

1. [un punt per frase, puntuada sobre 8]

- a) Ambigüitat estructural pura d'adjunció. El SP "el jutge" pot modificar el SV "entregar" en "entregar mostres de la droga que l'acusat venia" o només el SV "venia".
- b) Ambigüitat mixta (lèxica-estructural) d'origen categorial. El mot "las" pot ser pronom o determinant; el mot "pelas" pot ser substantiu o verb. De les quatre combinacions possibles, les dues mostrades són sintàcticament acceptables.
- c) Ambigüitat lèxica pura: "juicio" és un mot polisèmic que pot voler dir "enteniment, facultat de raonament" o "procés judicial". També es pot considerar "perder el juicio" com una unitat lèxica multimot que té dues interpretacions, una recta i l'altra idiomàtica.
- d) Ambigüitat lèxica (homografia). El mot "nou" té dues formes lèxiques possibles: [nou, adjectiu de quatre terminacions, masculí singular] (recent, jove, etc.) o [nou, substantiu/adjectiu invariable, masculí singular] (que té el número que va darrere del vuit). Ha de quedar clar que en aquesta posició, "nou" no és ni determinant ni numeral. Es podria discutir si l'estructura sintàctica canvia quan canviem un per l'altre.
- e) Ambigüitat mixta (lèxica-estructural) d'origen categorial. Aquesta frase té dos homògrafs: "qué" pot ser ("adjectiu") determinant interrogatiu o exclamatiu invariable, quan encapçala sintagmes nominals, o pronom interrogatiu o exclamatiu invariable, quan forma per ell mateix un SN. D'altra banda, "compras" pot ser substantiu femení plural o verb en 2a persona del singular del present d'indicatiu. Dues de les quatre combinacions possibles són sintàcticament acceptables.
- f) Ambigüitat mixta (lèxica-estructural) d'origen categorial. Tenim quatre mots homògrafs: dos "la" (pronom/determinant), "mata" (nom/verb), "droga" (nom/verb)". De les setze interpretacions, n'hi ha dues sintàcticament vàlides:
- 1: det n pron v  
2: pron v det n
- g) Ambigüitat lèxica per anàfora: el pronom relatiu "que" pot tenir dos antecedents: el SN "els alemanys" o el SN que l'inclou "els [salaris] dels alemanys" (amb el N "salaris" el·líptic). O bé, ambigüitat estructural d'adjunció: el sintagma (SC, sintagma complementitzador) "que són més alts" pot ser un adjunt dels SN "els alemanys" o "els [salaris] dels alemanys".
- h) Ambigüitat estructural per moviment de Qu. L'adverbi interrogatiu "on" apareix obligatòriament en la posició inicial, però prèviament al moviment podia estar modificant el SV "diu que l'armada espanyola va patir una gran derrota" o el SV "va patir una gran derrota".

2. [puntuada sobre 4: dos punts per a l'explicació, dos punts per a les aplicacions]

Si la distribució dels mots fóra al atzar, la freqüència de tots els parells de mots seria la mateixa i molt baixa. Però hi ha mots que tendeixen a estar junts (col·locacions, unitats lèxiques multimot, unitats terminològiques) més que l'atzar.

Per exemple, el mot "fondos" apareix davant del mot "estructurales" 63 vegades de les 203 vegades que apareix "estructurales", és a dir, unes 3 de cada 10 vegades, quan a l'atzar apareixeria 410 vegades per cada 925.461, és a dir, unes 4 vegades cada 10.000. Per tant, apareix quasi mil vegades més freqüentment que l'atzar.

Es pot demostrar que, a pesar de ser menys freqüents, "precio máximo" o "algunos sectores" també tendeixen a estar junts per damunt de l'atzar, pot ser per ser col·locacions pròpies del tema econòmic.

Un estudi de bigrames (parelles) com aquests pot servir:

- primàriament, per a identificar unitats terminològiques ("fondos estructurales", "Real Decreto", "política monetaria"), col·locacions ("hacer frente", "tomar posiciones"), o noms d'entitat ("Nueva York", "Rodrigo Rato", "Unión Europea") pròpies del text en qüestió.
- secundàriament, per a decidir automàticament, per a un mot que té diverses traduccions, quina és la traducció que "sona més natural" davant o darrere de la traducció d'una altra.

3. [puntuada sobre 3: un punt per exemple correcte]

Exemples castellà--català [havien de ser tots per al mateix parell de llengües i estrictament mot per mot]

Homografia:

Le traje un [sombbrero] -> Li vaig portar un [barret] OK

Le traje un [traje] -> Li vaig portar un [vaig portar] NO (vestit)

Polisèmia:

El [canto] de la sirena -> El [cant] de la sirena OK

El [canto] de la moneda -> El [cant] de la moneda NO (cantell, viu)

Concordança de gènere o nombre:

[La] indicació era [inequívoca] -> La [indicació] era [inequívoca] OK

[La] señal era [inequívoca] -> [La] senyal era [inequívoca] NO (el, inequívoc)

Anàfora:

Cualquier [indicación] es importante para quien la comprenda ->

Qualsevol [indicació] és important per a qui [la] comprega. OK

Cualquier [señal] es importante para quien la comprenda ->

Qualsevol [senyal] és important per a qui [la] comprega. NO

4. [puntuada sobre 8: quatre punts per regla, penalitzant fortament els qui proposen més de dues regles]

Dues regles són suficients (la resta va bé mot per mot).

R\_1: - detectar "determinant nom"

- propagar el nombre (sing./pl.) del determinant maori (te/ngaa) al nom català

- propagar el gènere (masc./fem.) del nom català al determinant català

R\_2: - detectar "determinant nom adjectiu"

- propagar el nombre (sing./pl.) del determinant maori (te/ngaa) al nom i a l'adjectiu catalans

- propagar el gènere (masc./fem.) del nom català al determinant i a l'adjectiu catalans.

5. [puntuada sobre 8: dos punts per avantatge o desavantatge. No calia decidir-se per un disseny! Només calia argumentar]

(a) Avantatges d'interlingua:

- és necessari menys mòduls nous quan s'afeg una llengua nova al sistema (només un d'anàlisi i un de generació).

- no calen experts bilingües per a construir mòduls de transferència (és poc versemblant que existisquen experts asturià-català o asturià-sard).

(b) Desavantatges d'interlingua:

- Vista la semblança sintàctica entre les llengües involucrades, sembla excessivament costós fer l'esforç de dissenyar una representació d'interlingua, fer l'anàlisi i la generació completa dels textos (lèxica, sintàctica, semàntica) quan una transferència morfològica completa i sintàctica parcial seria suficient.

(c) Avantatges de transferència:

- Les llengües són prou similars perquè un sistema de transferència morfològica completa i sintàctica parcial amb poques regles done resultats acceptables.

(d) Desavantatges de transferència:

- Per descomptat, cada vegada que s'afeg una llengua a un sistema amb N llengües s'han d'escriure 2N mòduls de transferència i calen experts bilingües per a construir-los tots.

Els desavantatges d'interlingua s'atenuarien si en comptes d'una representació interlingual semàntica (basada en nocions com agent, pacient, temps, etc.) fora més aviat de naturalesa lèxica. Fins i tot, podria ser similar a una llengua humana. El llatí clàssic no, perquè a pesar de ser l'origen de totes les llengües del sistema té una sintaxi ---verb final--- i morfologia --declinació--- molt diferents; la llengua artificial anomenada interlingua ---"le lingua international facile e de aspecto natural elaborate per linguistas professional como denominador commun del linguas le plus diffundite in le mundo"--- amb anotacions sintàctiques i marques de desambiguació podria ser una millor opció.

6. [puntuada sobre 10: 2.5 punts per excepció diferent, ben explicada i detectada]

El punt apareix en moltes construccions que no indiquen el final de una oració. Heus ací uns exemples i com detectar-los per a no segmentar.

- a) Punts de milers (1.259), milions (1.032.200), decimals saxons (3.4), números de telèfon francesos (01.10.23.87.49), dates (20.01.1999), etc:

Solució: Si detectem [xifra] "." [xifra] (sense blancs), no segmentem.

- b) Punts en mig de sigles: CC(.)OO., EE(.)UU., etc.

Solució: si detectem [majúscula] "." [majúscula] (sense blancs), no segmentem.

- c) Punts al final d'abreviatura de cortesia: Sr(.), Dra(.), Excm(.), etc. o d'altres que precedeixen noms propis (Avda., Pça.) o números (Tel.)

Estratègia de solució:

si detectem "." [blanc] [minúscula], no segmentem.  
si detectem "." [blanc] [número], no segmentem.

si detectem "." [blanc] [majúscula], què fem? Depèn de l'abreviatura. La decisió és impracticable sense vocabularis de noms propis.

d) Punts al final de sigles: CC.OO(.), O.N.U(.).

si detectem [majúscules] "." [majúscules] ".", no segmentar.

[atenció: fallaria en "El líder de U.G.T. Cándido Méndez"]

e) Punts en URLs i adreces de correu electrònic

Solució: la millor seria una regla general (patró o expressió regular) que detectara aquestes entitats i evitara segmentar-les. Possibles escapatòries:

si detectem [minúscula] "." [minúscula] (sense blancs), no segmentem.

f) Punts suspensius ("...").

Solució: no segmentar mai "..." si es troba.

g) Ordinals bascos, alemanys (1. Mai, 2003.ko)

Solució?

etc.