

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), concretades en forma de resultats d'aprenentatge:

- Gestió de la informació (G.5): destriar les fonts d'informació adequades en cada cas i organitzar-les de manera eficaç; saber valorar el grau d'enriquiment dels corpus necessari en funció de les hipòtesis a validar
- Coneixements d'informàtica (G.6): usar les eines informàtiques apropiades per a la creació i gestió de corpus personals
- Aplicació de coneixements a la pràctica (G.17):
 - usar la informació lingüística per a una bona gestió i explotació dels corpus
 - aplicar els conceptes teòrics i coneixements de la semàntica i pragmàtica a l'anàlisi de noves dades
- Utilització d'eines informàtiques aplicades a la mediació lingüística (E.11): emprar eines informàtiques i explotar corpus lingüístics per a finalitats aplicades relacionades amb la mediació entre llengües

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
 - o marcatge estructural
 - o enriquiment
 - o alineació

4. Avaluació

Avaluació ordinària

L'avaluació està basada en tres components, que són

- treballs dels seminaris de pràctiques
- examen sobre les lectures (amb valoració de l'aportació personal opcional als debats de les lectures)
- projecte personal de corpus per a la recerca pròpia

L'estudiant podrà optar per ser avaluat en totes tres modalitats o només en una o dues; si ho fa només en una, cal que la seva aportació estigui perfecta per arribar a 5 sobre 10; si ho fa en dues, pot arribar a un 8 com a màxim; si ho fa en totes 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 avaluació continuada durant el trimestre. Per poder-se acollir a l'avaluació continuada l'estudiant ha d'haver assistit al 80% o més de les classes de teoria i dels seminaris.

Els estudiants que vulguin aprofundir en la matèria en la línia Bolonya participant de projectes de l'IULA tenen diverses opcions, sempre avaluables com a increment de nota [implica que l'aprobat s'obté per la via convencional]. Els estudiants que tinguin altres inquietuds d'aquest estil, poden demanar hora per tal que analitzem conjuntament si hi ha alguna possibilitat de trobar-los encaix en algun projecte real que s'ajusti a les seves motivacions.

Si les circumstàncies ho permeten, es faran dues activitats voluntàries, avaluables com a increment de nota, condicionades a un nombre mínim d'inscrits i a disponibilitat d'espai. La primera serà un debat en forma de xat o fòrum sobre (algunes de) les lectures. La segona serà un seminari de programació en PERL.

Avaluació extraordinària

L'avaluació final de juny i l'extraordinària de setembre constaran d'un test sobre les lectures i d'un exercici pràctic de creació i explotació d'un corpus pesonal.

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.