



L'enginyeria lingüística en la societat de la informació



Maria Antònia Martí Antonín

Professora del Departament de Lingüística General i directora del Centre de Llenguatge i Computació (CLIC) de la Universitat de Barcelona
amarti@lingua.fil.ub.es



Joaquim Llisterri

Professor del Departament de Filologia Espanyola de la Universitat Autònoma de Barcelona
Joaquim.Llisterri@uab.es

Resum: Les necessitats que planteja la societat de la informació determinen en certa manera els objectius de la lingüística computacional en el tombant de segle: la revolució tecnològica ha donat lloc a canvis en profunditat en els sectors productius i professions relacionades amb el llenguatge i també en els sectors que, sense estar-hi directament relacionats, necessiten dotar la seva activitat del valor afegit d'una comunicació eficient i fluida amb clients i usuaris. Aquesta nova realitat imposa la necessitat de desenvolupar aplicacions de processament del llenguatge i de la parla que tindran un ampli ventall de possibles usuaris i que han de tractar amb centenars de milions de documents escrits o sonors, cosa que al seu torn implica la necessitat de disposar de recursos bàsics d'enginyeria lingüística que han de permetre el desenvolupament efectiu d'aquestes aplicacions. En aquest article es presenten les principals tecnologies lingüístiques, les relatives al tractament tant de la parla com del text escrit, i el seu paper dins del context europeu actual.

1. Comunicació i llenguatge en el marc de la societat de la informació

El llenguatge és el mitjà principal amb el qual els humans intercanviem informació. Cap altre sistema d'informació és tan eficaç per a compartir idees, transmetre coneixements, expressar emocions i crear, en definitiva, el que coneixem com a cultura. Així, no és estrany que es relacioni l'aparició del llenguatge amb la formació de grups socials cohesionats i amb la necessitat de disposar d'un procediment que permeti l'expressió del coneixement comú. *Llenguatge, societat i informació* són tres termes que apareixen estretament lligats des de l'inici de la civilització.

En el seu sentit més actual, el concepte de *societat de la informació* incorpora altres components, entre els quals cal destacar les tecnologies que permeten la difusió del coneixement compartit. No és la primera vegada que ens trobem davant de tècniques que faciliten la transmissió del saber; només cal recordar que el paper de les noves tecnologies en la pervivència de les llengües i les cultures ha estat sovint comparat amb el que en el seu moment va tenir la impremta (Danzin, 1992).



Les noves tecnologies ens posen a l'abast una gran varietat de dades i de modes d'interacció: a Internet podem accedir a sistemes d'ensenyament assistits per ordinador, comprar, establir noves relacions, conversar amb amics que són a l'altre cantó del planeta i accedir a tota mena d'informació. Es pot dir que l'accés a la informació s'ha "democratitzat": abans d'Internet, si ens volíem assabentar de la recerca que es desenvolupava en una universitat, s'havia de fer una estada al centre i romandre-hi un quant temps a fi de conèixer les línies de recerca i els investigadors implicats. Ara, podem accedir a qualsevol campus universitari, obtenir aquesta informació i fins i tot imprimir les seves publicacions, mitjançant una simple consulta a Internet, sense moure'ns del nostre lloc de treball.

Podem afirmar que es va creant un nou nivell de realitat, la realitat virtual, que ens permet realitzar interaccions comunicatives equivalents a les que tenen lloc en la realitat presencial, però deslligades de molts dels condicionaments que ens imposen els paràmetres espai-temps. Ara bé, el nou entorn comunicatiu que deriva de l'aplicació d'aquestes noves tecnologies planteja nous problemes: l'accés a la informació ha de ser eficaç, ràpid i senzill, i s'ha de preveure la possibilitat que els usuaris cometin errors. També és desitjable que en el mateix entorn s'integri la veu, la imatge i la llengua escrita.

L'aparició d'Internet com a xarxa de telecomunicacions i de la WWW canvia el concepte del que es considera informació: no està més ben informat qui té més dades, sinó qui disposa dels millors mitjans per a obtenir exclusivament les que necessita.

La recuperació d'informació, un problema que fins fa poc afectava col·lectius de professionals molt específics, ha passat a ser un dels problemes clau que haurà d'afrontar la societat de la informació. Internet constitueix un gran banc de dades que pot esdevenir inservible si no es disposa de sistemes de recuperació d'informació que satisfacin les demandes dels usuaris. A més, les institucions, els hospitals, les entitats bancàries i, en general, les empreses que generen una gran quantitat i varietat de dades també necessiten sistemes que els permetin gestionar-les de manera eficaç.

Cal, per tant, dissenyar sistemes informàtics que, davant d'una demanda de l'usuari, recuperin els documents que responen als seus objectius d'informació i que ignorin els irrellevants. A més, cal afrontar el problema del multilingüisme, tant en les demandes de l'usuari com en els textos que es recuperen.

Les necessitats que planteja la societat de la informació determinen en certa manera els objectius de la lingüística computacional en el tombant de segle: la necessitat de tractar les produccions lingüístiques sense restriccions constitueix l'aspecte clau del canvi que es produeix actualment. Les repercussions d'aquest nou enfocament afecten en profunditat la definició dels camps de recerca tradicionals. Per primera vegada cal desenvolupar aplicacions de processament del llenguatge i de la parla que tindran un ampli ventall de possibles usuaris i que han de tractar amb centenars de milions de documents escrits o sonors. Aquesta nova realitat imposa la necessitat de disposar de recursos d'enginyeria lingüística bàsics que han de permetre el desenvolupament



efectiu d'aquestes aplicacions.

Finalment, la revolució tecnològica ha donat lloc a canvis en profunditat en els sectors productius i professions relacionades amb el llenguatge. El món de l'edició i la traducció s'encaren a la necessitat de definir nous perfils professionals i de renovar les estratègies productives i comercials.

2. Les tecnologies lingüístiques

En el marc que acabem de descriure, les interaccions comunicatives no es poden concebre sense el complement d'unes tecnologies, anomenades "de la informació i la comunicació", que recolzen, en bona part, en l'ús de sistemes informàtics i de dades lingüístiques degudament formalitzades.

El processament dels textos i de la parla constitueix el nucli de les actuals tecnologies del llenguatge, on conflueixen la informàtica i la lingüística, fet que explica que en l'àmbit de la filologia s'utilitzi també el terme *lingüística informàtica*. Una de les accions del programa europeu Tecnologies per a la societat de la informació ([IST, Information Society Technologies](#)) en el [V Programa marc](#) reconeix explícitament el paper de les tecnologies lingüístiques en el desenvolupament de la societat de la informació, que inclou, entre les accions clau, la línia denominada Tecnologies per al llenguatge humà ([HLT, Human Language Technologies](#)).



Aquestes tecnologies constitueixen també la base del que s'anomena *enginyeria lingüística*, entesa com l'aplicació dels coneixements sobre la llengua en el desenvolupament de sistemes informàtics que puguin reconèixer, comprendre, interpretar i generar llenguatge humà en totes les seves formes (vegeu [Enginyeria lingüística. Com aprofitar la força del llenguatge](#)).

Les tecnologies lingüístiques comprenen una sèrie de tècniques relacionades amb el tractament informàtic del llenguatge, i se solen distingir, tal com es presenta a continuació, les que s'apliquen al tractament de la llengua parlada de les que són pròpies del tractament del text escrit.

3. Les tecnologies del text escrit

En el moment actual, l'accés a la informació ha de ser selectiu i guiat per objectius; si no és així, malgrat disposar d'una gran quantitat de dades al nostre abast, no s'arribaran a convertir en informació que puguem integrar com a coneixement. Aquest fet planteja la necessitat de definir



mètodes d'emmagatzematge i codificació de la informació a fi que posteriorment pugui ser degudament recuperada.

Actualment, en aquests entorns els documents s'indexen per formes o per paraules clau, sense tenir en compte les relacions entre les variants flexives dels mots o les relacions de significat. Tant les empreses editorials, com a creadores de continguts estructurats (diccionaris i enciclopèdies), com els sistemes de cerca a Internet o la gestió dels grans bancs de dades requereixen sistemes que permetin l'accés al contingut des de diferents perspectives segons les necessitats dels usuaris.



El desenvolupament d'aquests sistemes requereix disposar de recursos d'enginyeria lingüística que permetin optimitzar els resultats de les cerques. Les millores es poden produir en diferents sentits:

- Permetent expressar les cerques en llenguatge natural i sense restringir-les a un metallenguatge, normalment una combinació de paraules clau i operadors booleans, que cal aprendre.
- Ampliant l'espai de cerca mitjançant l'ampliació de la pregunta de l'usuari incorporant-hi les formes flexives i les paraules semànticament relacionades.
- Restringint els documents seleccionats a aquells que realment responen al sentit de les cerques.

Entre els [recursos d'enginyeria lingüística](#), destaquen:

- Els analitzadors morfològics, que permeten relacionar les variants flexives d'un mot amb el seu lema i assignen una descripció morfològica a les paraules dels textos.
- Els desambiguadors, que, davant d'una ambigüïtat en el nivell morfològic, sintàctic o semàntic, decideixen, a partir de dades estadístiques o sobre la base de regles lingüístiques, quina és la interpretació més correcta.



- Els analitzadors sintàctics, que identifiquen els grups de mots que funcionen com un tot, la qual cosa permet identificar les entitats del text.
- Les xarxes semàntiques, que organitzen el lèxic segons les seves relacions internes i permeten etiquetar semànticament els textos.
- Els corpus etiquetats morfològicament, sintàcticament i semànticament, que s'utilitzen com a fonts d'informació per a l'aprenentatge dels etiquetadors automàtics i com a consulta per a la recerca lingüística. La utilització de corpus va lligada a una sèrie de processos —com la codificació, l'etiquetatge i l'anàlisi lingüística— i eines —com ara els programes de concordances— que en faciliten l'ús i l'explotació.
- Lèxics computacionals que guien el procés d'anàlisi i interpretació semàntica del text.
- Lèxics multilingües que permeten el procés de traducció automàtica o bé ajuden la traducció manual.

Les tècniques associades a la gestió de continguts fan servir aquests recursos com a fonts d'informació per al processament de les dades. Les tècniques més significatives són l'extracció d'informació, la recuperació d'informació, la traducció automàtica i l'elaboració de resums.

L'objectiu últim és que els documents puguin ser indexats per conceptes, que es pugui accedir a la informació mitjançant interaccions en llenguatge natural i que puguem interactuar i recuperar la informació en la llengua que vulguem.

4. Les tecnologies de la parla

L'ús de la llengua oral com a mode de comunicació amb els sistemes informàtics permet, en primer lloc, alliberar les mans i la vista, cosa que facilita la recepció de la informació des de qualsevol punt i fa possible la mobilitat, de manera que es poden fer altres activitats simultànies. L'entrada de dades sobre productes en la gestió dels fons d'un magatzem, les operacions relacionades amb el manteniment i el control de la qualitat en entorns industrials, la recepció d'informació en entorns com cabines d'avió —amb una gran quantitat de sistemes de comandament i indicadors—, el control de la ràdio, del telèfon o de la climatització d'un cotxe i l'obtenció d'informació sobre itineraris —el prototip desenvolupat en el projecte [Vodis \(Advanced Speech Technologies for Voice-operated Driver Information Systems\)](#) ofereix aquestes possibilitats—, o bé [el dictat de documents](#) són algunes de les aplicacions per a les quals s'ha demostrat que utilitzar la llengua parlada és un avantatge important.

En segon lloc, emprar la parla per a accedir a la informació emmagatzemada en un sistema informàtic ens dona la possibilitat de fer servir el telèfon en lloc de la pantalla i el teclat tradicionals, amb tot el que això comporta pel que fa a rapidesa; cal pensar, a més, en la independència que aquest procediment proporciona quant a la distància i la disponibilitat d'un ordinador. Algunes aplicacions en aquest àmbit són els serveis d'informació sobre horaris de transports públics, les consultes a les pàgines grogues o altres directoris, la informació ciutadana, la banca telefònica o les centraletes automatitzades.

Finalment, cal no oblidar que guardar i rebre informació per mitjà de la parla és una alternativa pràctica quan s'ha de prescindir, per raons de dimensions, del teclat, com és el cas de les agendes electròniques —productes com [ArtCommander](#) faciliten l'entrada de dades per veu— i els comunicadors personals. La veu, juntament amb els llapis electrònics i les pantalles tàctils, és una



modalitat viable per a fer augmentar la funcionalitat d'aquests aparells.



Avui en dia disposem de programes comercials que fan possible el dictat automàtic —com ara els desenvolupats per [Dragon Systems](#), [IBM](#) o [Philips](#)—, la navegació per Internet —el [Free Speech Browser](#) n'és un exemple— i la recuperació d'informació mitjançant la parla. També és possible la lectura en veu alta d'un text en format electrònic, com podria ser un missatge de correu —per exemple, el projecte [EMU](#)— o el contingut d'una base de dades. És possible, a més, amb els sistemes de diàleg que incorporen síntesi, reconeixement i un cert grau de comprensió, obtenir informació o fer transaccions comercials per mitjà del telèfon.

Segurament són aquests darrers sistemes els que en l'actualitat desperten més interès, i per aquest motiu esmentarem algunes de les seves aplicacions.

Un dels sectors per als quals s'han desenvolupat sistemes de diàleg és el dels transports públics, ja que sovint les centraletes telefòniques no poden atendre totes les trucades dels usuaris, especialment pel que fa a la consulta d'horaris. Projectes com [ARISE \(Automatic Railway Information Systems for Europe\)](#) han desenvolupat prototips per a donar resposta a aquesta mena de peticions, concretament en el cas dels ferrocarrils, mentre que, per exemple, [ATIS \(Advanced Travel Information System\)](#) i [Pegasus](#) proporcionen informació sobre horaris de vols.

Hi ha també sistemes de diàleg complementats amb informació multimodal, entre els quals es pot esmentar [Waxholm](#), que ofereix informació sobre el transport marítim a l'arxipèlag d'Estocolm emprant la veu acompanyada d'un rostre generat també mitjançant un ordinador, [MASK](#), un quiosc que proporciona dades sobre circulació de trens francesos tant verbalment com per mitjà d'una pantalla i que també permet reservar bitllets, o [TOOT](#), dissenyat per a realitzar consultes vocals a la web i obtenir horaris de trens.

La informació ciutadana és un altre servei que es pot oferir de manera automàtica mitjançant el telèfon. [Voyager](#), per exemple, informa en anglès, francès i italià sobre adreces, números de telèfon i localització de restaurants, hotels, oficines de correus, estacions de metro, etc., a l'àrea de Boston. D'aquesta mateixa zona es pot rebre informació sobre restaurants amb el sistema [DINEX](#): un telèfon i un ordinador connectats a Internet permeten saber la localització, el preu, el tipus de targeta de crèdit acceptat i, en alguns casos, el menú, tot això fent preguntes en llenguatge natural com ara "Quins restaurants xinesos hi ha a prop de l'Hotel Plaza?". També hi ha aplicacions, com [Movieline](#) —centrada en la cartellera dels cinemes de Pittsburgh—, que ofereixen informació parlada.



En altres àmbits com el de la informació meteorològica destaquen sistemes com [Jupiter](#), que proporciona de manera automàtica i mitjançant el telèfon dades sobre el temps de més de cinc-centes ciutats del món.

Finalment, cal assenyalar la importància que va adquirint la integració dels sistemes de diàleg i la web, tal com es fa evident a [DINEX](#). Un dels projectes més ambiciosos en aquest camp és [WebGALAXY](#), orientat a la recerca d'informació a Internet mitjançant preguntes lliures, tractades per un reconeixedor de parla i per un sistema de comprensió del llenguatge natural, complementats per un procediment de navegació.

Una aplicació que cada vegada té una importància més gran són els sistemes de traducció automàtica de converses, tant telefòniques com cara a cara. Projectes com [Verbmobil](#), [JANUS](#) o [NESPOLE \(Negociating through Spoken Language in E-commerce\)](#), aquest darrer especialment orientat al comerç electrònic, posen de manifest el futur d'un camp que requereix simultàniament tecnologies pròpies del processament del llenguatge natural, com la traducció automàtica, i tecnologies de la parla, com el reconeixement, la síntesi i el diàleg.

La combinació d'aquestes tecnologies també és indispensable en l'ensenyament de llengües assistit per ordinador, tant en els sistemes presencials com en l'ensenyament a distància. Entre els primers, es poden esmentar projectes com [ISLE \(Interactive Spoken Language Education\)](#), aplicacions com les desenvolupades a [SRI International](#) o productes com els que comercialitza [Auralog](#), orientats a l'ensenyament de llengües estrangeres; per als infants que aprenen la llengua materna, eines com [LISTEN](#) faciliten la pràctica de la lectura. En l'ensenyament a distància, són útils, en el camp que ens ocupa, sistemes com [WebGrader](#), que fan possible avaluar la pronúncia per mitjà de la web.

També cal referir-se al paper de les tecnologies lingüístiques com a ajuda als col·lectius amb necessitats especials. Els [convertors de text a parla](#) o els sistemes que combinen un escàner amb un convertidor, com ara el ja clàssic [Kurzweil](#), fan possible la lectura als invidents. Les persones amb dificultats d'audició troben en la conversió de parla a text mitjançant tècniques de reconeixement un ajut important si disposen de telèfons que incorporin pantalles, i també poden fer ús de la lectura labial mitjançant un rostre generat artificialment, tal com mostra el projecte [Teleface](#). Les limitacions en la producció de la parla es poden resoldre amb sistemes de generació de missatges i, en el cas de persones amb problemes de mobilitat, el control de l'entorn domèstic mitjançant la veu augmenta notablement la independència.

5. La societat de la informació en el context europeu

Es podria dir que l'inici del [V Programa marc d'investigació i desenvolupament](#) (1998-2002) va situar el concepte de societat de la informació en una posició central per al desenvolupament de la Unió Europea.

Prenent com a punts de referència documents com el [Green paper on living and working in the information society: people first](#), es va arribar, el gener de 1999, a la decisió del Consell d'adoptar un nou programa anomenat Tecnologies per a la societat de la informació ([IST, Information Technologies Society](#)). El programa s'estructura en diverses accions clau, entre les quals n'hi ha una de dedicada [a les eines i els continguts multimèdia](#); en aquesta acció s'emmarca una línia centrada en les tecnologies lingüístiques ([HLT, Human Language Technologies](#)), els continguts de la qual descriurem molt breument a continuació.

Alhora, diversos països —entre els quals [Espanya](#)— han desenvolupat programes que tenen com a objectiu promoure la societat de la informació, complementant en l'àmbit nacional les iniciatives europees. Catalunya també té un pla per a fer més accessible a tothom la societat de la



informació, gestionat per la [Secretaria per a la Societat de la Informació](#) del [Departament d'Universitats, Recerca i Societat de la Informació](#).

La definició del paper de les tecnologies lingüístiques en la societat de la informació es va iniciar, en el context europeu, l'any 1997 amb el document [Human Language Technologies. Living and working together in the information society](#), en el qual es defineixen les grans línies que configuren el programa actual: multilingüisme, interacció natural i continguts actius. Com a resultat de consultes a diversos sectors, l'informe [Human language, human communication, human knowledge](#) precisava d'una manera més detallada els continguts futurs d'aquesta línia d'acció, per a arribar, finalment, a la publicació del programa de treball el 1999, actualitzat posteriorment per a [l'any 2000](#) (actualment, està en preparació el pla de treball del [any 2001](#)).

Un dels principals objectius del programa és ajudar a superar les barreres lingüístiques per a poder accedir amb més facilitat als continguts i als serveis digitals, com poden ser les transaccions a distància, la gestió distribuïda de coneixement corporatiu o els serveis d'informació en línia. Això requereix desenvolupar eines d'anàlisi, traducció, resum i generació de continguts lingüístics, i tecnologies —agents de cerca i de presentació de continguts— que permetin la recuperació i la presentació d'informació en diverses llengües. També es preveu en aquesta línia d'acció la transferència de tecnologies que semblen especialment interessants a una gamma més àmplia de llengües.

Des d'un punt de vista pràctic, el desenvolupament de projectes emmarcats en aquest àmbit ha de permetre l'automatització de les centraletes telefòniques i els serveis d'atenció al client, la identificació automàtica de la llengua i la verificació de la identitat de l'usuari a partir de la seva veu, el filtratge, la traducció i la resposta de correu electrònic, la gestió de l'accés a catàlegs multilingües o l'oferta de serveis comercials a la web amb una interfície oral. En el terreny de la llengua escrita, l'autoria conjunta de documents, el desenvolupament d'eines de correcció per a parlants no nadius, les eines de gestió de documents i de terminologia tècnica en diverses llengües o els programes d'ensenyament de llengües assistit per ordinador seran alguns desenvolupaments que el programa farà possibles. Finalment, també es planteja la creació de serveis a Internet que tinguin en compte les preferències lingüístiques dels usuaris i que permetin crear continguts publicats simultàniament en més d'una llengua i adaptats a contextos culturals diferents.

La segona línia d'acció, denominada genèricament *interactivitat natural*, se centra a aconseguir que els sistemes informàtics es puguin fer servir més fàcilment mitjançant l'ús espontani de la llengua parlada o escrita, de manera que superin les limitacions actuals en el nostre mode d'interacció amb els ordinadors. Es tracta, en conjunt, d'avançar cap al disseny d'interfícies més intuïtives, multilingües i multimodals que permetin, per exemple, la navegació per la web, el control de sistemes o el diàleg persona-màquina, gràcies a la comprensió d'informació parlada, escrita o, fins i tot, gestual.

Les tecnologies que calen per a fer possible aquestes aplicacions són, essencialment, el reconeixement i la síntesi de la parla d'alta qualitat, i la comprensió de la llengua oral en situacions d'interacció —podríem dir de conversa— entre un usuari i un sistema informàtic.

Finalment, una tercera línia d'actuació se centra en els continguts, especialment pel que fa a la gestió, cerca i recuperació d'informació multilingüe. Per a assolir aquest objectiu es preveu la utilització d'agents lingüístics intel·ligents i de tècniques que permetin identificar el tema dels textos, extreure'n dades i classificar-los automàticament, amb l'objectiu de crear, resumir, convertir i presentar informació en diverses llengües, tant en la seva representació escrita com oral, incloent-hi diàlegs i components multimèdia.

Es pot veure, per tant, que les línies d'acció europees emmarquen les tecnologies lingüístiques en el context de la societat de la informació, insistint especialment en els nous reptes que hem volgut destacar en aquest treball, centrats en el processament multilingüe de la informació, tant escrita com oral.



6. Perspectives professionals

La revolució tecnològica que ha significat la incorporació d'aquests nous sistemes de comunicació està provocant un canvi en profunditat en les professions i sectors productius que tenen el llenguatge com a eina de treball. L'ensenyament, la traducció, l'edició de textos i l'elaboració de diccionaris i enciclopèdies han experimentat canvis profunds, que exigeixen actualitzar els coneixements i els mètodes emprats pels professionals d'aquestes àrees.

Aquesta situació obre noves perspectives professionals als filòlegs i als lingüistes i, a la vegada, defineix un nou perfil de professional: a més del coneixement lingüístic cal adquirir coneixements d'informàtica i de lingüística computacional, tant en el camp del processament de la llengua escrita com de la llengua parlada, posant un èmfasi especial en la capacitat d'abstracció, el raonament lògic i la capacitat d'organització i estructuració de les dades.

Diferentment al que passava amb la filologia tradicional, el desenvolupament de sistemes de processament de la informació, del llenguatge i de la parla requereix la col·laboració d'especialistes d'àrees diferents. És per això que en el perfil del lingüista que vol centrar la seva activitat en el nou entorn de la societat de la informació és essencial una bona capacitat per al treball en equip i per a l'organització i gestió de projectes multidisciplinaris, on cadascuna de les tasques individuals s'orienta a la consecució d'un objectiu comú.

Bibliografia:

ABAITUA ODRIOSOLA, J.K. (1996) "Ingeniería de la lengua y normalización lingüística". A: FORCADA, V.M.; SAGER, J.C.; GIL DE CARRASCO, A. (Eds.) *Estudios computacionales del español y del inglés*. (pàg. 253-265) Madrid: Instituto Cervantes.
<http://sirio.deusto.es/abaitua/konzeptu/umist.htm>

BLECUA, J.M.; CLAVERÍA, G.; SÀNCHEZ, C.; TORRUELLA, J. (Eds.) (1999) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio.

COLE, R.A.; MARIANI, J.; USZKOREIT, H.; ZAENEN, A.; ZUE, V. (Eds.) (1997) *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press.
<http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>

DANZIN, A. (1992) *Vers une infrastructure linguistique européenne*. Rapport présenté par A. Danzin et le Groupe de réflexion stratégique pour la CCE (DGXIII). Versión española: *Hacia una infraestructura lingüística europea*. Informe presentado por A. Danzin y el Grupo de reflexión estratégica para la Comisión de las Comunidades Europeas (D.G. XIII). 31.03.1992.

EDWARDS, J.; KINGSCOTT, G. (1996) *Language Industries Atlas*. Amsterdam: IOS Press.
<http://158.169.50.95:10080/mlis/en/atlas/home.html>

GÓMEZ GUINOVART, J. (1998) "Fundamentos de lingüística computacional: bases teóricas, líneas de investigación y aplicaciones". A: CID LEAL, P.; J. BARÓ i QUERALT (Eds.) *Anuario SOCADI de Documentación e Información 1998*. (pàg. 135-146) Barcelona: Societat Catalana de Documentació i Informació.

GÓMEZ GUINOVART, J.; LORENZO SUÀREZ, A.; PÉREZ GUERRA, J.; ÀLVAREZ LUGRÍS, A. (Eds.) (1999) *Panorama de la investigación en lingüística informática*. RESLA, Revista Española



de *Lingüística Aplicada*, Volumen monogràfic.

GÓMEZ GUINOVAR, J.; PALOMAR, M. (Coords.) (1998) *Monografía: Lengua y Tecnologías de la Información. Novática, Revista de la Asociación de Técnicos de Informática*, 133 (Mayo-Junio). pàg. 3-45.

Green Paper on Living and Working in the Information Society: People First. European Commission, Directorate General V, Employment, Industrial Relations and Social Affairs. COM (96) 389, 22.07.96.

<http://europa.eu.int/ISPO/infosoc/legreg/docs/peopl1st.html>

Human language technologies. Living and working together in the Information Society. Discussion document. Luxembourg, DG XIII/E-5, July 1997.

<http://www.hltcentral.org/htmlengine.shtml?id=221>

Human Language, Human Communication, Human Knowledge. Summary Report. Luxembourg, DG XIII/E-5, July 1998.

<http://www.hltcentral.org/htmlengine.shtml?id=222.0>

Ingeniería lingüística. Cómo aprovechar la fuerza del lenguaje. Luxemburg: Anite Systems. Versión española a cargo del Observatorio Español de Industrias de la Lengua, Instituto Cervantes.

http://www.hltcentral.org/usr_docs/project-source/en/broch/harness.html

Ingeniería lingüística: un mayor uso del lenguaje en todos los sentidos. Luxemburg: Cray Systems.

http://www.hltcentral.org/usr_docs/project-source/en/whatis.html

KURZWEIL, R. (1998) "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding". A: STORK, D.G. (Ed.) *Hal's Legacy: 2001's Computer es Dream and Reality*. Cambridge, Mass.: The MIT Press.

<http://mitpress.mit.edu/e-books/Hal/chap7/seven1.html>

L'ingénierie linguistique. Délégation Générale à la Langue Française, Réseau international des observatoires francophones de l'inforoute et du traitement informatique des langues.

<http://www.culture.fr/culture/dgjf/riofil/enjeux.htm>

MARTÍ ANTONIN, M.A. (1999) "Panorama de la lingüística computacional en Europa". A: GÓMEZ GUINOVAR, J.; LORENZO SUÀREZ, A.; PÉREZ GUERRA, J.; ÀLVAREZ LUGRÍS, A. (Eds.) *Panorama de la investigación en lingüística informática*. RESLA, *Revista Española de Lingüística Aplicada*, Volumen monogràfic. pàg. 11-24.

MARTÍ ANTONIN, M.A. (Coord.) (1999) *Llengua catalana IV (Tecnologies de la llengua)*. Barcelona: Edicions de la UOC.

MARTÍ ANTONIN, M. A.; CASTELLÓN MASALLES, I. (2001) *La lingüística computacional*. Barcelona: Publicacions de la Universitat de Barcelona.

MORENO SANDOVAL, A. (1998) *Lingüística Computacional*. Madrid: Síntesis.

MOURE, T.; LLISTERRI, J. (1996) "Lenguaje y nuevas tecnologías. El campo de la lingüística computacional". A: FERNÁNDEZ PÉREZ, M. (Coord.) *Avances en lingüística aplicada*. Santiago de Compostela: Universidad de Santiago de Compostela, Servicio de Publicaciones Intercambio Científico (Avances en...; 4). pàg. 147-228.

OLIVE, J.P. (1997) "The Talking Computer Text to Speech Synthesis". A: STORK, D.G. (Ed.) *Hal's Legacy: 2001's Computer es Dream and Reality*. Cambridge, Mass.: The MIT Press.

<http://mitpress.mit.edu/e-books/Hal/chap6/six1.html>



Procesamiento del lenguaje natural. Número monográfico. *Philologia Hispalensis*. Vol. 11, núm. 2 (1997).

The EUROMAP Report. Challenge & Opportunity for Europe's Information Society. September 1998. DG XIII, Telematics Applications Programme.

http://www.hltcentral.org/usr_docs/Euromap-report/euromap97.zip

The Language Engineering Directory. Compiled by Language & Technology S.L. Madrid, 1996.

<http://158.169.50.95:10080/mlis/en/direct/home.html>

ZUE, V. (1999) "Talking with your computer", *Scientific American*, August 1999. pàg. 40-41.

<http://www.sciam.com/1999/0899issue/0899zue.html>

Enllaços relacionats:

★ V Programa marc d'investigació i desenvolupament de la Unió Europea:
<http://www.cordis.lu/fp5/home.html>

★ Secretariat de la Societat de la Informació (Generalitat de Catalunya):
http://www.genocat.es/csi/cat/societat_informacio/

★ Survey of the State of the Art in Human Language Technology:
<http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>

★ Centre de Llenguatge i Computació (Clic):
<http://clic.fil.ub.es>

[Data de publicació: maig de 2001]

Citació recomanada:

MARTÍ ANTONIN, Maria Antònia; LLISTERRI, Joaquim (2001). "L'enginyeria lingüística en la societat de la informació". *Digithum*, núm. 3 [article en línia].

DOI: <http://dx.doi.org/10.7238/d.v0i3.595>