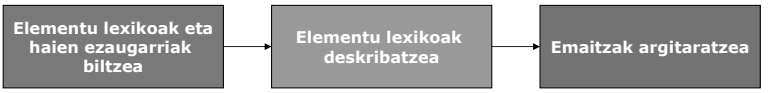
Zapata-kaxatik XML fitxategira euskal hiztegigintzan

Andoni Sagarna

Hiztegigintzaren hiru urrats nagusiak



Hiztegigintzaren hiru urrats nagusiak 1970eko hamarkadaren lehen urteetan

Elementu lexikoak eta haien ezaugarriak biltzea → **Elementu lexikoak deskribatzea** → **Emaitzak argitaratzea**

Kartoi mehezko fitxak

Fotokonposizioa

Fotomekanika

Inprimategia

Sistema informatiko handiak zituzten enpresa eta erakundeen laguntza 1975-1980

Unión Farmaceútica Guipuzcoana

1975

Elhuyar Taldeak egindako euskal testuen hustuketa euskarri digitalean jartzea.

EUTG

1980

SARASOLA ERRAZKIN, Ibon. 1982. *Gaurko euskara idatziaren maiztasun-hiztegia.*

1977ko euskarazko corpus bat prozesatzea.

Instituto Deusto

1980

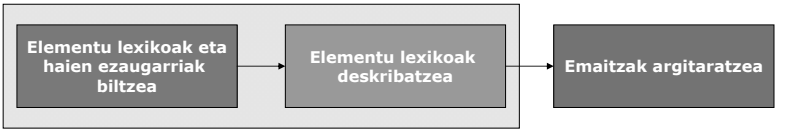
UZEIk egindako terminologia eleaniztunen hustuketak euskarri digitalean jartzea.

UZEIko lehen sistema informatikoaren definizioa
1980

- Enpresetan erabiltzen ziren datu-baseak zurrunak eta garestiak ziren
- Hardware "handia" eta garestia eskatzen zuten
- Informatikariek ezinezkoa ikusten zuten UZEIren lana informatizatzea

Klaudio Harluxet-en irudimenak ekarri zuen soluzioa

- Propio garatutako datu-base ez-konbentzionala
- Erregistro-egitura librekoa
- Fortran lengoaian garatua
- Makina "txiki" batek erabil zezakeena (Bull Mini6)



Testu-editoreak oso baldarrak eta garestiak ziren artean

Orotariko Euskal Hiztegia lantzeko sistema informatikoa
1984 - 2005

"El autor se convenció de que hoy no se puede hacer una obra de esta naturaleza a base de fichas y cuadernos, como fue el caso de Azkue. Se hacía preciso el recurso al ordenador."

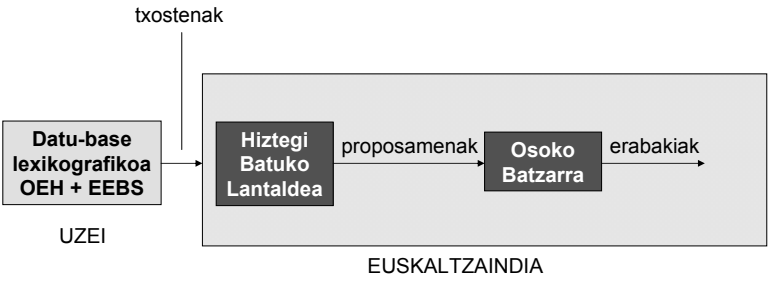
Aita Villasantek OEHren 1. Liburukiaren aurkezpenean

- Corpusa eta paperezko hiztegiaren originala ordenagailuz prestatzen dira.
- Euskal literaturako 300 obra teklatu bidez digitalizatzen dira
- Hasieran Wang sistema erabiltzen da, gero PCak (1999-2000tik aurrera)
- Corpusa ez da lematizatzen
- Hiztegiaren bertsio elektronikorik ez da argitaratzen

Egungo Euskararen Bilketa-lan Sistematikoa (EEBS)
XX. mendeko corpus estatistikoa
1987 - 2000

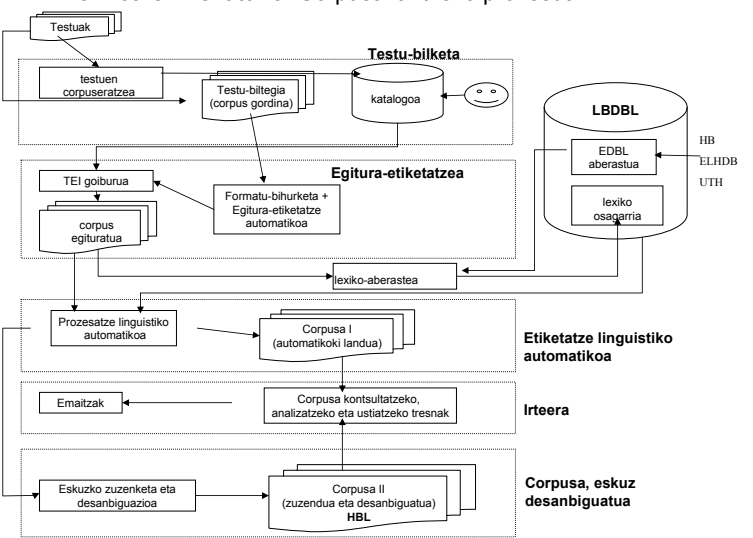
- Corpusaren %90 teklatua erabili gabe jartzen da euskarri digitalean
- Datu-base batean antolatzen da
- Corpus osoaren lematizazio erdi-automatikoa egiten da
- Kontsultarako Interneten jartzen da

Hiztegi Batua



landa-lan
20.21: ***landa-lan*** 1, Or *Mi* ("Lan txukunetan etzen ari izaten; geienez zamarientzat bastak egiten eta landa-lanetan"); ***kanpo(-)lan*** (eta *kanpolan*) IE 4: Mst ("Gizon pherestü, eta debot batek kanpolan egin behar dütianak, aitzinetik bere beithan ikhusten eta alkharren unduan egiteko berhezten dütü"), Elzb *Po* ("Landetan sorthurikan nahiz diren asko, / Ez dire hek eginak lanean hartzeko; / Bulharrak duzte flako; larrua xurixko; / Ez dute kuraierik kanpo-lanetako"), Arb *Igand* ("Dakizun bezala, nere auzoko etxean gizon bat bazen Igande-gabea, Meza-Bezperetako denboran kanpo lanetan trebeki ari zena"), JE *Bur* ("egun osoa barne ala kanpo lanetan iragan ondoan, jendeak, kabalak, gauzak zoin beren lekuetan, tenoretan eta izarian moldaturik, azkenik zoan bethi oherat"); eta OEH argitaratuan, Prop 1906 ("Haurrek egiteko dutena da, baratze eta alhor, hori da kanpo-lanetan, gure laguntzea") eta aip. Aran-Bago *ManMed*.
20.22: **a)** ikerketa-motari dagozkionak dira: ***landa-lan*** 1, *EHHistoriaz* ("Data hau euskal aztarnategiaz J. M. Barandiaranen landa-lanetan jardun ziren ikertzaile gazteen Doktoratze-tesi edo lan espezializatuen argitalpenetan oinarritzen da"), eta ***kanpo-lan*** 6: I. Ifurrieta 5 (adib., "kanpo-lana. (…) Ikasleak hirira irtengo dira", "Utzi ikasleei beren laguntzaileak gelara ekartzen, kanpo-lana berregiteko", "Prestaketa. Kanpo-lanari ekin baino egun batzu lehenago, ikasleek entzun edo irakurritako txisteak eramango dituzte ikasgelara"), *Bertsolaritza* ("Ikaspuntu kultural batetik gaurko bertsolaritzaren berbalioketak, berari buruzko kanpo-lan zabal bat egin ahal dadin giroa eta materiala prestatzen ditu"); **b)** cf. E.J. Zubillaga ("Emeretzigarren eunkian [siglo 19] bere erdialdetik gora Argentina'ra etorritako euskaldunak erderaz mintzatzen ez zekitenak, ta kanpo lanetan jarduntzen ziranak beren alaba-semeak euskeraz ikasi zuten apur bat, beren aita-amak beti euskeraz mintzatzen ziralako"), UZEI ("Asko erabiltzen da kanpo-lanetarako eta gure artean altzairuzko ere bai").
20.20: ***landa-lan***: Euskalterm 2; ***kanpo-lan***: Euskalterm 2 **//** Ez dugu aurkitu ap. AB38, AB50, DFrec, HiztEn, LurE **//** Eta *trabajo de campo* itzuliak: ***eremu-ikerketa***: Euskalterm 1; ***ikertxango***: AB38 1; ***lan bortxatuen esparru***: Euskalterm 1; ***lan sail***: AB38 1; ***landa-azterketa***: Euskalterm 1; ***laneko udaleku***: Euskalterm 1 (campos de trabajo); ***tokiatokiko lan***: AB50 1.
20.81: ***kanpolan***: HiruMila (trabajo de fuera), ElhHizt (trabajo de fuera, trabajo de campo), Casve *EF* (travail extérieur), HaizeG *BF* (travaux des champs), Lh *DBF* (travaux du dehors et plus communément travaux des champs), DRA (*kanpo-lan*: trabajo de fuera, y más generalmente trabajo del campo) **//** Ez dugu aurkitu ap. EuskHizt, EskolaHE, Lur *EG/CE* eta *EF/FE*, XarHizt, PMuj *DVC*.
20.83: Erdal *trabajo de campo* / *travail sur le terrain*; *travaux des champs* formen ordainak: ***landa-lan***: HiruMila, ElhHizt **//** HaizeG *FB*: Ø / *lur-lan*, *kanpo-lan* **//** Ez dugu aurkitu ap. Lur *EG/CE* eta *EF/FE*, XarHizt, Casve *FE*, T-L *LBF*, PMuj *DCV*.
20.9: **fr** (DLLF): Ø (ik. Le Petit Robert, adibideetan: *Faire une enquête sur place*); **it** (S. Carbonell): **Ø**; **en** (Collins): *fieldwork*; **de** (Langenscheidts): Ø.

Lexikoaren Behatokia: Corpusa lantzeko prozesua



Egitura-etiketatzea.

TEI P5 (XML).

Automatikoa: testuaren egitura-ezaugarriak (atalburuak, atalak, azpiatalak, paragrafoak, zerrendak, taulak, oin-oharrak, irudi-oinak, eta abar.); nabarmentze-ezaugarriak (tipografikoak, hau da, letra-estiloa eta komatxo edo kakotxak)

Eskuz lantzekoa: nabarmentze-ezaugarrien balioa (enfasia, aipua, atzerri-hitza, ohiz kanpoko adiera edo erabilera ironikoa, metahizkuntza, terminoa, izen berezia...)

Aurreprozesamendu linguistikoa

Corpusaren lexikoi osagarria elikatzea (EDBLn ez dauden eta corpusean atzeman diren maiztasun handiko lemak)

Aldaeren normalizazioa (<reg> etiketa)

Akats ortotipografikoen zuzenketa (<corr> etiketa)

Analisi linguistikoa