

METAL

Andoni Sagarna

Produktuaren historia

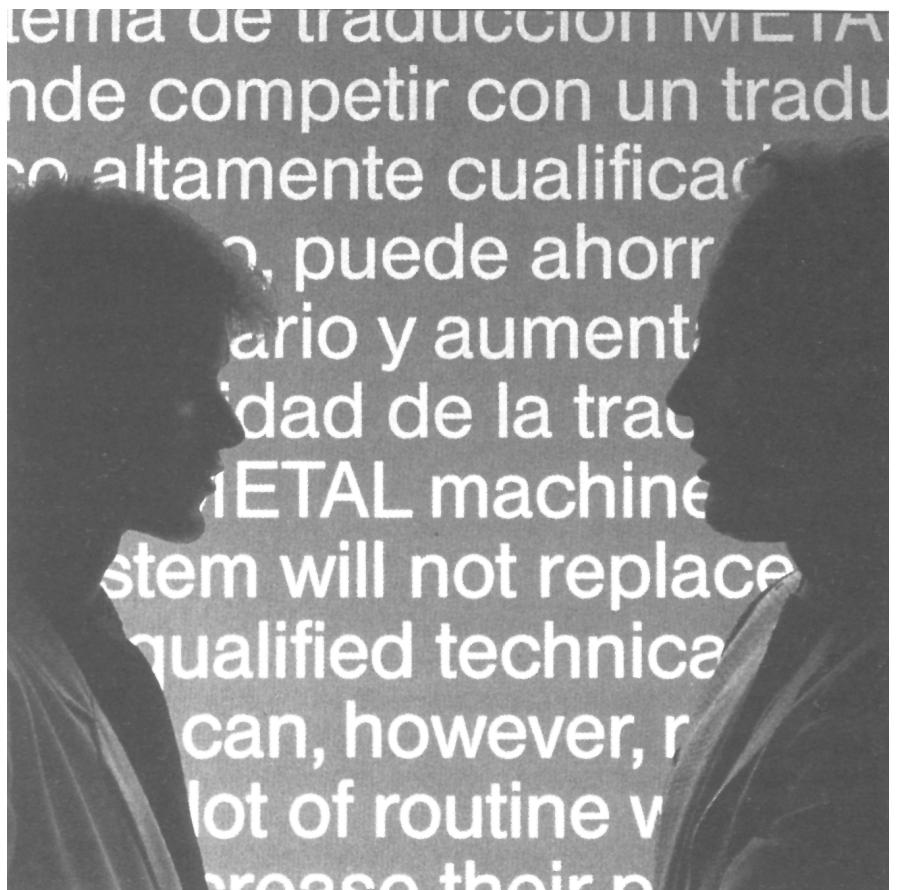
Jakina denez, Siemens enpresa alemana elektrizitatearekin zerikusia duen ororen fabrikatzaile erraldoia da. Besteak beste, zentral telefonikoak ere egiten ditu. Iharduera honek bultzatu zuen hain zuzen, hamar urteko ikerketaren ondoren, gaur egun merkatuan dagoen ordenadore bidezko itzulpen-sistematik aurreratuena den METAL garatzeko erabakia hartzea. Izan ere, telefoni sistema baten dokumentazioa osatzen zuten 100.000 orrialdeak ezin bait zituen alemanetik ingelesera itzuli arrazoizko epe eta kostu batzuen barnean. Zergatik ez, orduan, produktu bat garatuz beharri erantzun? Eta, zergatik ez abiatu beste norbaitek hasia zuen garapen batetik?

Horixe egin zuten; Texas-eko Unibertsitatean prestatzen ari ziren METAL (Machine Edited Text Aspiring Legibility) erosi zuten. Hamar urte geroago salgai dago alemanetik ingelesera itzultzeko sistema hau.

Kaleratu aurretik ongi probatu dute METALen funtzionamendua: lehenbizi, 1986az geroztik Siemens-en Munich-eko egoitzan, gero Zürich-eko Compulex eta Rendsburg-eko Schönau und Damsels-en itzulpen-bulegoetan, Nuremberg-eko Philips Kommunikations-Industrie-n, Villingen-eko, Mannesmann Kienzle-n eta Hildesheim-eko Unibertsitatean.

Oraindaino METALek alemanetik ingelesera itzultzeko aukera bakarrik

Siemens-en ordenadore bidezko itzulpen-sistema



ematen badu ere, alemanetik espainiera eta ingelesetik alemanera itzulpenak egiteko bertsioak aurtengo udaberrian probatzen hasteko moduan izango dira.

Laneko prozedura

Itzulpen-lanabes bat baino gehiago da METAL. Testu-masa handiko itzul-

pen tekniko errepikakorrak egiteko pakete integratu bat dela esan daiteke. Software honek bi ordenadore behar ditu: Unix sistema eragileaz lan egiten duen Siemens beraren erdi-mailako SINIX ordenadore bat testuaren formatua lantzeko, eta itzulpena burutzeko behar diren gramatika eta hiztegia dauzkan LISP makina bat. Makina biok Ethernet

sare batez elkartzen dira. Erabiltzaileak SINIX ordenadorearen terminal gisa konektaturiko PC batean lan egiten du.

Terminal honetan sartzen da iturburu-testua, dela disko batetik, dela irakurketa optikoz eta OCR (karaktereak ezagutzeko programa) batez. Terminaletik SINIXera bidaltzen da testua hark itzulpenerako presta dezan. Prestakuntza amaitutakoan LISP makinara bidaltzen da itzultzeko, eta lan hau burututakoan ostera SINIXera joaten da postedizioa egitera. Itzulpen automatikoan egin

behar izaten diren lexiko berriaren gehikuntza eta gramatika aldeko doikuntzak LISP makinaren pantailan egiten dira.

Bistan da METALen helburua ez dela itzulpena egitea soilik; dokumentua sartzeko eta formateatzeko arazoei ere erantzuten bait die. Horregatik jatorrizko dokumentuaren grafiko, taula, letra-mota, etab. errespetatu egiten ditu. Hau zurrunki egiterik balego, arazoa erraza litzateke, baina edozein itzultzailek dakien bezala, itzultzerakoan hitzen luzera eta ordena aldatu egiten dira.

Eragozpen hau gainditzeko, METALEk banandu egiten ditu lehenbizi formatua eta testua. Hainbat testu-prozesadoretan landutako testua onartzen du METALEk, hala nola, Word Star-en eta Word Perfect-en prestatuturikoa.

Behin formatua eta testua banandu ondoren, testua esaldi laburretan zatikatzen da eta LISP makinara igortzen, itzultzeko.

Makina honek, lehenengo eta behin, hitz ezezagunak bilatzen ditu testuan; hiztegia ez dauzkanak, alegia, eta hauek zerrendatu ondoren, erabiltzaileak sarrera lexiko berriak kodetu egin behar ditu erizpide linguistiko batzuen arabera, programak eskaintzen dion leiho-sistema lagungarriaz baliatuz.

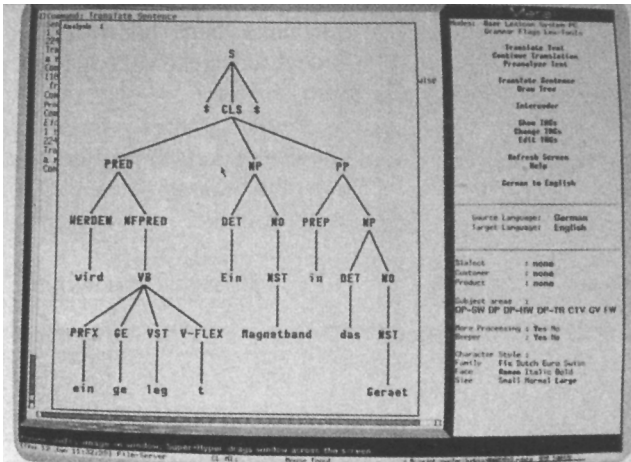
Aurreanalisiak hitz berrien agerpenak eta hauen testuinguruak erakusten ditu. Horrela, erabiltzaileak hitz horien erabilera zein den laster batean ikus dezake. Bidebatez, idaztokerrak detektatzeko bide bat ere bada, gaizki idatzitako hitzek gehienetan forma ezezaguna izango dutenez.

Oinarritzko hiru hiztegi erabiltzen ditu METALEk. Bi hiztegi elebarkar ditu, 50.000 sarrerakoa bakoitza, bata alemanezkoa eta ingelesezkoa bestea, eta bi hizkuntzotako hitzen arteko baliokidetzak hiztegi bat gainera. Hiztegi hauek hierarkizaturik daude: morfema gramatikalak ganean, hauen azpian lexiko arrunta eta beherago hiztegi tekniko orokorra. Horretaz gainera, hiztegi teknikoak daude (informatika, telekomunikazioak, medikuntza, etab.) moduluka antolatuturik.

Hiztegi tekniko berezitu desberdinak erabiliz termino bakoitza nola itzuliko denaren berri ematen duten glosario-fitxategi batzuk sortzen ditu lehentxeago aipaturiko aurreanalisiak.

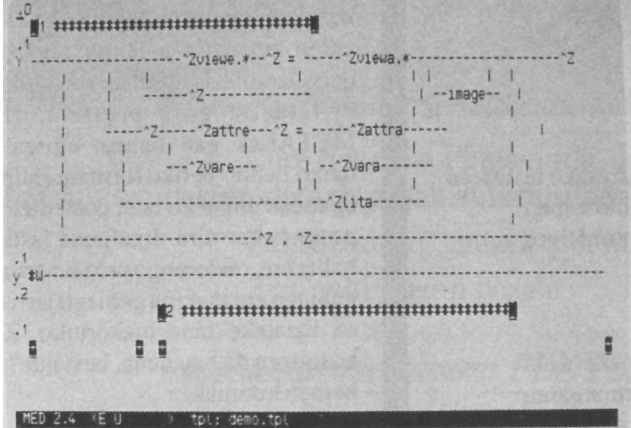
Bada hitz elkartuen fitxategi bat ere, hitz elkartu ezezagunak analizatu ondoren, osagaien esanahian oinarrituz behin-behineko baliokideak sortzen dituen. Batezbeste kasuen %70ean asmatzen duela esan daiteke. Zenbat eta testu teknikoagoa izan, orduan eta arrakasta gehiago lortzen du sistema honek.

Termino bat itzultzerakoan baliokidea hiztegi espezializatuenetan bilatzen da lehenbizi eta hauetan aurkitzen ez bada gero eta orokorragoetara jotzen du bilaketan. Nolanahi ere, erabiltzaileak alda dezake ordena hau nahi izanez gero.



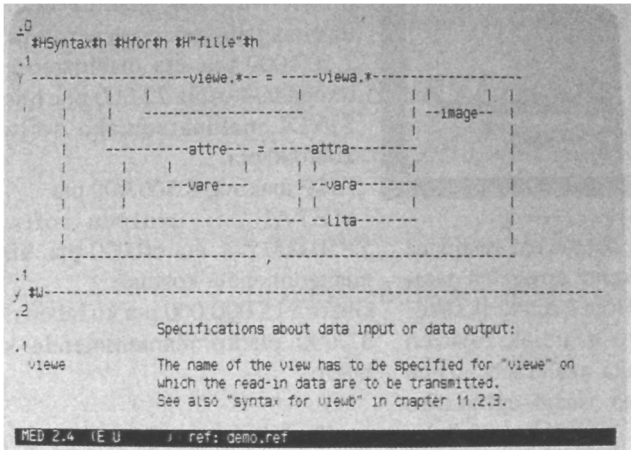
The diagram shows a hierarchical syntax tree for the German sentence "Ein Nest in dem das Nest". The root node is 'S', which branches into 's', 'CLS', and 's'. Further down, it branches into 'PRED', 'NP', and 'PP'. The 'PRED' node branches into 'HERDEW' and 'NFPRED', which further lead to 'wird' and 'VB'. The 'NP' node branches into 'DET' and 'ND', leading to 'Ein' and 'NST'. The 'PP' node branches into 'PREP' and 'NP', leading to 'in' and 'DET ND'. The 'DET ND' node further branches into 'PRFX', 'GE', 'VST', and 'V-FLEX', leading to 'ein', 'ge', 'tag', and 't'. The final node is 'Geraet'.

Perpauaren egitura analisifasearen ondoren



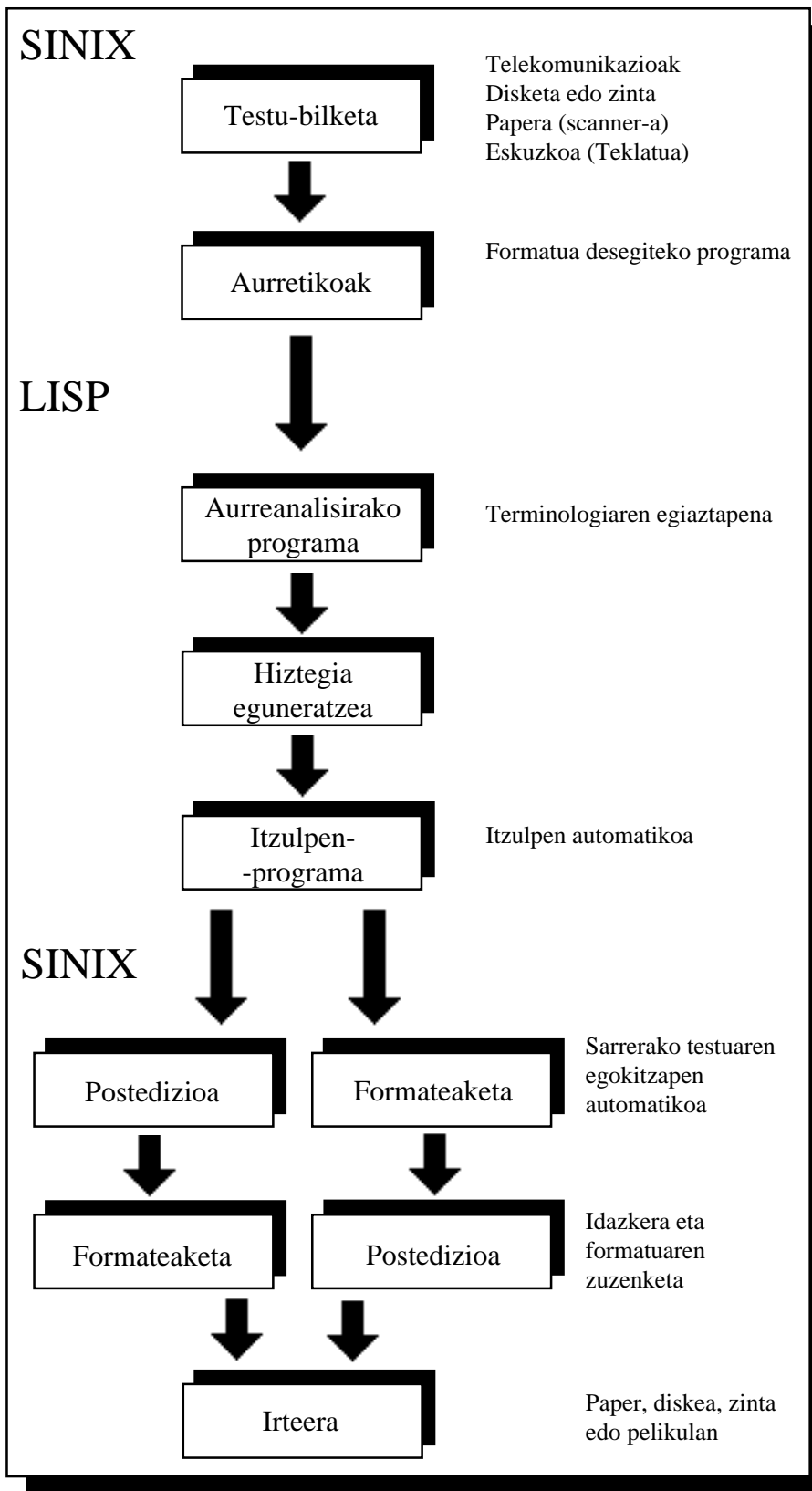
The screenshot shows a text alignment tool with two columns of text. The first column contains the words "Zuviewe", "Zattra", "Zvare", and "Zlita". The second column contains the words "Z", "Z", "Z", and "Z". Dashed lines connect the words in the first column to the corresponding "Z" in the second column, illustrating the alignment process.

Testua eta komposaketa banandu ondoren lortutako maskara



The screenshot shows a text alignment tool with two columns of text. The first column contains the words "Zuviewe", "Zattra", "Zvare", and "Zlita". The second column contains the words "Z", "Z", "Z", and "Z". Dashed lines connect the words in the first column to the corresponding "Z" in the second column, illustrating the alignment process.

Birformateatutako helburu-testua



METALek dauzkan hiztegiak ez dira liburu-forman ikusten ditugun hiztegien berdin-berdinak. Sarrera bakoitzak informazio morfoloikoa eta sintaktikoa ere badauka, berridazketa-erregelen bidez adierazia. Hitz berrientzat erregela lehenetsiak proposatzen ditu.

Itzulpena bera perpausen analisisaz arduratzen den hizkuntz erregelaren datu-baseak egiten du. Honek mailarik sakonenean txertaturiko esaldiak bilatzen ditu. Gero pixkanaka azalerantz joaten da, maila bakoitzean osagai-egiturazko erregelak esleituz. Azaleko mailako

adabegietara iritsitakoan, zuhaitz-egitura bat sortzen du perpaus osoarentzat.

Azken zuhaitza hautatu baino lehen, erregela bat baino gehiago aplikatzeko posibilitatea dagoen kasuetan probabilitatezko estrategia bat erabiltzen du. Honek memoria asko behar du: 120 Mb-etik gora, hain zuzen.

Zuhaitza lortu ondoren LISP makinak kasu-gramatikaren antzeko erre-presentazio-modu batean ipintzen ditu perpausak.

Sakoneko analisi-maila honetatik abiatuz sistemak irteerako zuhaitz bat sortzen du helburuko hizkuntzan. Erabiltzaileak badu honen adabegietako kodeak aldatzeko aukera, behar izanez gero.

Perpausak banan-banan aztertzen ditu sistemak eta lortzen duen itzulpena irteerako fitxategi batean gordetzen du postediziorako.

Kostua eta etekina

Zortzi orduko lan saioan 200 orrialde inguru itzultzen ditu METALek. Abiadura hau handia ala txikia irudi daiteke, baina itzulpen-oso (formateatzea barne) kontutan hartzen bada, nahikoa lasterra dela esan daiteke, zeren postedizio-lanpostuan diharduen langileak ezin bait ditu 40 edo 50 orrialdetik gora prestatu. Hortaz METALek gau batean egiten duen lanari behin-betiko formatua hurrengo egunean emateko bost postedizio-lanpostu behar dira. Itzulpena lastertuko balitz ere, ondorengo prozesuan langile gehiago eta makina gehiago jarri ezean ez litzateke lana bizkortuko. Zenbat kostatzen da hau dena, bestalde? Hona hemen kontuak:

SINIX MX 300 makina laserrezko inprimagailuarekin eta periferikoekin: 2.600.000 pta. eta makinaren mantentimendu-kostua 22.000 pta. hileroko. SINIX makinarentzako softwarea: 208.000 pta.

LISP makina: 6.500.000 pta.

METAL itzulpen-softwarea: 5.850.000 pta. eta 60.000 pta. hileroko mantentimendu-kostua.

Guztira 15.000.000 pta. ko inbertsioa eta 82.000 pta. ko mantentimendu-kostua hileroko.

Zeuk ikusi erostea komeni zaizun. □