

XUXEN euskararako zuzentzaile ortografikoa

1994/11/01 Elhuyar Zientzia ITURRIA: ELHUYAR ALDIZKARIA

Softwarea Elkarrizketak

Joan den irailean Xuxen euskararako zuzentzaile ortografikoaren aurkezpena egin zen Donostiako Koldo Mitxelena kulturunean. Bertan ikusitakoari buruz patxadaz hitz egiteko aukera izan genuen egileetako bat den Iñaki Alegriarekin. Hona hemen elkarrizketaren emaitza.

Elhuyar.- Zer da Xuxen programa?

I. Alegria.- Euskaraz idatzitako testuetarako zuzentzaile ortografikoa da **Xuxen** eta euskarazko testuen ortografia detektatu eta zuzentzea du helburu; hau da, huts tipografikoak eta ortografikoak detektatu eta zuzentzea. Helburu honi begira euskara batua da programak onartzen duena. Dokumentu bat edo gehiago azter ditzake exekutaldi bakoitzean. Dokumentuak hitzez hitz tratatzen dira eta hitzak ezagutzen dituen bitartean lanean jarraitzen du etengabe, baina hitzen bat ezagutzen ez duenean abisua eman eta geratu egiten da. Balizko hutsaren aurrean erabiltzaileak hainbat aukera izango du, besteak beste, ontzat ematea, erabiltzaileak zuzentzea, zuzenketarako programak eskaintzen dituen proposamenak eskatzea edo, berriro lema horretatik deribatutako forma guztiak ezagutu ditzan, hiztegi pertsonalean sarrera ematea (ikus 1. irudia).



1. irudia. Pantaila nagusia.

Argazkia ongi ikusteko jo ezazu PDF-ra

Elh.- Zenbat denbora behar izan duzue programa burutzeko? eta nortzuek hartu duzue parte?

I.A.- Programa prestatzea lan handia izan da nagusiki bi arrazoiengatik: euskarak duen berezko konplexutasun morfologikoagatik batetik eta euskararen morfologiari buruzko deskripzio sistematikoaren faltagatik bestetik.

Aurretik egin genuen analizatzaile morfologiko automatikoa aprobetxatuz azken hiru urteetan garatutako programa hau UPV/EHUko Informatika Fakultatea, UZEI eta Baionako Hizkia enpresaren arteko elkarlana izan da. Informatika Fakultateko taldeak prototipoaren garapena eta koordinazio-lana egin du, UZEIk berma linguistikoa eman dio eta Hizkiak produktu komertzialaren ardura hartu du. Macintoshera eta PCrako bertsioak prestatu dira. Lan honetan ezin da aipatu gabe utzi honako erakundeen laguntza: IVAP/HAEE, Gipuzkoako Foru Aldundiko Ekonomia Saila eta Euskadi/Akitania lankidetzaren programa.

Elh.- Zeintzuk dira Xuxen programaren diseinu-ezaugarriak?

I.A.- Euskarak duen konplexutasun morfoloikoa dela eta, hitz bat ondo dagoen ala ez erabakitzeke ezin da beste hizkuntza batzuetarako egiten den bezala hitz-zerrenda bat kontsultatu eta kito; lema batetik abiatuta sor daitezkeen forma zilegiak asko direnez, zerrenda izugarria izango bailitzateke. Adibidez, izen batetik abiatzen bagara, deklinabide-atzizki bakar bat erantsiz 135 legezko forma lor daitezke (elipsiak kontutan hartzen baditugu zenbaki hau izugarri igotzen da). Gainera, horrela jokatu izan bagenu, erabiltzailearen hiztegian lema bakarrik sartuz bere deklinabide osoa ulertu beharrian, **Xuxenen** gertatzen den legez, lema horri dagozkion forma guztiak sartu beharko lituzke banan-banan erabiltzaileak. Aurreko guztiarengatik hitz zilegiak ondo identifikatzeko analisi morfoloikoa egitea ezinbestekoa izan da.

Elh.- Analisi morfoloikoa ezinbestekoa izan dela aipatu duzu, zein nolako urratsak eman dituzue hori lantzeko?

I.A.- Analisi morfoloikoa 1.983an Helsinkiko Unibertsitateko Koskenniemi irakasleak proposatutako bi mailatako formalismoan dago oinarrituta. Formalismo hau hasiera batean suomierarako proposatu bazen ere, oso arrakastatsua izan da beste edozein hizkuntzatarako eta euskara bezalako hizkuntza eranskarietarako batez ere. Formalismo honen ezaugarri nagusiak honakoak dira: testuetan agertzen diren hitzak —azaleko maila— eta lexikoa —maila lexikoa— argi bereiztea, analisi zein sintesirako balio izatea, eta programa eta ezagumendu linguistikoa bereiztea. Informazio linguistikoa morfemez osatutako lexikoeak eta erregela morfofonologikoeak osatzen dute.

Lexikoak 60.000 sarrera baino gehiago du, datu-base batean gordeta eta 120 azpilexikotan banatuta. Sarrera bakoitzari jarraitze-klase bat egokitzen zaio bere atzean etor daitezkeen atzizkien multzoa definitzen duena. Morfemak biltzean gertatzen diren azaleko aldaketak adierazteko hogeita lau erregela morfofonologiko daude. Erregela hauetako bakoitzak karaktere baten txertaketa, ezabapena edo aldaketa bat noiz gertatzen den adierazten du. Adibidez, zortzigarren erregelari ondoko aldaketa deskribatzen da: lexikoko **k** letra azaleko **g** bihurtzen da baldin **k** letra “ko” atzizkikoa bada, eta aurretik dagoena **n** letraz bukatutako lema bat edo **l**, **m** edo **n** letraz bukatutako toki-izen bat baldin bada. Esate baterako *izan* eta *ko* morfemak biltzean *izango* forma sortzen da.

Abiadura handitzeko gehien erabiltzen diren hitzekin zerrenda bat osatu da, berorien analisi morfoloikoa saihestearren.

Elh.- Zuzenketarako proposamenak programak eskaintzen duen aukeretako bat da, zertan datza?

I.A.- Errore baten aurrean erabiltzailearen esku dago programari proposamenak eskatzea. Lan horretan errore tipografikoeak eta ortografikoeak tratamendu desberdina dute. Tipografikoetan



karaktere bat galtzea, txertaketa zein aldaera edo bi karaktere jarrairen arteko trukea hartzen da errearen iturritzat, eta alderantzizkoa eginez bilatzen dira proposatzeko hitz egokiak.



I. Alegria. Euskal Herriko Unibertsitateko Informatika Fakultateko irakaslea. Xuxen euskararako zuzentzaile ortografikoaren egileetako bat.

Euskararen ezagumendu eskasak, batasunaren azken aldaketen berri ez izateak edo erabilpen dialektalak eragindako erroreak ortografikoak edo tipikoak deitzen dira. Horrelakoak detektatu eta zuzendu ahal izateko **Xuxenek** azpilexiko eta erregela bereziak ditu. Adibidez, *haundi* forma lexiko berezi batean dago *handi* forma hobetsiarekin lotuta; “haundiaren” analisia egitean haundi+aren lortzen da, baina haundi errore bezala markaturik egotean *handi* bihurtzen da eta, ondorioz, *handi+aren*-ren sorkuntzaz *handiaren* sortuko da proposamen gisa. Erregela berezien artean h-ren galera eta x-s-z aldaketa deskribatzen dutenak daude. Horrela *zuaitezko* analizatzean automatikoki *zuhaitz+ko* lortzen da eta sorkuntzaz *zuhaitzeko* proposamena.

Elh.- Zein hizkuntz eredu erabili duzue?

I.A.- Euskararen flexioa kontutan hartuta, deklinabide-sistema ordenadorez erabilgarria eraiki behar izan da. Horretarako, Euskaltzaindiak proposatutako taulan oinarritu gara eta gure sistemara egokitu dugu; hau da, taula hori hartu eta lexiko-kategoria bakoitzari egokitzen zaizkion kasuak multzoka eratu ditugu. Horrela oinarri bakoitzari atzizki-multzo bakarra dagokio, har ditzakeen atzizkiez osatua.

Eratorpenean zenbait aurrizki eta atzizki landuta daude, baina arruntenak hiztegi-sarrera bezala daude. Hala ere, erabiltzaileak bere hiztegian sar ditzake hitz eratorri berriak. Hitz elkarketan ere, ohizkoena eta sistematizagarriena landu da oraingoz, Euskaltzaindiaren LEF Batzordeak markatutako irizpideen arabera. Aditz faktitiboa ere sistematikoki landuta dago

1992ko Euskaltzaindiaren gomendioa

Aditzari dagokionez, aditz laguntzailearen zein trinkoaren formak ezagutzen ditu **Xuxenek**, beti ere Euskaltzaindiak erabakiak. Forma neutroak, markatu gabeak nahiz hitanozkoak ezagutzen ditu.

Gramatikaren atalean arau-iturri bakarra Euskaltzaindia izan bada, bestela gertatzen da lexikoa lantzen hasi orduko. Puntu batzuetan emanak ditu kasuan kasuko gomendio eta erabakiak: *H* letra, *-a* berezkoa, zenbakien osaera eta idazkera, etab. Horiek jarraitu ditugu lexikoa osatzean, nahiz eta zenbakien kasuan oraingoz bi aukerak mantentzen ditugun (*hogeita bost* eta *hogeitabost* onartuz). Beste hainbeste gertatu da pertsona- eta leku-izenekin, bai eta maileguen idazkeran ere.

Oinarritzko hiztegia sortzeko, hau da, edozein lexikotan maizenik agertzen diren lemen zerrenda, gaurko beste iturrietara jo behar izan dugu: Ibon Sarasolaren *Hauta-Lanerako Euskal Hiztegia*, UZEIko Euskalterm datu-bankua eta EEBS datu-base lexikografikoa, Xabier Kintana eta besteren *Hiztegia 2000*, J.M. Etxebarriaren *Maiztasun- eta Prestasun Hiztegia*,

etab. Euskaltzaindiaren irizpideekin bat ez zetozenean, sarrerak “egokitu” egin dira; eta, Euskaltzaindiak erabaki ez dituenetan, Ibon Sarasolaren hiztegia izan da irizpide-iturri.

Oinarrizko hiztegia osatu nahirik, UZEIko EEBStik hainbat esapide, lokuzio eta forma konplexu hartu da. Siglak eta laburtzapenak ere UZEIko irizpideen arabera landu dira. Hiztegi arruntetik abiatuz, terminologiarraino iritsi behar izan da zenbaitetan. Ezinbestekoa izan da Euskalterm horrelakoetan.

Izen propioen zerrenda osatzeko (izen propioak hiztegi arruntetan ez badatoz ere), bi iturritara jo da: lehena Euskaltzaindiak proposatutako euskal pertsona- eta leku-izenen zerrenda izan da, baina munduko leku-izenen zerrenda eskuratzeko Elhuyar-era jo da.

Iturri guzti hauetatik edanda tamaina handiko hiztegia osatu dugu, testu arruntetako lexikoa, gutxienez, baduena. Hala ere, gai berezitetako terminologia norberak sartzeko aukera izango du bere hiztegi pertsonalean.

Elh.- Etorkizunari begira zer?

I.A.- Taldean analisi sintaktiko automatikoari aurre egin nahi diogu hurrengo urteetan. Hori eginez gero zuzenketa aurreratua egiteko aukera izango du etorkizuneko XUXENek. Beste aldetik hiztegi gintzan ere ari da lan egiten gure taldea, hiztegiei baliabide informatikoak aplikatzean askoz etekin handiagoa ateratzeko asmoz. Hala ere, analisi morfologikoan oinarrituta EUSLEM lematizatzaile automatikoa atera nahi dugu hemendik urtebetera edo.

Ezaugarri funtzionalak

Oinarri-makina

- Macintosh, 7.0 Sistema Eragilea eta gutxienez 4 Mbyte-ko RAM memoria.
- PC 386 edo ahaltzuzagoa, Windows 3 eta gutxienez 4 Mbyte-ko RAM memoria.

Testu-prozesadore bateragarriak
(testuaren formatea mantentzeko):

- Microsoft Word 4.0, 5.0, 5.1
- WordPerfect 2.0

Aukera nagusiak (2. irudia):

- Testu bat zuzentzea. Dokumentu editatu oso bat egiaztatuta eta zuzentzea.
- Testu-sail bat egiaztatzea eta zuzentzeko prestaketa. Ari den bitartean ez da pantaila aurrean egon behar. Bukaceran zuzenketa azkarrenako aukera.

- Zalantzen kontsulta. Editatu ahala hitz solteak kontsultatzea arbelaren bidez.
- Hiztegi pertsonalak osatzea lexiko bereziturako (3. irudia).

Hiztegi orokorra:

- Zabala eta eguneratua: 60.000 sarrera baino gehiago.
- Euskaltzaindiaren azken erabakiak kontutan hartuz.

Erregistroa	Edizioa	KID
ireki.3		KID
Zuzentzerako prestaketa...		HP
Zuzendu arbelakoa...		
Hiztegia ireki...		HH
Laguntza		H7
Irten		KD

2. irudia. Aukera nagusiak.

Hiztegi pertsonala

Hiztegi pertsonala... Informatiko.ida

Kategoria

- Izena
- Aditza
- Adjektiboa
- Siglak
- Besterik

Izen-mota

- Arrunta
- Pertsona
- Leku

Hiztegi pertsonala

Informazio logiko

Lema: bitol

Leguntza

Agurratza

Kendu

Itzi

1. irudia. Hiztegi pertsonala aberasteko pantaila.