

Esbós de solució de l'examen d'IAT del 8 de setembre de 2003 i explicacions sobre com ha estat corregit.

Mikel L. Forcada, 19 de setembre de 2003.

-----  
Com que els resultats de les dues últimes preguntes han estat molt baixos, potser perquè eren d'una dificultat excessiva (no ens ho semblaven quan les vam proposar) s'ha decidit que aquestes preguntes valguen la meitat, de manera que les preguntes 1-4 valen 2 punts cada una i la 5 i la 6, 1 punt cada una. Això reduïa molt el nombre de suspensos.  
-----

1. [Puntuada sobre 9 punts]

- (a) Ambigüitat estructural d'adjunció del sintagma "que fou fundat el 1504" (oració subordinada adjectiva de relatiu) que pot modificar el sintagma nominal "el poble" o "el teatre del poble". També es pot considerar un cas d'ambigüitat lèxica per anàfora del pronom relatiu "que" (que fa de subjecte de l'oració subordinada) que pot tenir qualsevol dels dos sintagmes nominals com a antecedents.
- (b) Ambigüitat estructural d'adjunció del sintagma preposicional "del piercing", que pot modificar el sintagma verbal "parlava amb la xica" (nucli "parlava") o el sintagma nominal "la xica" (nucli "xica").
- (c) Ambigüitat lèxica pura (polisèmia) del mot "vestuario" que pot voler dir "roba" (català "vestuari", interpretació 2) o "lloc on es canvia de roba" (català "vestidor", interpretació 1) i, per metonímia, l'equip que s'hi vesteix.
- (d) Ambigüitat estructural d'origen categorial. El sintagma adjectival "ben calents" pot modificar el sintagma nominal "dos cafés" o tot el sintagma nominal coordinat "dos iogurts i dos cafés".
- (e) Ambigüitat estructural d'adjunció. El sintagma preposicional "de fusta" pot modificar el sintagma nominal "llits per a xiquets" o només el sintagma nominal "xiquets".
- (f) Ambigüitat estructural d'adjunció de l'adverbi "ahir", que pot pertànyer a l'oració principal i modificar el sintagma verbal "llegir la nota que vas escriure" (nucli "llegir") o només el sintagma verbal "escriure".
- (g) [Aquesta ambigüitat és difícil d'explicar: per això, la pregunta s'ha puntuat benevolentment, considerant com a parcialment vàlides moltes solucions]. Ambigüitat complexa. D'una banda, hi ha una homografia: el mot "burro" és amb tota probabilitat un adjectiu en la interpretació figurada, la 2 (ja que s'hi podria dir "el muy burro") però un substantiu en la interpretació recta (la 1). Però, com que l'ús de "burro" com a adjectiu és clarament una habilitació del substantiu, també es podria considerar un cas de polisèmia. Però, d'altra banda, sembla que aquest és un cas d'ambigüitat entre la interpretació recta d'una expressió multimot i la figurada, encara que en aquest cas, l'expressió multimot siga variable: es podria donar la mateixa ambigüitat amb frases com "la bestia de la alcaldesa", i, en general, amb "la (animal o adjectiu valoratiu negatiu) de (persona)".
- (i) Es tracta d'una ambigüitat relacionada amb els antecedents dels dos pronoms nuls o PRO que fan de subjecte i objecte de l'oració subordinada d'infinitiu "menjar":

el pollastre està preparat per a PRO\_1 menjar [ PRO\_2 ].

En la primera interpretació PRO\_1 tindria com a antecedent "el pollastre" i PRO\_2 (si s'hi postula perquè es considera que "menjar" és sempre transitiu), allò que menja. En la segona

interpretació PRO\_1 fa referència a qui es menjarà el pollastre i PRO\_2 a "el pollastre".

---

(2) [Puntuada sobre 10] Procediment:

1. Construïm les dues possibles traduccions de la part de la frase sobre la qual dubtem, incloent el context que considerem rellevant (per exemple, si el fragment original és "ha escrit sobre nosaltres") provaríem "a écrit de nous" i "a écrit sur nous" si traduïm al francès o "has written about us", "has written of us" i "has written on us" si traduïm a l'anglès).
2. Buscaríem aquestes traduccions com a expressions (tota l'expressió entre cometes [2 punts!]) en la xarxa usant un buscador (p.ex. Google) i, opcionalment, limitant la recerca a documents en la llengua meta.
3. Elegiríem la traducció que més vegades apareix en la xarxa (aquella per a la qual el buscador troba més documents).
4. [2 punts!] Si cap de les traduccions es troba, busquem una expressió similar canviant un dels elements per un element anàleg (p.ex., "a écrit de vous"/"a écrit sur vous" o "has written about you"/"has written of you"/"has written on you") o més reduïda ("écrit de nous"/"écrit sur nous" , etc.) i comparem el nombre de vegades que apareix cada traducció.

Atenció: això, evidentment, no dóna la traducció més "normativa" sinó la més "usual", que podria fins i tot ser incorrecta per a alguna autoritat de la llengua meta corresponent.

---

(3) [Puntuada sobre 3] [les traduccions inadequades han de ser-ho!]

Exemples espanyol->català (de 5 de les 6 classes)

(a) Polisèmia:

- "desde que era pequeño sabía que mi destino era ser cantante"
- \* "des que era menut sabia que la meua destinació era ser cantant"
- C: "des que era menut sabia que el meu destí era ser cantant"

(b) Homografia dins la mateixa categoria:

- "vendo vendas para que los enfermeros venden a los soldados"
- \* "embene benes perquè els infermers venen els soldats"
- C: "venc benes perquè els infermers embenen els soldats"

(c) Ambigüitat estructural d'origen coordinatiu

- "vendo las cinco cosas en un paquete, las dos almohadas y las tres mantas juntas, por seis euros"
- \* "venc les cinc coses en un paquet, els dos coixins i les dues mantes juntes, per sis euros"
- C: "venc les cinc coses en un paquet, els coixins i les mantes junts, paquet per sis euros"

(d) Ambigüitat estructural d'adjunció d'un sintagma preposicional

- "como tengo Teleplús, compré el partido que el Deportivo ganó por doce euros"
- \* "com que tinc Teleplús, vaig comprar el partit que el Deportivo va guanyar de dotze euros"
- C: "com que tinc Teleplús, vaig comprar el partit que el Deportivo va guanyar per dotze euros"

[Els partits de futbol es guanyen "d'un gol", "de dos gols", etc. però mai "de dos euros". El traductor ha interpretat que era la diferència de gols (modificant "guanyar") i no el preu del "pay-per-view" (modificant "comprar") i ha usat la preposició "de" que cal en català.]

(e) Anàfora pronominal amb pronom explícit:

- "me han dado los resultados de los análisis de mis leucocitos y se los he dado al médico".

\* "m'han donat les anàlisis dels meus leucocits i els he donat al metge"

C: "m'han donat les anàlisis dels meus leucocits i les he donat al metge"

[no sembla adequada la traducció on el pacient dóna leucocits al metge]

(f) Ambigüitat estructural per moviment de Qu:

- "Cuando vieron el diagnóstico, a qué médico pensaron que irían?"

\* "Quan van veure el diagnòstic, a quin metge van pensar que anirien?"

C "Quan van veure el diagnòstic, a ca quin metge van pensar que anirien?"

["ir al médico" es diu "anar a cal metge" en català, però el sistema ha pensat que "a qué médico" pertanyia a l'oració principal "pensaron" i l'ha traduït com un objecte directe, però això no té sentit perquè no es pot pensar res a ningú.]

---

4. [puntuada sobre 5: 1 punt per la detecció de mots amb traduccions diferents, 1 per cada tipus d'operació (màxim 4 punts) i bonificació per les al·lusions a les limitacions]

Examinant una a una les paraules que apareixen en les frases és fàcil veure que el sistema fa moltes coses més que traduir mot per mot, ja que en contextos diferents apareixen traduccions diferents per a alguns mots:

the : el, la, los, las

red : rojo, roja, rojos, rojas

house: casa, alojan

houses: casas, aloja

saw : vi, sierra

seemed : parecía, parecían

Operacions que fa el sistema:

1. Resoldre homògrafs categorials: house (n/v), houses (n/v), saw (n/v) d'acord amb el context.
2. Fer concordar en gènere i nombre articles (the) i adjectius (red) amb el substantiu al qual modifiquen en estructures article-nom i article-adjectiu-nom. La informació sobre el nombre prové del nom en anglés i la informació sobre el gènere el gènere del nom en espanyol. El gènere i el nombre s'usen per a operacions posteriors.
3. Concordança de nombre entre el SN que fa de subjecte i el verb (p.e., en "seemed").
4. En construccions copulatives, fer concordar en gènere i en nombre l'adjectiu (atribut) amb el SN que fa de subjecte en construccions amb el verb "to be", però no amb el verb "to seem".

5. Reordenar algunes estructures com adjectiu-nom en nom-adjectiu.

Limitacions: de l'observat és difícil predir quines limitacions tindrà ja que totes les traduccions són acceptables excepte les que tenen el verb "seemed". Com que el que s'observa és compatible amb un sistema de transferència morfològica avançada, podria tenir les mateixes limitacions, però els exemples no són suficients per a assegurar que es tracte d'aquesta estratègia o d'una altra més avançada.

---

5. [La pregunta és molt oberta i la resposta pot no ser unívoca. S'ha valorat bàsicament que s'aborde el problema presentat i no altre i s'argumente correctament, seguint les línies indicades en la pregunta. El que segueix és un esbòs.]

[La majoria de les notes són zero perquè no s'ha arribat a abordar el problema concret que es plantejava.]

Arguments A FAVOR d'usar una anàlisi lleugera usant una "interlingua superficial" en lloc d'una "interlingua semàntica":

- Fa l'efecte que, en la majoria dels casos, la representació sintàctica dels rols semàntics entre les llengües romàniques és ben similar: en general, els subjectes van davant del verb, els objectes darrere, etc.

Aquest argument s'hauria de suportar amb dades estadístiques que demostraren que, per exemple, l'ordre i els àmbits de concordança dels mots en les llengües involucrades és, en general, molt similar, amb petits canvis locals, i els pocs errors no afectarien l'"objectiu 95%". De qualsevol manera, la experiència prèvia del grup amb els parells de llengües sembla argumentar a favor d'aquesta aproximació, que dividiria la "transferència morfològica avançada" en dues etapes, amb la "interlingua superficial" situada entre la RALO i la RALM; la primera etapa passaria a formar part de l'anàlisi i la segona, part de la generació.

- Una interlingua superficial (bàsicament lèxica, amb indicacions superficials quant als patrons detectats, la concordança interna i entre ells, etc.) és molt senzilla de generar amb una anàlisi morfològica seguida d'una desambiguació lèxica categorial i la detecció de patrons de paraules que necessiten un tractament especial. És molt més senzilla que una anàlisi semàntica completa i necessita dades lingüístiques molt més simples.

Arguments EN CONTRA d'usar una anàlisi lleugera usant una "interlingua superficial" en lloc d'una "interlingua semàntica":

- Encara que les llengües siguin sintàcticament similars, els casos de divergència sintàctica poden tenir un gran impacte. Per exemple:
  - l'estructura de les interrogatives sí/no, especialment divergent en francès ["est-ce que il fait chaud" (fr) / "fa calor" (ca)],
  - les divergències sintàctiques en verbs com "aimer" (fr) o "gostar de" (pt) ["j'aime le chocolat", "eu gosto do chocolate"] versus "piacere" (it), "agradar" (ca), "gustar" (es) ["mi piace il cioccolato", "me gusta el chocolate", etc.]
  - la traducció correcta dels pronoms (gènere, nombre) exigeix la determinació dels seus antecedents, i això no es pot fer amb una anàlisi superficial

Per a resoldre correctament aquests casos és necessària una

anàlisi més profunda, amb una representació intermèdia que represente els rols semàntics explícitament.

- Fins i tot sense entrar en un nivell d'anàlisi semàntica, la necessitat d'una anàlisi sintàctica completa (no parcial ni superficial) es manifesta en fenòmens com la concordança a llarga distància, que poden ser importants:

El postre que nos trajeron ayer de Murcia [...] estaba riquísimo.

LES POSTRES que ens van portar ahir de Múrcia [...] estavEN boníssimES.

Els arguments en contra s'haurien de suportar amb una evidència clara que els errors deguts a no tractar aquests problemes impedirien l'"objectiu 95%".

[ Argument no vàlids: "Les interlingües han de ser semàntiques". És cert que ho són històricament, però les interlingües han de ser sobretot "neutrals", és a dir, estar enmig de les llengües involucrades. ]

[ Errada típica: argumentar entre transferència i interlingua de manera genèrica. Moltes respostes es fiquen per aquest camí i això fa que no ataquen el problema, que és entre interlingües, no entre interlingua i transferència ]

- 
6. [puntuada sobre 2 punts: resposta correcta, parcialment correcta, o incorrecta]

El programa gestor de la memòria de traducció pot detectar automàticament quines unitats de traducció (UT) tenen variants molt similars en totes les llengües (dues en aquest cas) i anotar una espècie de "diferències" paral·leles entre elles (detectar UT paregudes és una cosa que ja fan les memòries de traducció; per tant, és completament factible).

Per exemple, la "diferència" entre les dues primeres UT

(ca:"Aquest és un missatge llarg", eu:"Hau mezu luze bat da")  
(ca:"Aquest és un missatge curt", eu:"Hau mezu labur bat da")

Seria

(ca:("llarg","curt"), eu:("luze","labur"))

És a dir, si en català canviem "llarg" per "curt" o al revés, en basc sembla que cal canviar "luze" per "labur" o al revés. Clar que amb un exemple és molt arriscat assegurar l'equivalència entre "llarg" i "luze" d'una banda i "curt" i "labur" d'altra, però, e fet, aquesta és exactament la diferència observada entre la tercera i la quarta UT de la memòria.

Aquesta anàlisi la pot fer el programa sense fer cap mena d'anàlisi lingüística. Si una diferència es repeteix molt en la memòria, vol dir que és molt probable que siga aplicable en altres casos.

En el cas que ens ocupa, la nova frase catalana "ha escrit un mot curt" es diferencia de la variant catalana de la 5a UT de la memòria exactament perquè canvia "llarg" per "curt". Aplicant la "diferència" més freqüent, tocaria només canviar, en la variant basca de la 5a UT "luze" per "labur". Així, obtindríem "Hitz labur bat idatzi du"; aquesta traducció seria, probablement, una millor pretraducció que la variant basca de la 5a UT ("Hitz luze bat idatzi du"). El sistema presentaria la traducció "construïda", destacant l'operació realitzada i el traductor la validaria.

Per què no fan això les MT comercials? Potser perquè és poc freqüent trobar "parells" tan clars com els de l'exemple. O potser sí que ho gan ja?