu ortografiko baten karakterrak era mintzatuan interpretatu ahal izateko, karakter bakoitza dagoen testuingurua kontutan hartu behar da, eta goragoko maila bakoitzean, hitz bakoitzarena ere bai. Euron ahoskapenerako, maila segmentalaz gain, azken gauzatze akustikoak intentsitate, iraupen eta intonazio aldatetako ere izan behar ditu hots bakoitzeko- maila suprasegmentala-, testuaren beraren adierazpenen araberrakoak (puntuazio markak, esklamazioak, etab.).

Ondoko irudian prozesaketa linguistikoa osatzen duten azpi-prozesuak ikus daitezke:

3. TAB-EN APLIKAZIOAK

Kalitate handiko TAB-en aplikazio potentzialak asko dira eta arlo desberdinetakoak. Hemen adibide batzuk emango ditugu:

- Informazioa emateko sistimak: Gaur egun telefonoaren bidez testu fitxeroetan sar daiteke. Dei askok elkar-trukatze maila handirik behar ez dutenez gero (%70 gutxi gora behera), sartze prozesua eta informazio hornidura automatizatzea merezi duke. Testu fitxeroak mezu hutsak izan daitezke (ekintza kulturalak, zinemak, antzokiak, etab.) edo erabiltzaileak bere ahotsez kontsulta ditzakeen datu-base handiak, ahots-errekonozimendua egin dezakeen tresna baten bidez, edo telefono-teklatura erabilita (DTMF sistimaren bidez).

- Hizkuntz ikaskuntza: Kalitate handiko TAB sistema bat ordenagailuaren bidezko ikasketa sistema bati egoki dakioke, hizkuntza bat ikasteko lanabes indartsua emateko. Edozein testuren sistesiak karakter sekuentzia baten ahoskera zein den jakitea ahalbide dezake, edo esaldi baten intonazio zuzena zein den, edo zein silabak duen enfasirik handiena (azentua), etab. Grabazio aplikazioekin eta ahots aztergai-luekin (espektogramak, intonazio kurbak, energia, etab. egiteko) konbinatuta, atzelikatzea eman dezake, ikasleak ahoskatutako esaldiaren ezaugarriak sintetizagailuaren esaldiaren ezaugarriekin alderatzea bidera dezakeelako.

- Pertsona gutxitu eta ezinduentzako laguntza: Ahots gabeziak arazo fisiko zein mentalek eraginda sortzen dira. Batzuetan gabezia hauek TAB sistema bati egokitutako teklatu batek (batzuetan beren-beregi diseinatua) ordezkatu ditzake. Stephen

Hawking astrofisikariak honela ematen ditu bere hitzaldiak. Bestalde, TAB sistimek laguntza handia ematen diete itsuei ordenagailuen erabilpenean. Karakterrak ezagutzeko tresna optiko baten bidez informazio idatzia erabil dezakete.

- Ahots ikerketa oinarrikoa: Testu-ahots bihurgailu baten garapenak ahots ekoizpenerako ereduak erabiltzea dakar. Eredu horien baliogarritasuna lortzen diren emaitzetan islatzen da (jaulkitzen den ahotsaren kalitatea), horregatik TAB ebaluatzeke lanabes gisa erabil daiteke. Gainera, TAB ereduaren parametro batzuen aldaketaren bidez sortzen diren efektu akustikoen azterketa xehea egiteko lanabes malgua ere bada.

- Ikerketa linguistikoa: arau eta ereduaren balidazioa: Sintesisirako erabiltzen diren eredu linguistikoak entzuleek ebaluatuak izan daitezke, eta emaitzak eredu horiek zuzendu eta hobetzeko erabil daitezke. Euskararen kasuan, hizkuntzaren zenbait alderditan hartutako normalizazio erabakien onartze maila ebaluatzeke erabil daiteke (ebakitze arauak, intonazioa, azentua, etab.).

- Berba egiten duten liburuak eta jostailuak: Jostailu egileek TAB sistimak erabiltzen dituzte hitz egiten duten pampina eta liburuak egiteko. Aisialdiaren esparruan sistema hauek gero eta garrantzi handiagoa daukatenez gero, luze barik, euskara plaza honetara ere jalgiko dela pentsa dezakegu.

- Ahots prozesuen monitorizazioa: Sistesirako sistimak neurri eta kontrol sistemetan egoki daitezke, era honetara ahozko informazioa ematen dute. Askotan, ahozko informazioa eraginkorragoa izaten da idatzia baino.

- Multimedia, gizaki-makinen arteko komunikazioa: Kalitate handiko TAB sistimen garapena gizaki eta ordenagailuen arteko komunikazio osoagorako ibilbidean beharrezko urratsa da. Multimedia ibilbide honetako hasierako oinarri bat besterik ez da.

4. EUSKARA ALORREKO ARAZO LINGUISTIKOAK

Une honetan euskararen esparruan daukagun arazorik larrienetakoa estandar ezak eragiten duena da. Euskaltzaindiaren ahoskera batzordeak, segmentalen alorrean, proposamena egin bazuen ere, (Euskaltzaindiak, 1994) Akademiak ez du oraindik erabaki tinkorik hartu euskara batuaren fonologiari dagokionez. Argi dago, bestalde, eredu fonologiko estandar ez eukitze hau oso larria dena, TAB alorrean ezezik, euskararen irakaskuntza, komunikabide alorrean eta euskara mintzatua erabili behar den edozein esparrutan ere. Hala ere, pentsatu behar da ezen, hizkuntza naturalean maila desberdinak egonda, erabakietan nolabaiteko malgutasuna eta aukera bikoitzak eman badaitezke ere, TAB-en alorrean, teknologiaren gaurko egoera kontutan hartuta, ahal denik eta arau zurrunenak behar direla, izan ere, aukera bikoitzek sistema astundu baino ez dute egiten eta salbuespenak oso mugatuak izan behar dira. Bestalde, euskal fonetika eta fonologiak egin diren ikerketak ez dira asko eta egin direnak nahikoa barreiatu daude han-hor-hemen, ikerketok aurreratu ahala erabakiak helduko direlako daukagu.

Segmentaletan jazotzen diren arazoak era bitakoak dira; alde batetik, grafema bakoitzari dagokion hotsa zein den jakitea, eta bestetik, hots elkarketan soluziobideak.

Bokalen grafemetan, euskara batuari dagokionez, ez daukagu arazorik; haatik, kontsonanteetan zenbait arazo dugu. Frikari (z, s, x) eta afrikatuen (tz, ts, tx) ebaikerari buruz, badirudi joera nagusia denei eustea dela. Arazo nagusiak, ostera, “j” eta “h” grafemei dagozkien hotsak erabakitzean ditugu, hala nola “ge” eta “gi” sekuentziei dagozkienetan ere. “h”ren kasuan ez ebakitzea izango litzateke une honetako joera nagusia, maila jasoari dagokionez, behinik behin (Euskaltzaindia, 1994). “j” grafiari dagokiokien hotsa, Euskaltzaindiraren Ahoskera Batzordearen proposamenari jarraikirik, /y/ izango litzateke hasieran: jan, jakin, josi, etab. /yan/, /yakin/ eta /yosi/ ebakiko lirateke. Batzordeak iradokitzen duen moduan, atal honetan datu-base txiki bat egin liteke salbuespenak biltzeko. Gauza bera egin liteke “ge” eta “gi” sekuentzien salbuespenekin, baleude.

Hots elkarketetan kitatu barik ditugun arazoak ere era bitakoak dira. Batetik, hitz barruko elkarketak ditugu; esate baterako, goranzko eta beheranzko diptongoak, bokal bikoitzak; alor honetan /h/a ez egitekotan, sor daitezkeen bigarren mailako diptongoak egin behar diren (nahi, ehun, e.a.), eta bokal bikoitzak edo bakunak egin behar diren (lehen, mahats, e.a.) erabaki beharko litzateke. Palatalizazioen eremuan ez dago argi zenbateraino palatalizatu behar den hizkera estandarrari begira, badirudi ezen, “in” eta “il”en kasuetan palatalizazioa proposa daitekeela; hala ere, arau automatiko bat ezartzean, makinak beti egingo luke salbuespenik gabe. Beronek arrisku batzuk dakarzkigu mailebu berrietan batez ere, har bitez kontutan “kiniela”, “kinina”, “milibare”, “miligramo”, e.a. Herskari taldeekin (kt, pk, ...) segitu behar den jokabidea ere ez dago oso argi, hots, “erretore/erretore”, “dodore/doktore”, eta abarretan zer egin behar den ez da argiro azaldu.

Bestetik, hitzen arteko kontsonante elkarketak ditugu. Berben amaierako /k/ eta /t/ hurrengo berbaren hasierako kontsonantearekin elkartzean jazotzen diren fenomenoak. Herskari ahoskabeak ahostunekin elkartzean, talde fonologiko berberaren barruan, lehena galdu egiten da eta ahostuna ahoskabetu egiten da: “lagunatira”, “neutaukat”, etab. Herskari ahoskabeak frikariekin eta afrikatuekin elkartzean, herskaria galdu egiten da eskuarki, zenbait barietateetan, frikaria afrikatu ere egiten da. Herskariaren degeminazioa hemen ere sartu beharko litzateke (honetaz Txillardegik: 1992 eta Gaminde: 1993 ikus daitezke). Amaierako /n/ bat /m/ zein /n/-gaz elkartzen denean, berau galdu egiten da; gauza bera dugu albokari batekin elkartuz gero: “esalezake”, “egoliteke”, etab. Galdegaigunean agertzen den kategoriaren bat “tz”z amaitua denean, eta hurrengo aditzean leherkari bat badago, afrikatua frikatizatu egiten da: “zorroz dago”, “hoz dago”, etab.

Azken hauen barnean “ez” aditzondoa eta aditzaren arteko erregelak aplikatzea edo ez sartu beharko litzateke. Arazo guzti hauek ez dira berriak eta konponbideak maiz irudikatu dira (Txillardegik, 1985, 1992; Euskaltzaindia, 1994); hala ere, egun konponkizun edo nahiago bada erabakizun irauten dute eta TABerako ezinbes-tekoak dira.

Segmentalen esparrua konplexu samar agertzen bazaigu ere, suprasegmentalena, hots, azentu eta intonazioena, ezin konplexuago irudikatzen zaizkigula esan behar dugu. Alor hauetan, erabaki eza ezezik, ikerketa eskazia ere nabarmendu beharrean gaude, intonazioari dagokionez bereziki.

Azentuaren alorra, gaindiroki besterik ez bada ere, nahikoa ondo ezagutzen dugu, azentu-molde desberdinen ezaugarriei dagokienez, behintzat. (Txillardegi, Hualde eta Gaminderen lanak, besteaz beste ikus daitezke). Hizkuntzaren eremuan oso molde desberdinak aurki ditzakegu; bai azentua txertatzeko araei dagokienez, bai azentuaren gauzape fonetikoari dagokionez, izan ere, doinu-azentudun barietateak ere aurki daitezke.

Euskal azentu molde desberdinak sailkatzeko, gaur egun ezagutzen dugunaren arabera, ondoko irizpide nagusiak proposa ditzakegu (Sailkapen zehatzago baterako Gaminde eta la. 1996 ikus daiteke):

1. ± Balio kontrastiboa: Barietate batzuetan azentuaren egongunearen bidez, hitzak bereiz daitezke (tx'stu/txistú, básó/bas—, etab.), deklinabidearen singularra, mugagabea eta plurala (giz—nak/g'zonak, giz—nari/g'zonari...) etab. Era honetako barietateetan azentu-arau orokorraren bidez azentuatzen diren hitzak ez-markatutzat hartuko genituzke eta arau orokorrari jarraikirik egiten ez direnak markatutzat edo salbuespentzat jo beharko genituzke.

2. Azentua txertatzeko eremua: Barietate batzuetan, azentua txertatzeko eremua sintagma eta talde klitikoak izan daiteke, edo talde prosodikoa gurago bada; beste batzuetan, ostera, azentua txertatzeko eremua erroa bera izaten da.

3. Azentua txertatzeko norabidea eta silabaren kokagunea: Irizpide honen arabera, azentua txertatzeko silabak nondik kontatu behar diren hartzen da kontutan; bi direlarik aukerak, [---> hasieratik kontatzen hasita, eta <---] amaieratik kontatzen hasita. Behin nondik hasten den kontatzen jakinda, zein silabatan ezarri behar den azentua jakin behar dugu; irizpide honen arabera, aukerak honako hauek dira: [1, [2, 1] eta 2].

4. ± Doinu azentua: Berau azentuaren gauzate fonetikoari dagokio, mendealdeko barietate batzuetan gertatzen den doinu-azentua, indar azentuaren aldean, irizpide sailkatzaile garrantzitsutzat jo behar dugu.

Euskara batuari begira proposamen bi egin dira, Txillardegi (1985, 1992) eta Hualde (1994); xehetasunak gora behera, proposamen biok nahikoa antzekoak dira, eta bietan arrazoi linguistiko eta soziolinguistiko ugari eman dira. Proposamen hauen kontra ez dirudi ezer dagoenik, arazo bakarra ereduaren ikerketa sakonean datza, izan ere, nahikoa proposamen orokorrak dira, egun dauden moduan aplikagarriak izan ahal izateko.

Honelako proposamen bat hartuta ere, hitz markatuen datu-baserako zein den zerrenda finkatu barik dago, azentua txertatzeko eremua zein izango den ez dago

argi. Pentsa bedi azentua txertatzeko eremutat erroa hartzen bada, silaba bakarrek unitateetan azentua ez dela lekutzen, adibidez: lúrra, lúrra da, lúrrarekin, lúrrerako, etab. Aditzetan, oina edo partizipioa har daitezke eremutat, emaitzak oso desberdinak izanik: hártu/hartú, hártu dute/hartú dute, hár dezakete/har dézakete. Azentu bigarrenkarien jokaera ere finkatu barik daukagu eta aurrekoarekin lotuta dagoela esan dezakegu; talde prosodikoren amaieran azentu bigarrenkaria txertatu behar dela proposatu da, baina oraingoz ez da argiro finkatu zelan antolatzen diren talde prosodiko horiek.

Intonazioaren alorrean daukagun arazorik handiena ikerketa eza dela esan dezakegu, izan ere, oso gutxi ezagutzen dugu bariedade desberdinetan egiten dena. Gauden uneon ezinbesteko deritzagu euskararen bariedade desberdinetan egiten diren intonazio eredu nagusiak ezagutzeari. Bestalde, ezin ahantz ditzakegu azentu eta intonazioaren arteko loturak. Honi buruz, doinu-azentudun bariedadeetan lan batzuk egin badira ere (Hernáez eta la., 1995; Gaminde eta la., 1995; Uriarte, 1995 eta Elordieta: 1995), oraindik alde ilun asko ikerkizun geratzen dira.

5. HARTU DIREN BEHIN-BEHINEKO ERABAKIAK

Euskaraz aipatu berri ditugun arazo linguistikoak kontutan hartuta, gure ibilbidean behin behineko erabaki zenbait hartu behar izan ditugu. Badakigu erabakiok arriskutsuak izan daitezkeena zenbait kasutan, hala ere, ezin begira egon gaitezke harik eta erabaki ofizial sendoak eduki arte, izan ere, beronek kalte handiagoa ekar lezake, euskara hizkera-teknologietatik guztiz baztertzearainoko arriskua, alegia; edo merkatu-arrazoiak direla eta, kanpoko zenbaitek euren eredu linguistiko sasi-euskaldunak inposa litzakete.

Gure erabakietan guztiz atxeki gatzazkio Euskaltzaindiaren Ahoskera Batzordearen proposamenari, ofiziala ez den arren. Gainera sistimaren parametroak beti alda daitezkeela kontutan hartu behar da. Azentu alorrean, Txillardegi (1992) eta Hualderen (1994) proposamenak geure egin ditugu orohar, eta uneon salbuespen zerrenda eta puntu ilunen ikerketan dihardugu.

Azkenik, intonazioari dagokionez, euskararen eremu osoko ikerketa burutzen ari gara eta oraingoz ondoen ezagutzen ditugun bariedadeen arabera parametroak eman zaizkio.

6. HURRENGO URRATSAK

Etorkizunerako ikusten ditugun beharizanetan, alde batetik, garrantzitsuenetarikoak analisi sintaktiko automatikoa hobetzea da. Analisi sintaktiko automatiko honen funtzioen artean, azentuen kokaguneak zehaztea dago. Era berean, hots elkarretetan ezagutu behar diren talde prosodiko desberdinen mugaketa automatikoa egingo luke. Azkenik, esaldien fokoa automatikoki ezagutuz gero, sistimaren into-

nazioa hobetuko litzateke, izan ere, azentu-tonudunen kokapenerako beronen zereginaren garrantzia sarritan nabarmendu da.

Atal linguistikoan, dagokiokeenak hartu behar dituen erabakiez gain, euskararen azentu moten ezagutzan sakontzea ezinbestekotzat jotzen dugu, alor honetan bertan, hitz markatuen datu-basea egin beharko litzateke. Haatik eta ezpairik gabe, intonazioaren esparrua da daukagun erronkarik premiazkoena. Batetik, euskararen barietate desberdinetako intonazio moldeak ikertu eta ezagutu behar ditugu, eta bestetik, eredu nagusien azterketan sakontzea litzateke beste helburu bat. Guzti honek euskararen prosodia osoari buruzko gramatika egitea ahalbidetuko digu, luze barik euskara estandarren eredu osoagoa egiteko.

OHARRA: Lan hau UPV 051.354-EA107/95 proiektuaren barruan kokatzen da.

BIBLIOGRAFIA

- Elordieta, G. (1995). Accent, tone and intonation in Lekeitio Basque (eskuizkribua). *Euskaltzaindia* (1994). Euskaltzaindiaren XIII. Biltzarra. *Euskera*, 39.
- Gaminde, I. (1993). Irakurmena eta euskal fonologiaren arazo batzuk. *Tantak*, 9:121-126.
- Gaminde, I. (1995). Bizkaieraren azentu-moldeez. Bilbo: Labayru.
- Gaminde, I. eta Hualde, J. I. (1995). Euskal azentu-ereduen atlasarako: zenbait isoglosa. *ASJUn* agertzeko.
- Gaminde, I., Hernáez, I., Etxebarria, P. eta Gandarias, R. (1995). Zeanuriko intonazioaz. *ELE*, 16.
- Gaminde, I., Hernáez, I., Etxebarria, P. eta Gandarias, R. (1996). Euskal fonologiaren didaktikarantz: intonazioaren azterketa. *Revista de Psicodidáctica*, 1: 121-138.
- Hernáez, I. (1995). *Conversión de texto a voz para el euskara basada en un sintetizador de formantes*. Doktorego tesia. EHU-UPV
- Hernáez, I., Gaminde, I., Etxebarria, P. eta Gandarias, R. (1995): Curvas de F0 en euskara: primera aproximación a la obtención de modelos para conversión de texto a voz. *Procesamiento del Lenguaje Natural*, 17: 272-289.
- Hualde, J. I. (1994). *Euskal azentu ereduen sailkapenerako*. Euskaltzaindiaren XIII. Biltzarrean aurkeztutako txostena.

- Hualde, J.I. eta Bilbao, X. (1993). The prosodic system of the basque dialect of Getxo: a metrical analysis. *Linguistics*, 31: 59-85.
- Hualde, J.I., Elordieta, G. eta Elordieta, A. (1993). Focalización y prosodia en vascuence vizcaíno. *ASJU XVII* (3), 731-749.
- Hualde, J. I. eta Elordieta, G. (1994). *Lekeitioko azentu eta intonazioaz*. (Esku-izkribua)
- Kaplan, G. eta Lerner, E.J. (1987). Realism in synthetic speech. *IEEE Spectrum*, April 1987: 32-37.
- Mariño, J. B., Nadeu, C. eta Llisterri, J. (1987). *Síntesis automática del habla. Inteligencia artificial: conceptos, técnicas y aplicaciones* (pp.157-165). Barcelona: Marcombo, Boixareu Editores.
- Txillardegi (1985). *Euskal Azentuaz*. Donostia: Elkar.
- Txillardegi (1987). Azentuari buruzko proposamen bat. *ELE*, 2.
- Txillardegi (1992). Ahoskera baturantz. *ELE*, 10.
- Uriarte, J. A. (1995). *Fruizko hizkera: azterketa linguistikoa*. Doktorego tesia. Deustuko Unibertsitatea.
- Witten, Y. H. (1982). *Principles of computer speech*. London: Academic Press.