

Informàtica Aplicada a la Traducció  
 Examen final  
 1 de febrer de 2002  
 Segona part (2 hores, 40% de la nota)

**Instruccions:**

- En aquesta part de l'examen podeu consultar qualsevol material que hàgeu portat però no podeu compartir material ni parlar amb ningú.
- Si teniu algun dubte, alceu la mà i aniré a atendre-us.

**Qüestions:**

1. Indiqueu quina classe d'ambigüitat presenten aquestes frases (justifiqueu breument però completament la vostra resposta):
  - (a) "La policia va entregar mostres de la droga que l'acusat venia al jutge" (1: El jutge comprava a l'acusat cocaïna de gran puresa; la policia en tenia mostres. 2: La policia va entregar les mostres al jutge).
  - (b) (Atracament a punta de pistola en una fruiteria) "¡Venga, venga! Cuatro sacos de patatas y las pelás" (1: "Si quieres te llevas las patatas pero pelártelas te las pelás tú"; 2: "Pelás? Hace más de un año que en caja sólo hay euros, colega").
  - (c) "El abogado perdió el juicio" (1: "L'advocat no va guanyar el cas"; 2: "L'advocat es va tornar boig").
  - (d) "Mentre esperàvem a la parada d'autobusos, va arribar el nou" (1: "L'autobús número 9"; 2: "Dels autobusos de la flota, el que s'ha adquirit més recentment").
  - (e) "Ven a nuestro supermercado y verás qué compras" (1: "Et sorprendràs"; 2: "Decidiràs").
  - (f) "La mata la droga" (1: "El vegetal la porta a un estat alterat de consciència"; 2: "El fàrmac l'assassina").
  - (g) "Els salaris dels espanyols s'han equiparat als dels alemanys, que són més alts" (1: "Els alemanys guanyen més diners"; 2: "L'estatura dels alemanys és més gran").

(h) "On diu que l'armada espanyola va patir una gran derrota?" (1: "En quina part del llibre?" 2: "En quin lloc del món van perdre la batalla?").

2. En un corpus de textos en castellà sobre economia de 925.461 mots estudiem quan apareixen mots conjuntament. En concret, i per posar un exemple, estudiem parells de mots gramaticalment vàlids on el primer mot apareix unes 400 vegades en total en el corpus i el segon mot hi apareix unes 200. Fixeu-vos en la taula 1 de freqüències d'aparició d'alguns parells. A pesar que tant el primer mot com el segon mot de cada parell tenen freqüències similars, en algun cas les freqüències d'aparició conjunta són molt elevades i en uns altres casos són molt més reduïdes. Podríeu explicar la causa d'aquesta variació? Per a quina aplicació de la informàtica a la aplicació podrien servir els resultats d'un estudi numèric com aquest?
3. Elegeix una llengua origen (català, espanyol, anglès, francès o alemany) i una llengua meta (català, espanyol, anglès, francès o alemany) i dóna tres exemples de frases que es poden traduir acceptablement *mot per mot* però tals que si canviem *un mot* de les frases per un altre de la mateixa categoria, la traducció *mot per mot* resulta incorrecta. En cada una de les frases, la raó lingüística per la qual la segona traducció és incorrecta ha de ser diferent.
4. Estudieu els següents sintagmes nominals en maori (una llengua polinèsia parlada en Nova Zelanda):
  - (1) *Te whare .*  
 Art. def. sg. casa .  
 La casa.
  - (2) *Ngā whare .*  
 Art. def. pl. casa .  
 Les cases.
  - (3) *Te whare nui .*  
 Art. def. sg. casa gran .  
 La casa gran.

Primer mot	Segon mot	Expressió
<i>fondos</i> (410)	<i>estructurales</i> (203)	<i>fondos estructurales</i> (63)
<i>precio</i> (415)	<i>máximo</i> (202)	<i>precio máximo</i> (2)
<i>algunos</i> (403)	<i>sectores</i> (211)	<i>algunos sectores</i> (1)
<i>hacia</i> (409)	<i>ellos</i> (204)	<i>hacia ellos</i> (0)
<i>otra</i> (411)	<i>crisis</i> (203)	<i>otra crisis</i> (0)

Taula 1: Taula de la pregunta 2

- (4) *Te whare nui o*  
Art. def. sg. casa gran de  
*te aroha* .  
Art. def. sg. amor .  
La casa gran de l'amor.

- (5) *Ngā whare nui* .  
Art. def. pl. casa gran .  
Les cases grans.

Com en els exemples, en maori la majoria dels noms i adjectius són invariables. Imagineu que sou part d'un equip de desenvolupament d'un sistema de traducció automàtica del maori al català (o al castellà) basat en l'estrategia que hem anomenat a classe transferència morfològica avançada<sup>1</sup>. Especifica completament *dues* regles (indicant possibles reordenaments i operacions per a assegurar la concordança) que permeten donar la traducció correcta de les oracions de dalt i de les següents. No us preocupeu de la contracció preposició–article.

- (6) *Ngā whare nui o te aroha* (les cases grans de l'amor)
- (7) *Te hau o te aroha* (El vent de l'amor)
- (8) *Ngā pukapuka o te whare* (El llibre de la casa)
- (9) *Ngā ingoa o te pukapuka nui* (Els noms del llibre gran)

<sup>1</sup>És a dir, llig les oracions mot a mot d'esquerra a dreta i fa l'anàlisi morfològica de cada mot, prova de detectar la seqüència més llarga de mots que concorda amb alguna seqüència de categories lèxiques que té en el seu catàleg, processa la seqüència, i continua immediatament després de la seqüència processada.

- (10) *Te ingoa o ngā whare* (El nom de les cases)

5. Es vol construir un sistema de traducció automàtica que tradueixi entre qualsevol dues llengües del grup format pel portugués, el gal·lec, el català, l'espanyol i l'italià. A més, es requereix que es puguin afegir fàcilment altres llengües com l'occità, el sard o l'asturià. No es busca la perfecció sinó més aviat traduccions en brut ràpides i fàcils d'entendre o de corregir (és a dir, amb pocs errors). Tenint en compte les llengües implicades, argumenteu a favor i en contra d'usar un sistema d'interlingua o un sistema de transferència, indicant en cada cas com haurien de ser les representacions intermèdies usades.
6. Les memòries de traducció usen els signes de puntuació per a segmentar les oracions abans d'alinejar els textos. En particular, l'aparició d'un punt (".") és moltes vegades un bon indicador del final d'una oració, però no sempre. Quan no? Doneu almenys *quatre* excepcions diferents a aquesta regla i descriguen, en cada cas, una regla senzilla que permeti decidir amb seguretat raonable quan ens trobem en cada una d'aquestes excepcions, usant el mínim possible d'anàlisi lingüística del text.