

EL CATALÀ I LES TECNOLOGIES DE LA LLENGUA

Gemma Boleda, Montse Cuadros, Cristina España-Bonet i Lluís Padró

Centre de Recerca TALP

Departament de Llenguatges i Sistemes Informàtics

Universitat Politècnica de Catalunya

Maite Melero, Martí Quixal i Carlos Rodríguez

Fundació Barcelona Media

Universitat Pompeu Fabra

1. INTRODUCCIÓ

Cada cop són més habituals els productes tecnològics que incorporen un cert grau de processament computacional de la llengua. Des del senzill predictor de paraules del telèfon mòbil fins a un traductor automàtic integrat al navegador o al correu electrònic, les interfícies tecnològiques intenten apropar-se cada cop més a la manera més natural que tenen les persones d'interactuar: la llengua, ja sigui oral o escrita.

El desenvolupament d'aquestes tecnologies és sovint costós, especialment pel que fa a les eines i els recursos específics per a cada llengua. Per tant, la incorporació d'una nova llengua a un producte basat en tecnologia lingüística suposa una inversió que no sempre resulta rendible per a les empreses. En el cas de llengües com el català, les administracions, les universitats i els centres de recerca són clau a l'hora de facilitar l'accés a aquests tipus d'eines i de recursos amb l'objectiu que el desenvolupament d'aquests productes tingui una rendibilitat comparable a la de llengües amb un mercat més ampli.

En aquest article, la visió de l'estat actual de les tecnologies de la llengua per al català i de les qüestions controvertides que hi estan relacionades procedeixen de les dades obtingudes en la Primera Jornada de Processament Computacional del Català

(JPCC),¹ que va tenir lloc al març de 2009. Aquesta Jornada tenia com a objectius: a) millorar la comunicació i la col·laboració entre els diferents grups de recerca, empreses i institucions que desenvolupen eines i recursos computacionals per al català; b) trobar maneres d'aprofitar de manera eficient els recursos existents, i 3) donar visibilitat a la recerca en el tractament computacional del català.

El segon apartat d'aquest article presenta una panoràmica general dels sectors socials involucrats en la recerca, el desenvolupament i l'ús de les tecnologies del català. Seguidament, en el tercer apartat, es plantegen una sèrie de qüestions polèmiques que afecten aquest sector tecnològic. Finalment, es presenten les conclusions.

2. SECTORS SOCIALS IMPLICATS EN LES TECNOLOGIES DEL CATALÀ

A més de la ciutadania, que és el destinatari final d'aquest servei, es poden distingir tres agents bàsics en l'àmbit de les tecnologies del català: la comunitat investigadora, l'empresa i l'Administració. Tots tres sectors en són alhora productors i

¹ Remetem el lector al lloc web de la JPCC per accedir a tota la documentació generada al voltant de la Jornada i que pot complementar la informació presentada en aquest article:
<http://sites.google.com/site/jornadacatala/>

usuaris, tot i que en mesures i de maneres diferents. En aquest apartat resumirem algunes dades sobre tots tres sectors, tot i que ens centrarem en el primer.

Les dades sobre els grups de recerca que donarem a continuació estan basades en les presentacions que es van fer en la JPCC. Pensem que, tot i no ser exhaustives, són prou representatives de la recerca i el desenvolupament de la tecnologia per al català. En l'annex presentem una taula que resumeix les característiques d'aquests grups, l'activitat, la localització i l'adreça de contacte.

A Catalunya hi ha més d'una desena de centres de recerca dedicats al processament del llenguatge i al País Valencià, tres. A la resta de territoris de parla catalana (Illes Balears, Andorra, Catalunya Nord, Franja de Ponent i l'Alguer), tot i que hi ha entitats que hi treballen com a mínim en el vessant dels recursos (Gabinet de Terminologia de la Conselleria d'Educació del Govern Balear) i empreses que fan servir algun tipus de tecnologia lingüística (IB3, *El Periòdic Andorrà*, etc.), no hi ha grups de recerca que treballin en aquest sector tecnològic.

Des d'una perspectiva sociològica, cal observar que les tecnologies de la llengua són clarament interdisciplinàries, de manera que tant els grups de recerca del món de les enginyeries (informàtica i telecomunicacions) com els d'humanitats (lingüística i traducció) hi dediquen esforços. Els grups presents en la JPCC mostren un equilibri considerable respecte d'aquest paràmetre, amb sis grups de les enginyeries i vuit de les humanitats.

Tots els grups treballen o bé en la recerca teòrica o bé en l'aplicada, i un ho fa específicament en la transferència de tecnologia, tot i que tots els grups s'involucren en aquesta tasca de diverses maneres. El finançament dels projectes procedeix tant de l'Administració

autonòmica com de l'estatal i europea. A més dels concursos específics per a la recerca, la majoria dels grups col·laboren amb empreses i l'Administració pública.

Pel que fa als aspectes tècnics, cal tenir en compte que una divisió bàsica en el tractament computacional de la llengua és la modalitat, ja sigui oral o escrita. En general, hi ha més recerca en la modalitat escrita que en l'oral i, en el cas del català, es corrobora aquesta tendència: deu grups tracten només text; tres s'ocupen de totes dues modalitats, i només un es dedica exclusivament a la veu.

En la modalitat de la veu, les àrees bàsiques en què s'investiga són la síntesi o producció i el reconeixement de la parla. En la modalitat de text, la recerca bàsica es concentra en la creació d'eines de processament general (segmentadors, etiquetadors morfològics i sintàctics, etc.) i en la creació de recursos (gramàtiques i models, diccionaris, lèxics i corpus). Pràcticament tots els grups participen més o menys en les dues tasques, tot i que la feina es divideix entre enginyers (eines) i lingüistes (recursos).

Quant a les aplicacions, el centre de la recerca és la traducció automàtica: nou dels catorze grups hi dediquen esforços. Segueixen en l'interès la correcció i el resum automàtic (dos grups) i després tasques diverses com ara l'extracció d'informació, la generació automàtica de textos o la detecció de paràfrasis.

Pel que fa a les empreses productores de tecnologia lingüística que incloguin el català, en trobem de dedicades a la traducció automàtica (AutomaticTrans,² Transducens³ o Translendum⁴), a la correcció i revisió de textos (Barcelona Media,⁵ Maxigramar⁶ o

2 <http://www.automatictrans.es/>

3 <http://transducens.dlsi.ua.es/>

4 <http://www.translendum.com/>

5 <http://www.barcelonamedia.org/>

6 <http://www.maxigramar.com/>

Thera⁷), a la recuperació i l'extracció d'informació (Barcelona Media, Inbenta⁸ o Thera), a la síntesi de parla (Barcelona Media, Loquendo,⁹ Telefónica I+D¹⁰ o Verbio¹¹) i al reconeixement de veu (Telefónica I+D o Verbio).

Així mateix, hi ha un bon nombre d'empreses i ens públics que treballen en la creació i el manteniment de recursos lingüístics, bàsicament lexicogràfics i terminològics (Thera, Assessorament Lingüístic de la Corporació Catalana de Mitjans Audiovisuals, Institut d'Estudis Catalans i TERMCAT), i alguns serveis lingüístics d'entitats administratives que treballen en l'explotació i la creació d'eines i de recursos diversos com ara els multicercadors, les memòries de traducció i les guies per a la redacció en català (a la Universitat Oberta de Catalunya, la Universitat de Barcelona o la Comissió Jurídica Assessora de la Generalitat de Catalunya, per citar-ne alguns).

Altres entitats que també col·laboren i contribueixen per aconseguir la normalitat del català en l'àmbit de la tecnologia, a més de les esmentades, són institucions com ara l'Acadèmia Valenciana de la Llengua o entitats del sector de la comunicació i la tecnologia com ara Softcatalà¹² o VilaWeb¹³. La implicació de tots aquests agents, com els hem anomenat al principi, es va poder constatar en les inscripcions de la JPCC, on al voltant d'un terç dels inscrits (en total, 166) provenia d'empreses, de l'Administració i d'altres entitats públiques i privades, tot i que la Jornada estava dirigida principalment a la comunitat investigadora.

Gràcies a la feina de tot aquest col·lectiu,

7 <http://www.thera-clic.com/>

8 <http://www.inbenta.com/>

9 <http://www.loquendo.com/es/>

10 <http://www.tid.es/>

11 <http://www.verbio.com/>

12 <http://www.softcatala.org/>

13 <http://www.vilaweb.cat/>

el català disposa de la majoria d'eines i aplicacions de tecnologia lingüística bàsiques per a qualsevol llengua, tot i que no totes són de lliure distribució ni tan sols per a propòsits de recerca (vegeu l'apartat següent). Podem destacar, per exemple, recursos com corpus anotats (CTILC¹⁴ i CICA, desenvolupats per l'Institut d'Estudis Catalans; AnCora,¹⁵ desenvolupat pel CLiC), eines de processament genèriques (FreeLing,¹⁶ de la Universitat Politècnica de Catalunya, o LINLAP, de Barcelona Media) o sistemes de traducció automàtica a diverses llengües (com Apertium¹⁷ o el traductor de *Google*).

En conjunt, doncs, podem dir que la situació de la tecnologia lingüística aplicada al català posa de manifest la varietat i l'abundància de productes tecnològics existents i la vitalitat de la comunitat desenvolupadora. Malgrat tot, com veurem en l'apartat següent, encara es percep una manca d'aprofitament dels recursos i una manca d'implicació de determinats àmbits de la societat civil.

3. QÜESTIONS CONTROVERTIDES EN L'ÀMBIT DE LES TECNOLOGIES DEL CATALÀ

3.1. *Com es podria fer per ampliar la presència de la llengua catalana en els serveis i productes tecnològics? I per augmentar la transferència de la tecnologia que es desenvolupa en els centres de recerca?*

Hi ha el consens generalitzat que la salut computacional del català és força bona; de fet, els recursos i les eines disponibles són molts més que els que li correspondrien per

14 <http://ctilc.iec.cat/>

15 <http://clic.ub.edu/ancora>

16 <http://www.lsi.upc.edu/~nlp/freeling/>

17 <http://xixona.dlsi.ua.es/apertium/>

demografia o poder polític, i la situació de gairebé normalitat en aquest sentit és excepcional per a una llengua sense estat.

Això no obstant, hi ha una problemàtica específica que dificulta la difusió de les eines computacionals del català, tant per als objectius de recerca com per a les aplicacions finals. Aquesta problemàtica es materialitza en els aspectes que tot seguit esmentarem.

En primer lloc, la baixa demanda social de recursos tecnològics del català del món de l'empresa i els serveis (per exemple, sectors com ara la banca o la sanitat), en què el català no és llengua vehicular real. Això fa que el desenvolupament privat de recursos en català sovint es faci per «voluntarisme». Tot i així, podria donar-se el cas que les tecnologies lingüístiques mateixes fessin el paper d'actor en la recuperació sociolingüística del català. Això succeiria si les eines en català fossin prou útils i atractives. En aquest cas, l'usuari les escolliria d'una manera espontània.

En segon lloc, la desconexió entre el desenvolupador i l'usuari final. En aquest sentit, des de certs àmbits s'apunta que les aplicacions distribuïdes amb llicències de codi obert (per exemple, Apertium)¹⁸ podrien facilitar aquesta connexió gràcies a les comunitats que s'estructuren al voltant d'aquest tipus de codi. Aquestes comunitats comparteixen i amplien les aplicacions segons els interessos i les necessitats que tinguin, i poden aglutinar tant grups de recerca, com usuaris i usuàries, o empreses orientades a serveis.

Un punt crucial és també la dificultat en l'accés a certes tecnologies i recursos per problemes de llicència. Aquí de nou les llicències de lliure distribució podrien tenir un paper important. En el cas que un recurs no es pogués fer de lliure distribució per

problemes de propietat o de *copyright*, com ara un corpus amb textos que són propietat d'autors o editorials, caldria facilitar-ne com a mínim l'ús per a objectius de recerca. En aquest cas, cal que les condicions d'ús estiguin ben especificades i que l'accés als recursos sigui àgil i sense obstacles burocràtics. Actualment, es dona una situació molt complicada, amb molta diversitat de llicències i condicions d'ús. Així, no és estrany el cas que, per accedir a un recurs que és lliure per a la recerca, calgui esperar gairebé un any perquè cal signar un conveni entre les institucions implicades.

Especialment nocives són la duplicació i la dispersió d'esforços en la construcció de corpus, eines, etc. El català és una llengua relativament petita; els recursos lingüístics són costosos, i el finançament, escàs. La impressió general és que no ens podem permetre com a comunitat malbaratar esforços d'aquesta manera. Cal, doncs, millorar la coordinació i fomentar la cooperació entre grups de recerca i també amb empreses.

És també patent la relativament poca interacció entre lingüistes i informàtics a l'hora de crear i compartir aplicacions i recursos. Cal més interdisciplinarietat, i això es veu dificultat per les estructures rígides en l'organització acadèmica de la universitat, com també en la manca de centres de recerca híbrids.

Finalment, cal esmentar una certa indiferència dels investigadors en tecnologies del llenguatge per la qüestió de la llengua. Això fa que llengües com l'anglès, que són les que disposen de més recursos, actuïn com a atractius dels esforços investigadors, desequilibrant cada vegada més la proporció de recursos entre les llengües. Per combatre aquest fenomen, cal afavorir la presència de recursos textuals per al català, com ara corpus anotats, en competicions internacionals per a tasques computacionals.

¹⁸ <http://xixona.dlsi.ua.es/apertium>

3.2. Quines estratègies de finançament són necessàries? Quina relació hi ha d'haver entre l'origen del finançament i les condicions d'accessibilitat i de distribució dels recursos?

Hi ha l'opinió estesa que calen polítiques d'incentivació de la innovació a les empreses. Les polítiques actuals, però, estan orientades a projectes massa grossos només assumibles per grans empreses, com Microsoft. Les accions de l'Administració podrien ser més efectives si es fessin polítiques orientades a PIME.

Així mateix, per afavorir el codi obert, cal que en les convocatòries de projectes competitiu, les entitats que sol·liciten els ajuts hagin d'especificar les condicions d'ús dels recursos que es generaran, amb l'objectiu que els avaluadors ho puguin tenir en compte a l'hora de concedir el projecte i de controlar-ne el progrés. Aquesta demanda es podria fer en l'àmbit estatal a fi d'incorporar-la a la nova Llei de la ciència.

Una qüestió afegida en relació amb la distribució lliure dels recursos generats pels projectes de recerca és que cal preveure els costos de l'empaquetament final de l'aplicació. Aquests costos, com també la preparació de llicències i el mecanisme de distribució, s'haurien de preveure en la definició inicial dels projectes.

3.3. Com s'hauria d'articular la comunitat que formen les persones vinculades d'una manera o altra a l'àmbit del processament del llenguatge natural i de la parla de la llengua catalana?

Seguint el model de la Linguatca de la llengua portuguesa,¹⁹ es podria crear un portal web de les tecnologies del català que centralitzés tota la informació referent als recursos, les eines, els grups i les empreses que hi treballen, etc. Respecte al suport

necessari per crear el portal i, sobretot, mantenir-lo, hi ha diferents alternatives. Es pot aprofitar el marc d'algun projecte ja existent, ja sigui en l'àmbit de l'Administració local o estatal (per exemple, amb el programa AVANZA), amb l'objectiu que els poders públics financin aquest manteniment. També hi ha la possibilitat alternativa de crear-lo col·lectivament en forma de cooperativa wiki. Hi ha també iniciatives properes com el projecte europeu CLARIN,²⁰ que ja inclou el català, i que pretén facilitar als investigadors en ciències socials l'explotació de recursos lingüístics.

Amb l'objectiu de possibilitar iniciatives com la del portal web, fóra interessant comptar amb una associació que donés estructura a la comunitat vinculada a l'àmbit de les tecnologies del català, a l'estil francès (ATALA)²¹ o l'espanyol (SEPLN)²². Una associació permetria donar veu al col·lectiu de recerca i alhora fer de nexa tant de les empreses com de l'Administració i els usuaris i usuàries finals.

4. CONCLUSIONS

Atesa la quantitat i qualitat de grups de recerca i d'altres entitats públiques i privades, com també els productes, les eines i els recursos existents, és palesa la vitalitat tecnològica del català. Cal, però, incidir en alguns aspectes que dificulten la transferència tecnològica entre grups, i entre la recerca i el desenvolupament de productes i serveis.

Pel que fa a la col·laboració i l'aprofitament dels recursos, cal una acció comuna i activa de la comunitat, però també suport institucional. En particular, per establir una política més oberta de distribució i més facilitat d'ús de la

¹⁹ <http://www.linguatca.pt/>

²⁰ <http://www.clarin.eu/>

²¹ <http://www.atala.org/>

²² <http://www.sepln.org/>

tecnologia lingüística, cal un suport decidit dels poders públics, tant autonòmics com estatals i europeus, ja que estableixen les condicions de finançament dels projectes de recerca i de transferència de tecnologia.

És necessari també articular iniciatives per intensificar la comunicació i la col·laboració entre els diferents agents de la comunitat tecnològica catalana, tant entre els grups de recerca com amb l'empresa, ja sigui desenvolupadora o usuària. La Primera Jornada del Processament Computacional del Català ha estat una iniciativa en aquesta línia, però cal anar més enllà amb la creació d'un portal web que centralitzi les eines i els recursos, i possiblement també amb la constitució d'una associació que aglutini tota la comunitat implicada.

5. ANNEX

<p>Grup de Veu i Llenguatge (Barcelona Media Centre d'Innovació)</p> <p>Objectiu: desenvolupar nous productes i serveis, i potenciar la recerca i el desenvolupament (R + D).</p> <p>Línies de treball: extracció d'informació, correcció i revisió de textos, traducció automàtica, síntesi de veu, llengua de signes i tecnologies per a la docència de la traducció i la lingüística.</p> <p>URL: http://www.barcelonamedia.org/linies/7/ca</p>	<p>Línies de treball: generació multilingüe de llenguatge natural i altres continguts, resum automàtic, paràfrasi, traducció, lexicologia computacional i aprenentatge automàtic orientat a l'adquisició de recursos lingüístics.</p> <p>URL: http://www.recerca.upf.edu/taln</p>
<p>Grup de Processament del Llenguatge Natural, Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (Universitat Politècnica de Catalunya)</p> <p>Objectiu: investigar en el camp del processament computacional del llenguatge natural, com també crear recursos lingüístics.</p> <p>Línies de treball: recerca en processadors lingüístics, adquisició, integració i explotació de coneixement lexicosemàntic, anàlisi semàntica des de diferents punts de vista (WSD, SRL o NLU), aprenentatge automàtic aplicat al PLN, resum automàtic i traducció automàtica.</p> <p>URL: http://www.lsi.upc.edu/~nlp</p>	<p>Servei Lingüístic (Universitat Oberta de Catalunya)</p> <p>Objectiu: fer un ús intensiu de la tecnologia disponible per optimitzar els processos productius (rendibilitzar costos i aportar qualitat i homogeneïtat).</p> <p>Línies de treball: l'explotació i la millora dels sistemes de traducció automàtica, i la creació d'eines i recursos de suport als processos de correcció i traducció.</p> <p>URL: http://www.uoc.edu/serveilinguistic</p>
<p>Grup de Tractament Automàtic del Llenguatge Natural, Departament de Tecnologies de la Informació i les Comunicacions (Universitat Pompeu Fabra i Barcelona Media Centre d'Innovació)</p> <p>Objectiu: fer recerca i aplicacions en alguns camps específics del processament automàtic del llenguatge natural, en el marc teòric de la Teoria Sentit-Text (TST).</p>	<p>FlexSem, Departament de Filologia Francesa i Romànica (Universitat Autònoma de Barcelona)</p> <p>Objectiu: dur a terme recerca aplicada al processament automàtic del llenguatge natural des d'una perspectiva de formalització lingüística i crear recursos lingüístics.</p> <p>Línies de treball: prosòdia, patologies de la parla i intercomprensió entre llengües romàniques; les vinculacions i vehiculacions semàntiques del lèxic amb aplicabilitat en la creació de diccionaris electrònics, i mòduls de processament lingüístic.</p> <p>URL: https://masters.uab.es/flexem</p>
	<p>Centre de Llenguatge i Computació (Universitat de Barcelona)</p> <p>Objectiu: desenvolupar recursos de tecnologia lingüística (corpus anotats, lexicons, analitzadors) sobre la base de les aplicacions basades en el processament del llenguatge, amb cura dels fonaments lingüístics i metodològics.</p> <p>Línies de treball: processament massiu de textos, anàlisi morfològica i sintàctica, corpus anotats, analitzadors basats en aprenentatge automàtic, resolució de la coreferència i identificació de paràfrasis.</p> <p>URL: http://clic.ub.edu</p>
	<p>Institut Universitari de Lingüística Aplicada (Universitat Pompeu Fabra)</p> <p>Objectiu: crear recursos lingüístics i dur a terme recerca en lingüística de corpus i tecnologies lingüístiques.</p> <p>Línies de treball: demolingüística, discurs especialitzat, documentació digital, enginyeria lingüística, extracció d'informació, fonologia aplicada, lexicografia, lexicologia, lingüística forense, neologia, sociolingüística, terminologia i variació lingüística.</p> <p>URL: http://www.iula.upf.edu</p>
	<p>Grup de Recerca en Tecnologies Mèdia Enginyeria i Arquitectura La Salle (Universitat Ramon Llull)</p> <p>Objectiu: treballar la multidisciplinarietat centrada</p>

<p>en la innovació en noves tecnologies multimèdia i multimodals.</p> <p>Línies de treball: síntesi de la parla expressiva, creació d'avatars (cares parlants i llengua de signes catalana), reconeixement de la parla, reconeixement audiovisual d'emocions i, finalment, disseny i enregistrament de corpus de veu.</p> <p>URL: http://www.salle.url.edu/~iriondo</p>	<p>URL: http://gps-tsc.upc.es/veu</p>
<p>Grup de Recerca Interuniversitari en Aplicacions Lingüístiques, Grup Interuniversitari (Universitat Autònoma de Barcelona, Universitat de Barcelona i Universitat Oberta de Catalunya)</p> <p>Objectiu: desenvolupar, avaluar i aplicar recursos lingüístics de gran cobertura, creant i fent ús de gramàtiques, lèxics, corpus i memòries de traducció.</p> <p>Línies de treball: gramàtiques de dependències i HPSG (<i>Head-Driven Phrase-Structure Grammars</i>), recursos lèxics, corpus (anotació i explotació), adquisició d'informació (esquemes de subcategorització i restriccions de selecció) i memòries de traducció.</p> <p>URL: http://grial.uab.es</p>	<p>Grup Transmedia Catalonia, Departament de Traducció i d'Interpretació (Universitat Autònoma de Barcelona)</p> <p>Objectiu: experimentar per mitjà de tecnologies com <i>l'eye-tracking</i> o el reconeixement de parla aplicades a la traducció audiovisual i a l'accessibilitat.</p> <p>Línies de treball: traducció audiovisual (doblatge, subtítolació i veus superposades), accessibilitat als mitjans (audiodescripció, subtítolació per a sords i audiosubtitulació) i accessibilitat en la docència.</p> <p>URL: http://www.fti.uab.cat/transmediacatalonia</p>
<p>Pattern Recognition and Human Language Technology, Institut Tecnològic d'Informàtica (Universitat Politècnica de València)</p> <p>Objectiu: reconèixer patrons i desenvolupar l'aprenentatge automàtic i la interacció multimodal.</p> <p>Línies de treball: traducció de textos i veu (automàtica o assistida), reconeixement de la parla, reconeixement de textos manuscrits (automàtic o assistit), biometria i recuperació d'imatges interactiva.</p> <p>URL: http://prhlt.iti.es</p>	<p>Projecte institucional d'investigació <i>IVITRA</i> (Universitat d'Alacant)</p> <p>Objectiu: dur a terme recerca pel que fa a les eines relacionades amb la traducció automàtica i amb les aplicacions per al tractament de corpus multilingües i paral·lels.</p> <p>Línies de treball: elaboració de corpus del català antic, textos paral·lels (traduïts) i totes les eines relacionades amb el seu tractament.</p> <p>URL: http://www.ivitra.ua.es</p>
<p>Grup de Tractament de la Parla, Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (Universitat Politècnica de Catalunya)</p> <p>Objectiu: investigar en el camp del processament computacional de la llengua, com també crear recursos lingüístics, amb un interès especial cap a la traducció automàtica.</p> <p>Línies de treball: reconeixement automàtic de la parla, conversió de text a veu, traducció automàtica oral i textual, creació de recursos lingüístics (orals i textuals) i el processament de la parla i àudio en interfícies multimodals.</p>	<p>Transducens, Departament de Llenguatges i Sistemes Informàtics (Universitat d'Alacant)</p> <p>Objectiu: investigar i crear aplicacions que emprin tecnologies de processament del llenguatge natural.</p> <p>Línies de treball: l'educació assistida per ordinador, la indexació i els llenguatges de marcat en biblioteques digitals, i la inferència gramatical a partir de mostres estocàstiques i les seves aplicacions; traducció automàtica, anotació fonètica de textos per a la lectura en veu alta, i eines per al processament de textos paral·lels i per a la generació de memòries de traducció.</p> <p>URL: http://transducens.dlsi.ua.es</p>