

## INFORMATIKA ETA HIZKUNTZA

<b>Videotex sistemaren etorkizun hurbila</b> <i>Andoni Sagarna</i>	I
<b>Ordenadorea lagun, testu-zuzenketan</b> <i>Xabier Artola</i>	IV
<b>Mikroinformatika eta zenbait komunikazio-arazo</b> <i>A. Arruabarrena, G. Buldain, L. Gardeazabal, E. Gomez &amp; J. Gonzalez Abascal</i>	VIII
<b>Lengoaia naturalaren ordenadore bidezko aplikazioak</b> <i>Kepa Sarasola</i>	XII

---

# VIDEOTEX SISTEMAREN ETORKIZUN HURBILA

---

Andoni Sagarna

---

### Zer da Videotex Sistema

Zenbait datu-basetan kokatutako informazioa telefono bidez konektatutako terminaletatik kontsultatzea ahalbidetzen duen sistema informatikoa da.

Datu-baseko informazioa zerbitzari deritzon ekipu batek bidaltzen du telefono-sarera. Erabiltzaileak informazioa kontsultatu ezezik zerbitzariarekin elkarrizketa-moduan ere lan egin dezake.

Elkarrizketa honetan zenbait kasutan transakzioak ere burutu daitezke; erosketa-aginduak emateko esate baterako. Beste zenbait eragiketa burutzea ere posible izaten da, hala nola mezularitzaz, posta elektronikoz edo softwarearen telekargaz baliatzea.

Erabiltzaileak 1.200 baud-eko abiaduraz jasotzen du informazioa eta 75 baud-etan igortzen. Zerbitzariak eskaintzen duen informazioa zuhaitz-forman eratutako orrialde-multzo baten tankerakoa izaten da. Erabiltzaileak zuhaitzaren adarrak korrituz edo orrialde jakinak eskatuz kontsultatzen du informazioa.

### Norma desberdinak Europan

Videotex sistema Britainia Handian sortu zen Prestel normarekin 1982an eta berehala izan zuen arrakasta, baina aurrerapenik handiena Frantzia egin du azken urteotan. Azken datuen arabera, 3.300.000 erabiltzailetik gora daude jadanik herrialde horretan.

Sistema hauek beren artean ez dira bateragarri. Norma bakar batera heltzeko egin diren ahaleginek ez dute emaitza osabeterik sortu. Gaur egun lortua den normalizazio-mailaren arabera diferentzia nabarmenak daude sistemen artean.

Alderdi batetik alemanek sortutako CEPT-1 norma dugu. Honako hauek dira bere ezaugarriak:

- terminalaren pantaila 40 edo 80 zutabetan eta 24 errenkadatan banatzen da.
- karaktere alfanumerikoak eta marrazkiak errepresenta daitezke koloreak erabiliz.

Norma honetan oinarritzen dira Alemaniako Errepublika Federaleko, Austriako, Norvegiako, Suitzako eta Espainiako sare publikoak.

Frantzia erabiltzen den normari CEPT-2 derit-  
zo eta CEPT-1 delakoak baino mosaiko larriagoa  
(40x24) eta kolore gutxiago ditu. Sistema hau aipat-  
zerakoan Antiope, Teletel eta Minitel izenak ere  
erabiltzen dira.

Prestel sistemak Britainia Haundian, Belgikan,  
Danimarkan, Holandan, Suedian eta Italian aurkitzen  
dira.

Espanian sare publikoak CEPT-1 norma hauta-  
tu badu ere, badira CEPT-2 normaz baliatzen ari  
diren erakunde pribatuak (Lankide Aurrezki gure  
artean).

## Espainiako Estatuan arazoa nola dagoen

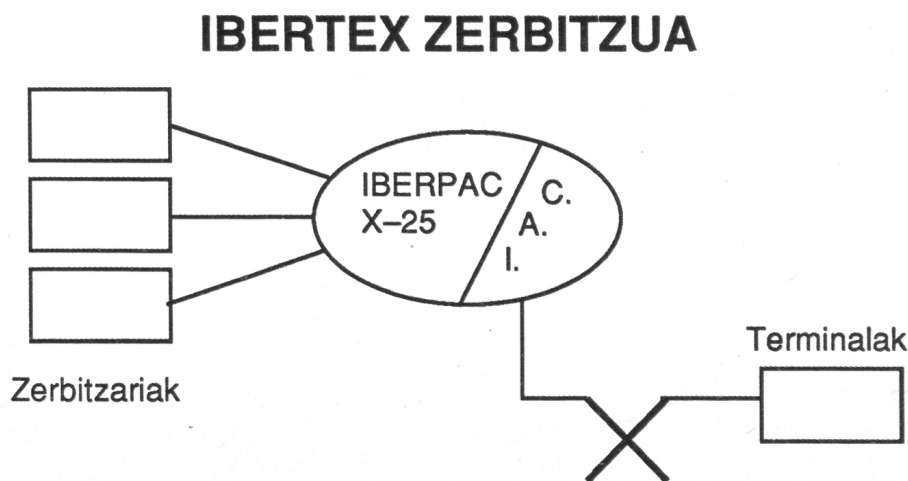
Badirudi Espainiako Estatuan iluntasuna izan  
dela orain arte behintzat nagusi. Telefonikak agindua  
zuen 1982rako sare publikoa prestatua izango zuela,  
baina ez zuen bere promesa bete; ezta gutxiagorik  
ere.

1987ko uztailean definitu ditu azkenean bere  
erizpideak eta oraingoz sare publikoak behin-  
behineko funtzionamendua besterik ez du.

Espainiako Telefonikak Ibertex izeneko Video-  
tex zerbitzu publikoa jarri du martxan. Zerbitzu honek  
CEPT-1 norma erabiltzen du. Honek, Ibertex-ek  
eskaintzen dituen zerbitzuez baliatu nahi duten zer-  
bitzariak eta erabiltzaileak norma hori bete behar  
dutela esan nahi du. Ekipoen arteko bateragarritasuna  
bermatzeko, Telefonikak pasarazitako kalifikazio  
teknikoko probak gairatu izana eskatuko zaie, bai  
terminalari, bai zerbitzariari.

Telefonikak berak ere saltzen ditu terminalak,  
baina erabiltzaileak aukera ditzake beste batzuk,  
Telefonikak homologatutakoak izanez gero.

1. Irudiak erakusten du zerbitzariak eta erabilt-  
zaileak Ibertex zerbitzuaren bidez nola konektatuko  
diren.



1. irudia.

Zerbitzariak IBERPACekin konektatzen dira  
eta 2400, 4800 eta 9600 baud-eko abiaduraz lan  
egiten dute. IBERPAC Telefonikaren paketeak trans-  
mititzeko sarea da, Frantzia TRANSPAC denaren  
baliokidea. Zerbitzarien eta IBERPACen arteko  
konexioa X-25 sarearen bidez egiten da. Honek  
erabiltzaileak zerbitzua erabilera-denboraren arabe-  
ra soilik ordaintzen duela esan nahi du eta ez denbo-  
raren eta distantziaren arabera.

Erabiltzaileak Telefonikari ordainduko dio zer-  
bitzuaren kostua eta honek zerbitzariari ordainduko  
dio, informazioagatik ordaindu beharreko zerbitzua  
denean.

Erabiltzailea sare telefoniko konmutatuaren eta  
C.A.I. (Centros de Acceso Ibertex) delakoen bidez  
sartuko da sistemara.

Orain arte iluntasun handiak zeuden Espainian  
videotex sistemak izango zituen ezaugarriez. Ba-  
zirudien Telefonikak bere sistema muntatuko zuela  
eta beste inork ez zuela sistemak jartzeko esku-  
biderik izango. Hala ere, Telefonikaren moteltasunak  
luzetsita, batzuek ekin egin diote zerbitzuak  
eskaintzeari (Banco de Santander, Lankide Aurrez-  
kia, etab.), normen aldetik nork bere aukera egin  
duelarik. Lankide Aurrezkiak CEPT-2 norma auke-  
ratu zuen, esate baterako.

Momentu honetan gauzak argiago daudela di-  
rudi: Telefonikak Ibertex sistema eskaintzen du eta  
honen abantailaz baliatu nahi duenak (X-25 sarearen  
erabilera, erabiltzailearen fakturak bateratzearen, h.d.  
kiosko-sistema deritzonaz baliatzearen abantaila),  
bai zerbitzariaren aldetik, bai terminalaren aldetik,  
Telefonikak homologatutako ekipoak erabili behar  
ditu, Telefonikari erositakoak izan ala ez.

Spritel-ek antolatutako hitzaldietan zenbait  
hizlarik, honako planteamendu hau egin zuen:

"Videotex sistema Es-  
painian hedatuko bada ezin  
zaizkie erabiltzaileei era-  
gozpenak jarri, normak di-  
rela eta ez direla. Bateragarri-  
tasun-arazoak zerbitzari-  
mailan konpondu behar dira.  
Bestalde, hor daude  
Europako herrialde desber-  
dinetako normen arteko  
diferentziak. Beraz batera-  
garritasun-arazoak kon-  
pondu beharra bestela ere  
azalduko da (eta eleanizta-  
sun-arazoena ere berdin).  
Horren irtenbidea normen  
bihurketa-sistemetan aurki-  
tu behar da."

Berez, hitzaldi hauetan zerbitzari-ekipamenduak aurkeztu dituzten etxe gehienek norma desberdinekin lan egiteko aukera azpimarratu zuten.

Etxe askok videotex-terminal berezituak eskaintzen dituzte, eta hauetako batzuk multistandard-ak dira, baina garbi ikusten da videotex-aren erabiltzaile potentzialen artean multzo ugari eta interesgarri bat (Espainian 200.000tik gora) mikroordenadore pertsonalen jabeek osatzen dutela. Hauek videotex-a erabiltzeko ez dute beste terminal bat erosi beharrik. Aski dute modem egoki bat mikroordenadoreari jartzea eta komunikazio-software egokia erabiltzea. Dagoeneko horrelako produktuen eskaintza badago merkatu espainiarrean. Adibidez, Kortex etxearena eta Bilboko CEINSAk saltzen duen Vidicom-en Rafi izeneko modem-a.

Beste erabiltzaile batzuek, berriz, ASCII terminalak dituzte. Hauek ere balia daitezke videotex-az. Jakina, beren terminaletan ezin jaso ahal izango dituzte koloreak eta grafikoak, baina informazio profestionalean kasu askotan hori ez da hain inportantea izaten.

## **SPRITELen balio erantsiko sarea**

SPRI sozietateak Komunitate Autonomoan telematikaren erabilera sustatzeko eta industria telematikoaren garapena indartzeko SPRITEL programa antolatu du. Helburu hauekin zenbait ekintza-mota martxan jartzeko asmoa du: erabiltzaileei denboraldi batean terminalak doan uztea, Kontsultategi bat jartzea, erakusketak eta mintegiak antolatzea, zerbitzariak muntatu nahi dituztenei dirulaguntzak eta kreditu erosoak eskaintzea, eta gauza guzti hauekin batera oso interesgarria izan daitekeen balio erantsiko sare bat eratzea.

SPRITEL sare honen bidez Gasteizen, Bilbon eta Donostian egongo diren sarrera-puntu (telefono-zenbaki) batzuetara deituz erabiltzaileak IBERPAC X-25 sarrera konektatu ahal izango dira eta horretaz gainera SPRITELek jarriko duen norma-bihurgailu bati esker beste norma batzuen arabera funtzionatzen duten bertako edo atzerriko sistemetara ere bai.

## **Abiaraztea zaila**

Frantzian videotex-ak izan duen arrakastaren ondoan Espainia eta Hego Euskal Herria oso atzera daude, zalantzarik gabe. Egoera honen errudunik handienak Espainiar Administrazioa eta Telefonika dira, panorama argitzeko ezer gutxi egin dutelako. Alabaina, miresgarria izango omen den 1992. urte horrek eta aukerak galtzen ari ziren enpresek bultzatuta, gauzak argixeago daudela dirudi.

Beste arazo bat merkatua nola abiarazi da. Askotan aipatu izan da sorgin-gurpil bat dagoela hor:

erabiltzailerik ez dagoenez ez dago zerbitzurik eta, jakina, zerbitzurik gabe ezin izan erabiltzailerik.

Hitzaldi batean hizlari batek esan zuen bezala, arazo hori ez da berria; beste hainbeste gertatzen zen telebistarekin ere eta kasu hartan badakigu nola hautsi zen sorgin-gurpila: TVE bere telebista-saioak ematen hasiz, noski.

Hau egia izan daiteke, baina norbaitek zera galdetuko du: ea aurreneko urteek sosik ere emango ez dutela jakinda ekimen pribatuko proiekturik sortuko den. Batzuk prest egon daitezke, zerbitzu-mota hau irudia lantzeko bidetzat hartuko dutelako, erabiltzaile-talde itxiekiko harremanak hobetzeko egokia izango delako, etab.

Horretaz gainera, erakunde publikoen laguntzek garrantzi handia izan dezakete.

Seguraski videotex-ak gure artean datozen urteetan hazkunde handia izango duela esatea ez da arriskatu egia izango hazkundera termino erlatibotara mugatzen baldin badugu behintzat.

## **EUSKALTERM Euskal Terminoloji bankua**

EUSKALTERM Terminoloji Bankua, Lankide Aurrezkiaren egitura informatikoaren barnean integraturik aurkezten den terminologi kontsultazerbitzua da.

Honako arlo hauek kontsulta daitezke:

- |                  |               |
|------------------|---------------|
| * Ekonomia       | * Finantzak   |
| * Administrazioa | * Enpresa     |
| * Estatistika    | * Zirkulazioa |
| * Matematika     | * Fisika      |
| * Kimika         | * Biologia    |
| * Medikuntza     | * Teknologia  |
| * Filosofia      | * Psikologia  |
| * Linguistika    | * Zuzenbidea  |
| * Arteak         | * Kirolak     |

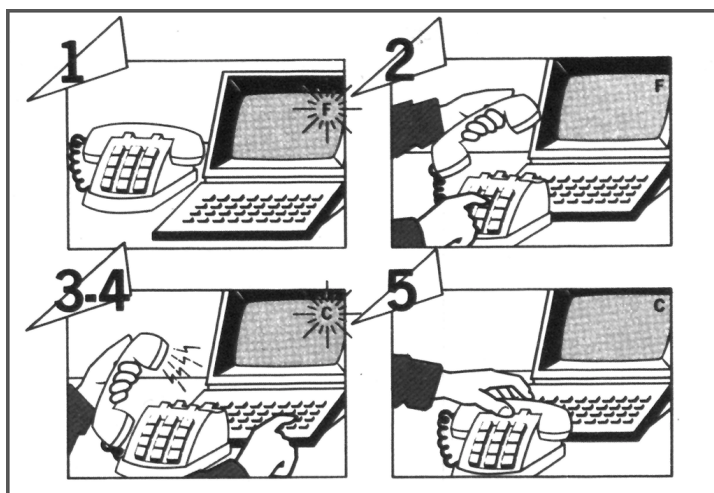
Guztira 100.000 kontzeptu baino gehiago, ondorengo hizkuntza hauetan: Euskara, espainiera, frantsesa eta ingelesa.

UZEIk landutako terminoloji banku hau euskara gaur eguneko gaietan erabili behar duen edonorentzat baliozkoa da.

Konexioa bi modutara egin daiteke:

- a) Minitel-aren bidez, hau da, telefonoari konektatutako terminal txikiaren bidez.
- b) Mikroordenadorearen bidez, hau da, modem baten bidez.

## MINTEL-AREN BIDEZKO KONEXIOA



1. Terminala piztu.
2. Telefonoa altxa eta seinalatutako zenbakia markatu.
3. Dei-tonua(k) hartzen da, altua eta baxua.
4. "KONEXIO/AMAIERA" tekla sakatu.  
"KONEXIOA EGINA" agertzen da pantailan.
5. Telefonoa eseki.

# ORDENADOREA LAGUN, TESTU-ZUZENKETAN

Xabier Artola\*

*Edozein testuk –honek ere bai, noski– izan du argitara aurretik zuzentze-prozesu bat edo beste. Denok ezagutzen ditugu zerbait inprimarazitakoan inprimategitik bueltan etorri ohi diren behin-behineko proba direlako horiek; behin-betiko alearen inprimaketari ekin aurretik egileak edo, kontu handiz erreparatu eta txukun-txukun zuzendu beharrekoa alegia. Hala ere, askotan –beti ez esatearren– zer edo zer itzuri egiten zaio zuzentzaile horren begiratuari, eta nekez aurkituko dugu inolako hutsik gabeko testurik.*

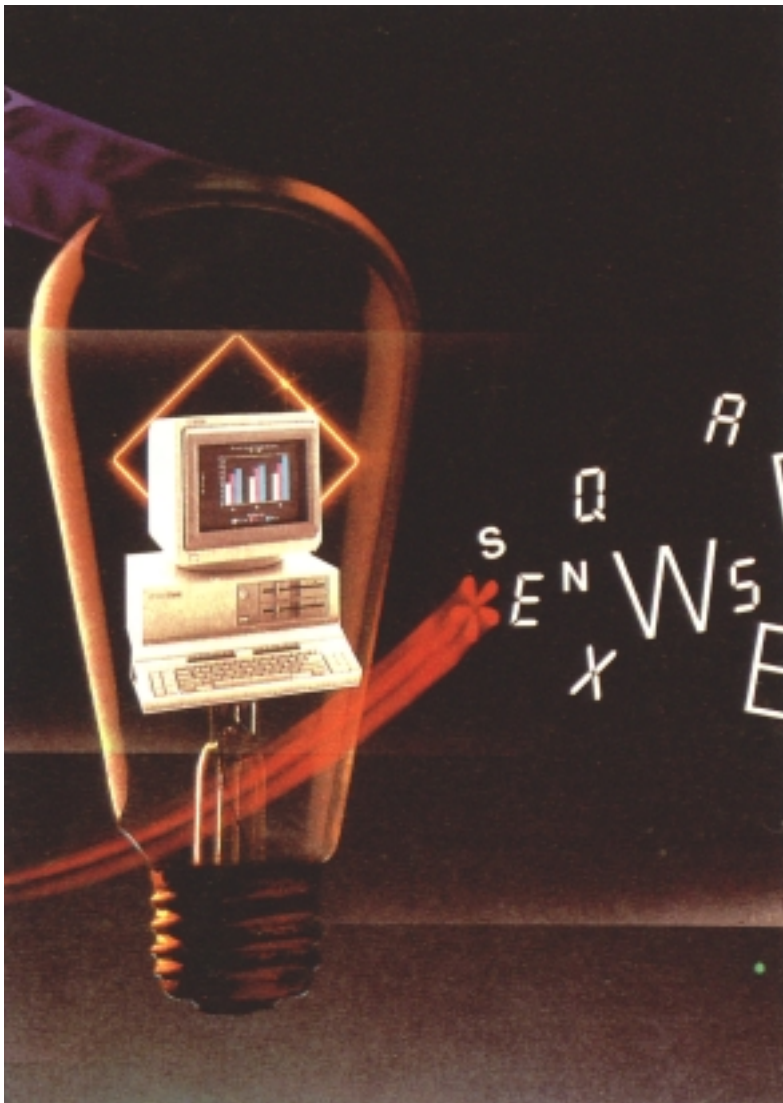
Gaur egun, gero eta gehiago, testuak konputagailuratu egiten dira testu-prozesaketarako programa baten laguntzaz, programa horiek ondorengo zuzenketa- eta orrazte-lanetarako izugarritzko erraztasun eta laguntzak eskaintzen dituztelarik.

Euskarazko testugintzan, edozein hizkuntzako arazoez gain (tekletze-erroreak eta abar) beselako zenbait ere ageri dira zuzenketaren alderditik begiratuta. Horra hor, besteak beste, testua makinanatu edo inprimategian konposatzen duen pertsona alfabetatu gabea –edo erdalduna– izateak ekarri ohi dituen era guztietako hutsegiteak, hizkuntzaren egungo batasun-egoera bereziari dagozkion arazoak, azken urteotan euskara hainbat eta hainbat eremu berritara zabaldu delarik horrek berekin dituen problema guztiak, eta abar eta abar.

Guzti hori dela eta, oraindik era askotako "irregulartasunak" ageri dira euskarazko testuetan. Gero eta gutxiago, hala ere. Ba bait dirudi arauak ez jakiteari edo axolagabekeriari zor zaizkion hutsak gero eta urriago direla. Baina zuzenketaren premia hor dago eta hor egongo da.

Eta ordenadoreak zertan lagun dezake? Lehen esan dugu gaur eguneko testugintzan ordenadorea gero eta maizago erabili ohi dela testu-prozesaketarako programek eskaintzen dituzten erraztasunak direla eta. Erraztasun horien artean besteak beste, hauek ditugu: testuaren formatua egokitzeko, testuzatiak batetik bestera aldatzeko, testu osoa berridazten ibili gabe testu-puska berriak txertatzeko eta abarretarako laguntzak. Horrez gain, zenbait hizkuntzatan –batez ere ingelesez, jakina, baina baita

\* Xabier Artola Informatika-Fakultateko irakaslea da.



gertuagoko erdaretan ere– idazten dugunaren ortografia automatikoki egiaztatzeko aukera ematen duten sistemak agertzen ari dira. Sistema horietako batzuk urrutirago joan nahi dute, puntuazio, sintaxi eta estiloaren aldetiko zuzenketa ere laguntzak eskaintzen dituztelarik.

Beraz, testu-prozesaketarako programen muenetan zuzenketa ortografikoa egiteko aukera izango dugu hemendik aurrera, baina bistakoa da laguntza hori emango duen sistema ez dela hizkuntza guztietarako; hizkuntza bakoitzerako bertsio desberdina eduki beharko duela baizik. Eta euskararentzat, nere ustez, hemen egiten ez badugu jai daukagu, puska baterako behinik behin. Aurki izango ditugu –dagoeneko badira– laguntza hori eskainiko diguten sistemak, beren erdarazko jatorrizkoetan aukera hori ematen badute ere.

oazen bada zuzenketa automatizatu nahi honek zer-nolako arazoak aurkezten dituen ikustera. Oraingoz ortografiaren alorrera mugatuko gara.

Bi motatako programak edo sistemak daukagu: alde batetik, **ortografi egiaztatzaileak**, hau

da, testua emanik testu horretan oker idatzita ageri diren hitzak ezaguterazten dizkigutenak, ondoren geronek zuzendu beharko ditugularik. Bestetik, **ortografi zuzentzaileak** ditugu edo ortografi egiaztaketa egiteaz gain zuzenketa ordenadorearekiko elkarrekintzan egiteko aukera emango digutenak, horretarako oker idatzitako –edo okertzat jotako– hitzaren ordezko posibleak ere proposatuz.

Arlo honetan estreinako ikerlanak 1957koak dira. Lehen produktu bukatua SPELL izeneko programa (1971) da.

Aurren-aurreneko programen zeregin bakarra, testu bateko hitz desberdinen zerrenda (agerpen-maiztasunaz ordenatua, jeneralean) ematea zen. Gero pertsonaren batek pazientzia handiz zerrenda hori aztertu egingo zuen eta oker idatzitako hitzak topatu (kontuan izan okerrak zerrendaren azken aldean agertuko zirela beren maiztasun txikia zela medio). Ondorengo programak hitzen nolabaiteko azterketa egiten hasi dira (digrama eta trigrama azterketan oinarriturik batez ere, hau da, letra-bikote eta letra-hirukote desberdinek hizkuntza bakoitzean dituzten maiztasun desberdinak kontutan hartuta kalkula bait liteke hitzaren berezitasun-indize bat) oker idatzita egon litezkeenak bakarrik zerrendatuko dituztelarik. Baina gaur egungo programen berezitasunik behinena, hiztegia erabiltzen duten

programak izatearena da. Hau da, hitz bat zuzen idatzita dagoen ala ez jakiteko, hiztegiara jotzen da: hitza han aurkitzen bada ontzat ematen da eta bestela ez.

Hiztegiaren erakuntza oso garrantzi handikoa da sistema hauetan. Izan ere, hiztegiaren neurria erabaki funtsezkoa gertatzen bait da: hiztegi horretan zer sartu eta zer ez?. Gehiegi pentsatu gabe onena "dena" sartzea dela ematen du. Baina berehala konturatuko gara hori egitearen arriskuaz: hitz zaharkitu asko edukiko ditu hiztegi horrek, alferrik gehienak, eta oker idatzitako hitz arruntak ontzat emateko posibilitatea handiagoa izango da. Euskara, adibidez:

'aueta' hitza hiztegiaren izanik

**aueta** (BN-gar), **augeta** (BN-sal) serenata, alborda / aubade, sérénade (Azk.)

'auetako' hitza onartu egin beharko du, 'hauetako' idatzi ordez hutsegitea dela askoz seguruago izanik ere.

Hiztegi urriegiaren akatsak ere nabarmenak dira, ongi idatzitako hitzak (hiztegiaren ez daudela eta) txartzat jotzeko duen arriskuarekin.



Honetaz gain, sistema hauen eraginkortasunari begira bistakoa da egin beharreko lanik kritikoenetako bat hiztegi bilaketa horixe dela, eta beraz, hiztegiaren neurria eta antolaketa oso faktore garrantzizkoak gertatzen dira. Bilaketan denbora irabazteko usuen erabilitako estrategia, maiztasunik handieneko hitzen tratamenduan datza: azterketa estatistiko baten bitartez hitz horiek zeintzuk diren jakingo dugu eta hiztegiaren egin beharreko bilaketa bi mailatan bereiztuko da: lehenik ea testuko hitza maiztasunik handieneko horien artean dagoen begiratuko da (bilaketa hau aguro burutuko da, hainbeste hitz ez dira eta), eta ez badago (eta orduan bakarrik) hiztegi orokorrera joko da. Maiztasun handieneko hitzen azpizitzi horretan 250etik 500era bitarteko hitz-kopurua eduki liteke, testuko hitzen %50 gutxi gorabehera horien artekoa suertatzea espero delarik.

Orainarte, ortografi egiaztatzaileetara mugatu gara. Baina gaur egun salgai dauden programa gehienek elkarrekintzazko zuzenketarako laguntza ere ematen dute: ortografi zuzentzaileak ditugu. Hauen berezitasuna, lan egiteko eran datza. Programak ortografiaren egiaztapena egiten duen bitartean erabiltzailea pantaila aurrean ari da. Programak hutsen bat topatu orduko, pantailan hitza nabarmen erazi eta erabiltzaileari hitz horri zer egin behar dion galdetuko dio. Orduan erabiltzaileak aukera desberdinak ditu: hitza zuzendu dezake edo sistemari hitz horren ordezkio posibleak eman diezazkion eskatu eta gero horien arteko bat –zuzena, noski– hautatu. Hitza bere horretan uzteko aukera ere ez zaio ukatuko. Gainera, sistema gehienek maneiatzen dute erabiltzailearen hiztegi delako bat, non erabiltzaileak sistemak eza gutzen ez duen eta berak erabili ohi duen zenbait hitz sar dezakeen. Handik aurrera berriro agertutakoan, sistemak ez dizkio okertzat hartuko.

**G**uzti honen aurrean, bada zenbait hilabete euskararentzako zuzentzaile ortografikoa lehen helburutzat duen proiektuari ekin zaiola.

Proiektu honetan partaide APIKA informatika-eremuko zerbitzu-enpresa, UZEI eta Donostiako Informatika-Fakultateko lengoia naturalaren prozesamenduaren alorrean diharduen taldea dira. Esan bezala, talde honen lehen asmoa ortografi zuzentzaile elkarreragilea ordenadorearen laguntzaz euskaraz idazten duen orori eskaintzea litzateke. Garbi gera bedi oraingoz ortografiaz besterik ez garela ari eta, beraz, hitz bat (zuriguneen arteko karaktere-segida) ontzat emateko karaktere-segida horri erreparatzen zaiola, hitz horren testuinguruaz ezertarako jabetu gabe. Nere eritziz, euskarazko testuetan ageri diren huts asko eta asko honela zuzentzea ez da posible, sarri sintaxiari edo besteri dagozkion hutsak dira eta. Esaterako, *iraksakuntza*, *aztelehena*, edo *emen* idatziz gero okertzat joko dira, baina ez da inolako hutsik harrapatuko *zu esan didazu bezala*, *gaur astelehena da* bezalako esalditan; hitzak banan-banan hartuta onesteko modukoak bait dira. Horrelakoei antzeman ahal izateko, hiztegi-informazio ortografiko soilaz aparte bestelako informazio asko da beharrezkoa; perpausaren analisi morfosintaktikoak emandakoa besteak beste. Utz dezagun bada hau geroxeagorako; ez bait da lan makala, nahiz eta egunen batean horri ere heldu beharko zaion.

Gorago proiektuan partaide den lengoia naturalaren tratamenduaren alorreko taldea aipatu dugu. Informatikaren munduan erabili ohi diren lengoia artifizialak direla medio, hizkuntza arruntari buruz hitz egiteko lengoia natural delako deitura bitxia sortu da. Hizkuntzaren ulermena eta hizkuntzaren sorkuntza automatikoaz diharduen arlo honen garrantzia gero eta handiagoa da. Ordenadoreekin lengoia naturalez (euskaraz gure kasuan) komunikatu ahal izateak berez duen garrantzia du alde batetik, eta hizkuntzaren beraren ezagumendu teorikoari ordenadoreen logika eta sistemetara egokitze horrek egiten dion ekarpena bestetik.

Lengoia naturalaren prozesamenduaren lan-eremu honetarako ere zenbait tresna eta sistema orokor egonagatik, hizkuntza bakoitzak behar du

bere azpiegitura berezia: oinarri-oinarrizkoak edozein hizkuntzatan analizatzaile morfologiko eta sintaktiko automatikoak dira. Ondoren etorriko dira semantikaren eta pragmatikaren arazo nahasiagoak.

Euskara bezalako flexio-maila handiko hizkuntza batek, problema morfologiko bereziak aurkezten ditu bere tratamendu automatikoari aurre egin nahi zaionean. Problema horiek ebatziz gero ordea, morfologi mailako analitik beretik lortzen den informazioa askoz ere aberatsagoa gertatzen da morfologia ximpleagoa duten beste hizkuntza batzuetan baino. Informazio hori garrantzi handikoa da berriz analisiaren ondorengo mailetan, hau da, hizkuntzaren sintaxia eta semantika aztertu nahi direnean. Gurea bezalako hizkuntzetan analisi morfologikoa da hizkuntzaren tratamendurako edozein sistemak ebatzi beharko duen lehen arazoa.

Bestetik, zuzentzaile ortografikoak behar duen hiztegiaren neurrien garrantzia aritu gara lehenago. Bistakoa da euskara bezalako hizkuntzetan hitz bakoitza bere flexio guztiekin (eta deklinabide-mailako flexioez besterik ez gara ari) hiztegitatuz gero hiztegia ikaragarri haziko dela, bertan bilaketak egiteko denbora ere luzeago izanik. Beste hizkuntza

batzuetan (pentsa adibidez ingelesaren morfologian) arazo hau gutxietsi egin dute eta batzuetan hitz-forma guztiak hiztegian sartu izan dituzte. Morfologia konplexuagoko hizkuntzetan ordea, problema honi behar den bezala heldu behar zaio eta hiztegian lemak besterik ez dira edukiko, tratamendua zailago gertatuko bada ere.

Esandakoaren arabera, hitz baten zuzentasun ortografikoa egiaztatzea ez da hitz hori hiztegian dagoela ziurtatze soil izango. Izan ere, hitz osoak ez bait du hiztegian zer egonik. Hitzaren erroa lehen hiztegian dagoela gehi lema horri lot lekizkiokeen atzizkien segida duela atzetik izango da egiaztatu beharrekoa, hitza ontzat emango bada. Beraz hitzaren zuzentasuna egiaztatzeke hitzaren analisi morfologikoa (oso-oso ez bada ere) burutuko da.

Ikusten denez, analizatzaile morfologikoa eta zuzentzaile ortografikoaren arteko lotura izugarri estua da eta batabestearen azpiproduktu gisa ikus daiteke. Bi horiek izan nahi lukete bada, aipatutako talde honen lehen emaitzak; euskararen tratamendu automatikoaren bidea urratzeko lanean abiapuntu hain zuzen. ●



# MIKROINFORMATIKA ETA ZENBAIT KOMUNIKAZIO-ARAZO

A. Arruabarrena, G.Buldain, L.Gardeazabal, E.Gómez, J.González Abascal\*

Informatika, ordenadorea, mikroprozesadoreak, ... arrunt bihurtzen ari diren hitzak ditugu. Teknika horietan oinarriturik, hamaika aplikazio daukagu jadanik kalean. Ondorengo orrialdeetan, aplikazio berezi bat azalduko dugu; normaltzat hartzen ditugun ahalmen fisikoak murrizturik dituzten pertsonentzako laguntza teknikoaren garapena, hain zuzen ere.

## Komunikazio-arazoak

Pertsonen arteko harremanetarako erabiltzen diren bideen artean ahotsa dugu garrantzitsu edo erabiliena, idazmena osagarria delarik. Baina batzuetan bi espresio horiek itxita egoten dira. Hori gertatzen da, adibidez, garun-paralisia nozitzen duten zenbait pertsonengan, beste akatsen artean mintzamen edo idazmenerako behar den gorputzaren mugimendukontrola ez edukitzeagatik. Horregatik, beren bizitza erabat hertsia da; besteekiko dauzkaten harremanak zaildurik bait daude oso. Ezin bait dute mintzatu edo idatzi eta, kasu larriengan, ezta mugimendu kontrolaturik egin ere.

Duela bi urte Informatika-Fakultate eta ASPACEren arteko proiektu bati hasiera eman zitzaien, pultsadore bakar baten bitartez kontrola litekeen komunikaziorako tresneria berezia garatzeko asmoz. Proiektu hori bi adarretan zegoen banaturik: alde batetik hardware komertzialaren egokitzapenean eta besteetik komunikaziorako hardware bereziaren eraikuntzan.

Komunikazio-arloa oso zabala denez, gure lana hiru atalatan banatu dugu:

- testu-edizioa: komunikagailuak, hau da, idazteko edo bere nahiak komunikatzeko erabil daitezkeen tresnak.
- ordenadoreen erabilpena, tresna hauek beren beharrei egokituz.
- inguru-kontrola, bizitza arruntean egiten dugun milaka ekintza sinpleen laguntza.

Hauxe izan da gure helburu nagusia: arlo horietan gaitasun eza nozitzen duten pertsonen ahal den autonomiarik handiena eskaintzea; beren mugimenduen kontrolik ez daukatenei zuzendurik batez ere.

## Komunikagailuak

Komunikagailu elektronikoa, pantaila batean esaldiak idazten dituen tresna da. Esaldi edo mezuak idazteko, erabiltzaileak pultsadore bakar bat erabiltzen du, ez bait dezake teklatu arrunta

kontrola (bestela, idazmakina erabiliko luke, adibidez, bere komunikaziorako). Karaktere bat idatzi ahal izateko, komunikagailuak karaktere-multzoa banan-banan eskainiko du bere pantailan. Erabiltzailea zai egongo da berak idatzi nahi duen karakterea pantailan agertu arte, eta orduan, berak kontrolatzen duen pultsadorea sakatuko du. Horrela, beste karaktere bat gehituko du idazten ari den esaldian.

Pantaila bi zatitan banaturik dago. Goiko partea, esaterako, esaldiak idazteko libre dago eta beheko partea karaktere-aukeraketa gauzatzeko da. Karaktere-multzoa egitura matrizialean kokaturik dago. Komunikagailuak, lehenbizi, lerroz lerro eskainiko ditu karaktereak. Pultsadorea lehenengo aldiz sakatzen denean, berriz, karakterez karaktere hasiko da eskaintzen aukeraturako lerroaren barnean. Prozesu honi ekorketa deitzen diogu. Horregatik, karaktere bat idazteko bi aldiz sakatu behar da pultsadorea: leroa aukeratzeko lehenengoan eta karakterea bigarrenean. (1. irudia)



1. irudia. Komunikagailu eramangarriaren egitura orokorra. Pantaila bi zatitan banaturik (testua idazteko eta karaktere-aukeraketa burutzeko) guztia pultsadore batez kontrolaturik.

Esaldiak letraz letra osatzen dira, beraz, idazketa-prozesu hau nahikoa geldoa da. Prozesua azkartzeko, karaktereak ez daude alfabetoaren arabera kokaturik; beren erabilpen-maiztasunaren arabera baizik. Adibidez, euskaraz e bokala f konsonantea baino askoz erabiliagoa da, eta hori medio, matrizearen lehenengo posizioetan ageriko da, azkarrago idazteko. (2. irudia).

Karaktere bat idazteko itxoin behar den denbora, hauxe da:

$$T = (LZ + ZZ) * T_{itx}$$

\* Donostiako Informatika-Fakultatea



Ø A I T K G P  
 E N O D S J C  
 R Z B H X V W  
 U L M F T Q

**erabilpen-maiztasunaren taula**

ø	0.141607	O	0.045926	G	0.022305	F	0.001945
A	0.132010	Z	0.045012	S	0.021705	C	0.000756
E	0.111920	U	0.042258	H	0.016129	V	0.000334
I	0.075234	K	0.041597	M	0.009102	Y	0.000151
N	0.072017	D	0.026052	P	0.007797	W	0.000077
R	0.068016	B	0.025694	J	0.003208	Q	0.000046
T	0.063202	L	0.022727	X	0.003011		

2. irudia. 4x7 matrize batentzako, egitura matrizar egokiena euskara idatziaren letra bakoitzaren maiztasuna kontutan harturik.

non LZ karakterea dagoen lerro-zenbakia, ZZ zutabe-zenbakia eta  $T_{ix}$  komunikagailuak erabiltzen duen eskaintza-denbora diren.

Adibidez, irudian agertzen den karaktere-multzoa erabiliz, n bat idazteko behar den denbora hau da:

$$(2 + 2) * T_{ix} = 4 * T_{ix} \approx 4 s$$

( $T_{ix}$  = segundo bat, hartzen baldin badugu) n karakterea bigarren lerroko bigarren letra bait da.

Komunikagailuaren funtzionamenduan badago oso garrantzitsua den parametro bat:  $T_{ix}$  eskaintza-denbora, hau da, komunikagailuak pultsadorea ea sakatzen den itzaroten duen denbora, erabiltzaileak momentu horretan "aktibatuturik" dagoen karakterea aukera dezan. Pultsadorea sakatzen ez bada, hurrengo karakterea aktibatuko du. Parametro horrek erabiltzailearen kontrolpean egon behar du zeren eta hasieran ezin bait du komunikagailua oso azkar maneiatu. Gero, ikasketa-prozesua aurrera doan heinean, denbora hori asko labur daiteke. Horregatik, komunikagailua  $T_{ix}$  kontrolatzen duen funtzio batez horniturik dago, erabiltzaileak makina bere beharretara egoki dezan.

Letraz aparte, erabiltzaileak kontrol-funtzio batzuk ere erabil ditzake testu-idazketa erosoagoa izan dadin: karaktere bat ezabatzeke, testuan zehar kurtsorea mugitzeko, maiuskula/minuskula aukeratzeko, eta abar. Edizio-funtzio hauek ere sekuentzialki aukeratu behar dira, beste karaktereak bezala (irudian letra grekoz adierazita daude).

Azaldutako karaktere eta funtzio guztiak ez dira pantaila batean sartzen. Horregatik, bi sailetan (edo gehiagotan) daude kokaturik. Lehenengoan, irudian agertzen denean, letrak eta zenbait kontrol-funtzio agertzen dira. Bigarreanean, berriz, zenbakiak, karaktere ortografikoak eta kontrol-funtzio gehiago. Beraz, sail batetik bestera pasatzeko kontrol-funtzio bat izan behar da. Aukeratzuz gero, pantailan agertzen diren karaktereak aldatu egiten dira, beste sailekoak ager daitezen.

**Testuen aurreidazketa**

Testuen idazketa azkartzeko badago zenbait aukera eta hauen artean predikzioa edo aurreidazketa dago. Teknika hau zertan den argitzeko, jar dezagun adibide bat.

Eman dezagun orain arte honelako testu bat idatzi dugula:

**Mesedez, ireki le\_**

Hori irakurrita, ezer gehiago idatzi gabe denok asma deza-kegu ireki behar dena zer den. Zergatik ez du, bada, komunikagailuak berak hitza bukatzen?. Ideia hau inplementatzeko eman dezakegun lehenengo pausoa, komunikagailua hiztegi batez hornitzea izango da, non hitzak eta beren erabilpen-maiztasuna aurkituko bait dira. Orduan, komunikagailuak bere hiztegiaren barnea miatuko du posible diren hitzak bilatzeko. Letraz letra idazteko aukeraz gain beste aukera bat eskainiko du: hitzen hautaketa.

Goiko esaldia bukatzeko, adibidez, bere hiztegitik hartutako hitz hauek eskain ditzake (le-z hasten direlako eta bere hiztegian le-z hasten direnen artean maiztasunik handiena dutelako):

**leioha lehen leioha lehengusua**

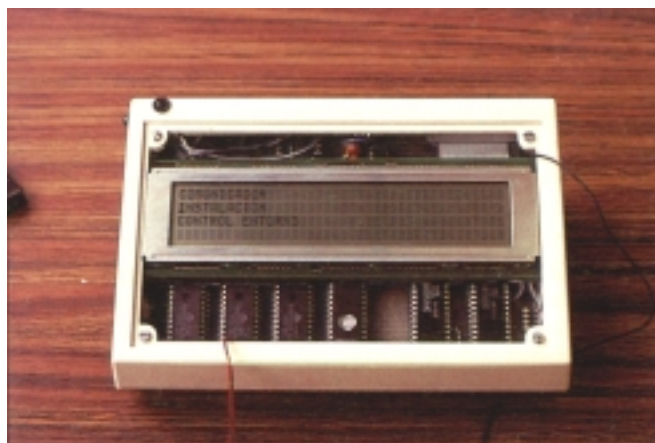
Horrela, erabiltzaileak lau hitz horien artean aukeratu ahal izango du, eta pultsadorearen sakatze batez hitz osoa idatziko du. Berak idatzi nahi duen hitza hor aurkitzen ez bada, lehenengo bidari jarraituko zaio, beste letrak banaka idatziz.

Teknika hau are eta eraginkorrago egiteko, adimen artifizialaren erremintak erabil ditzakegu. Hauen bitartez, aurreidazketa burutzeko ez dugu hitzen maiztasuna bakarrik kontutan hartuko; baizik eta esaldiaren zentzua ere bai. Horrela izanik, eta lehenengo adibideari jarraituz, komunikagailuak idazteko aukera bat bakarrik eskainiko luke: leioha. Ez bait du zentzurik lehen, leioha edo lehengusua irekitzeak.

**Mezu estandarrak**

Arestian azaldu dugun esaldiaren eraikuntza letrak letra egin da, prozesu honen bitartez edozein esaldi idazteko gauza izan dadin. Badago, hala ere, "esaldi eginak" azkar idazteko beste aukera bat, hots, karaktere bat aukeratzekoan esaldi oso bat pantailan agertzea.

Adib. A aukeratzten da eta  
*Piztu Argia !*  
 agertzen da pantailan



Komunikagailua horrela erabili nahi bada, esaldiak bi mailatan antolatuz aurkituko dira, esaldi-hiztegi bat osatuz: gaiak eta esaldiak. Lehenengo karaktere-aukeraketaz, "gaia" aukeratzeko da eta bigarrenetan esaldia bera. Modu honetan, komunikaziorako testu estandarrek oso era eraginkorren idatz daitezke, baina esaldi-multzo tinko bati loturik. Esaldi-multzo hau kargatu egin behar da hasieran komunikagailuan eta pertsona konkretu bati egoki dakioko.

## Mezuen aurkezpena

Eraikitako testu guztiak komunikagailuaren pantailan agertzen dira. Pantaila hau LCD (kristal likidozko pantaila) displaya izango da eta beharrezkoa bada, koloretakoa. Baina badaude beste aukerak ere. Eraikitako esaldiak inprimagailu batera bidal daitezke, testuak etxean edo eskolan paper gainean uzteko. Ahots-sintetizadorea ere erabil daiteke irteera bezala, egindako esaldiak "ahoskatuak" izan daitezen (behar bada kalean erabiltzeko) eta abar. Posibilitate guzti hauek erabiltzailearen esku daude, testu bat idatziz gero behar den bezala ateratzeko.

## ORDENADOREEN ERABILPENA

Komunikazio hitza zentzu zabalean hartzen badugu, orain arte aipatutako komunikazio-prozesuaz gain beste komunikazio-beharrak ere baditugu. Adibidez, gure inguruan dauzkagun tresnekiko komunikazioa. Tresna horien artean ordenadorea dugu, gaur egun, garrantzitsuenetako bat.

Gure aginduak ordenadorera bideratzeko erabiltzen dugun ohizko bitartekoa teklatura da. Baina, idazmakina bezala, teklatura maneiatzeko behatzen mugimendu-kontrol zehatza behar dugu. Kontrol hori ezinezkoa suertatzen denean eta eskuen edo

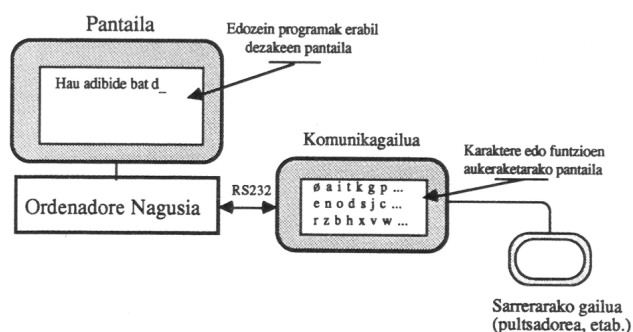
kontrolatzeko, bi gauza egin behar dira:

- komunikagailua ordenadorearen sarrera bati konektatu (normalean RS-232-C sarrera/irteerari seriean)
- ordenadorean programa egoiliar bat egikaritu, horren bitartez konputagailuak teklaturik espero dituen karaktereak komunikagailua konektaturik dagoen sarreratik irakur ditzan.

Horrela egiten bada, ordenadorearen pantailara bidalitako karaktereak, bere teklaturik etorriko balira bezala ageriko dira eta komunikagailuaren pantaila teklaturaren karaktereak aukeratzeko erabiliko da. Karaktere bat aukeratu gero, ez da, lehen bezala, komunikagailuaren pantailara bidaliko; ordenadorera baizik.

Ordenadorean egikaritzen ari den programa konkretu batek teklata bat baino gehiago batera sakatzea eskatzen badu, hots <ctrl c> eta abar, orain banan-banan bidaliko dira, hau da <ctrl> eta <c>. Aipatutako programa egoiliarrek bi karaktereak lotuko ditu karaktere bat osatzeko.

Horrela, pultsadore bakar batez kontrolaturik, edozein programa erabil daiteke makina honen ahalmen guztia elbarriei eskainiz. (3. irudia)

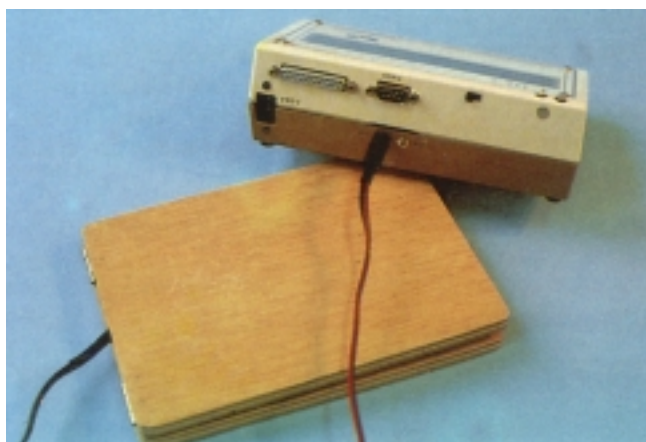


3. irudia. Komunikagailua ordenadore pertsonalaren teklatura ordezkatzeko. Komunikagailuaren pantaila teklaturaren karaktere-aukeraketarako erabiltzen da, testuak ordenadorearen pantailan agertuz.

## INGURUAREN KONTROLA

Azken zatian, kontrolatutako higidurarik ez daukaten pertsonentzako ordenadorearen erabilpen berria komentatu dugu, pultsadore bat bakarrik erabiliz. Gure inguruarekiko dauzkagun harremanak bideratzeko badaude beste behar batzuk ere. Batzuetan oso ekintza sinpleak dira; esaterako gelako argia piztu edo itzali, leiho edo ate bat ireki edo itxi, pertsiana bat igo edo jaitsi, telefonoa hartu edo zenbaki bat markatu, eta abar. Mota honetako ekintza guztiak betetzeko beharrezkoa da, nahitana-hiez, goian aipatutako higidura-kontrola. Gurpilezko aulki bati (edo ohe bati) loturik dagoen pertsonak ezin ditu erabat arrunt ditugun eragiketa horiek burutu eta are gutxiago bere eskuak kontrolatu.

Aurkezten ari garen bezalako sistema baten bitartez, eta izpi infragorrien igorle/hartzaile batzuetan oinarriturik, agindu guzti horien aukera errealitate bihurtzen da. Egunero erabiltzen ditugu infragorrien igorleak gure etxeko gailuak kontrolatzeko. Adibidez, telebistako kanalak aldatzeko, edo garajeko atea



ukabilen kolpe batez bakarrik baliu gaitzekenean, aukera bakar bat dugu: atzipen-sekuentzian oinarritutako sarrera-sistema bat erabiliz ausazko atzipena erabiltzen duen sarrera-sistema ordeztea. Eta hori izan da arestian testuak idazteko erabili dugun sistema: karaktereak multzoka edo banan-banan eskaintzea, pultsadore bakar batez kontrolatu ahal izateko.

Izan ere, karaktere bat idazteko komunikagailuaren eskaintzari itxoin egin behar diogu. Beraz, komunikagailua ordenadorerako sarrera-gailu bezala erabili nahi badugu, azken hau

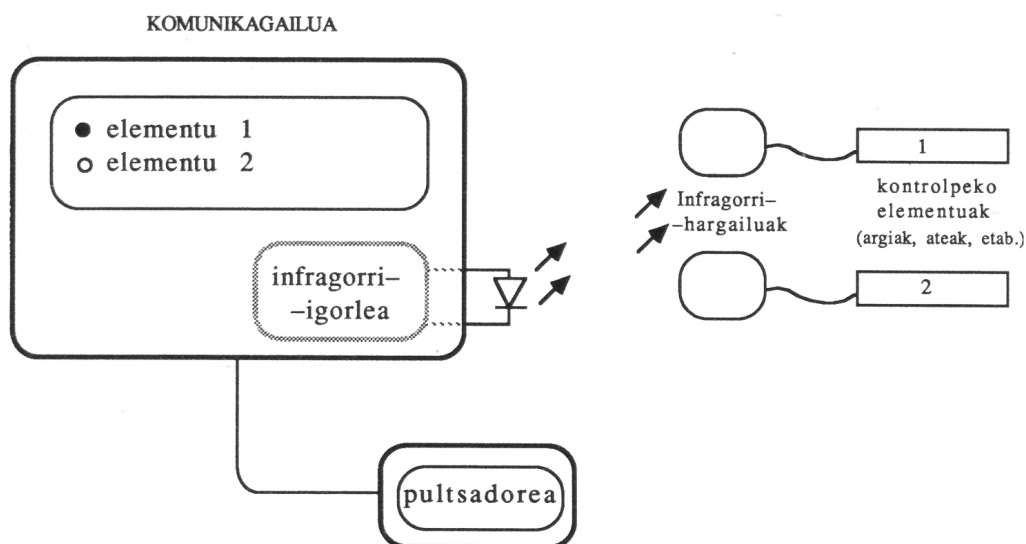
irekitzeko, eta abar. Egin nahi dena, beraz, argitan, ateetan, telefonoan eta abarretan, hargailu bat kokatuz igorle batez kontrolatzea da. Horrelako sistema erabiltzeko baldintza bat bakarrik eskatzen da: teklatu bat kontrolatu ahal izatea. Pertsona hauei bere bizitzako zenbait arlotan ahal den autonomiarik handiena eskaini nahi badiegu, "ohizko" teklatura egokiago zaien sistema batez aldatu behar da, hots, aukeraketa sekuntzialaz.

Azaltzen ari garen tresnan bertan, hirugarren posibilitate hau jar dezakegu. Berarekin lan egiten baldin bada, kontrolpean dauden gailuak pantailan ageriko dira, gero ekorketa sekuntzialaz aukeratu ahal izateko. Bat aukeratzen denean, komunikagailuak izpi infragorrien bitartez kode bat bidaliko dio hargailuari eta azken honek dagokion sistema aktibatuko du: argia, atea, telefonoa, etab.

Kontrol hau gauzatzeko, etxebizitzatan edo ikastetxetan zenbait aldaketa egin behar da, hots, hargailuak kontrolatu nahi diren aparatutan kokatu behar dira. (4. irudia)



RAM, sarrera/irteera serie eta paralelo, modem-a), lehenengoan ez bezala, ordenadore pertsonalaren teklatura -PC batekoa hain zuzen- ordezkatzeko sistema ere diseinatu da.



4. irudia. Inguru-kontrola infragorri-igorle baten bitartez. Pantailan kontrolpean dauden gailuak agertzen, dira pultsadore erabiliz bat aukeratu ahal izateko.

## KOMUNIKAGAILU ERAMANGARRIA

Artikulu honetan azaldu ditugun hiru funtzioak edozein makinatan eraiki daitezke egokitze-prozesu baten ondoren; pultsadore bat konektatuz eta dagokion softwarea garatuz.

Baina komunikazio-sistema eramangarria diseinatu nahi badugu (adibidez gurpilezko aulkian kokatzeko edo kalean erabiltzeko), bide konkretu bat urratu behar da: diseinu berezi eta egokia sortzearena. Hasieran esan dugunez, bi bideak landu dira proiektu honetan.

Alde batetik, merkatuan dauden bi makina erabili dira. Bata, Sharp kalkulagailua eta bestea, Tandy ordenadore pertsonala. Lehenengoan, 6 K RAM duen kalkulagailua, komunikagailu simple bihurtzen duen softwarea garatu da. Oso makina txikia izanik (poltzikokoa) egokia da adibidez kalean erabiltzeko. Bigarrenean, ahalmen handiagoa duen makinak (32 K

Bestetik, arazo hauei dagokien diseinu berezi bat sortu da -JAL2- esandako hiru helburuekin: testu-edizioa, teklatuaren emulazioa eta inguru-kontrola. Horretarako, 64 K RAM, sarrera/irteera serie eta paralelo, 16 kanaleko infragorri-igorlea eta abar duen sistema bat osatu da, 8 biteko mikropresadorez kontrolaturiko sisteman. Azken diseinu honek posibilitate gehiago ditu besteekin konparatuz (gailu batean komunikazio-funtzio guztiak biltzen bait ditu) eta eman behar diren garapenarako abiapuntu bezala hartu da. (5. irudia)

## ETORKIZUNA

Komentatu dugun kasua muturreko kasua izan da; pultsadore bat bakarrik kontrola bait daiteke. Baina badira bitarteko kasuak ere; higidura-ahalmen gehiago dagoenean. Kasu hauetan ere, posible da pultsadore bat baino gehiago komunikagailura konektatzea edo zenbait teklatu berezi, handiagotuak, ordenadorera konektatzea eta abar. Hartuko den irtenbideari,



baina hori ez da horrela. Lengoaia ulertzea oso prozesu konplexua da eta gaur eguneko ordenadoreak urrun ikusten ditu giza adimenaren ahalmen linguistiko eta orokorrak.

Telebistako kirol-komentarista baten azalpena, 093 telefonoan zer ordu den esatean duen leloa eta "100 metro" nobela euskaraz egon arren, bakoitzean erabiltzen diren hitzak, esateko erak eta esanahiak erabat desberdinak dira. Euskaraz dakien edonork ez lituzke hiru kasuotako maila desberdinak bereiztuko (hirurak euskaraz daude eta), baina ordenadore bidez ulertzen saiatzen bada, desberdintasunak berehala nabaritu ditu. Berez ordenadoreak ez dira hizkuntzaren zailtasun guztiei aldi berean aurre emateko gauza. Emaizta probetxugarriak lortzeko, ordenadorearen lana domeinu espezifiko eta mugatu batean kokatu behar da. Etorkizunean, aplikazio mugatuzko sistemak bilduz, lor litezke ahalmen handiagoko sistema berriak, baina oraingoz martxan dauden aplikazio-motak helburu espezifikoetarako dira eta beren arteko loturak hutsean daude.

1. taulan Lengoaia Naturalaren gaur eguneko aplikazio-mota posibleak biltzen dira. Bakoitzarentzat produktu erabilgarriak noraino heldu diren azaltzen da. Taulan bi aplikazio-multzo nagusi bereizten dira: alde batetik gizaki eta ordenadorearen arteko komunikazioa errazten dutenak, eta bestetik giza komunikaziorako aplikazioak. Taulako aplikazio konkretuak beste modu batera sailkatzen dira: martxan daudenak, ia erabiltzeko moduan daudenak eta oraindik proiektu-mailan daudenak.

Ondoren aplikazio-motak, bakoitzaren deskribapena eta aplikazio zehatzak azalduz deskribatuko diren aplikazioen nazioarteko maila utzi ondoren, Lengoaia Naturalak gure inguru hurbilean lortu dituen oihartzun eta fruituak bilatuko ditugu, eta Euskara berriro plazara jaldi dadin bide bat proposatzen da bukaeran.

### *Datu-baseen galdeketa-sistemak*

1. taulak erakusten duenez, dagoeneko merkatura iritsi diren aplikazioak bi motatakoak dira: datu-baseen galdeketa-sistemak (bai ordenadore erraldoietan, bai mikroetan) eta itzulpen automatikoa. Sail honetan aztertuko dugun lehenengo mota horretan, datu-baseak erabiltzea erosoagotu egin nahi da.

Datu-base batean gai konkretu batez datu asko sartzen dira formatu zehatz baten arabera, geroago datu horien artetik zenbait baldintza betetzen dutenak erraz eta azkar atera ahal izateko. Adibidez: enpresa handi bateko enplegatuen datu-basean berehala jakin daiteke bost urtetan zeintzuek lan egin duten edo sail jakin bateko langileen zerrenda azkar lor daiteke. Baina datu-baseari galderak egin ahal izateko, bere lengoaia berezia ezagutu behar da. Betebehar hau dela eta, datu-baseen erabilpena askoz murriztagoa izaten da. Beraz oso interesgarria da galderak lengoaia naturalez egin ahal izatea, baina horrelakoetan sistemak berak galderak datu-basearen lengoiara itzuli beharko ditu. Datu-baseei buruz azaltzen dugun guzti hau, beste edozein aplikaziotarako programez ere esan daiteke.

	Aplikazio-motak	Definizioa	Azken urteotako egoera	Enpresa eta salgai nagusiak
Ordenadorearekin komunikaziorako	Datu-baseen galdeketa-sistema	Galderak lengoaia naturalez egiten dira	Lehenengo salgaia 1981.ean Ehunen bat sistema merkaturatua	Cornegie Group (Language Craft) Symantec (Question & Answer) Texas Instruments (Natural Link)
	Elkarrizketarako interfaceak	Sistema konplexuak elkarrizketa-moduan erabiltzeko bihurtzekoak	Prototipoak (ez dira guztiz operazionalak): Diru-inbertsionarako ahokulariak Irakaskuntzarako sistemak	Zenbait ikerkuntz proiektu SOPHIE MYCIN eta beste sistema adituak
Giza komunikaziorako	Testuen eduki-araketa	Mezu-prozesaketa, datu bat bilatzeko edo laburpen bat egiteko	Prototipoak: Bankuen telexen prozesaketak Itsasuntzietako mezuak	Carnegie Group Cognitive Systems Smart AI
	Testu-edizioa	Ortografi eta idazkera-erroreak bilatzeko, noizbait zuzenketa posibleak erakutsiz	Ortografi zuzentzaileak salgai, baina esanahia eta sintaxia ez dute aztertzen.	IBM, Open ACCESS Macintosh, ...
	Itzulpen Automatikoa	Testu baten itzulpena Lengoaia Natural batetik bestera ordenadore bidez	Nahiz eta guztiz zuzenak ez izan, zenbait sistema erabilgarri eta errentagarri. Lehenengoa 1960.tik.	SYSTRAN, ALPS, Logos Weidner METEO, METAL
	Ahozko idazmakina	Ahozko esaldiak testu idatziak bihurtu	Oraindik garapen-mailan Pertsona jakin baten hitz solteak ulertzen dira.	Zenbait ikerkuntz proiektu IBM, MIT Carnegie-Mellon Edinburg

1. taulan Lengoaia Naturalaren aplikazioak.

Alde ikaragarria dago ordenadore erraldoietarako eta mikroetarako egindako interface-en artean; bai prezioz (milioiak eta hamarnaka mila pezeta inguru hurrenez hurren) eta bai ahalmenez. Hizkuntzaren tratamendua erabat zabalago eta sakonagoa izateaz gain, sistema handietan erabiltzailearentzako laguntza eta erraztasun handiagoa dago. Erabiltzaile anitzi erantzun dakioke eta datu-base ahaltsuagoak atzitzeko aukera eskaintzen da. Mikroetan kokatutako interface-ak, guztiz desberdinak dira. Nahiz eta Lengoia Naturalaren Prozesamenduko ikertzaile gehienek mikroetako pakete hauek gutxies-teko joera ukan, merkatuan bada zenbait sistema interesgarri. Erosleen erantzuna aztertzeke dago oraindik.

Ordenadore handietarako sortu zen lehenengo sistema, 1981. urtean Artificial Intelligence Corporation-ek kaleratutako "Intellect" izan zen. 1984. urteaz geroztik ordenadore-enpresa nagusiek berori egokitu edo antzeko sistema berriak asmatu dituzte (Mathematica, IBM bere ekipoentzat, BBN eta IBS Digital-en VAX sistementzat). 1985.ean Carnegie-Mellon-eko unibertsitateko "Carnegie Group" izeneko talde ausartak, "Language Craft" izeneko hizkuntz tresnen multzoa atera zuen. Beronek VAX sistemetarako lengoia naturalazko interface-ak eraikitzeke aukera eskaintzen die programa-injineruei; adimen artifiziala erabiltzeke bereziki.

Mikroen munduan Lengoia Naturalaren aitzindaria, Mexiko Berriko Excalibur Technologies enpresaren "Savvy" sistema izan zen. Bere ondorengoez bezala ez zuen hizkuntz analisirik egiten; zenbait gako-hitz aurkitzean geratu egiten bait zen. Zabalkunde handirik ez zuen lortu 1984.era arte; orduan jaio bait zen IBM-PC-rako RBase paketearen "Clout" sistema. Symantec-en "Question & Answer" sistemak arrakasta ederra izan du 1986. urteaz gero. Sistema hau analisi sintaktikoa bigarren mailarako lagatzen duten "gramatika semantiko"etan oinarritzen da. Galderak oso zailak ez badira, emaitza harrigarriak lortzen ditu. Mikroetarako merkatuko azken eskaintza, Texas Instruments enpresaren "Natural Link" paketea da. Azken honetan erabiltzaileak ezin du galdetu edonola, berari aurkezten zaion menu moduko pantaila batean hitzak edo esaldi-zatiak hautatu behar ditu azaldu nahi duen galdera osatzeko. Bere hitzak idatziz hautatu behar ditu, aurkezten zaion menuan azaldu nahi duen galdera eraikitzeke. Horrela esanda, menu hutsa dela dirudi, baina bere atzetik dagoen analizatzaile linguistikoa antzeko beste sistemen mailakoa da. Pakete honen ezaugarri onena gardentasuna da: erabiltzaileak ondo daki zeintzuk esaldi ulertuko diren eta zeintzuk ez.

### ***Itzulpen Automatikoa***

Itzulpen Automatikorako ere bada zenbait produktu merkatuan salgai. Horietako batek berak ere ez ditu testu literarioak itzultzen. Guztiek itzulpen teknikoan dihardute, non anbiguitasuna murriztagoa den.

Testu itzuliak beteko duen helburuaren arabera, bi multzo desberdin bereizten dira: Informazio-masa handitik edukiaren ideia orokorra ateratzen dutenak, eta zabalkuntza handia izango duten informazio zehatzak itzultzen dituztenak. Lehenengo multzorako adibidez, hizkuntza arrotzean argitaratzen den guztiaren berri ukan behar duen ikerlariak beharko

lukeena dugu. Guztia ondo itzultzea denboraz edo diruz oso garestia litzateke, eta gainera zati asko ez litzaizkioke interesatuko gero. Guztiz zuzena ez den eta merkea den itzulpenaz erraz jakin baina ahalko luke benetan interesatzen zaion partea zein den, gero zati hori zehatz-mehatz itzultzeko. Oro har, denbora eta dirua irabaziko lirateke. Beste aldean, zabalkuntza handiko informazio zehatzen adibide gisa, etxetresna elektronikoko baten erabilpenerako azalpenak ditugu. Testu horien zehaztasun eta ulergarritasuna, salgaiaren arrakastarako giltza izango dira. Beraz, kalitate handikoa izan beharko du itzulpenak.

Itzulpenaren automatizazioa ez da beti erabatekoa. Bere mailaren arabera ondoko sailkapena egiten da:

- Erabateko itzulpen automatikoa. Lan osoa makinaren bidez burutzen da, giza laguntzarik gabe. Errealitatea baino ametsa gehiago da gaur egun.
- Giza laguntzaz buruturiko ordenadore bidezko itzulpena. Lanaren arduraduna makina da, baina fase desberdinetan laguntzak eska ditzake; hitz baten adiera zuzena hautatzeko edo esaldi baten analisisa nondik hasi behar den erakusteko adibidez.
- Ordenadorez lagunduriko giza itzulpena. Lanaren arduraduna pertsona da, baina hiztegi berezitan hitz bat edo beste bilatzeko edo testuaren formatua txukuntzeko, ordenadoreaz baliatzen da. Agian itzulpenaren zati handi bat ia laguntzarik gabe ordenadoreak egingo du, baina testua egokitzeke aurreprozesaketa edota emaitza zuzentzeko postedizioa ohizkoak izaten dira.
- Datu-Banku Terminologikoak. Hiztegi berezituak erabiltzeko aukera hutsa da ordenadoreak kasu honetan eskaintzen duena. Dena dela, hau ez da laguntza txikia oso testu teknikoak itzuli behar direnean, batez ere hiztegi inprimatuak baino askoz ere eguneratuago egoten direlako.

Montrealeko TAUM taldeko METEO sistema da emaitzarik arrakastatsuen lortu duena. 1977. urtean hasi zen parte meteorologikoak ingelesetik frantsesera itzultzen, testuaren %80 guztiz zuzena zelarik. Egunero oso antzekoak ziren itzulpen aspergarri hauek egiteke itzultzaileak bilatzea zaila zen, nahiz eta soldata ederrak eskaini. Urte hartatik hona lana egunero burutzen da METEO-ren laguntzaz. Hamaika saio egin da geroztik sistema honen diseinua beste gai batzuetara zabaltzeko, baina ezin izan da horren biribila den beste gairik aurkitu. TAUM taldeak berak hegazkinetarako eskuliburuak itzultzeko saioak egin zituen, baina hasierako emaitza itxaropentsuek piztutako ametsak laster itzali ziren.

METAL sistema ingeles/aleman itzulpenetarako salgai dago 1985. urteaz geroztik. 1961.ean hasi ziren diseinatzen Texas-eko Unibertsitatean. Zenbait aldiz baztertu eta berrartua izan ondoren, 1980.az gero Siemens enpresa da babesle bakarra. 1986. urteaz gero Bartzelonan badago alemana/espainiera bikotera egokitzen ari den talde bat.

ALPS, Weidner eta LOGOS sistemak erabili egiten dira gaur egun; European batez ere. Lehenengo biak Mormoien Elizaren eraginez sortutakoak dira, bere testu sakratu ugariak errazago itzul zitezten. Hirugarrena berriz, Vietnam-eko gerran erabilitako armen eskuliburuak itzultzeko jaio zen.

INPUT TEXT:	RAW MACHINE OUTPUT:	RAW MACHINE OUTPUT:
<p>HYDRAULIC PRESSURE IN-LINE RELIEF VALVE</p> <p>(See figures 2-4 and 2-5)</p> <p>30 Identical interchangeable hydraulic pressure in-line relief valves (in-line relief valve) are provided for each ac hydraulic pump and for the dc hydraulic pump. The in-line relief valve are located in the hydraulic service center. Those for the No. 1, No. 1A, and No. 1B ac and dc hydraulic pumps are on the left side next to the No. 1 service center assembly. The in-line relief valve for the No. 2 ac hydraulic pump is incorporated in the No. 2 service center assembly.</p> <p>31 The in-line relief valves are poppet-type, spring-loaded to the closed position. A pressure of 3450 psi impinging on the poppet is sufficient to overcome the opposing spring force, and the poppet will move from its knife-edge seat.</p>	<p>CLAPET DE DECHARGE INCORPORE DE PRESSION HYDRAULIQUE</p> <p>(Voir les figures 2-4 et 2-5)</p> <p>30 Les clapets de décharge incorporés interchangeables identiques de pression hydraulique (clapets de décharge incorporés) sont prévus pour chaque pompe hydraulique ca et cc par la pompe hydraulique cc. Les clapets de décharge incorporés sont situés dans le compartiment hydraulique. Ceux pour les pompes hydrauliques ca et cc no 1, no 1A et no 1B sont du côté du bloc collecteur no 1. Le clapet de décharge incorporé pour la pompe hydraulique ca no 2 est intégré au bloc collecteur no 2.</p> <p>31 Les clapets de décharge incorporés sont champignon, sont rappelés par ressort à la position fermée. Une pression de 3450 psi s'exerçant sur le clapet-champignon est suffisante pour vaincre la force de rappel du ressort et le clapet-champignon se déplacera de son siège en couteau.</p>	<p>CLAPET DE DECHARGE INCORPORE DE PRESSION HYDRAULIQUE</p> <p>(Voir les figures 2-4 et 2-5)</p> <p>30 <i>Des</i> clapets de décharge incorporés interchangeables identiques de pression hydraulique (clapets de décharge incorporés) sont prévus pour chaque pompe hydraulique ca et cc par la pompe hydraulique cc. Les clapets de décharge incorporés sont situés dans le compartiment hydraulique. Ceux <i>des</i> pompes hydrauliques ca et cc no 1, no 1A et no 1B sont du côté du bloc collecteur no 1. Le clapet de décharge incorporé pour la pompe hydraulique ca no 2 est intégré au bloc collecteur no 2.</p> <p>31 Les clapets de décharge incorporés <i>du type</i> champignon, sont rappelés par ressort à la position fermée. Une pression de 3450 psi s'exerçant sur le clapet-champignon est suffisante pour vaincre la force de rappel du ressort et le clapet-champignon <i>s'écartera</i> de son siège en couteau.</p>

TAUM-AVIATION sistemak egindako itzulpena. Ezkerreko zutabeen jatorrizko ingelesezko testua, erdian sistemaren emaitza eta eskuinean gizakume batek zuzendu ondoko itzulpena. Lau baino zuzenketa gehiago ez dago.

SYSTRAN Institutua 1970. urteaz gero Itzulpen Automatikorako tresna-saltzaile nagusia izan da. NASA, Europako Ekonomi Elkartea, General Motors eta Xerox dira bere bezero-rik ezagunenak. Europako Ekonomi Elkarreak egokitzapen neketsua behar izan zuen (100.000 hitzeko hiztegia definitu behar bait zuen) frantses/ingeles itzulpena ahalbideratzeko. Egun 20 itzultzailek erabiltzen dute sistema hau Luxemburg-en, hilabetean milaren bat orrialde ingeles-frantses, frantses-ingeles eta ingeles-italiera bikoteetarako itzultzen dutelarik. Kanadako General Motors-ek eskuliburuak itzultzen ditu ingelesetik frantsesera. 130.000 hitzeko hiztegia definitu ondoren, itzultzaileen lana 3 edo 4 aldiz arinagoa zen, eguneko 1000 hitzeko mailaraino helduz. SYSTRANen oinarri informatikoa, guztiz atzeratua dago; 1960.eko hamarkadako teknologia erabiltzen bait du.

Europako Ekonomi Elkarreak, SYSTRAN sistema bere itzulpen-beharrak betetzeko tresna aski ahaltsu ez zela ikusita, EUROTRA proiektu berria abiarazi zuen 1978.ean. Europako 9 hizkuntza nagusiak hartuta (euskara, katalana eta antzeko hizkuntzak ez daude) edozeinetatik beste edozeinetara itzultze-ko gauza izango zen, espezifikazioaren arabera. Hasieran finkatu ziren epeak ez dira bete hizkuntza guztietarako (frantsesa, ingelesa, alemana eta danierarako soilik). Hizkuntzalari eta informatikarien arteko proportzioa guztiz desorekatuta dago lehenengoan alde. Arrazoi hauegatik, zenbait behatzaileren ustez proiektu honetatik ezin daiteke, epe laburrerako behintzat, fruitu zehatzik espero.

Grenobleko GETA taldeak 1961. urteaz gero dihardu eremu honetan. Hasieran errusieratik frantseserako itzulpenak burutu ziren. Gero ingelesa, alemana eta arabiera edo malaysiera bezalako beste hizkuntza batzuk ere aztertu dira. Azken

urte hauetan, Frantziako gobernuaren diru-laguntzak direla medio adimen artifizialeko lengoia eta tresnekin sistema birmoldatzen ari dira.

Itzulpen Automatikoak garrantzizko papera du Japoniako bostgarren belaunaldiaren proiektuan. Bertan adimen artifizialak eta itzulpen automatikoak elkarrekin lan egiten dute. 1985.era trilioi bat yen gastatuta zegoen ikergai honetan. Orduan Europa eta Estatu Batuetan 12 talde ikertzaile biltzen ziren iharduera honetan eta aldi berean Japonian 18 ziren. Fujitsu, Hitachi eta Toshiba erraldoiak bereziki interesatuta daude eskuliburuaren itzulpenaz, eta beren taldeak osatu dituzte.

#### **Testuen eduki-araketa**

Aplikazio-mota honetan, testua barruan daukan datu bat bilatzeko edo laburpen bat lortzeko aztertzen da. Oraindik ez dago merkatuan era honetako sistemarik salgai, baina agertzeaz daude batzuk. Cognitive Systems enpresaren Atrans pake- teak bankutako telex-en informa- zioak irakurtzen ditu. Antze-ko sistema bat garatzen du Carnegie-Mellon-eko taldeak. Cognitive Systems-ek Estatu Batuetako Kostazaintzarako egiten duen sisteman, untxiei buruzko mezuak hartu eta munduko itxasuntxiei buruzko datu-basea eguneratzeko erabiltzen du.

#### **Testu-edizioa**

IBM, Macintosh eta bestelako PC arruntetan ingelesa, frantsesa, espainiera eta beste hizkuntza nagusientzako ortografi zuzentzaileek bete dituzte urte batzuk merkatuan. Laborategietan bukatzeaz dauden pakete berriek, idazkera- eta sintaxi-

erroreak ere zuzenduko dituztela dirudi. Nahiz eta errore guztiak harrapatu ez, laguntza ederra eskaintzen dio eskutitzak edo bestelako txostenak idatzi ohi dituenari.

### **Testuen eduki-araketa**

Aplikazio-mota honetako sistemek, ordenadore eta gizakiaren arteko komunikazio eroso ahalbideratzen dute. Galdera eta erantzunez osatutako elkarrizketa ulertu ahal izateko, partaideen planak eta helburuak aztertzeo tresnak beharrezkoak dira. Hiztun bakoitzak momentu bakoitzean zer dakien eta zer nahi duen zehaztu behar da eta gainera ezagumendu horiek dinamikoki eguneratu behar dira elkarrizketa aurrera joan ahala.

Honelako sistemak implementatzen zailak dira. Helburu orokorrekorik ez da salgai egongo urte luzetan, baina badira aplikazio zehatzei lotuta dauden batzuk. Gehienak adimen artifizialeko erabilpenei lotzen zaizkie.

Ordenadorez Lagunduriko Irakaskuntzarako SOPHIE sistemak, gaizki dabilzan zirkuitu elektronikoak diagnostikatzen laguntzen dio ikasleari. Elkarrizketa osoa lengoiaia naturalez egiten da.

MYCIN sistema adituak, elkarrizketa baten bidez lorzen ditu diagnostikatuko duen gaixotasunaren sintomak.

### **Ahozko idazmakina**

Azken aplikazio-mota honek zailena dirudi, baina lortuz gero aldaketa ikaragarriak sortuko lituzke. Zailena, hizkuntza idatzia ulertzeko arazoei ahozko hizkuntzaren anbiguetateak eransten zaizkiolako da: hitzak ez dira guztiz bereizten hitz egiterakoan, esaldietako hasiera eta bukaera erdikoa baino intentsitate txikiagoz ematen dira eta gainera seinale fisikoen zaratak ohizko oztupoak izaten dira.

Ikertalde ospetsuenek (IBM, Carnegie-Mellon, MIT, ...) saio neketsutan dihardute *ahozko idazmakina* lortzeko, baina beren lorpena oraindik ez dago gertu. Gaur egun uler daitezke makinaz pertsona zehatz batek emandako hitzak (esaldiak ere bai noizbait), baina beste pertsona batenak ulertu ahal izateko berriro hezi behar da makina, hiztun berriak aurrez prestatutako testu bat irakurri eta hitzak espreski bereiztuz. Onartzen den hiztegia, mugatua da. Transkripzio-erroreen kopurua handia izaten da.

### **Euskal Herriko egoera**

Aurretik azaldutako aplikazio guztiak ingeleserako eginak dira. Salbuespen batzuk badira; alemaniera, frantsesa, japoniera eta daniera lantzen dituztenak batez ere. Espainiera ere hizkuntza interesgarri bihurtu da azkenaldi honetan; herri aurreratuetako produktuentzat espainieradun erostunen kopurua handia delako batez ere. Egun Bartzelonan itzulpen automatikoari buruz hiru proiektu zabaldu dira, beste proiektu erraldoien sukurtsal moduan: METAL (Siemens-ena alemaniera/ingeleza itzulpenak burutzen dituen), EUROTRA (Europako Ekonomi Elkarteara) eta FUJITSU japoniarra. Madrileko IBMk MENTOR proiektuen ingeles, espainiera eta hebraierarako sistema bat garatzen du. Frantsesa egoera hobean dago;

ingeles/frantses, alemaniera/frantses eta errusiera/frantses bikoteak aztertuak izan bait dira Kanadako TAUM-METEO sisteman, Grenobleko GETAn eta EUROTRA barruan. Talde hauen esperientzia Espainiakoena baino askoz ere luzeago eta sakonagoa da.

Katalanerako oraindik ez dago proiektu sakonik, baina analisi morfologikoa guztiz implementatua dute, ordenadore bidez automatikoki egin ahal izateko. Bestetik Bartzelonan kokatu diren itzulpen automatikorako hiru proiektuen itzalpean katalanerako tresna berriak garatu nahi lituzkete bertako langile katalanzale batzuk.

Oraingo hamarkada hau hasi zen arte, Euskal Herrian ez zen ezer entzun ikerkuntzaren eremu honetaz. Orduetik hona burutu diren lanak ondorengoak dira:

- UZEIko Euskal Term. Datu-base terminologikoa, non UZEIren hiztegi berezitu guztiak erraz kontsulta daitezkeen.
- Joseba Abaitua hizkuntzalariak Manchester-en aurkeztu zuen tesia. Euskararen morfologia eta sintaxirako gramatika lexikofuntzional bat proposatzen du. Egun Bartzelonako FUJITSUren itzulpen automatikorako proiektuan ari da lanean.
- Donostiako Informatika-Fakultatean 7 partaidetako talde bat ari da bide berri hauek urratu nahian. Hasiera CAPRA proiektuaren eskutik etorri zen. Proiektu horretan, ordenadore bidez ordenadorearentzako programak idazten irakatsi nahi da. Bere barruan bi doktorego-tesi gorpuzten dira. Batean problemaren enuntziatuak lengoiaia naturalez automatikoki ulertzeko sistema bat eraiki zen eta bestean ikasle eta ordenadore-tutorearen arteko komunikazioa lengoiaia naturalez burutzen da.

Sistema gaztelaniarako egin da, baina euskarari ere egokitu zaio zenbait modulu. Aurten proiektu berria abiarazi da UZEI Institutuarekin eta APIKA informatika-enpresarekin batera. Proiektu honen helburuak bi dira: euskaraz idatzitako testuentzako zuzentzaile ortografikoa eta euskararako analizatzaile morfologikoa orokorra burutzea.

### **Euskararen etorkizunerako bidea**

Datu-baseen galdeketa-sistemak, ordenadorez lagunduriko irakaskuntz sistemak edo itzulpen automatikoa euskaraz ikusi ahal izateko, azpiegitura guztia egiteko daukagu. Informatikariak eta hizkuntzalariak trebatu behar dira arlo honetan, gero taldelanean eta aplikazioen artean dagoen lehenetsunari jarraituz azken helburuetara iritsi ahal izateko.

Oinarri-oinarrizko tresnak honako hauek lirateke: hiztegi informatizatuak, zuzentzaile ortografikoa, analizatzaile morfologikoa automatikoa eta analizatzaile sintaktikoa. Hizkuntz tratamendu errazetarako sistemak eraikitzea litzateke bigarren pausoa. Testu luzeak ez direnerako, esaldi-mailako elkarrizketak ulertu eta gidatuko dituzten sistemak asma daitezke, gai mugatu baten barruan eta zenbait muga linguistikorekin. Aplikazio ugari aurki liteke muga hauen barruan (datu-baseen galdera-erantzun moduko sistemen bidez batez ere), baina baita Ordenadorez Lagunduriko Irakaskuntza edo Sistema Adituen bidez ere. Bigarren pauso paraleloa, Itzulpen Automatikoa litzateke.