

Programa-guia d'activitats d'aula d'Informàtica Aplicada a la Traducció*

Mikel L. Forcada
Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant, E-03071 Alacant

Curs 2007–2008

Índex

1	Què estudiarem en aquest curs?	1
2	Ordinadors i programes	2
3	Internet	3
4	Textos i formats	3
5	Usos de la traducció automàtica	5
6	Ambigüitat	5
7	Com funciona la traducció automàtica?	7
8	Avaluació de la traducció automàtica	9
9	Memòries de traducció	10
10	Bases de dades lèxiques	10

1 Què estudiarem en aquest curs?

Act. 1.1 El nom d'aquesta assignatura és Informàtica Aplicada a la Traducció. Si mireu el pla d'estudis, veureu que el descriptor de l'assignatura és

“Accés als instruments de treball per al suport a la tasca de traductor. Traducció automàtica i semiautomàtica i integració de sistemes”

Com a primera aproximació, responeu les preguntes següents:

1. Quin creieu que és el significat dels conceptes que apareixen en el descriptor? En concret, què s'entén per *traducció automàtica* i *semiautomàtica*?
2. Com creieu que es pot aplicar la informàtica a la traducció? Quina part del treball la realitzaria l'ordinador (automàticament) i quina part realitzarien els professionals? Doneu-ne detalls.
3. Heu usat ja algun programa dels que es poden aplicar a la traducció? Quin?
4. Què penseu que heu de conèixer sobre altres disciplines o assignatures per a aprendre aquesta assignatura?
5. Què us agradaria aprendre en aquesta assignatura?
6. Com us agradaria que fóra?

*© 2007 Universitat d'Alacant. Aquest material pot ser distribuït, copiat i exhibit si el nom de l'autor es mostra en els crèdits. Les obres derivades han de distribuir-se sota els mateixos termes de llicència que el treball original. Més detalls: <http://creativecommons.org/licenses/by-sa/3.0/deed.ca>. Podeu demanar els fonts LaTeX a l'autor (mlf@ua.es).

2 Ordinadors i programes

Act. 2.1 Ací teniu algunes descripcions informàtiques (algunes, quasi prehistòriques!). És probable que no entengueu completament què vol dir cada paraula, però si creieu que enteneu alguna part, podríeu provar a definir-la en el grup menut i explicar-la després al grup gran. A poc a poc les anirem desxifrant entre tots.

1. La màquina que té Toni a casa (comprada a finals de 1992) és un 386 de 40 MHz, amb 4 MB de RAM i un disc fix de 203 MB; l'adaptador de vídeo és SuperVGA. Té dues unitats de disquet, una de tres polzades i mitja i l'altra de cinc i quart, i ratolí, per descomptat. A més, li ha instal·lat un lector de CD-ROM de velocitat quàdrupla, una placa de so SoundBlaster i un mòdem de 2400 bps. Té instal·lats els sistemes operatius DOS 6.2 i Windows 3.1.
2. La màquina del despatx de Petra (de l'any 1998) és un Pentium II a 400 MHz, amb 64 MB de RAM i un disc dur de 20 gigabytes. Té una unitat de disquet de 3,5 polzades i un lector de CD-ROM 40×. No s'hi ha instal·lat ni Windows ni DOS ni OS/2, sinó una versió antiga de Linux.
3. Martí es va comprar el 2004 un ordinador en MegaTrastos: té una placa base ASUS A7s DDR 333, processador AMD 2600 xp, 256 MB de RAM DDR a 333 MHz, unitat de disquet de 3"5 de densitat alta, disc dur de 40 GB a 7200 rpm Seagate, teclat, ratolí PS2, targeta gràfica ATI 7000 amb 64 MB de RAM i eixida de TV, Unitat enregistradora de CD LG 52×-24×-52×, i una unitat lectora de DVD LG a 16×.
4. Per a executar els programes de traducció automàtica de l'Organització Panamericana de la Salut ENGSPAN i SPANAM, cal un Pentium, 64 MB de RAM, la xarxa d'àrea local Novell Netware o Windows NT/2000, una unitat de CD-ROM i 50 MB d'espai al disc dur per sis-

tema. Els programes treballen amb documents RTF, ASCII, ANSI, SGML, XML i HTML.

5. Requisits d'Acrobat Reader 6.0 per a Windows: processador Intel Pentium, Microsoft Windows 98 Second Edition, Windows Millennium Edition, Windows NT 4.0 amb *Service Pack 6*, Windows 2000 amb *Service Pack 2*, Windows XP Professional o Home Edition, Windows XP Tablet PC Edition, 32MB de RAM (64MB recomanats), 60MB d'espai lliure en el disc.
6. Requisits del navegador Mozilla 1.5 per a Windows: Windows 95, 98, o 98SE, Windows Millennium Edition, Windows NT 4.0, Windows 2000, o Windows XP; processador de la classe Intel Pentium (233 MHz o més ràpid recomanat); 64 MB de RAM; 26 MB d'espai lliure al disc.
7. Per a instal·lar el sistema operatiu GNU-Linux és necessari un Pentium a 200 MHz amb 1,5 GB de disc dur, 64 MB de RAM i una targeta gràfica de 4 MB.
8. Els CD-R que has comprat no valen, perquè només tenen 650 MB. Per a guardar tot necessitem un d'aquells de 700 MB. Si tens un CD-RW, millor, no siga que ens equivoquem.
9. Com que el fitxer té 2 MB, no cap en un disquet. Farem un CD-ROM perquè te'l pugues emportar.
10. La impressora que tinc a casa és d'injecció; té una resolució de 720 × 360 dpi i un alimentador de 40 fulls.

Act. 2.2 Ara que ja sabem una miqueta més sobre ordinadors, responeu les qüestions següents:

1. El programa SPANAM, es podria executar en l'ordinador de Toni? I en el de Petra? Per què?
2. Es pot instal·lar GNU-Linux en l'ordinador de Petra?
3. Es pot instal·lar Acrobat Reader 6.0 per a Windows en l'ordinador de Petra?
4. Es pot instal·lar Mozilla 1.5 per a Windows en l'ordinador de Martí?

5. Pot un ordinador basat en un processador *Pentium* executar programes escrits per a un processador 386?
6. Quanta informació cap en un disquet de 3,5 polzades de densitat alta (HD)? 144 MB? 720 kB? 1,44 MB?
7. Quantes combinacions diferents es poden fer amb 5 bits?
8. Quants *bytes* hi ha en un *megabyte*?
9. Quin dispositiu és necessari per a connectar l'ordinador de casa nostra a Internet per via telefònica?
10. Quina velocitat del processador és més típica en un PC actual? 30 MHz? 300 MHz? 3000 MHz?
11. Quina capacitat té un disc dur típic actualment? 100 GB? 100 MB? 100 kB?

Act. 2.3 Parlem ara de *programes*:

1. Què vol dir la paraula *programa* en la vida quotidiana?
2. Com definiríeu un programa d'ordinador?
3. On resideix un programa abans d'instal·lar-lo en el vostre ordinador? Com s'hi guarda?
4. On resideix després d'haver-l'hi instal·lat?
5. On ha de ser per a poder ser executat? Quan canvia de lloc? Qui el canvia?

Act. 2.4 I ara, parlem de *fitxers*:

1. Què és un fitxer?
2. On resideixen els fitxers en un ordinador?
3. Què pot contenir un fitxer d'ordinador?
4. En un disquet podríem tenir els fitxers un darrere de l'altre, sense organitzar-los, perquè no hi caben molts, però, com s'organitzen els fitxers en un disc fix o en un CD-ROM, on en caben molts més?

3 Internet

Act. 3.1 Abans d'una explicació d'algunes nocions bàsiques sobre Internet, és molt important saber què coneixeu sobre el tema. Per això, digueu:

1. Què creieu que és Internet? Podríeu proposar una definició provisional per a millorar-la posteriorment en classe?
2. Per a què pot servir Internet? Més concretament, en què creieu que pot servir a una persona que es dedica professionalment a la traducció?
3. De quina classe són els documents típics d'internet? En quin format estan escrits?
4. Com s'especifica en Internet on és un *recurs* (servei o document) concret? Quines parts té aquesta especificació? Què especifica cada part?
5. Com es busca en Internet un recurs quan no sabem on és (si és que hi és)?
6. Què és l'*E-mail*? Quins altres serveis similars a l'*E-mail* coneixeu?
7. De quantes maneres ens podem connectar a Internet des de casa? En què es diferencien? S'ha de pagar?

4 Textos i formats

Act. 4.1 Tal com vam veure en el primer bloc d'activitats, quan algú vol utilitzar l'ordinador per a fer una traducció, en algun punt ha de manipular o generar un *text informatitzat*. Indiqueu com fa per a generar aquest text i quines eines informàtiques —programari, maquinari— usa (procureu donar noms i definicions tan precises com pugueu). No oblideu que els textos es poden generar també sense haver de teclejar-los.

Act. 4.2 Com qualsevol tipus de dades, els textos informatitzats s'emmagatzemen com a seqüències de bits (bé: agrupats de 8 en 8, és a dir, com a seqüències d'octets o de *bytes*).

Hi ha moltes maneres diferents d'emmagatzemar textos informatitzats, és a dir, d'organitzar en octets la informació que contenen.

De fet, normalment s'ha d'especificar de quina manera concreta s'ha emmagatzemat un text, ja que cada programa espera que els textos estiguen organitzats d'una o de diverses maneres determinades i també els genera organitzats d'una o de diverses maneres determinades.

Us proposem que reflexioneu una mica sobre els aspectes següents:

1. Els textos contenen, a més d'altres informacions, els *caràcters* amb què els idiomes formen els mots.
 - (a) Com s'emmagatzemen els caràcters en forma d'octets?
 - (b) Per què passa de vegades que no podem veure bé tots els caràcters d'un text? (per exemple, veiem bé els caràcters no accentuats que la llengua del document té en comú amb la llengua anglesa, però no veiem bé altres caràcters)
 - (c) A més dels sistemes d'escriptura alfabètics com els nostres, hi ha sistemes d'escriptura amb conjunts de caràcters molt més grans, com ara els sil·làbics, amb centenars de símbols —*devanagari* (Índia), *hangul* (coreà), *hiragana* i *katakana* (japonés), etc.—, o els ideogràfics, amb milers de símbols —xinés, *kanji* japonés, etc. Com s'emmagatzemen els textos escrits en aquests sistemes?
 - (d) I si en un mateix document es mesclen diversos sistemes d'escriptura?
2. Però els textos informatitzats, a més dels caràcters que formen les paraules, contenen molta informació addicional.
 - (a) Quins elements d'informació addicional solen portar els textos informatitzats a més de les seqüències de caràcters que formen els mots?
 - (b) Per a què serveix aquesta informació addicional? És diferent aquesta informació segons quina siga l'aplicació del text informatitzat?
 - (c) Com s'emmagatzema aquesta informació en forma d'octets? És possi-

ble usar caràcters per a representarla, de manera que es podrien veure amb un editor de textos bàsic? Com?

- (d) Imaginem que ens han encarregat traduir un text informatitzat. En la llengua d'origen és costum posar en *cursives* tant els mots estrangers ("*Sprachgefühl*") com els termes nous quan es defineixen per primera volta ("Un *octet* és..."), se sagna la primera línia de tots els paràgrafs, i els números de secció porten un punt al final ("1.1. Introducció"), però en la llengua d'arribada els termes nous van en negretes ("Un **octet** és ..."), se sagna la primera línia de tots els paràgrafs excepte la del primer paràgraf d'una secció, i els números de secció no porten punt al final ("1.1 Introducció"). És adequat emmagatzemar els textos atenent només a la presentació visual? Com seria més adequat emmagatzemar-los?
 3. De vegades *el mateix text* s'ha de presentar de maneres diferents: complet en un navegador, resumit en la pantalla d'un telèfon mòbil, imprès com a document de gran qualitat, llegit en veu alta a un invident, etc. Com fem per a no tenir tantes versions com mitjans per a cada document?
 4. A més dels descrits, enumereu d'altres problemes associats a les diferències en els esquemes d'emmagatzematge dels textos informatitzats.
- Act. 4.3** Com ja sabeu, els processadors de textos més usuals segueixen, en la mesura del possible, un disseny *wysiwyg* (anglès: *what you see is what you get*, "el que veieu és el que obtindreu"): la presentació es basa en una o diverses finestres, cada una de les quals mostra una secció de l'estat actual d'algun dels documents de text informatitzats que estem creant i modificant (els documents que tenim *oberts*). El text es mostra tan paregut com siga possible a la versió impresa que se'n produirà, quant a format, tipus de lletra, etc. Aquest

disseny fa que la persona escriptora tendisca a centrar-se en els atributs *visuals* del text, ja que confia que una bona *presentació* transmetrà a les persones lectores l'estructura *lògica* que la persona escriptora té en el seu cap per al document.

Però un disseny de documents guiat *únicament* per la presentació té inconvenients molt importants. Fixeu-vos en la següent situació problemàtica:

Joaquim ha decidit que els títols de secció de l'informe anual que li han encarregat estaran en Helvetica de 14 punts, negreta i els de subsecció en Arial de 12 punts, negreta cursiva. A Marina, la seua directora, no li agraden així i li'ls ha fet canviar a Lucida Sans de 14, negreta i Lucida de 12, negreta sense cursives. Com que l'informe ha d'estar acabat per a demà de matí, Joaquim es queda a l'oficina fins a les 11 de la nit, canviant un a un els tipus de lletra del títols de seccions i subseccions. A l'endemà, de matí, Marina li passa un document amb una secció més que s'ha d'inserir entre la 4 i la 5. Joaquim no pot anar a esmorzar: ha de canviar els números de seccions i subseccions a partir de la 5 i repassar si s'ha de canviar alguna referència que es faça des d'una part del text a una secció pel seu número.

Tenen els processadors de textos actuals solucions perquè Joaquim no passe per aquest calvari una i altra vegada? Quina relació té això amb el que hem discutit en l'activitat 4?

5 Usos de la traducció automàtica

Act. 5.1 En el primer bloc del curs vam avançar una definició preliminar de *traducció automàtica*. Abans de tractar els aspectes relatius a l'automatització, convé que ens plantejem què entenem per *traducció* i reflexionem

un poc sobre aquest concepte. Podríeu provar a definir-lo amb tot el detall que pugueu (prepareu-ne una definició per a llegir-la en veu alta), després de pensar una miqueta en situacions on s'usa el mot *traducció*.

Act. 5.2 Quin interès pot tenir la traducció automàtica? Quins usos i camps d'aplicació se us acut que pot tenir?

Act. 5.3 Indiqueu, preliminarment, però amb tot el detall que pugueu, quines característiques del treball de traducció de textos penseu que fan difícil la seua automatització.

Act. 5.4 És necessari que un sistema de traducció automàtica sempre faça traduccions *perfectes* de qualsevol text per a ser útil? Si no, quin nivell d'*imperfeció* es podria tolerar? Per què? Doneu exemples i penseu en situacions concretes.

Act. 5.5 Convé fer en aquest punt una *tempesta d'idees* sobre els sistemes de traducció automàtica:

1. En què consisteix la tasca?
2. Quin aspecte creieu que hauria de tenir (per a qui l'usa) un sistema de traducció automàtica? Com hauria d'estar dissenyat perquè fóra fàcil d'usar?
3. Com creieu que funciona? Quines tasques bàsiques fa? De quines parts es compon?
4. En quina informació es basa cada una de les parts?

Quan acabem l'activitat hauríem de ser capaços de fer un diagrama de blocs preliminar d'un sistema genèric de traducció automàtica.

6 Ambigüitat

Act. 6.1 *L'ambigüitat* de les llengües naturals és una de les característiques que fa que la traducció automàtica siga especialment difícil. Podríeu definir breument el concepte d'*ambigüitat*? Per què el llenguatge humà és *ambigu*? Per què l'*ambigüitat* dificulta la traducció?

Act. 6.2 Considereu les frases ambigües següents:

1. *Vaig veure Joan parlant amb Maria i li vaig dir que no vingués avui a casa.* (A qui ho vaig dir?)
2. *Porta quaderns i llibres vells per a cremar en la ximenera.* (Els quaderns, vells també?)
3. *A Catalunya, la millor estació és aquesta.* (L'estiu? L'estació ferroviària de Sants?)
4. *Qui diu que va venir?* (1: "Qui diu això?"; 2: "Diu que va venir... qui?")
5. *Les finestres de la casa que va pintar Joan són grans.* (Què va pintar Joan, les finestres de la casa o la casa?)
6. *És molt amic de Joan.* (Qui?)
7. *Porta les claus de l'armari gran* (les claus que l'obrin o les claus que hi ha allà?).
8. *Viu la festa* (1: "Ell participa de la festa"; 2: "Participa de la festa!"; 3: "Vaig veure la festa").
9. (en)¹ *I saw her duck under the table* (1: "Vaig veure enu ànec sota la taula"; 2: "Vaig veure com s'ajupia (per a amagar-se) sota la taula")
10. (en) *Time flies like an arrow* (exemple clàssic amb tres interpretacions possibles: busqueu-les).
11. (en) *"I saw the girl with the telescope"* (un altre clàssic).
12. (en) *"Sue went to put the key under the doormat. When she lifted it up, a cockroach quickly scampered across the path"* (it és doormat o key?) (Arnold et al. 1994).

Els exemples contenen ambigüitats de molts tipus diferents.

Fins i tot entre dues llengües molt similars com l'espanyol i el català, l'ambigüitat pot estar associada a l'existència de més d'una traducció possible (l'elecció de la interpretació incorrecta pot donar lloc a un error de traducció). Fixeu-vos en els casos següents:

13. *Nadie conocía el destino del avión secuestrado*

¹Els exemples usen els codis ISO-639-2 (http://www.loc.gov/standards/iso639-2/php/code_list.php) per a indicar les llengües diferents del català.

T1: *Ningú no coneixia el destí de l'avió segrestat* (anaven a morir)

T2: *Ningú no coneixia la destinació de l'avió segrestat* (Washington?)

14. *Este vino de Jerez para el trabajo*

T1: *Aquest vi de Xerès atura el treball* (té massa alcohol i no hi ha manera de treballar)

T2: *Aquest va venir de Xerès per al treball* (falten obrers i va venir de molt lluny)

15. *Les pidió que fueran como él les había enseñado*

T1: *Els va demanar que anaren com ell els havia ensenyat* (en autobús)

T2: *Els va demanar que foren com ell els havia ensenyat* (honestos i sincers)

16. *Te vendo un coche*

T1: *T'embene un cotxe* (amb benes)

T2: *Et venc un cotxe* (perquè necessite diners)

17. *Almohadas y mantas amarillas*

T1: *Coixins i mantes grogues* (les mantes només)

T2: *Coixins i mantes grocs* (les mantes i els coixins)

18. *Como tenía puntos de Teleplús, compré el partido que el Lucentum ganó por doce puntos*

T1: *Com que tenia punts de Teleplús, vaig comprar el partit que el Lucentum va guanyar de 12 punts* (Tau 87, Lucentum 99)

T2: *Com que tenia punts de Teleplús, vaig comprar el partit que el Lucentum va guanyar per 12 punts* (I encara em queden 30 punts en el saldo del satèl·lit)

19. *Me han dado los análisis de mis almendros y se los he tenido que enseñar al inspector de agricultura*

T1: *M'han donat les anàlisis dels meus ametlers i li'ls he hagut d'ensenyar a l'inspector d'agricultura* (li he ensenyat els ametlers)

T2: *M'han donat les anàlisis dels meus ametlers i li les he hagut d'ensenyar a l'inspector d'agricultura (li he ensenyat les anàlisis)*

20. *El camarero trajo el postre y se fue; lo miré con deseo*

T1: *El cambrer va portar les postres i se'n va anar; les vaig mirar amb desig (feia temps que no havia menjat arròs amb llet)*

T2: *El cambrer va portar les postres i se'n va anar; el vaig mirar amb desig (si caminava així, com seria al llit?)*

21. *¿A qué médico dijeron que irían?*

T1: *A quin metge van dir que hi anirien? (Pregunte pel metge a qui van explicar la intenció d'anar a algun lloc)*

T2: *A ca quin metge van dir que anirien? (Pregunte pel metge que van dir que visitarien)*

Proveu de classificar aquestes ambigüitats i les de la llista anterior usant algun esquema que estiga motivat lingüísticament.

Act. 6.3 Esbosseu possibles estratègies (automatitzables) per a abordar en un sistema real de traducció automàtica els tipus bàsics d'ambigüitat que han aparegut en l'activitat anterior.

Act. 6.4 Indica què fa que els textos ambigus següents siguin especialment difícils de tractar en un sistema de traducció automàtica, encara que majoria de les persones puguen elegir la interpretació correcta:

1. (en) "The soldiers shot at the women and I saw them fall" ("Els soldats van disparar a les dones i *els* vaig veure caure" —els soldats— o "...i *les* vaig veure caure" —les dones—).
2. "Va agarrar les claus de la cadira" (hi ha claus que òbriguen cadires? tenen pany les cadires?)
3. "Les assessories fiscals ajuden molt quan s'han de preparar les declaracions de la renda però així i tot les odie".

4. (conversa entre dues dones) "Vas tenir relacions amb el teu home abans de casar-te amb ell?" "Jo no, i tu?" "Jo sí, però no sabia que acabaria casant-se amb tu".

Quin tipus d'ambigüitat representa cada cas? Com podria un sistema de traducció automàtica resoldre l'ambigüitat?

Act. 6.5 Un cas interessant succeeix quan es tradueix una frase que és ambigua en la llengua origen i resulta que les traduccions corresponents a cada una de les possibles interpretacions són idèntiques. Això s'anomena *free ride* (passi gratuït). En podríeu posar algun exemple?

7 Com funciona la traducció automàtica?

Act. 7.1 Una aproximació preliminar —i bastant rudimentària— a la traducció automàtica és l'anomenada *traducció mot per mot*: el sistema llig el text original mot a mot i d'esquerra a dreta, substitueix cada mot original per un mot equivalent en llengua meta² i escriu els mots un a un i en el mateix ordre en el text meta, de manera que l'ordre dels mots es conserva (aquesta aproximació l'anomenarem en classe "model 0"). Indiqueu alguns problemes d'aquesta aproximació, inspirant-vos en les traduccions mot per mot següents, i indiqueu com es podrien resoldre en una estratègia més avançada de traducció automàtica (que anomenarem "model 1").

1. (es) El oso apareció tarde → (ca) **El gosso va aparèixer trigui*
2. (en) The computer expert's large table is full → (ca) **El ordinador expert gran taula és ple*
3. (es) El satélite enviaba una buena señal pero sin datos específicos. → (ca) **El satèl·lit enviava una bona senyal però sense dades específics*

²El millor possible, el més freqüent, etc.

4. (es) Menos mal que sólo murieron sesenta y cinco personas → (ca) *Menys malament que només van morir seixanta i cinc persones

Act. 7.2 Indiqueu alguns problemes no resolts pel “model 1”, proposat durant la realització de l’activitat anterior i feu un esbós de les possibles estratègies de solució que s’haurien d’incloure en un hipotètic “model 2”. Per a inspirar-vos, fixeu-vos en les frases següents i les traduccions produïdes per un sistema “model 1”.

1. (a) (cat.) No volem més problemes → (esp.) ?*No volamos más problemas.*
(b) (esp.) No salen de casa → (cat.) **No salin de casa*
2. (traduccions d’un “model 1” de l’espanyol al català):
(a) Una almohada → *Un coixí*
(b) Una almohada cómoda → *Un coixí còmode*
(c) Una buena almohada → **Una bona coixí*
(d) Una almohada muy cómoda → **Un coixí molt còmoda*
(e) La almohada que me compraste es muy cómoda → **El coixí que em vas comprar és molt còmoda.*
3. (traduccions d’un “model 1” de l’anglès al català):
(a) A house → *Una casa*
(b) A car → *Un cotxe*
(c) Red houses → *Cases vermelles*
(d) A large house → *Una casa gran*
(e) The young expert → *L’expert jove*
(f) The professor’s house → *La casa del catedràtic*
(g) The young professor’s car → *El cotxe del catedràtic jove*
(h) The young professor’s large car → **El catedràtic jove cotxe gran*
(i) The physics professor’s car → *El cotxe del catedràtic de física*
(j) The young physics professor’s car → *La física jove catedràtic cotxe*

Act. 7.3 Com ja hem vist, els sistemes de traducció automàtica entre dues llengües amb sintaxi molt diferent necessiten fer reordenaments. Imagineu un sistema de traducció automàtica que tradueix sintagmes nominals del català al basc i usa regles per a reordenar seqüències de categories gramaticals segons s’ha discutit en l’activitat anterior: d’esquerra a dreta, reordenant la seqüència més llarga detectada i sense operar dues voltes sobre el mateix mot.

Fixeu-vos en els exemples següents, on s’indica la traducció produïda pel programa i, on és incorrecta (*), la traducció correcta:³

1. *La casa*: etxea
2. *La casa vermella*: etxe gorria
3. *La casa de la dona*: emakumearen etxea
4. *La casa vermella de la dona*: emakumearen etxe gorria
5. *La casa de la dona jove*: *emakumearen etxea gazte (corr.: emakume gaztearen etxea)
6. *La casa vermella de la dona jove*: *emakumearen etxe gorria gazte (corr.: emakume gaztearen etxe gorria)

Quines són les regles actives en cada frase? Per què es produeixen les traduccions incorrectes? Justifiqueu les vostres respostes.

Act. 7.4 Fixeu-vos en l’exemple de l’activitat 7. Les limitacions observades són les d’un sistema que no fa anàlisi sintàctica i que, per tant, és incapaç d’identificar sintagmes i manipular-los com a tals. La solució consisteix a fer l’anàlisi sintàctica i aplicar a les frases analitzades regles que en transformen la sintaxi. Per exemple, les noves regles per als sintagmes nominals de la pregunta 4 es podrien escriure⁴:

$$R_1: \text{tr}([_{SN} SN_1 [_{SP} \text{de} SN_2]]) = [_{SN} [_{SP} \text{tr}(SN_2) \text{gen}] \text{tr}(SN_1)]$$

$$R_2: \text{tr}([_{SN} \text{art} n]) = [_{SN} \text{tr}(n) \text{tr}(\text{art})]$$

³Vocabulari: emakume n, dona; etxe n, casa; gazte adj, jove; gorri adj, vermell, -a; -a art, el, la; -ren gen, marca de genitiu (de).

⁴Usant la representació clàssica amb claudàtors [...] en comptes d’arbres.

$$R_3: \text{tr}([_{SN} \text{ art } \mathbf{n} \text{ adj }]) = [_{SN} \text{tr}(\mathbf{n}) \text{tr}(\mathbf{adj}) \text{tr}(\mathbf{art})]$$

on $\text{tr}(\dots)$ representa “la traducció de ...” i els claudàtors $[_{X\dots}]$ indiquen l’estructura sintàctica produïda per l’analitzador o l’enviada al generador.

1. En el nou model, quines serien les traduccions automàtiques al basc de les frases incorrectes de l’exemple anterior?
2. Queden correctes?
3. Si en basc *fill n* es diu *seme*, quines serien les traduccions de “la casa vermella del fill jove de la dona” i “la casa del fill de la dona jove”?
4. Quines regles s’hi han aplicat? (fixeu-vos que és possible que en aquestes frases hàgeu d’aplicar més d’una volta la mateixa regla).

Act. 7.5 Indiqueu, si podeu, alguns problemes que encara no resol el “model 2” proposat en l’activitat 7 i que s’haurien de resoldre en un model més avançat de sistema de traducció automàtica. Podeu usar les oracions següents com a inspiració:

1. (en) I like roasted peanuts \rightarrow (ca) **Jo agrade cacaus torrats* (però: I buy roasted peanuts \rightarrow *Jo compre cacaus torrats*).
2. (ca) Porta les claus de la porta verda \rightarrow (en) ?*Bring the keys from the green door*
3. (en) Cleaning fluids can be toxic \rightarrow (ca) ?*Netejar líquids pot ser tòxic*
4. (en) The ship sank \rightarrow (ca) **El vaixell afonà* (però: We sank the ship \rightarrow [*Nosaltres*] *Vam afonar el vaixell*).

Act. 7.6 Els sistemes descrits fins ara realitzen la traducció en tres fases, que es corresponen amb tres mòduls o subprogrames ben definits: l’*anàlisi* o extracció de les característiques rellevants per a obtenir una representació que simplifica la traducció, la *transferència* on s’apliquen les transformacions necessàries per a obtenir una representació anàloga però referida a la llengua meta, i la *generació*, on es

genera un text meta a partir d’aquesta representació. Imagineu ara que tenim un sistema de traducció automàtica d’aquesta classe que tradueix en qualsevol direcció entre tres llengües.

1. Quants mòduls d’*anàlisi*, de *transferència* i de *generació* té?
2. Quants mòduls més hem d’escriure si volem afegir una quarta llengua a totes les direccions de traducció?
3. Quants experts monolingües bilingües necessitem per a construir aquests mòduls?
4. Com es podria evitar aquesta complexitat?

8 Avaluació de la traducció automàtica

Act. 8.1 Imagineu que esteu considerant la possibilitat d’usar un sistema de traducció automàtica en el vostre treball o assessorant una empresa que està considerant adoptar-ne un, i hi ha més d’una opció. Indiqueu com faríeu —quins criteris d’avaluació usaríeu— per a decidir quin sistema adopteu. Quins aspectes del sistema avaluaríeu?

Act. 8.2 A més de quan es vol decidir l’adopció d’un sistema, en quines altres situacions pot tenir interès l’avaluació de la traducció automàtica?

Act. 8.3 Considereu el concepte d’*avaluació predictiva* de la qualitat de les traduccions, que podem definir com segueix: un procés d’avaluació organitzat de tal manera que, amb les dades recollides, puguem predir raonablement el comportament d’un sistema de traducció automàtica en situacions noves respecte de les que s’han avaluat. Expliqueu com caldria organitzar un procés d’avaluació predictiva i quins coneixements hi són necessaris.

Act. 8.4 Molts esquemes clàssics d’avaluació es basen en comparar directament la qualitat de la traducció automàtica i de la traducció humana. Quins problemes pot tenir aquesta visió de l’avaluació?

9 Memòries de traducció

Act. 9.1 Moltes vegades, un equip de professionals de la traducció ha de traduir textos que són molt similars en naturalesa i contingut a molts altres que ja han estat traduïts anteriorment pel mateix equip. Fer-ho sense usar la informació ja existent en les traduccions prèvies és, a més de tediós per repetitiu, extremament poc eficient.⁵ S'anomena *memòries de traducció* al programari que permet aprofitar la informació existent en traduccions realitzades anteriorment per a assistir la persona que ha fer una nova traducció.

1. Com han d'estar els textos anteriorment traduïts per a poder ser útils? Cal alguna mena de preparació?
2. Què passa si, quan es va a traduir un fragment del document nou, no se'n troba un d'exactament igual en les traduccions prèvies? L'hem de traduir completament a mà?

Act. 9.2 Quina hauria de ser l'aparença d'un programa de memòries de traducció des del punt de vista de la persona usuària, perquè siga fàcil d'usar?

Act. 9.3 Indica quines conseqüències pot tenir per al treball d'un equip de professionals de la traducció la gestió i l'ús compartit amb altres equips de grans memòries de traducció.

3. Quins avantatges té l'organització de la informació en bases de dades respecte dels mètodes tradicionals (per ex., fitxes de cartolina)?

Act. 10.2 Entre les bases de dades esmentades en l'activitat anterior hàgeu esmentat les bases de dades que ens ocupen; és a dir, les *terminològiques* o *lèxiques*.

1. En què consisteixen?
2. Per a què serveixen?
3. Què contenen els registres d'aquestes bases de dades?
4. Com convé que estiguen organitzades?

10 Bases de dades lèxiques

Act. 10.1 Aquest bloc està dedicat a un tipus de *bases de dades*. Segur que heu sentit parlar de *bases de dades* més d'una vegada. Digueu:

1. Què és una *base de dades*? Com s'organitza la informació en una base de dades?
2. Per a què vol les bases de dades un traductor o una traductora? En què consisteixen les bases de dades que pot usar qui es dedica a la traducció?

⁵Tot i que, malauradament, encara es fa molt!