

MEMÒRIA FINAL DE LA LLICÈNCIA D'ESTUDIS

**ELABORACIÓ DE DICCIONARIS ESPECÍFICS PER AL
TRADUCTOR BABYLÓN ANGLÉS-CATALÀ**

**Autor: Jaume Casas Casanellas
Curs escolar: 2002-2003**

1. Introducció
 - 1.1 Antecedents del tema objecte del treball
 - 1.2 Explicació del tema
 - 1.3 Objectius que es pretenen assolir

2. Treball dut a terme
 - 2.1 Disseny del pla de treball
 - 2.2 Metodologia emprada
 - 2.3 Descripció dels recursos utilitzats

3. Resultats obtinguts
4. Conclusions
5. Relació dels materials continguts en els annexos
6. Bibliografia

1. Introducció

Els traductors de paraules informatitzats i en concret el Traductor Babylon Anglès-Català –l'actualització i millora del qual ha estat l'objectiu d'aquest treball- no permetia d'efectuar consultes en àmbits de terminologia específica. La seva utilitat pràctica quedava restringida als entorns "normals" de traducció i comprensió de l'anglès-català. El vocabulari que aquesta eina era capaç de traduir (uns 250.000 termes anglesos) no contemplava en cap cas els entorns de treball professionals específics ni els neologismes que s'han anat incorporant els darrers anys al català. Això en dificultava l'ús per part d'àmbits professionals diversos, però també per una branca de l'ensenyament tan important com són els cicles formatius.

Amb aquest treball s'ha pretès fer el Diccionari Traductor Anglès-Català avinent a un ventall d'usuaris molt més ampli. En concret s'han implementat diccionaris específics i de neologismes que el fan idoni per a tothom qui es mou dins del món tècnic-professional i estudiantil: professors, informàtics, biòlegs, metges, empresaris, alumnat de cicles formatius, alumnat en general, etc.

Els diccionaris que es pretenia implementar eren set, concretament els d'**informàtica i tecnologies de la informació, comptabilitat, neologismes, electrònica, anatomia, antropologia i lingüística**. Val a dir que s'han implementat tots set a més del diccionari específic de **genètica**.

El cos del treball consta aproximadament de 12.000 nous termes (uns 34.000 termes comptant les flexions de nombre i verbals). Aquests vuit diccionaris, a més del seu ús amb el programa Babylon, probablement podran en un futur aplicar-se a d'altres àmbits com ara a la confecció de diccionaris tècnics per als "eReaders" (llibres en format electrònic), com per exemple els de format LIT de Microsoft, PDF d'Adobe, HTML, etc. Això darrera possibilitat, però, no ha estat l'objectiu d'aquest treball, havent-se limitat a l'implementació dels diccionaris per al seu ús amb les eines de traducció i comprensió que ens aporta l'entorn Babylon.

1.1 Antecedents del tema objecte del treball

Les llengües minoritàries necessiten tots els mitjans possibles per defensar-se i contrarestar l'atropellament de les llengües majoritàries. Si un catalanoparlant treballa sobre un text anglès, aprendrà vocabulari anglès. Si, a més, obté la traducció de les paraules desconegudes en castellà, aprendrà nova terminologia castellana, com s'esdevé actualment d'una manera generalitzada. Hi haurà, doncs, una doble influència: l'anglesa per una banda i la castellana per l'altra. En canvi, si el catalanoparlant obté la traducció dels termes anglesos desconeguts en català, adquireix nou vocabulari català. És una manera de positivament el català. Aquestes eines tan senzilles d'usar fan que l'usuari s'involucri positivament en l'ús del català, i el vagi incorporant quasi inconscientment al seu bagatge intel·lectual. Lògicament, en tractar-se d'una eina que actua directament sobre la pròpia llengua (com tots els diccionaris) és molt efectiva. Simplificant més, quan un catalanoparlant ha de llegir un fullet d'instruccions en anglès en format informàtic i n'ha d'obtenir la correspondència en castellà, resta automàticament o "negativitzat" l'aprenentatge i ús de la llengua catalana. Si al llegir-la en anglès, compta amb una traducció del vocabulari desconegut en català, ja hi haurà quelcom que sumarà -positivarà- el coneixement del català. És tot això -que em penso que no he explicat pas d'una manera especialment entenedora- el que fa que aquest treball, que té aquest darrer efecte, és especialment important per a la millora i manteniment del català, en tota una colla d'àmbits, però molt especialment en el de l'ensenyament, que és el que ens interessa més ara i aquí.

Donat que el català ja disposa d'un diccionari bàsic anglès-català amb unes 300.000 entrades que treballa amb les eines de Babylon i que es pot aconseguir gratuïtament a l'adreça d'Internet <http://www.babylon.es>, l'única cosa que hem fet ha estat ampliar-lo bo i afegint-li diversos diccionaris específics –vuit en total- que fa que pugui ser emprat per un nombre major d'usuaris, especialment de professionals de les noves tecnologies i de l'electrònica, pels metges, pels estudiants de cicles formatius, etc.

1.2 Explicació del tema

Els traductors de paraules informatitzats actuals i en concret el Traductor Babylon Anglès-Català -l'actualització i millora del qual ha estat l'objectiu del treball fet- no permeten efectuar consultes en àmbits de terminologia específica. La seva utilitat pràctica queda restringida als entorns "normals" de traducció i comprensió de l'anglès-català. El vocabulari actual que és capaç de traduir aquesta eina (uns 250.000 termes anglesos) no contempla en cap cas els entorns de treball professionals específics ni els neologismes que s'han incorporat els darrers anys al català. Això en dificulta l'ús per part d'àmbits professionals diversos, però també per una branca de l'ensenyament tan important com poden ser els cicles formatius.

Amb aquest treball s'ha pretès fer el Diccionari Traductor Bàsic Anglès-Català avinent a un ventall d'usuaris molt més ampli. En concret s'han implementat diccionaris específics i de neologismes que el fan idoni per a tothom qui es mou dins del món tècnic-professional i estudiantil: professors, informàtics, metges, biòlegs, empresaris, alumnat de cicles formatius, alumnat en general, etc.

Els diccionaris anglès-català que s'han implementat han estat vuit en total, concretament els d'**informàtica i tecnologies de la informació, comptabilitat, neologismes, electrònica, anatomia, antropologia, lingüística i genètica**.

El diccionari de neologismes conté no només la traducció al català de cada terme anglès sinó també una definició i explicació exacta de l'àmbit específic on s'aplica. També una bona part del diccionari d'informàtica i tecnologies de la informació conté definicions aclaridores.

El cos del treball consta aproximadament de 12.000 nous termes (uns 34.000 termes comptant les flexions de nombre i verbals), sent el diccionari traductor de neologismes el més ampli, amb una mica més de 4.000 termes. Aquests diccionaris informatitzats podran aplicar-se probablement en el futur a d'altres àmbits com ara la construcció de diccionaris tècnics per als "eReaders" (llibres en format electrònic), com per exemple els de format LIT de Microsoft, PDF d'Adobe, HTML, etc.

El treball es presenta en un CD-ROM i també imprès. Dins del mateix CD-ROM s'inclou una versió de prova de l'eina Babylon (amb una validesa d'un mes des del moment de l'instal·lació) amb l'actual diccionari bàsic anglès-català a més dels vuit diccionaris específics que han estat l'objectiu d'aquest treball, de manera que pugui fer-se una valoració del treball de forma global i particularitzada, a més de permetre de fer-ne ús immediatament. També s'inclouen els vuit diccionaris en format PDF per si es vol fer alguna consulta en format electrònic.

1.3 Objectius que es pretenen assolir

El treball fet -implementació de diccionaris específics al traductor informatitzat Babylon Anglès-Català- té uns objectius molt precisos i de caire eminentment pràctic:

- a. Completar el Diccionari Traductor Bàsic Babylon Anglès-Català informatitzat amb diversos diccionaris específics i de neologismes catalans de manera que sigui una eina de treball idònia per a entorns més amplis del que ho és fins ara: el de l'electrònica, el de la informàtica i les tecnologies de la comunicació, el de la medicina, el comercial i financer, etc.
- b. Permetre la consulta ràpida -en línia o no- de qualsevol paraula anglesa (terminologia específica) desconeguda en qualsevol context informàtic en què ens la trobem -pàgines d'Internet, processadors de paraules (wordPerfect, Word, etc.), jocs d'ordinador, etc.- i obtenir-ne la traducció immediata al català.
- c. Comprensió ràpida de qualsevol text anglès, principalment quan es treballa en contextos terminològics molt específics (vegeu l'apartat a).

d. Increment de vocabulari anglès i millora del coneixement d'aquesta llengua, principalment pel que fa a entorns amb terminologia específica i neologismes.

e. Increment del coneixement de vocabulari català específic i neologismes catalans acceptats pel Termcat.

f. Actualitzar els serveis pràctics del traductor Babylon Anglès-Català per a la seva utilització en entorns professionals, per exemple, en molts àmbits de la formació professional actual.

g. Aplicació immediata a tots els nivells de l'ensenyament ja que en molts centres el Diccionari Bàsic Babylon ja s'està utilitzant.

2 Treball dut a terme

El treball ha consistit a millorar i arrodonir un producte que ja s'estava usant: el Diccionari Traductor Babylon Anglès-Català. Aquest diccionari general es va fer per aconseguir que el català disposés d'una eina informàtica que permetés una interacció instantània entre l'anglès i el català, fent que qualsevol persona amb uns coneixements elementals de l'anglès, com és el cas de la major part de l'alumnat actual de l'ESO, del batxillerat i dels cicles formatius, pogués accedir a tota classe de continguts que es troben en llengua anglesa a la xarxa Internet i a d'altres entorns com poden ser els jocs, els llibres, revistes i diaris electrònics, etc. d'una manera el més senzilla i pràctica possible: n'hi ha prou de prémer un botó del ratolí damunt d'una paraula anglesa desconeguda per obtenir-ne una traducció instantània al català. I també pot aconseguir-se el mateix resultat bo i entrant qualsevol paraula anglesa a l'interfície del Babylon.

Aquest treball ha implicat la confecció de vuit diccionaris específics. Concretament són els d'**informàtica i tecnologies de la informació**, el de **comptabilitat**, el de **neologismes**, el d'**electrònica**, el de **anatomia**, el d'**antropologia**, el de **lingüística** i el de **genètica**.

Els corpus d'aquests vuit diccionaris -un total aproximat de 12.000 termes (uns 34.000 termes si comptem les flexions de nombre i verbals)- poden usar-se tan per ampliar les possibilitats de l'actual diccionari de traducció Babylon anglès-català (objectiu primordial del treball), com per, en un futur i amb un treball previ sobre el corpus, estendre les possibilitats de traducció en d'altres entorns prou concrets com poden ser els de la lectura anglesa de llibres electrònics en formats codificats com són els del Microsoft Reader, els PDF d'Adobe, i d'altres actuals i futurs. El més important és comptar amb els corpus degudament informatitzats per poder-los aplicar als camps de programari on puguin ser més beneficiosos per al català.

2.1 Disseny del pla de treball

Ha resultat molt senzill seguir la planificació feta en començar el període de la llicència. L'ordre d'elaboració dels diccionaris ha estat el que vaig preveure a la memòria feta per obtenir la llicència d'estudis i que es pot veure a continuació:

Finals de gener del 2003:

- Diccionari de Comptabilitat
- Diccionari d'electrònica
- Diccionari d'Anatomia
- Diccionari d'Antropologia

Finals d'agost del 2003:

- Diccionari d'Informàtica i tècniques de comunicació

- Diccionari de lingüística
- Diccionari de Neologismes
- Diccionari de Genètica

Aquest darrer diccionari –diccionari de genètica- no constava a la memòria prèvia però he tingut temps d'implementar-lo. En total, doncs, han estat vuit els diccionaris implementats.

2.2 Metodologia emprada

Per a l'entrada de termes s'han usat dos mètodes diferents per bé que amb resultats idèntics. Per una banda, s'ha utilitzat una eina de l'entorn Babylon anomenada "Builder" l'ús de la qual és l'entrada de termes de manera que siguin reconeguts pel programa de gestió de diccionaris de Babylon. L'extensió dels fitxers confeccionats amb el programa "Builder" és "gls" si bé no és més que un fitxer de text que pot carregar-se amb qualsevol processador de textos (Word, wordPerfect, llibreta del Windows, etc.). La segona manera emprada per a l'entrada de dades ha estat simplement usant un processador de textos. A aquest fitxer s'han hagut d'introduir diverses marques per separar les entrades simples de les entrades corresponents a les diverses flexions de nom, verb, adjectiu, etc., així com marques especials per separar l'entrada anglesa de la traducció catalana de cada terme i també etiquetes HTML. Llavors, només ha calgut guardar els fitxers amb format "txt" i importar el fitxer de text amb el "Builder" de manera que es convertís de nou amb un fitxer "gls" reconeixible pel programa Babylon de gestió de diccionaris.

Els dos mètodes emprats han donat com a resultat fitxers que poden ser tractats amb qualsevol processador de textos i, per tant, manipulats amb macros per fer-ne un ús futur per a la confecció d'altres eines sobre la base dels corpus dels diferents diccionaris.

Una vegada obtinguts els fitxers "gls" dels diferents diccionaris específics reconeixibles pel motor de confecció de diccionaris de l'entorn Babylon, s'han transformat a fitxer de diccionari Babylon (extensió "bgl") que son autoinstal·lables amb un doble clic del ratolí (sempre i quan a l'ordinador s'hi trobi instal·lat el programa Babylon).

2.3 Descripció dels recursos utilitzats

Val a dir que he disposat de diverses obres de referència publicades pel Termcat, per d'altres departaments de la Generalitat, i per diverses universitats. També hi ha oberts diversos fronts d'investigació sobre la terminologia de la llengua tan a nivell català (CREL, IULA, etc.) com a nivell europeu, aquests darrers normalment duts a terme per les universitats dels diferents països, per bé que en les investigacions a nivell europeu en molt casos s'obvia el català, limitant-se a les llengües estatals (anglès, castellà, francès, etc.).

Sempre i quan ho han permès els drets d'autoria, s'ha usat tota la informació ja recopilada i simplement s'ha passat al format necessari de l'entorn Babylon. En cas d'usar eines molt concretes, com ara el diccionari de neologismes del Termcat, s'ha citat sempre la font original, tal i com s'acostuma a fer. En cap cas s'ha usat material amb drets d'autor vigents (per exemple, el material publicat per Enciclopèdia Catalana). Hi ha hagut un treball de recerca de terminologia específica basada molt especialment en textos específics anglesos i catalans, com per exemple, textos sobre plans comptables i financers, sobre anatomia, sobre les noves tecnologies, etc.

Quan als neologismes, a part del fons de neologismes del Termcat, s'han tingut en compte totes les publicacions de l'IULA (Institut Universitari de Lingüística Aplicada) de la Universitat Pompeu Fabra, molt especialment els *Papers de l'IULA* publicats anualment sobre els neologismes documentats a la premsa en català, a partir de l'any 1989 fins el 2001. Dos exemples d'aquests papers publicats, referits als anys 1989 i 1998, poden ser els següents:

Neologismes documentats a la premsa en català (1989)

IULA/INF002/96
Núm. de fitxes: 3.530
Diaris buidats: Avui (AV) i el Diari de Barcelona (DB)

Neologismes documentats a la premsa en català (1998)
IULA/INF026/99
Núm. de fitxes: 3.929
Diaris buidats: Avui (AV), El Punt (PD) i la revista El Temps (ET)

D'altres fonts d'investigació que s'han emprat en la confecció del treball són els productes, informatitzats o publicats en paper, del CREL (Centre de Referència en Enginyeria Lingüística) que aportaven informació i/o terminologia específica. El projecte *Wordnet o EuroWordNet* en anglès, català, basc i castellà ha estat de força ajuda.

El *WordNet* de la Universitat de Princeton (en anglès) ha estat objecte de consulta per a la recerca de sinònims de terminologia, a causa de la ingent quantitat d'accepcions i sinònims que conté.

Una de les diverses fonts del diccionari d'informàtica i tecnologies de la comunicació ha estat el *Recull de termes de l'OS/2*, una recopilació de l'equivalència anglesa-catalana del programari d'aquest sistema operatiu, de domini públic, molt especialment pel que fa a frases fetes referides al món del programari en català.

3. Resultats obtinguts

El resultat final que s'ha obtingut amb la incorporació de terminologia específica al Diccionari Traductor Babylon Anglès-Català no és altre que fer-lo avinent a un ventall més ampli d'usuaris. I, especialment en el camp de l'ensenyament-aprenentatge, actualitzar-lo per fer-ne una eina encara més idònia per treballar les llengües anglesa i catalana. Amb aquest treball de confecció de diccionaris específics es pretén dotar tots els centres d'ensenyament amb una eina optimitzada amb neologismes i elements específics que els alumnes puguin usar, a les aules d'informàtica i també a casa, i familiaritzar-se finalment amb l'ús d'un diccionari traductor informatitzat que els doni els significats en llengua catalana. També s'aconsegueix que l'alumnat de cicles formatius trobi més fàcil accedir a terminologia catalana pròpia d'allò que estan estudiant.

També podem esperar que molts usuaris, tècnico-professionals o no, dintre de la societat catalana que actualment estan usant aquesta eina en castellà, puguin trobar els avantatges de poder usar-la en català, gràcies probablement a la riquesa que comporta la incorporació de diccionaris específics a la mateixa.

I, per acabar, també s'espera que els corpus creats, tant el general com el dels diccionaris específics objectiu d'aquest treball, puguin usar-se per desenvolupar noves eines que facilitin l'ús del català dintre del món de les noves tecnologies, tant a l'escola com a la societat en general.

4. Conclusions

Una vegada acabat un treball tant ampli com aquest, només em ve a la ment la utilitat que pot tenir si es fa avinent la seva existència a la societat catalana. És tanta la informació existent a Internet, que no resulta fàcil trobar eines com aquestes sinó se'n sap res de la seva existència. Per aquesta raó, crec que seria molt raonable que aquests diccionaris específics, juntament amb el diccionari general, es possessin a disposició de tothom en una web catalana, com ara l'EDU365 o l'XTEC. També caldria fer-ne una ressenya a la premsa, si més no a l'especialitzada, i donar-la a conèixer de manera expressa als instituts i col·legis de Catalunya. Tal i com està ara, només la faran servir els qui ja estan familiaritzats en l'ús del programa Babylon al poder descarregar els diccionaris desde la web de Babylon on es trobaran una

vegada haurà estat entregat aquest treball. De moment, per a tots aquells a qui pugui interessar, els diccionaris es troben tots recopilats a la següent pàgina web: <http://www.sinera.org>.

5. Relació dels materials continguts en els annexos

- 1.- Diccionari imprès de Biologia
- 2.- Diccionari imprès de Tecnologies de la informació
- 3.- Diccionari imprès de Genètica
- 4.- Diccionari imprès de Comptabilitat
- 5.- Diccionari imprès d'Anatomia
- 6.- Diccionari imprès d'Antropologia
- 7.- Diccionari imprès d'Electrònica
- 8.- Diccionari imprès de Lingüística
- 9.- CD-ROM que inclou els 8 diccionaris en format BGL (executable de Babylon), en format PDF, en format GLS (es poden obrir com si fossin fitxers de text o tractar-se amb el programa Builder de Babylon que també s'inclou en el CD), memòria final en format PDF, resum de la memòria final en format PDF, breu explicació impresa i en PDF de com ha anat tot el treball de l'any sabàtic, el programa Babyloncat, versió 4, demo.
- 10.- Memòria final impresa
- 11.- Resum de la memòria final imprès.
- 12.- Explicació impresa de com ha anat tot el treball de la llicència.

6. Bibliografia

- 1.- Diccionari d'Anatomia. Fundació Barcelona. Termcat.
- 2.- Lèxic d'electrònica. Departament d'indústria i energia. Generalitat de Catalunya.
- 3.- Diccionari de Compabilitat. Fundació Barcelona. Termcat.
- 4.- Diccionari d'Antropologia. Fundació Barcelona. Termcat.
- 5.- Diccionari de biologia cel·lular. Fundació Barcelona. Termcat.
- 6.- Diccionari de Lingüística. Fundació Barcelona. Termcat.
- 7.- Diccionari d'Antropologia. Fundació Barcelona. Termcat.
- 8.- Diccionari de Neologismes del Termcat a Internet.
- 9.- Diccionari d'Internet. Diccionaris de l'Enciclopèdia.
- 10.- Societat de la informació. Noves tecnologies i Internet. Termcat.
- 11.- Diccionari de Genètica d'Internet. Recull de termes.
- 12.- Papers de l'IULA a Internet.