

Dacco: Diccionari Anglès-Català de Codi Obert

Projecte Dacco

<http://www.catalandictionary.org>

Carles Pina i Estany - carles@pina.cat

Linda Oxnard - catalandictionary@gmail.com

Abril 2008

Resum

Dacco, Diccionari Anglès-Català de Codi Obert, és un diccionari lliure. Dacco ja té una història consolidada, així com una sèrie de recursos. Disposa de més de 30.000 paraules amb diversa informació com ara exemples, notes d'ús, enllaços, fotos... El projecte Dacco ofereix diversos recursos lingüístics, interessants tant per estudiants de català com d'anglès. En aquest article es comentarà què ofereix Dacco, la seva història, les seves motivacions i la importància de disposar de recursos culturals de lliure accés.

1 Introducció

Dacco és un acrònim de “Diccionari Anglès-Català de Codi Obert”. Va ser el primer diccionari lliure de català-altres llengües. Dacco va ser oficialment alliberat el 2003, tot i que va començar a nivell personal el 2001. Des de llavors ha anat evolucionant dia a dia amb empena i constància.

Actualment consta d'unes 30.000 entrades en els dos idiomes, exemples, notes d'ús i d'altres recursos paral·lels com són vídeos de diferents dialectes anglesos i catalans, programes per accedir al diccionari, fitxers PDF i molt més.

Dacco ja és una eina molt bona per cobrir les necessitats del dia a dia d'usuaris tant angloparlants com catalanoparlants que necessiten d'un diccionari bilingüe o dels altres recursos que Dacco ofereix.

En aquest article es parlarà de sobre què és Dacco, l'història d'aquest i com es treballa per aconseguir un bon resultat. A nivell tècnic es parlarà de com està organitzat Dacco i com es generen les diferents sortides.

1.1 La importància del català a Europa

El català a vegades és considerat llengua de segona, fins i tot per alguns catalans. Un dels motius és que no hi ha cap estat que doni suport a la

llengua de forma oficial, exceptuant Andorra que és un principat força petit.

No obstant això cal recordar que el català, amb uns 5,5 o 6 milions de parlants, és parlat aproximadament pel 2% de tot Europa, i té tants parlants com ara el finlandès, el suec o el noruec que a vegades es consideren més prestigioses.

En l'àmbit de les noves tecnologies, un altre tret diferencial del català i de la cultura catalana és, segurament, el fet d'haver aconseguit el primer domini cultural: el .cat.

2 Què és Dacco

2.1 El diccionari

Dacco, tal i com s'ha comentat anteriorment, és un acrònim de “Diccionari Anglès-Català de Codi Obert”. Dacco està orientat a usuaris de qualsevol dels dos idiomes.

Dacco té quatre diccionaris:

- Diccionari anglès-català per usuaris catalanoparlants
- Diccionari català-anglès per usuaris catalanoparlants
- Diccionari anglès-català per usuaris anglòfons
- Diccionari català-anglès per usuaris anglòfons

Aquesta diferenciació és molt important ja que fins l'aparició de Dacco hi havia molt bons diccionaris català-anglès/anglès-català per catalanoparlants degut a la gran demanda però, en canvi, no hi havia gaires bons diccionaris català-anglès/anglès-català per anglòfons. I, tot i que comparativament hi ha pocs usuaris, sí que hi ha estudiants de català anglesos i americans que es trobaven davant d'una dificultat més a l'hora d'aprendre l'idioma.

A part de la poca qualitat dels diccionaris, també hi havia el problema d'accés als diccionaris: els diccionaris anglès-català més coneguts per catalanoparlants són difícils d'aconseguir fora de Catalunya. I altres diccionaris, com són el Routledge[2] català-anglès són massa bàsics i de difícil accés.

La base de Dacco són uns fitxers en format XML amb totes les entrades del diccionari, a partir dels quals es generen els fitxers PDF que són utilitzats per altres eines.

2.2 Mètodes de consulta

Per tal de facilitar al màxim l'accés a les dades del diccionari, des del projecte Dacco s'han preparat diversos mètodes de consulta com són:

- Formulari via Web a l'adreça <http://www.catalandictionary.org>.

- Motor de cerca pel Firefox i l'IE.
- iGoogle gadget. Un gran nombre d'usuaris avui en dia utilitzen aquest enginy. Té més usuaris que altres gadgets de mitjans de comunicació importants catalans.
- qdacco. Un programa que utilitza els fitxers XML sense necessitat de connexió i multiplataforma.

2.3 Altres eines

Dacco no és només un diccionari, sinó que també recull altres eines lingüístiques com ara:

- Conjugador de verbs. Molts usuaris utilitzen aquest conjugador de verbs catalans.
- Una selecció de vídeos per escoltar dialectes anglesos i catalans transcrits.
- CatFon. Un programa que fa la transcripció fonètica de qualsevol paraula o frase catalana.
- Phrasal verbs. Un llistat de més de 2.000 phrasal verbs, properament sincronitzat amb el diccionari mateix.

3 Història de Dacco

Linda Oxnard estudiava filologia catalana a la Universitat de Sheffield, Anglaterra quan durant els seus estudis de doctorat, mitjançant unes enquestes va intentar esbrinar quins eren els principals problemes dels estudiants anglesos a l'hora d'aprendre català. Un dels problemes que va detectar va ser la falta de bons diccionaris en català orientats a anglòfons. Els estudiants comentaven que els diccionaris bilingües eren poc útils i força estudiants treballaven amb diccionaris anglès-castellà o anglès-francès per després traduir cap al català. Es queixaven de la mala qualitat dels seus materials per l'estudi, sobretot comparat amb altres llengües. Hi havia bastants diccionaris monolingües pels estudiants d'anglès catalanoparlants (p. ex. The Oxford Advanced Learner's Dictionary[5], the Cambridge Advanced Learners Dictionary[4] and the Longman Dictionary of Contemporary English[6]), però no hi havia cap diccionari monolingüe per estudiants de català anglòfons. Els estudiants primer treballaven amb diccionaris bilingües molt bàsics i després amb diccionaris monolingües (com ara el DIEC[3]), però res cobria les seves necessitats durant l'etapa intermitja. La Linda va tractar d'omplir aquesta necessitat amb Dacco incloent-hi els exemples i notes d'ús que trobaven a faltar a diccionaris com el Routledge.

És per això que el 2001 va començar a crear el diccionari. El 2003 va alliberar la primera versió: amb unes 500 entrades per llengua, unes quantes però realment insuficient per un ús real del diccionari.

El seu marit James Macgill, qui no té coneixements de català però sí d'informàtica i de programari lliure, la va ajudar des del primer dia. Va ser l'any 2003 quan es va alliberar la primera versió de forma pública i de mica a mica més gent s'hi va afegir.

La gent que col·labora a Dacco és molt heterogènia: gent catalana o angloparlant, d'entre 20 i 80 anys; de diferents dialectes, per exemple angloparlants de Nova Zelanda, Austràlia, Amèrica, Canadà o Anglaterra; diferents punts de vista... és un diccionari plural que intenta abastar tota la realitat lingüística.

També cal destacar que el diccionari és descriptiu i no normatiu. Intenta incloure les paraules que els usuaris necessiten, és per això que una de les fonts són les paraules d'ús cotidià de les dues llengües. Es construeix, sobretot, un diccionari útil.

Igual que molt programari lliure, el projecte no cal que tingui un rendiment econòmic de tal forma que es poden dedicar esforços a aspectes que econòmicament no serien viables però que són útils per els usuaris. Dacco té com a objectiu oferir un diccionari lliure i útil als usuaris.

4 Origen de les entrades

Les entrades no es copien de cap altre diccionari o lloc on ja puguin existir. És justament la gent que fa servir el diccionari que, conscient (enviant voluntàriament) o inconscientment (per exemple buscant les paraules a la Web) aporten les noves entrades i que són recollides i discutides amb els col·laboradors, o els mateixos col·laboradors envien noves entrades.

Sí que es van incorporar les entrades d'alguns dels repertoris terminològics oferts per Termcat com a "terminologia oberta". Tanmateix, des del projecte Dacco, es va notificar a Termcat el fet que volia incorporar els seus repertoris i confirmar que no hi havia cap problema. Els de Termcat estan molt oberts a qualsevol comentari o suggerència i, molt amablement, ens van confirmar que no hi havia cap problema.

La introducció d'una paraula al diccionari consta de tres passos:

- Trobar una paraula no existent a Dacco. Aquesta paraula és enviada a la llista de correu, ja sigui utilitzant el formulari per enviar noves paraules de la Web, utilitzant el llistat de paraules buscades i no trobades, el sistema de seguiment d'incidències JIRA [1] o altres fonts.
- A la llista de correu es parla informalment del significat de la paraula i de quina n'és la traducció, notes, exemples, dialecte i altres informacions útils. Alhora que es té en compte que hi ha diferents dialectes,

edats, parlants,...

- Finalment la paraula és afegida als fitxers XML.

Tot el procés és transparent i fàcilment comprensible. Hi ha un fàcil accés a les entrades pendents d'afegir així com a les entrades acabades d'afegir. D'aquesta manera s'assegura que hi ha molta visibilitat als canvis. Això es fa mitjançant el sistema JIRA, un sistema de gestió d'incidències que s'ajusta molt a les necessitats de Dacco. D'aquesta manera, fins i tot sense apuntar-se a cap llista de correu es pot seguir l'evolució del diccionari. Es fa un diccionari lliure i a la vegada es potencia la cultura lliure.

5 Llicències de Dacco

Dacco com a projecte és lliure, però no totes les parts tenen la mateixa llicència. Les parts principals són els fitxers XML, els fitxers PDF i qdacco:

5.1 Fitxers XML

Els fitxers XML estan alliberats sota la llicència LGPL 2.0. Això permet que tothom en pugui fer ús, dins de projectes lliures o no lliures. Es vol potenciar l'ús d'eines amb Dacco, així com potenciar la cultura catalana en tants llocs com sigui possible.

5.2 Fitxers PDF

Els fitxers PDF tenen llicència Creative Commons Atribució-Compartir Igual. Creative Commons és una llicència molt adequada per fitxers que no són codi font, i la variant utilitzada permet fer qualsevol cosa sempre i quan hi hagi atribució al projecte original i es comparteixi amb la mateixa llicència o amb una compatible.

5.3 qdacco

Aquesta aplicació està alliberada amb llicència GPL 3, la versió més recent de GPL i garanteix que sempre seguirà sent lliure alhora que es pot aprofitar el codi per altres usos, seguir-lo millorant, i en definitiva assegurar-ne la llibertat.

6 Característiques especials de Dacco

Dacco té una sèrie de característiques que el diferencien d'altres diccionaris. Aquestes són la seva llibertat, el nombre d'entrades, els quatre diccionaris i la informació addicional que disposa a les seves entrades.

6.1 Llibertat

Dacco és un dels primers diccionaris bilingües lliures. Això el fa especialment atractiu. És possible utilitzar-lo en qualsevol àmbit, hi ha grans possibilitats d'adaptació, no hi ha gairebé restriccions i s'està sempre obert a modificacions i millores. Igual que la llibertat en el programari que ja hi estem acostumats però aplicat a un diccionari.

6.2 Nombre d'entrades

El gran nombre d'entrades de Dacco (unes 30.000 en els dos idiomes) el fan també molt atractiu. És un diccionari ja madur i complet, que es pot utilitzar en el dia a dia sense problemes, cosa que no treu que segueixi creixent i millorant.

6.3 Quatre diccionaris

Dacco té quatre diccionaris: adaptacions per la gent catalanoparlant o bé per anglòfons de l'anglès-català i català-anglès. També té altres diccionaris més simplificats per gent que tot just està aprenent les bases dels idiomes.

6.4 Altre informació, camps semàntics

S'intenta aprofitar al màxim les capacitats del medi on es troba el diccionari, el medi digital: ordinadors, normalment connectats a Internet. Així doncs hi ha enllaços a fotos, enllaços a vídeos, exemples, notes d'ús o altres recursos per ajudar a entendre el significat de les paraules. No hi ha les limitacions d'espai que hi podria haver a altres medis.

Per exemple, traduir la paraula “ensaïmada” és difícil (podria ser “pastís mallorquí en forma d'espiral”). Si tenim una foto ho aclareix força més.

Cada paraula també conté, en els fitxers XML, la freqüència que apareix a Google i el camp semàntic (per exemple, “poma” està al camp semàntic “fruita”, el qual està al camp semàntic “menjar”). Això permet construir petits diccionaris o vocabularis per temes específics. S'ha vist que aquests llistats de paraules agrupades per temes són populars als estudiants de les dues llengües, i, si es miren els logs de Dacco, fins i tot es veuen estudiants no catalans que aprenen anglès que utilitzen aquests llistats de paraules com a eina d'aprenentatge, construint el seu propi diccionari bilingüe a partir de les llistes de Dacco com a punt d'entrada.

El fet de marcar les paraules dins un camp semàntic permet incorporar llistes de nou vocabulari específiques de temes en concret i treballar amb elles amb les eines de Dacco, per exportar-les temps després cap a altres llistats. Per exemple, s'està en procés d'incorporar una extensa llista d'ocells en Anglès-Català-Llatí. Ja que la llista és treballada i millorada dins de Dacco,

les dades poden ser exportades un altre cop cap als projectes que havien originat les llistes.

7 Futur

L'esquema XML utilitzat a Dacco ha anat evolucionant al llarg del temps per incorporar les noves característiques que eren necessàries. En els propers mesos, s'espera fer un depurat dels XML i produir un esquema XSD formal. D'aquesta manera es facilitarà la creació d'altres eines que importin i exportin dades i oferir un marc de treball a altres diccionaris bilingües. També es dissenya una base de dades que contindrà les entrades per permetre, fins i tot, generar diferents fitxers XML en funció de com han de ser fets servir.

Una altra part del treball de refactorització és facilitar als usuaris finals afegir noves entrades, canviar les entrades existents i en general millorar el diccionari. Igualment, els canvis seran accessibles més ràpid tot i mantenir el procés de revisió per seguir tenint una alta qualitat i tenir canvis correctes.

8 qdacco

Una de les eines de Dacco és qdacco: un programa que consulta directament els fitxers XML de Dacco per tal d'accedir al diccionari sense connexió. Proporciona una interfície d'usuari més còmode, per a consultes puntuals, que els fitxers PDF. També té algunes característiques que el diferencien d'altres programes, com ara:

- És un programa multiplataforma. Funciona a plataformes Unix (entre elles Linux) i Microsoft Windows.
- Està completament integrat amb els fitxers de Dacco i utilitza tots els seus recursos (exemples, notes d'ús, enllaços a pàgines Web...
- Permet enviar paraules que falten o suggeriments al projecte Dacco per a la millora d'aquest.
- Integració amb el sintetitzador de veu Festival: pot llegir-nos paraules per aprendre a pronunciar-les.
- La versió principal és gràfica però avui en dia també inclou una versió text: es pot utilitzar sense entorn gràfic o bé integrar-lo amb altres sistemes de forma molt fàcil.
- Està disponible als repositoris de Debian.
- Està llicenciat amb GPL 3.

L'idea inicial de qdacco era ser el programa de referència de Dacco, per donar una empenta més al projecte en sí i poder aprofitar tots els recursos que ofereix.

9 Conclusions

Dacco és un diccionari lliure útil amb més de set anys que ofereix una gran quantitat d'eines i informació a molta gent.

La cultura lliure es troba en un estat una mica més prematur que el programari lliure, que té més acceptació i història. Dins la cultura lliure hi ha projectes insígnies com la Wikipedia, Wiktionary, Creative Science... Tot i així encara hi ha moltes àrees on cal créixer, una d'aquestes àrees són els diccionaris oberts, ja siguin bilingües o monolingües. Esperem que Dacco estigui obrint camí i que properament hi hagi altres diccionaris oberts, ja sigui aprofitant l'experiència de Dacco o part de la seva infraestructura.

També cal destacar que hi ha almenys dos projectes de diccionaris lliures catalans, que són:

- GPL Deutsch-Katalanisches Wörterbuch: <http://www.aldeaglobal.net/diccionari/index.php> . És un diccionari català-alemany amb llicència GPL.
- Wiktionary (<http://ca.wiktionary.org/wiki/Portada>, http://en.wiktionary.org/wiki/Main_Page) a diferència de Dacco està més focalitzat en definicions que en traduccions. Algunes entrades tenen traduccions cap a altres llengües però no és el seu objectiu principal ni el seu millor valor. Una de les claus que diferencien Dacco i Wiktionary és com es presenten les traduccions. A Wiktionary, les traduccions són enllaços a altres termes i definicions en altres llengües -i d'aquesta manera es promou que l'usuari explori d'una pàgina a una altra pàgina. Per l'altre costat, Dacco, se centra en la creació d'entrades autònomes amb unes extensives notes d'ús en la mateixa llengüa que l'usuari, així com exemples i frases clau. És possible que en un futur hi hagi col·laboració entre els projectes adaptant entrades.

10 Llicència de l'article

Aquest article es distribueix sota una llicència Creative Commons Reconeixement-Compartir amb la mateixa llicència 2.5 Espanya. Veieu <http://creativecommons.org/licenses/by-sa/2.5/es/deed.ca> per més informació.

Referències

- [1] <http://www.atlassian.com/software/jira/>.
- [2] *Catalan Dictionary: English-Catalan, Catalan-English*. Routledge, London, UK, 1994.
- [3] *Diccionari de la Llengua Catalana de l'Institut d'Estudis Catalans*. Enciclopèdia Catalana / Edicions 3 i 4 / Edicions 62 / Editorial Moll / Publicacions de l'Abadia de Montserrat, 1995.
- [4] *Cambridge Advanced Learner's Dictionary*. Cambridge University Press, Cambridge, UK, 2005.
- [5] *Oxford Advanced Learner's Dictionary*. Oxford University Press, Oxford, UK, 2005.
- [6] *Longman Dictionary of Contemporary English*. Pearson ESL, London, UK, 2006.