

Anàlisi Lingüística i Corpus (21522)

Titulació/estudi: Grau en Llengües Aplicades

Curs: tercer

Trimestre: segon

Nombre de crèdits ECTS: 4

Hores de dedicació de l'estudiant: 100

Tipus d'assignatura: Obligatòria

Professor GG: Lluís de Yzaguirre i Maura

Llengua de docència: català

1. Presentació de l'assignatura

Aquesta assignatura aborda la consulta de corpus per a l'anàlisi del llenguatge, el treball amb diferents tipus de corpus i eines de consulta i l'alternativa entre *grammar driven linguistics* i *corpus driven linguistics*. Els estudiants realitzaran tasques d'exploració lingüística mitjançant la constitució i marcatge de corpus.

2. Competències que s'han d'assolir

D'entre el conjunt de competències que s'assoliran en acabar el grau, en aquesta assignatura es treballen les següents, generals (G) i específiques (E):

- Gestió de la informació (G.5)
- Coneixements d'informàtica (G.6)
- Aplicació de coneixements a la pràctica (G.17)
- Utilització d'eines informàtiques aplicades a la mediació lingüística (E.11)

Resultats d'aprenentatge:

- Avaluació de l'adequació d'un corpus determinat a les necessitats d'una recerca
- Localització de casuística en un corpus en base a
 - a) familiarització amb el seu etiquetari o
 - b) ús d'expressions regulars (corpus no enriquits)
- Obtenció de textos per a la creació, increment o actualització de corpus
- Creació, gestió i explotació d'un corpus personal per a la recerca o la docència (fins a 0,5Mw)
- Transferència de formats propietaris a estàndards públics dels textos a incorporar i viceversa amb els resultats obtinguts
- Avaluació dels beneficis eventuais d'una migració entre estàndards d'etiquetatge

3. Continguts

- L'anàlisi lingüística, entre la filosofia i l'empirisme
- Lingüística basada en corpus: la teoria, la pràctica i les aplicacions
- Utilització de corpus institucionals o personals
- Tipus de corpus i estàndards
- Tècniques d'interrogació i REGEX
- Creació de corpus i estratègies de processament
 - marcatge estructural
 - enriquiment
 - alineació

4. Avaluació

L'avaluació és continuada i té com a requisit indispensable l'assistència als *Seminaris* i la realització del treball pràctic que s'hi encarrega. L'assistència al *Grup Gran* no és obligatòria, però s'exigirà atenció i participació als estudiants que hi assisteixin i se n'obligarà a marxar els qui distreguin altres estudiants de manera reiterada.

L'avaluació està basada en tres components, que són

- treballs dels seminaris de pràctiques i la corresponent coavaluació (màxim 6 punts)
- examen sobre les lectures (màxim 3 punts)
- projecte personal de corpus per a la recerca pròpia (màxim 3 punts)

El primer component és obligatori i no recuperable (al juliol). L'estudiant que deixi d'assistir a dos seminaris o deixi de coavaluar-ne dos quedarà exclòs de l'avaluació continuada (ergo, haurà de repetir l'assignatura el curs vinent). Cal arribar a un mínim de 3 sobre 6 en aquest component per aprovar l'assignatura, encara que amb els altres dos components s'ultrapassi el cinc. Els altres dos components de l'avaluació són recuperables (al juliol). Tot això supeditat a la normativa oficial que es recull al final. La recuperació es farà al juliol.

L'estudiant podrà optar per ser avaluat en tots tres components o només en un o dos, almenys el dels seminaris; si ho fa només en un, cal que la seva aportació sigui (quasi) perfecta per arribar a 5 sobre 6, que és l'aprovat; si ho fa en dos, pot arribar a un 9 com a màxim; si ho fa en tots tres, el seu sostre de qualificació és el 10 amb opció a matrícula. Doncs, aquesta proposta d'avaluació està dissenyada pensant en un prototip d'estudiant que treballarà en dues de les tres línies proposades. En el cas dels estudiants que treballin en les tres línies i sumin una nota superior a 10, la nota final serà de 10 i l'excedent servirà per discriminar la persona mereixedora de la matrícula d'honor.

Els sis punts que s'obtenen amb els seminaris es basen en l'avaluació dels professors i en la coavaluació, que es fa entre els propis estudiants. Un sorteig emparella al final de cada seminari els assistents, que s'han d'enviar el treball fet durant el seminari. Cada assistent avalua el treball del company seguint una pauta detallada que se li proporcionarà. El resultat d'aquesta coavaluació serà públic, la persona coavaluada la podrà recórrer i finalment serà ponderada pels professors per harmonitzar el pes de l'avaluació amb els objectius de l'assignatura. Es pretén que, en avaluar el treball dels companys, cada estudiant pugui reflexionar autocríticament sobre els seus propis assoliments.

Els tres punts que s'obtenen amb els exàmens (tipus test) sobre les lectures són a raó de cinc dècimes per lectura amb un mínim de tres lectures i un màxim de sis, que s'aniran distribuint al llarg del trimestre, fins a la darrera setmana lectiva; en el cas de la recuperació (de juliol), es farà un sol examen amb totes les lectures i l'estudiant respondrà les que haurà treballat.

Els tres punts del projecte personal de corpus exigeixen haver recopilat, codificat, homogeneïtzat i estandarditzat els textos, haver-ne extret llistats alfabètics, freqüencials i de concordances i documentar tot el procés en un text explicatiu de la finalitat que es persegueix i les hipòtesis que es voldrien obtenir (la durada del curs no permet passar d'un simulacre). Els estudiants disposaran del mes de març per fer aquest treball.

5. Metodologia: activitats formatives

El curs té un valor de 4 ECTS, que impliquen 100 hores de treball de l'estudiant, 25 realitzades dintre de l'aula i la resta dedicades al treball individual. S'imparteixen dos tipus de sessions dintre de l'aula:

- Grup Gran (GG): sessions dedicades a la presentació dels fonaments teòrics, la discussió de les lectures de referència i tutories conjuntes per a l'orientació prèvia i valoració posterior del treball dels seminaris i altres activitats .
- Seminaris (S): sessions destinades al treball pràctic sobre corpus lingüístics

El treball individual fora de l'aula se centrarà sobretot en les lectures de referència i en el desenvolupament d'un projecte personal de corpus per a la recerca pròpia.

6. Bibliografia bàsica de l'assignatura

BIBER, Douglas; CONRAD, Susan; REPPEN, Randi (1998). *Corpus linguistics: investigating language structure and use*. 1a. edic. Cambridge: Cambridge University Press.

DE YZAGUIRRE, Lluís; RIBAS, Marta; VIVALDI, Jorge; CABRÉ, M.^a Teresa (2000). "Some Technical Aspects About Aligning Near Languages". Dins Gabrilidou, M. *et al* (ed.) (2000). *Second International Conference on Language Resources and Evaluation LREC-2000. Proceedings*, vol I. Atenes, 31 maig-2 juny 2000. Atenes: National Technical University of Athens Press. 545-548.

MCENERY, Tony; WILSON, Andrew (2001). *Corpus linguistics: An Introduction*. 2a. edic. Edinburgh: Edinburg University Press.

SINCLAIR, John (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

TORRUELLA, Joan; LLISTERRI, Joaquim (1999). "Diseño de corpus textuales y orales". Dins Blecua, J. M.; Clavería, G.; Sánchez, C.; Torruella, J. (1999). *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología, Departamento de Filología Española, Universidad Autónoma de Barcelona, Editorial Milenio. 45-77.