

Els bancs de dades terminològiques. Organització per a la llengua catalana

M.T.Cabré
TERMCAT, Centre de Terminologia

Els bancs de dades terminològiques. Organització per a la llengua catalana.

M. Teresa Cabré i Castellví

TERMCAT, Centre de Terminologia

RESUM:

Els bancs terminològics van començar a aparèixer durant la dècada dels anys seixanta. En aquest moment inicial copiaven el model dels bancs documentals. Avui dia el model que s'ha configurat és específic per a la terminologia i al servei de les necessitats dels usuaris de termes.

Encara que puguem parlar de diversos tipus de bancs de termes en funció de criteris diferents, com l'objectiu que pretenen assolir o la seva actitud respecte de la llengua, la descripció de qualsevol banc terminològic segueix necessàriament tres elements: l'entrada, l'emmagatzematge i la sortida. El banc de dades que s'està muntant al TERMCAT, Centre de Terminologia per a la llengua catalana, es caracteritza per una entrada estructurada automàticament procedent de dues fonts (fitxes manuals i obres publicades), per un emmagatzematge consistent en tres fitxers, dos de treball i un de difusió i per una sortida de format divers segons els tipus de consulta i les formes d'accés.

Per a l'organització del banc s'han tingut en compte les condicions suggerides per Infoterm a través de la xarxa TERMNET, a fi de possibilitar en un futur l'intercanvi d'informació.

PARAULES CLAU: banc de dades terminològiques, terme, estandardització, normalització, classificació.

Els bancs de dades terminològiques. Organització per a la llengua catalana

1. Els bancs de termes

Un banc de dades terminològiques és, segons Rondeau, un diccionari automatitzat de nocions i denominacions que permet donar respostes ràpides a demandes terminològiques puntuals o sistemàtiques.

L'origen dels bancs de termes⁽¹⁾ hem de buscar-lo en les necessitats plantejades per la traducció automàtica. Per traduir cal un gran diccionari d'equivalències. Així va néixer als 60 el primer banc de termes, el DICAUTOM (Dictionnaire automatisé), després EURODICAUTOM, de la CEE.

La dècada dels seixanta es caracteritza per una puixant demanda de termes estandarditzats i una pobra resposta de mitjans. Les eines amb què es treballa en terminologia són arcaiques i clarament insuficients. Els fitxers manuals només permeten una lectura lineal de la informació. Els progressos científics i tècnics i l'increment del comerç internacional reclamen eines més sofisticades i especialitzades. La informàtica comença a sortir del terreny de les dades numèriques i a entrar en el de la llengua. Els bancs de termes sorgeixen precisament pel creuament d'aquests dos factors: la proliferació de termes i el desenrotllament de la informàtica.

Així neixen durant els anys seixanta el TEAM de l'empresa Siemens a Múnic, fonamentalment adreçat a la traducció; el LEXIS de l'Oficina Federal de Llengües del govern de la RFA i, l'any 1970, el primer fitxer automatitzat del Banc de Terminologia de Quebec.

Durant la dècada dels setanta es creen més bancs terminològics a diversos països: UNIIKI a Moscou, NORMATERM a França, TERMDOK a Suècia, DANTERM a Dinamarca, el banc de terminologia de Canadà (BTC) i el BTUSB de la Universitat Simón Bolívar de Caracas. Més tard, l'any 1981, el banc de l'UMIST a Manchester. Tots aquests bancs tenen objectius precisos i diversos i abasts diferents.⁽²⁾

Els primers bancs terminològics representen un tanteig en aquest camp. La primera generació reflecteix la vacil·lació sobre quin havia de ser el format més adequat. Així doncs, els bancs de termes copien el model dels bancs documentals i són concebuts com a magatzems de paraules. No cal dir que un banc lexicogràfic i un banc terminològic són ben diferents entre si; són diferents pels objectius que persegueixen, pels usuaris que els consulten, per les dades que contenen i fins i tot per l'ordenació que privilegien. Un banc lexicogràfic és un magatzem de paraules, normalment ordenades alfabèticament, amb un conjunt d'informació lingüística relativament simple i al servei dels lexicògrafs. Un banc de termes és un muntatge lexicoc conceptual, d'assignacions terme-concepte, ordenades temàticament, amb

especificacions variades i al servei de diverses professions: traductors, intèrprets, redactors tècnics, locutors, docents, científics, tècnics, comerciants, etc.

El format dels bancs terminològics actuals és cada cop més afinat i precís i pensat específicament per a aquest tipus de bancs. L'orientació que es perfila actualment és la de muntar bancs especialitzats, descentralitzant la informació, amb accessos ràpids i directes, sense necessitat que un terminòleg faci d'intermediari.

2. Organització específica dels bancs de termes

El treball terminològic consisteix en un conjunt d'operacions que poden ser inventariades del mode següent⁽³⁾:

1. Recull i enregistrament dels termes assignats a conceptes dins d'una àrea o camp específics.
2. Anàlisi, proposta de creació i estandardització d'un sistema conceptual per a cada camp o àrea.
3. Anàlisi, discussió, proposta i estandardització d'una denominació per a un concepte o d'un concepte per a una denominació.
4. Descripció dels conceptes, a través de definicions, i estandardització de les definicions.
5. Enregistrament de les dades terminològiques: termes, definicions o explicacions, contextos, conceptes, relacions, equivalències, fonts, valor sociolingüístic, etc.

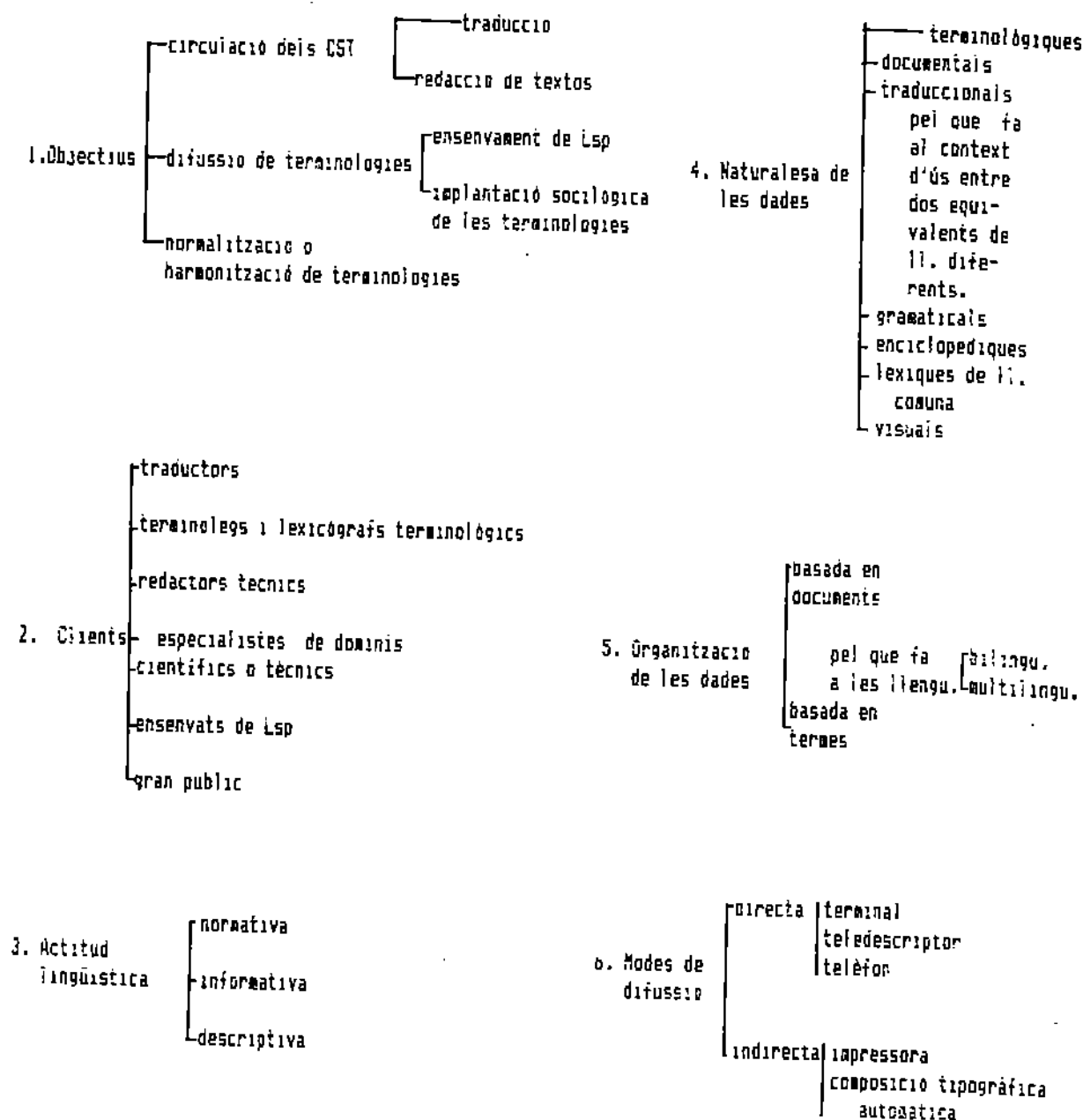
El conjunt de totes aquestes operacions pot desembocar en la constitució d'un banc de dades informatitzat consultable de forma més àgil i amb capacitat de poder ser posat al dia continuament. El muntatge d'un banc de termes ha de ser programat en tres blocs o sectors

1. L'entrada de la informació
2. L'emmagatzematge de les dades
3. La sortida de la informació

Fons ----> entrada--emmagatzematge--sortida -->usuaris

3. Tipus de bancs de termes

També en aquest terreny ha estat Rondeau qui ha fet la classificació més completa dels tipus de bancs terminològics a partir de criteris diferents: els objectius que persegueixen, els usuaris que se'n serveixen, l'actitud lingüística del banc, els tipus de dades contingudes, el nombre de llengües que inclou, l'organització de les dades i les vies de difusió de la informació.



Felber (1982) fa distinció de dos tipus fonamentals de bancs segons l'estructuració de la informació i els elements de partida:

1. Els bancs-diccionari
2. Els bancs-vocabulari

Els bancs-diccionari, al servei bàsicament dels traductors, s'organitzen a l'entorn d'un terme, codificat en la seva àrea respectiva. Contenen equivalències d'aquest terme en altres llengües.

Els bancs-vocabulari parteixen del concepte, codificat a partir d'un thesaurus, i contenen una sèrie de dades terminològiques de cada concepte: relacions amb d'altres conceptes veïns, definició i/o explicació, contextos, equivalències, etc.

La majoria de bancs existents, però, són mixtos, contenen un vocabulari i un diccionari. El vocabulari permet accedir a la informació a partir de la denominació, la qual és especificada conceptualment amb indicació de les equivalències en d'altres llengües.

4. Alguns dels problemes dels bancs terminològics. La cooperació internacional

Els bancs de dades terminològiques que funcionen actualment presenten una diversitat molt difícil de vèncer, malgrat els esforços d'INFOTERM per harmonitzar-los de forma que esdevinguin compatibles o comparables.

Cada banc té objectius propis i organitza la informació en funció d'aquests objectius de la forma més còmoda i ràpida possible.

Una mostra clara de la diversitat és la classificació conceptual dels termes. La gran majoria de bancs usen un thesaurus o una classificació pròpia i són molt pocs els que o bé no classifiquen conceptualment els termes o bé usen una classificació universal, encara que adaptada a les seves necessitats idiosincràtiques.⁽⁴⁾ Aquesta dispersió provoca un problema molt greu de fiabilitat de les equivalències. Com es pot constatar que un terme d'una llengua equival a d'altres termes en d'altres llengües? Sembla evident que només garantint que es refereixen a un mateix concepte.

Per unificar els plans i els treballs i fer-los esdevenir comparables, Infoterm impulsa des del 1976 TermNet, xarxa internacional d'organismes que treballen en Terminologia. En l'actualitat estan proposats tres programes de treball⁽⁵⁾:

1. Programa 1. Desenvolupament de la Teoria terminològica
2. Programa 2. Cooperació per posar a purt les terminologies i enregistrar-les en suport automatitzat
3. Programa 3. Cooperació per recollir, enregistrar, tractar i difondre dades i informacions terminològiques.

El programa 2 de TermNet és el que hauria de donar sortida als problemes de compatibilitat. En aquests moments s'està treballant en la creació d'un format estàndard per enregistrar les dades en un suport automàtic del qual ISO ja ha tret un document (DIS 6156) titulat "Format d'un échange sur bande magnétique des données terminologiques ou lexicographiques".

5. El banc de dades del TERMCAT

El TERMCAT és el Centre de coordinació del treball terminològic per a la llengua catalana. Va ser creat el maig de 1985 pel Departament de Cultura de la Generalitat de Catalunya i l'Institut d'Estudis Catalans. Va posar-se a treballar sistemàticament el novembre de 1985(6).

El conveni de creació li assigna la funció de crear un banc de dades terminològiques consultable per diverses vies. El banc que s'està posant en marxa ha de servir fonamentalment els interessos de la normalització lingüística i abordar-ne només el seu vessant terminològic.

Els criteris de Rondeau sobre classificació de bands exposat anteriorment són difícils d'adequar als objectius del banc del TERMCAT. Podríem dir que el banc del TERMCAT serveix uns interessos d'una banda més generals i d'altra més específics per al cas català. Així, per exemple, encara que puguem afirmar que el seu objectiu bàsic és la normalització terminològica, també posa en circulació terminologies diverses; encara que els seus usuaris seran preferentment els especialistes, traductors, redactors, tècnics i docents especialitzats, també persones interessades per la llengua en general podran consultar-lo i segurament ho faran. Les dades que contindrà no seran totes normatives; també n'hi haurà de no normalitzades, convenientment valorades i bloquejades per a la difusió entre el gran públic. No més inclourà dades terminològiques, adequadament referenciades segons els codis establerts pel banc documental del TERMCAT en vies de constitució. No contindrà dades de la llengua general, ni paraules ni usos gramaticals de les paraules de la llengua comuna. Serà plurilingüe: català, castellà, anglès, francès, alemany i italià. Podrà donar informació directament per pantalla, si s'hi està connectat, i indirectament per telèfon o per escrit. Serà la base de diverses edicions, fetes directament per impressora o editades per fotocomposició, encarregades de difondre la informació sobre els termes d'una matèria, nous en la llengua, rebutjables, aprovats, etc.

Per analitzar més específicament les característiques del banc del TERMCAT ens basarem en els tres aspectes subratllats abans per Rondeau: l'entrada, l'emmagatzematge i la sortida.

L'entrada.- Està organitzada a partir de dues fonts diferents:

1. Fitxes terminològiques manuals elaborades per equips de recerca d'una àrea determinada. Són entrades pels mateixos autors. (7)
2. Obres ja editades i convenientment preparades. Són entrades pels terminalistes del TERMCAT o dels centres associats per a l'entrada. (8)

El treball de preparació de tot el material consisteix a:

1. Revisió prèvia de les fitxes manuals que han d'entrar en màquina. Control d'adequació dels termes. Control de compleció de la fitxa. Pertinència de les informacions.
2. Revisió prèvia de les obres editades. Selecció d'informació. Distribució de la informació. Reajustaments terminològics de les entrades i subentrades. Eliminació de la informació no pertinent. Compleció de la informació obligatòria.

L'emmagatzematge.- Les dades entrades a l'ordinador són organitzades en quatre fitxers; un de principal que conté els termes i totes les seves informacions, i tres de secundaris, que inclouen dades sobre els autors del treball terminològic, els organismes terminològics i les àrees de classificació temàtica (9). Els documents i obres de referència seran codificats al banc documental.

El fitxer principal està organitzat a partir del terme i preveu les informacions següents per a cada terme (10):

Situació terminològica (provisional/definitiva)
Font
Valor gramatical
Àrea i subàrees
Marca de ponderació
Il·lustració
Creació de la fitxa: codi i data
Definició. Font. Situació
Context. Font. Situació
Equivalències. Situació
Remissions
Descriptors semàntics
Nota

No totes les informacions previstes són considerades obligatòries. Aquelles de què no es pot prescindir i sense les quals l'ordinador no acceptarà fitxa són les següents:

- terme
- situació terminològica
- font
- valor gramatical
- àrea temàtica
- marca de ponderació
- dades sobre la creació de la fitxa

La resta d'informacions pot no ser-hi. Això no obstant, una fitxa serà considerada completa si, a més de les informacions obligatòries, inclou també la definició i les equivalències.

Hi ha altres graus d'obligatorietat de les informacions. Així, per exemple, si hi ha definició és obligatòria la font i la situació provisional o definitiva de la definició; si hi ha àrea temàtica, la seva situació; i si hi ha context, la font i la situació del context. Encara més, si el terme és considerat complementari d'un altre terme la fitxa no serà acceptada sense la remissió explícita en aquest terme principal i sense el concepte de la remissió: sinònim, abreviatura, variant gràfica, sigla o nomenclatura.

Les fitxes definitives o completes seran aptes per a la difusió. Les provisionals o incompletes formaran part de fitxers de treball.

La sortida.- El fitxer de difusió serà apte per ser consultat des de l'exterior i usant formes diverses d'accés: pantalla, telèfon, impressió, edició, etc.

El sistema IBM/38 permet d'accedir a la informació a partir de qualsevol de les informacions incloses en la fitxa automatitzada i carregar la consulta amb totes les dades que es vulgui. Només a mode d'exemple presentarem algunes de les possibilitats de sortida:

1. Qüestions puntuals

- Informació sobre si un terme és al banc
- Nombre de termes del banc
- Nombre de termes d'àrees específiques
- Situació sociolingüística d'un terme
- Sinònims d'un terme
- Nomenclatura científica d'un terme
- Definició d'un terme
- Equivalències en altres llengües
- Fitxa completa d'un terme, etc.

2. Llistes terminològiques

Termes d'una àrea o subàrea

Termes definitius

Termes difosos pel TERMCAT

Llistes bilingües, trilingües, plurilingües, des de qualsevol

llengua de partida i seguint qualsevol ordre

Termes als quals falta informació complementària

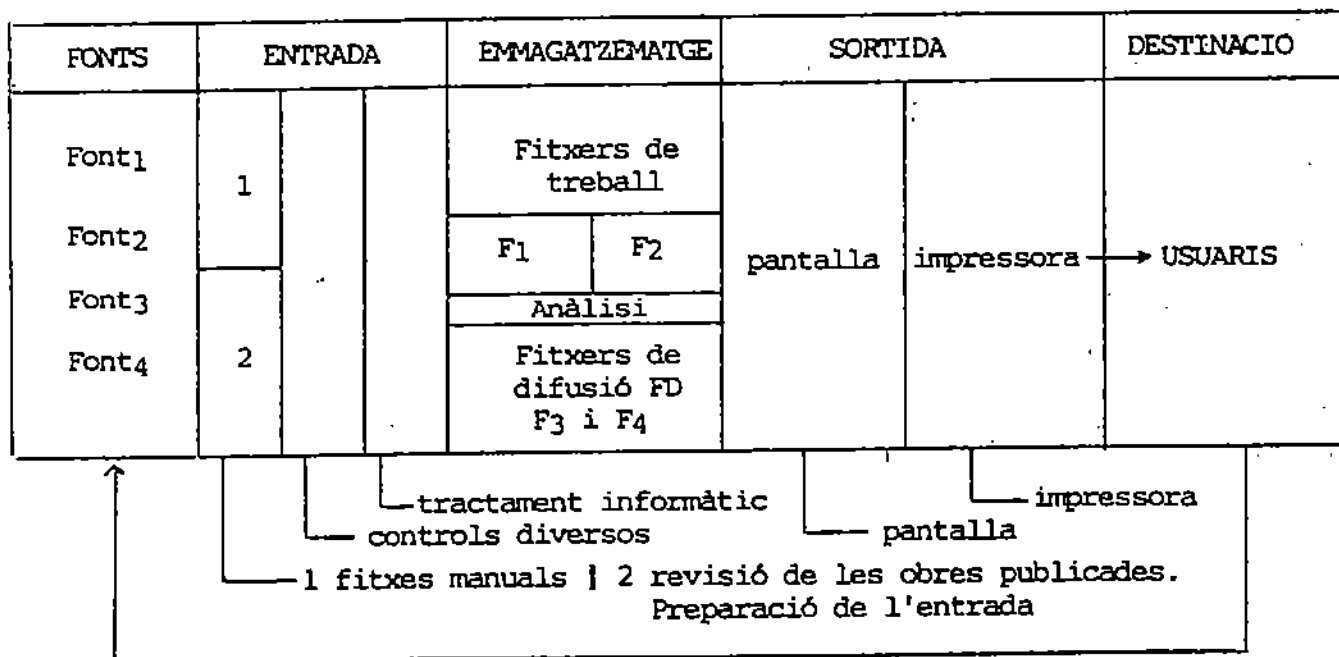
Termes de nivells especials de llengua

Neologismes aprovats

Termes rebutjables, etc.

L'organització del banc del TERMCAT podria ser representada pel quadre següent:

Organització de la informació



Font1 = Terminalistes del TERMCAT

Font2 = Terminalistes de centres associats al TERMCAT

Font3 = Equips de treball

Font4 = Altres bancs de dades

Controls diversos: Terminòlegs del TERMCAT. Depuració de fitxers. Anàlisi de la fiabilitat de les fitxes.

Tractament informàtic: Fitxa automatitzada.

F₁ Fitxer de Termes no aptes per a la difusió

Termes amb marca 6
Fitxes incompletes
Tots provisionals

F₂ Fitxer de Termes per analitzar

Termes amb marca 3 i 4
Tots provisionals
Fitxes per completar

FD Fitxers de difusió: Termes amb marca 1 o 2
F₃ i F₄

Tots definitius
Fitxes completes amb definició i equivalències

F₃: Termes amb les informacions obligatòries
Marca de ponderació 1 o 2
Definitius

F₄: Fitxes totalment completes
Marca de ponderació 1 o 2
Definitives

Impressora: llistat i edicions

6. Situació actual i perspectives

El banc del TERMCAT tot just acaba de constituir-se.

Allò que podem dir que ja és acabat és el següent:

Disseny general del banc
Disseny de l'entrada
Disseny de la sortida
Estructura de la informació emmagatzemada
Programació de l'entrada: altes, baixes i modificacions

S'esta treballant actualment en:

La programació de la sortida
Les formes de sortida. Tipus de sortida
La presentació dels diversos tipus de sortida
L'edició per impressora

El ritme d'entrada ara per ara és lent. En aquest moment s'estan entrant

El ritme d'entrada ara per ara és lent. En aquest moment s'estan entrant els vocabularis de comerç i banca i els de medicina. A partir del mes de març es començarà a entrar les obres de tecnologia i, a més, es contractarà terminalistes només per entrar dades en màquina. Pensem que en començar l'any 1987 tot el material publicat ja pot ser introduït. Alguns treballs seran passats directament en suport automàtic: (diccionari de la construcció, thesaurus tèxtil, diccionari d'informàtica) si aconseguim de fer-los compatibles per mitjà d'una interfície o d'una conversió.

També esperem de tenir en màquina immediatament el Diccionari de la Llengua Catalana (entrada i valor gramatical) cedit per Ll. Izaguirre, professor del Departament de Llengua Catalana de la UB, i més tard el Diccionari complet i tots els bilingües, cedits per la Fundació Enciclopèdia Catalana. Això permetria d'assignar automàticament la marca de ponderació als termes, amb la qual cosa el procés d'entrada esdevé més ràpid.

Pensem que a començaments de 1987 el banc contindrà al voltant de 30.000 termes propis i d'altres procedents de bancs de dades exteriors que seran classificats d'acord amb el sistema de classificació temàtica del banc del TERMCAT.

FE D'ERRATES

Full núm. 78	línia 29	On diu tanteig ha de dir tempteig
Full núm. 83	" 10	On diu en ha de dir a
Full núm. 86	" 11	S'ha d'eliminar "fitxes completes amb definició i equivalències"
Full núm. 86	" 15	On diu Fitxes totalment completes ha de dir Fitxes totalment completes amb definició i equivalències
Full núm. 87	" 10	On diu Izaguirre ha de dir Yzaguirre

NOTES

- (1) Considerarem sinònimes les denominacions següents: banc de dades terminològiques, banc de termes i banc terminològic
- (2) Per a més informació sobre aquests i d'altres bancs de termes vegeu INFOTERM Series, 4. KG Verlag
- (3) Seguim bàsicament Rondeau (1983)
- (4) NORMATERM, DANTERM, BTQ, BTC, EURODICAUTOM i TERMDOK han fet la seva pròpia classificació. El banc de l'Oficina de llengües de la RFA i el de Siemens han adaptat classificacions estereotipades. El banc de la FAO no conté classificació
- (5) Infoterm és Centre Internacional d'Informació Terminològica. Va ser promogut el 1971 per la UNESCO dins del programa intergovernamental UNISIST, encaminat a fomentar la cooperació per a l'intercanvi d'informacions científiques i tècniques a nivell internacional, regional i nacional. És finançat per la UNESCO, el govern austríac, la Cambra de Comerç d'Àustria i l'Institut Austríac de Normalització
- (6) Per a més informació poden consultar els documents de presentació del TERMCAT.
L'adreça del TERMCAT és: carrer del Carme, 47. 08001-Barcelona
- (7) Els models normalitzats de fitxes de buidatge i terminològiques són a disposició dels equips de treball al TERMCAT
- (8) El TERMCAT disposa de documentació pràctica que estandarditza el treball terminològic de preparació de l'entrada i de recerca en general
- (9) La classificació temàtica del Banc del TERMCAT és una adaptació específica del thesaurus del Banc de Termes del Govern del Canadà. El desenrotllament de cada una de les àrees està encara pendent de revisió i completació per part d'experts de cada camp
- (10) El TERMCAT disposa de documentació sobre el model de fitxa automatitzada i d'instruccions sobre l'entrada de les dades a l'ordinador

- FEDOR de DIEGO, A. 1981. La importancia de un banco de Terminología en la transferencia de tecnología. Dins Banquisimeto, 1, Venezuela
- FELBER, H. (1980) A plan for data management within TermNet. Dins Terminological Data Banks. Infoterm Series, 5, K G. Saur
- FELBER, H. (1981) The terminological data elements as derived from the General theory of Terminology and their recording in machine-readable form. Infoterm. Dins Terminologies for the eighties. München-New York, London, Paris, K 6. Saur 1982 (Infoterm Series, 7)
- FELBER, H. (1982) Computerized terminography in TermNet: the role of terminological data banks. Dins SNELL, B (ed.) Term Banks for Tomorrow's world. London Aslib, 1983
- FELBER, H. i GALINSKI, Ch. (1981) L'action de TermNet au plan international pour l'enregistrement des données terminologiques sous une forme assimilable par les machines. Infoterm
- FORTIN, J.M. 1982 La Banque de Terminologie du Québec et la francisation des Terminologies. Québec. Policopiat
- HENNING, J.M. La Banque de Terminologie de l'Université de Clermont-Ferrand II. Paper fotocopiat
- Infoterm series, 4. World Guide to terminological Activités. K 6. Verlag 1977
- RAVENTÓS, D. Proyecto del Banco de datos terminológicos de la Universidad. Simón Bolívar (BTUSB). s/d
- RONDEAU, G. (1983) Introduction à la terminologie. Québec. G. Morin
- Report on the Second international meeting of experts on "Guideling for the recording of terminological data for machine processing" Infoterm, 1983
- SAGER, J.C. i Mc NAUGHT, J. Specifications of a linguistic Data Bank for the U.K. UMIST Manchester, 1980
- Feasibility Study of the establishment of a terminological data bank in the U.K. UMIST Manchester, 1981

SAGER, J.C. i Mc NAUGHT, J. Selective survey of existing linguistic data banks in Europe UMIST, 1980

SCHULZ, J. A Terminology Data Bank for translators (TEAM). Dins Meta, vol. 25 núm. 2, juny 1980. Les Presses de l'Université de Montréal

TermNet News núm. 12, 1985 Special issue on the Nordic countries