

Veu i Llenguatge

Barcelona Media

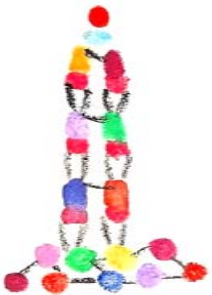


Jornada del Processament Computacional del Català

26 DE MARÇ DE 2009, PALAU ROBERT (BARCELONA)

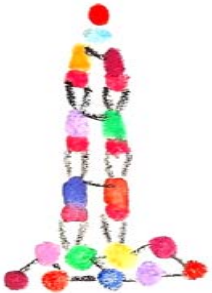
Organització

- *Grup de Lingüística Computacional*
 - *Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra*
- *Grup de Veu i Llenguatge*
 - *Barcelona Media, Centre d'Innovació*



Línies de recerca (I)

- *extracció d'informació*
 - *lingüística*
 - *semàntica*
- *correcció de textos*
 - *escrits per nadius*
 - *escrits per aprenents de segones llengües*
- *traducció automàtica*



Línies de recerca (II)

- *síntesi de veu*
 - *síntesi per corpus i síntesi paramètrica*
 - *prosòdia i expressivitat*
- *llengua de signes*
 - *corpus de la llengua de signes catalana*
 - *traducció a la llengua de signes*
- *llengua antiga*



Línies de recerca (III)

- *tecnologies per a la docència*
 - *de llengües*
 - *de traducció*
- *docència de tecnologia per a la traducció*



Projectes (relacionats amb el català)

- *amb finançament competitiu*
 - *ALLES: Advanced Long-distance Language Education System (IST 2001-05)*
 - *eTitle: European Multilingual Transcription and Subtitling Services for Digital Media Content (eContent 2004-05)*
 - *MeLLANGE Multilingual eLearning in Language Engineering (Leonardo 2004-07)*
 - *ARQUITEXT. Arquitectura integrada para el tratamiento avanzado de textos (MEC 2004-07)*



Projectes (relacionats amb el català)

- *amb finançament intern (UPF)*
 - *CUCWeb: constitució del Corpus d'Ús del Català a la Web*
 - *BancTrad: constitució d'un corpus de traduccions*
- *amb contractes*
 - *El Corrector: Corrector ortogràfic i gramatical per a la llengua catalana*
 - *Desenvolupament d'un sintetitzador de veu bilingüe català-castellà*
 - *Revisor d'estic corporatiu (CCMA)*
 - *Extracció d'informació (TID)*



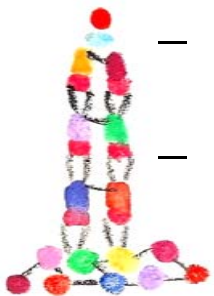
Eines i recursos

- **CatCG**
 - *Gramàtiques per a la desambiguació morfològica, l'anàlisi sintàctica, i la correcció de continguts*
 - *Llicència per a recerca i ús comercial*
- **COTiG/LINLaP**
 - *Arquitectura de processament lingüístic*
 - *El programari és codi lliure, però no tots els recursos lèxics ho són*



Eines i recursos (II)

- *CUCWeb*
 - *Corpus d'ús del català a la web*
 - *CC 3.0 Reconeixement-No comercial*
- *Cerevoice CAT*
 - *Veü per a la síntesi de parla en català*
 - *Llicència per a recerca i ús comercial*
- *Materials per al tractament de la veu*
 - *corpus i recursos lèxics*
 - *Llicència per a recerca i ús comercial*



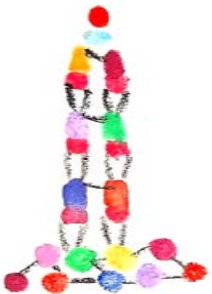
Eines i recursos (III)

- *METIS ES-CA*

- *Sistema de TA amb corpus monolingües*
- *Llicència per a recerca i ús comercial*

- *BancTrad*

- *Corpus alineats de traduccions des de i cap al català (amb castellà, anglès, alemany i francès)*
- *Llicència per a recerca i ús comercial*



Futur

- *seguir en les línies de treball encetades*
- *interès especial en la tecnologia útil per als usuaris finals*
 - *en entorns empresarials, educatius...*
 - *per a usuaris individuals o corporatius*

