

Le scritture del catalogo: uno sguardo alle esperienze straniere

di Barbara Poli

Questo intervento riprende ed amplia un analogo intervento presentato nel 2003 al seminario di studio organizzato a Venezia dalla Fondazione Querini Stampalia con il titolo "Da Babele ad Alessandria? Catalogazione e alfabeti non latini"¹. Mi è sembrato infatti interessante andare a verificare a distanza di tempo quale evoluzione ci fosse stata nel campo del trattamento catalogafico dei materiali in scritture non latine, sia in ambiti in cui la questione era già stata affrontata da molto tempo, sia in realtà in cui il tema era magari meno sentito. Il quadro che ne emerge è complesso e sfaccettato, e senza alcuna pretesa di esaustività cercherò qui di offrirvene alcuni esempi utili, spero, per ulteriori approfondimenti.

Per quanto riguarda il mondo anglosassone, negli ultimi anni vi sono stati alcuni rilevanti cambiamenti. Nel 2006 RLG (Research Libraries Group) entra a far parte di OCLC e ciò comporta dal 2007 la dismissione di RLIN (Research Libraries Information Network) come catalogo separato, con il riversamento di tutti i dati in WorldCat. RLIN includeva già da molti anni registrazioni bibliografiche in caratteri non latini (dal 1983 il gruppo CJK – Chinese Japanese Korean, dal 1986 il cirillico, dal 1992 l'arabo), e così pure i cataloghi di OCLC.

Il multilinguismo è presente sotto varie forme in OCLC, e vorrei analizzarne alcuni aspetti.

In WorldCat sono rappresentate più di 470 lingue e dialetti (erano 400 nel 2003) ed è possibile rilevare una crescita costante dei materiali in scritture non latine; a titolo di esempio riporto le cifre relative a quattro lingue che insieme costituiscono circa il 3,6% dei record presenti nel catalogo:

lingua	2003	2008
Arabic	285.319	379.692
Chinese	1.118.276	1.692.843
Japanese	1.263.231	1.118.276
Korean	207.083	259.198

Per affrontare questa varietà OCLC ha predisposto vari strumenti.

BARBARA POLI, Fondazione Querini Stampalia Onlus, Castello 5252, 30122 Venezia, e-mail bpoli@querinistampalia.org.

Ultima consultazione siti web: 6 marzo 2009.

1 <<http://www.aib.it/aib/commiss/mc/rabchio1.htm>>.

Bollettino AIB, ISSN 1121-1490, vol. 49 n. 1 (marzo 2009), p. 47-53.

*Language sets*² è un servizio che offre alle biblioteche la possibilità di acquistare “pacchetti” di documenti in lingua non inglese (non necessariamente in lingue che usano scritture non latine) per rendere «easy and affordable to serve your culturally diverse community» e «meet the needs of diverse users with popular materials in 17 languages: Arabic, Chinese, French, German, Gujarati, Hindi, Japanese, Korean, Malayalam, Marathi, Panjabi, Russian, Spanish, Tagalog, Telugu, Urdu and Vietnamese» (erano 12 lingue nel 2003).

Ciò consente di superare il fatto che non sempre è possibile avere in tutte le biblioteche specialisti delle varie lingue e culture che siano in grado di interagire con editori e fornitori e di selezionare in modo appropriato i materiali da acquisire; anche la catalogazione viene velocizzata, perché libri e DVD arrivano «with MARC records, ready to serve» e i dati con le nuove localizzazioni vengono naturalmente resi disponibili in WorldCat attraverso interfacce in lingue e scritture diverse.

A questo proposito, OCLC ha sviluppato *FirstSearch*³, il sistema di ricerca in WorldCat che consente l'utilizzo di sette diverse interfacce: «English, French, Spanish, Japanese, Korean, and Chinese (traditional and simplified)» con la precisazione che «WorldCat record display in Chinese, Japanese, Korean and Arabic vernacular for items cataloged in those languages».

Spostandoci sul versante del *back office*, e nello specifico della catalogazione, scopriamo che *Connexion*⁴, la principale interfaccia di catalogazione in OCLC, consente di creare o modificare record «using worldwide scripts» e precisamente «languages written in Arabic, Bengali, Chinese, Cyrillic, Devanagari, Greek, Hebrew, Japanese, Korean, Latin, Tamil, and Thai scripts».

Il sistema consente una certa flessibilità di ricerca e/o inserimento: per esempio è possibile ricercare le registrazioni bibliografiche attraverso oltre 90 indici inserendo «either script search terms or romanized (Latin-script equivalent) search terms», dato che gli indici di WorldCat includono sia i dati in scritture non latine sia i loro equivalenti trascritti o traslitterati in alfabeto latino.

È inoltre possibile «create bibliographic records with Latin script data only, non-Latin script data only, or any combination of multiple scripts. If you include both Latin and non-Latin script, the client provides paired fields that have the same tag number. The top field of the pair is for non-Latin-script data followed by a field for the corresponding romanized data». Naturalmente le registrazioni possono essere esportate nei vari sistemi locali convertendole attraverso il formato MARC 21 (che utilizza il MARC-8 o l'UTF-8 Unicode character set) e selezionando quale scrittura debba andare nei vari campi del record e quale invece debba essere inclusa nei campi 880 (*Alternate graphic representation*).

Connexion fornisce anche alcuni strumenti per convertire automaticamente nelle scritture originali le registrazioni catturate in WorldCat che presentino solo i dati traslitterati in alfabeto latino; i *transliteration tools* attualmente disponibili consentono il passaggio dalle forme romanizzate ad arabo, cirillico, greco ed ebraico o viceversa.

Per quanto riguarda invece la forma del nome degli autori cinesi, *Connexion* offre l'accesso ai circa 130.000 record di autorità del *Chinese name authority file* gestito dal Joint University Librarians Advisory Committee (JULAC) di Hong Kong, in cui è possibile ricercare e derivare i nomi da inserire nei record bibliografici. Naturalmente

2 <<http://www.oclc.org/languagesets/default.htm>>.

3 <<http://www.oclc.org/firstsearch/default.htm>>.

4 <<http://www.oclc.org/connexion/default.htm>>.

non è tutto facile come sembra e i problemi non mancano, come dimostra l'esigenza di istruzioni particolareggiate per i catalogatori, a causa delle diverse situazioni in cui si trovano le varie scritture⁵ e la presenza di gruppi di discussione per i vari ambiti linguistici, come l'OCLC-CJK (listserv discussion of OCLC Chinese, Japanese, and Korean functionality).

Anche per la Library of Congress molte cose sono cambiate. Fino al 2006 il CDS Cataloging Distribution Service di LC distribuiva in archivi separati le registrazioni catalografiche relative a materiali in alcuni alfabeti e scritture non latini, ma dal novembre 2006 si annuncia che non verrà più effettuata la distribuzione separata dei record JACKPHY (Japanese, Arabic, Chinese, Korean, Persian, Hebrew and Yiddish), inclusi invece in MDS-Books All, ossia l'archivio contenente i record di tutti i documenti trattati da LC. La decisione viene spiegata da un lato con una motivazione tecnica: la dismissione del sistema RLIN e la fusione di RLG con OCLC richiedono nuove procedure di trattamento dei record JACKPHY. Dall'altro lato, c'è anche una concreta motivazione di compatibilità economica: i ricavi delle sottoscrizioni JACKPHY non sono sufficienti a garantire la produzione di file separati nel nuovo ambiente di lavoro.

Per quanto concerne la catalogazione, utili informazioni si trovano nella sezione del sito⁶ che la Library of Congress dedica alle questioni catalografiche legate al trattamento dei materiali in alfabeti non latini. Si tratta di documenti molto tecnici che però evidenziano la politica catalografica dell'ente. I caratteri non latini vengono gestiti sempre attraverso i campi MARC 880 (*Alternate Graphic Representation*), che contengono registrazioni parallele rispetto alle forme romanizzate, a lungo le sole ammesse per la ricerca.

Ma proprio quest'anno c'è uno sviluppo interessante per quanto riguarda i record di autorità, qualcosa che sta accadendo proprio adesso. La questione è trattata nel *White paper: Issues related to non latin characters in name authority records*⁷ che riferisce dell'accordo fra British Library, biblioteche canadesi, LC, OCLC ecc. per consentire l'inserimento di caratteri non latini nei record di autorità distribuiti da NACO (Name authority cooperative program).

Il punto fondamentale è che si è deciso di non usare il MARC 21 *Authority Model A: Vernacular and transliteration* (in cui i campi 880 registrano informazioni parallele con lo stesso contenuto dei campi regolari ma in una scrittura diversa), bensì il *Model B: Simple multiscript records* (che può contenere varie scritture ma generalmente non ripete le informazioni). Usando il *Model B* le forme in caratteri non latini vengono inserite nei campi 4XX (*See from*) ossia trattate come rinvii che identificano «unauthorized forms of headings and other variants not chosen as an authorized form».

Il *White paper* riferisce infatti che la discussione sull'individuazione di una forma "preferita" in caratteri non latini è stata rinviata, perché avrebbe potuto ritardare gravemente l'avvio del progetto e si è invece preferito l'obiettivo a breve termine di consentire l'uso dei caratteri non latini, anche a rischio di conflitti fra forme del nome.

Con un approccio molto pragmatico, si è deciso quindi (aprile 2008) di riversare nel LC/NACO *authority file* tutte le forme in caratteri non latini derivate dai campi intestazione di WorldCat, il che presenta l'indubbio vantaggio di riutilizzare il lavoro della cooperazione (circa 500.000 nomi), ma il serio svantaggio della mancanza di unifor-

5 <<http://www.oclc.org/support/documentation/connexion/client/international/default.htm>>.

6 <<http://www.loc.gov/catdir/cps/nonlatingeneral.html>>.

7 <http://www.loc.gov/catdir/cps/nonlatin_whitepaper.html>.

mità dovuta all'assenza di procedure comuni e condivise. Ciò rende necessaria l'individuazione di buone pratiche o, per dirla con le parole del documento: «policies for allowing non-Latin references in authority records need to be established».

Il punto di partenza per questo lavoro è la consapevolezza che nella prassi catalografica seguita fin qui da LC le forme romanizzate dei nomi rappresentano le sole forme di intestazione autorizzata per il catalogo basato sulla lingua inglese, le sole soggette a pieno controllo di autorità, mentre le forme nelle scritture originali svolgono una funzione identificativa complementare, offrendo accessi non controllati. Ciò ha reso ammissibile lo sviluppo di prassi difformi, con una ulteriore segmentazione fra le lingue che utilizzano scritture da destra a sinistra (gruppo HAPY Hebrew, Arabic, Persian, Yiddish) e le altre (gruppo CJK Chinese Japanese Korean), a causa della difficoltà di gestire dati multidirezionali (come le qualificazioni). Gli alfabeti greco e cirillico, inoltre, non sono ancora supportati.

Il documento conclude identificando tre possibili linee di lavoro:

- inserire i riferimenti *non-Latin* come si trovano nelle pubblicazioni, senza ulteriori qualificazioni;
- inserire i riferimenti *non-Latin* seguendo la struttura della forma traslitterata controllata;
- adottare approcci diversi secondo le scritture.

Il tutto da discutere dopo i primi sei mesi di lavoro nel nuovo archivio, e quindi a partire dalla fine del 2008. Vedremo cosa accadrà. Nel frattempo, è interessante consultare l'immane sezione FAQs per i catalogatori, *Non-Latin Script Data in Name Authority Records: Frequently Asked Questions*⁸, ma anche alcune istruzioni per la ricerca dalla parte del lettore: *Displaying and Searching Non-Roman Characters in the Online Catalog (Unicode)*⁹.

Tornando invece nella vecchia Europa, dopo aver ricordato che il CJK sia di OCLC sia di RLIN era stato sperimentato in Europa ma poi non adottato proprio a causa dei diversi sistemi di trascrizione per cinese e giapponese in uso in Europa e negli USA, occorre parlare delle fondamentali esperienze legate a CJK Allegro. Alla fine degli anni Ottanta, la Staatsbibliothek di Berlino, la Bodleian Library e l'Università di Heidelberg sviluppano un particolare adattamento di Allegro, il programma di gestione integrale dei dati bibliografici in uso in molte biblioteche tedesche. Allegro CJK gestisce gli ideogrammi e dal 1997 consente l'accesso in linea ai cataloghi cinese e giapponese di queste biblioteche. Si tratta di cataloghi separati, quindi non integrati nel catalogo generale; si vedano per esempio i cataloghi della Bodleian Library¹⁰.

A fianco di questi cataloghi esiste però oggi una significativa novità: due cataloghi collettivi.

Lo European Union Catalogue of Japanese Books¹¹, accessibile dal 2005, è gestito dal NII (National Institute of Informatics) di Tokyo, che collabora con le biblioteche britanniche già dal 1991. Comprende anche i dati di alcune biblioteche di Germania, Svezia, Svizzera, Belgio e Norvegia e contiene oltre 187.000 registrazioni bibliografiche cui sono collegate 267.000 copie.

8 <<http://www.loc.gov/catdir/cpso/nonlatinfaq.html>>.

9 <<http://catalog.loc.gov/help/unicode.htm>>.

10 <<http://www.bodley.ox.ac.uk/dept/oriental/allegro.htm>>.

11 <<http://www.niicat.eu/>>.

Lo UK Union Catalogue of Chinese Books¹², creato grazie alla possibilità offerta da Allegro di importare e convertire dati da vari formati, riunisce le registrazioni di British Library, Bodleian Library, Brotherton Library (Leeds), e delle biblioteche universitarie di Cambridge, Durham, SOAS (Londra) e Edimburgo. Si tratta di un catalogo cumulativo, contenente oltre 297.500 record che rappresentano circa 242.000 diversi titoli (è quindi presente un certo numero di duplicazioni di notizie).

Venendo alla Francia, la biblioteca che da sempre rappresenta una molteplicità di lingue e scritture nelle proprie collezioni è naturalmente la Bibliothèque Nationale de France, il cui dipartimento “Orient et collections orientales” riunisce le opere relative a tre aree: Asia, Europa balcanica e orientale, Medio Oriente. Molte opere sono in lingua originale: «arabe, arménien, chinois, coréen, grec, hébreu, japonais, persan, polonais, russe, slovaque, tchèue, tibétain, turc, vietnamien, langues finno-ougriennes, de l’Inde, et slaves du Sud».

Inevitabilmente, nel corso degli anni il trattamento dei materiali in alfabeti e scritture non latini ha subito vari cambiamenti ed evoluzioni, dando luogo ad una stratificazione catalografica che rende complessa la ricerca di questi materiali, come spiega la sezione del sito¹³ che la BNF dedica a questa tematica. Si possono riconoscere fondamentalmente tre situazioni:

- opere entrate prima del 1960: la maggior parte è in OPALE Plus, ma spesso alcuni elementi della descrizione non sono traslitterati e sono visualizzati come @. Occorre completare la ricerca sui cataloghi cartacei;
- opere entrate dopo il 1997: catalogate e traslitterate/trascritte in notizie che talora comprendono anche i caratteri originali;
- opere entrate dal 1960 al 1996: cataloghi digitalizzati in pdf accessibili dal sito¹⁴.

Per quanto riguarda gli archivi di autorità, è molto interessante notare che nei record di autorità in OPALE Plus viene in genere riportata la forma accettata sia nella scrittura originale non latina sia nella sua traslitterazione o trascrizione secondo le norme ISO.

Sempre in ambito francese, rivestono una certa importanza le riflessioni teoriche che hanno preceduto la realizzazione della BULAC, la Bibliothèque Universitaire des Langues et Civilisations presentata a Parigi il 23 ottobre 2008.

Nel gennaio 2002 la Direction de l’enseignement supérieur, Sous-direction des bibliothèques et de la documentation del Ministère de l’Education Nationale presenta il documento *Catalogage des documents en caractères non-latins: rapport du Groupe de travail*¹⁵, redatto da Danièle Duclos-Faure. Il documento è finalizzato alla nascita del “Pôle des langues et civilisations du monde” che ha lo scopo di riunire collezioni multilingui e multialfabeto disperse presso vari istituti e in situazioni catalografiche molto diverse (cartaceo o informatizzato, in caratteri originali o traslitterati), obiettivo realizzato appunto con l’apertura della BULAC¹⁶. Questo rapporto sulla catalogazione dei documenti in caratteri non latini analizza la varietà di situazioni in cui si trovano i materiali nei vari istituti e delinea le raccomandazioni per il loro tratta-

12 <<http://www.bodley.ox.ac.uk/rslpchin/oldindex.htm>>.

13 <http://grebib.bnf.fr/html/caracteres_nonlatins.html>.

14 <http://www.bnf.fr/pages/zNavigat/frame/catalogues_num.htm?ancre=imp-num.htm>.

15 <<http://www.sup.adc.education.fr/bib/Acti/fcnl/titre.htm>>.

16 <<http://www2.bulac.fr/>>.

mento nel nuovo catalogo. A fondamento di tutto, un assunto irrinunciabile: «le catalogue doit être multi-écriture».

Le condizioni tecniche che lo renderanno possibile sono le seguenti:

- adozione della norma di codifica caratteri UNICODE UTF 16;
- utilizzo di norme di traslitterazione reversibili (traslitterazione carattere per carattere), quando esistono, per le scritture alfabetiche e sillabiche, mentre per le scritture consonantiche sarà possibile eseguire con procedure automatizzate solo il passaggio dalla romanizzazione alla scrittura originale;
- utilizzo di un formato di catalogazione (bibliografica e di autorità) che autorizzi la doppia registrazione delle aree in scritture originali e in caratteri latini;
- la costruzione di archivi di autorità legati agli archivi bibliografici che riuniscano, per gli autori e i titoli uniformi, tutte le diverse forme e grafie possibili per permettere l'accesso alle notizie bibliografiche «*quelles que soient la langue et l'écriture des ouvrages et la forme utilisée par le lecteur pour interroger le catalogue*».

Il risultato, dal punto di vista del lettore, dovrebbe essere questo: «un lecteur interrogeant le catalogue par le nom d'auteur en caractères latins <Puskin> accèdera à l'ensemble des ouvrages de cet auteur figurant au catalogue qu'il s'agisse d'ouvrages en russe, d'ouvrage traduits ou d'ouvrages bilingues. Le résultat sera identique s'il interroge par la forme du nom en cyrillique ou par une forme courante française <Pouchkine> ou encore conforme à la transcription anglo-saxonne». Avendo fatto la prova, possiamo dire che non è ancora così, visto che interrogando con forme diverse proprio dal nome Puskin si ottengono risultati diversi, ma la strada ormai è tracciata.

E nella presentazione del catalogo della BULAC¹⁷ si afferma in modo chiarissimo che la principale caratteristica del catalogo è proprio quella di offrire «des notices en plusieurs écritures: notamment écritures latines et latines diacritées, chinoise, japonaise, coréenne, arabe et persane», lista che sarà progressivamente estesa mano a mano che verranno sviluppati i supporti informatici necessari.

Concludo mostrando le pagine di accesso a due cataloghi che da sempre fanno convivere lingue e scritture diverse, anche se non tutte le scritture possibili. Uno è il catalogo della Biblioteca nazionale di Grecia, anzi della Εθνική Βιβλιοθήκη της Ελλάδος che visualizza insieme caratteri greci e latini. L'altro è il catalogo della biblioteca di Alessandria d'Egitto¹⁹, che accanto alle interfacce in arabo, inglese e francese ne propone una meno scontata in giapponese.

Spero che questa panoramica, necessariamente incompleta, possa essere utile per la riflessione che dobbiamo fare in Italia. Nei prossimi mesi saremo impegnati nell'adozione delle REICAT (Regole Italiane di Catalogazione), che in vari punti trattano il tema dei sistemi di scrittura diversi dall'alfabeto latino e prescrivono che le informazioni che devono essere riportate in lingue che utilizzano tali sistemi «si danno, quando è possibile, sia nella *scrittura originale* sia nella forma *traslitterata o trascritta in alfabeto latino*» (REICAT o.6.3). Ciò significa che dovremo porci molto concretamente, molto operativamente il problema di far convivere scritture diverse nei cataloghi, e in primo luogo nel catalogo di SBN.

17 <<http://www2.bulac.fr/acces-au-catalogue/presentation-du-catalogue/>>.

18 <<http://www.nlg.gr/default.htm>>.

19 <<http://www.bibalex.org/English/index.aspx>>.

Alberto Petrucciani ha aperto questa sessione ricordando che l'alternativa, ossia che ciascuno acceda solo alle informazioni nella lingua e nella scrittura che conosce, è una prospettiva che dà i brividi. A questa prospettiva dobbiamo contrapporre un'altra, che è poi nient'altro che il nostro mestiere, ossia fare in modo che la mediazione catalografica serva a far incontrare libri e lettori, indipendentemente dalla lingua e dalla scrittura del libro, indipendentemente dalla lingua e dalla scrittura del lettore. Sappiamo che non è solo una questione tecnico-catalografica, ma soprattutto culturale e politica.

Catalogue scriptures: a glance at foreign experiences

Since the early Eighties, when the large collective North-American catalogues started to insert bibliographical records in non-Latin alphabets, cataloguing of material in languages that use other scriptures has posed a number of problems that are still far from being solved. Without any claims to being exhaustive, the article reviews some foreign catalogues and the ways in which the various scriptures are managed and made to live and work together.