

La condivisione della conoscenza attraverso la mappatura di concetti giuridici

Ginevra Peruginelli

*Istituto di teoria e tecniche
dell'informazione giuridica
Consiglio nazionale delle ricerche
ginevra.peruginelli@ittig.cnr.it*

Strategie per l'interoperabilità fra thesauri

Il contesto

La possibilità di accedere a risorse informative eterogenee è oggi accresciuta notevolmente per la disponibilità sulla rete mondiale di vaste collezioni provenienti da fonti diverse, di natura istituzionale, commerciale, ed anche privata. Vi sono anche ragioni di tipo economico che spingono ad una condivisione di dati e servizi in modo tale che risorse informative predisposte per un particolare contesto siano recuperabili per altre applicazioni. In pratica ciò che è richiesto è di partire da una interrogazione e poter trovare tutta la documentazione corrispondente alla ricerca effettuata, indipendentemente dal metodo di indicizzazione e dalla terminologia usata.

Di qui la necessità di affrontare efficacemente i problemi derivanti dall'utilizzo di diversi strumenti di indicizzazione, di vocabolari controllati (siano essi thesauri, schemi di classificazione, stringhe di soggetto, tassonomie, liste di autorità, ontologie) che possono differire notevolmente fra loro quanto alla lingua in cui sono espressi, al loro scopo e al livello di astrazione e dettaglio. Il presente contributo si concentra in particolare sull'interoperabilità fra thesauri nell'area giuridica ed analizza la possibilità di creare una mappatura fra termini nel tentativo di costruire corrispondenze le più esatte possibili

fra i concetti espressi dai termini stessi. Questo è oggi un argomento assai dibattuto, che tocca da vicino problematiche complesse, di tipo semantico e concettuale così come di natura tecnologica, dal momento che la realizzazione di sistemi per il recupero di informazioni indicizzate utilizzando diversi vocabolari controllati si basa su protocolli e software in grado di gestire applicazioni per l'accesso a collezioni eterogenee.

Le questioni legate alla corrispondenza fra termini sussistono anche per la ricerca a testo libero; si tratta infatti in questo caso di operare un confronto fra i termini di un vocabolario utilizzato per la ricerca e quelli presenti in determinate collezioni.

Resta il fatto che i thesauri, pur essendo strumenti disegnati con il consenso e la collaborazione di esperti nella predisposizione di un insieme stabilito di termini (con opportune relazioni fra loro) relativi ad un particolare ambito di conoscenza o addirittura a più aree disciplinari, presentano molto spesso fra loro differenze significative che rendono difficile la loro interoperabilità.

Numerose sono le iniziative volte alla predisposizione di linee guida per la costruzione e l'interoperabilità di thesauri¹ e allo sviluppo di progetti per la soluzione dei problemi legati all'allineamento di vocabolari controllati. Ne sono un

esempio i progetti CARMEN,² RENARDUS,³ MACS,⁴ HILT.⁵ Le esperienze in tema di predisposizione di metathesauri, realizzati fondendo i termini propri di più thesauri ed operando dei collegamenti alle fonti originali sono oggi abbastanza frequenti anche se non molto conosciute. Un esempio significativo di questo tipo è il metathesaurus, realizzato nel campo della medicina, denominato Unified Medical Language System (UMLS).⁶ Prima di entrare nel merito della specificità del mondo del diritto e delle funzionalità di accesso a materiale giuridico di diversa provenienza e indicizzazione, è opportuno soffermarsi su alcuni aspetti fondamentali che riguardano l'interoperabilità fra thesauri indipendentemente dal dominio di cui trattano. Infatti nell'operazione di allineamento, accesso e scambio dell'informazione giuridica su vasta scala ricorrono problematiche comuni ad altre discipline, anche se la terminologia del diritto ha precise particolarità che condizionano notevolmente le possibilità di scambio e comunicazione.

La comparazione di termini e concetti: cenni sulle possibili strategie per l'interoperabilità dei thesauri

L'operazione di fusione, da intendersi come vera e propria fusione

fisica o predisposizione di corrispondenze fra termini per accedere a collezioni indicizzate utilizzando più thesauri, richiede un confronto fra voci appartenenti a fonti terminologiche diverse. Ciò comporta la soluzione dei problemi dovuti alle differenze esistenti nella semantica dei termini stessi e alla loro conseguente parziale mancata convergenza: tutto questo rende l'interoperabilità un'impresa certamente complessa dal punto di vista linguistico e scientifico, nonché onerosa per le risorse intellettuali e tecnologiche necessarie alla sua realizzazione.

Gli studi e gli esperimenti in questo settore mostrano che esiste una certa analogia delle questioni legate alla gestione di thesauri multilingue con quelle derivanti dalla esigenza di far corrispondere thesauri diversi di una stessa lingua:⁷ ciò è comprensibile, date le differenze esistenti, nell'universo culturale e più in generale nel mondo della conoscenza, nel concepire i concetti, nel definirne le innumerevoli possibili sfumature di significato e quindi nell'esprimerli utilizzando una determinata terminologia.

Il dibattito internazionale sulla mappatura dei thesauri copre diversi aspetti: i modelli di interoperabilità, i possibili scenari e contesti in cui avviene il confronto, le strategie sui livelli di accettabilità di tali corrispondenze, gli strumenti tecnologici necessari per le applicazioni. Ancora scarse sono invece le indagini sui risultati di ricerca in termini di precisione e richiamo quando si opera su collezioni indicizzate con vocabolari controllati di tipo diverso che si intende far convergere.

In linea generale i contesti in cui è richiesta una interoperabilità possono raggrupparsi in situazioni caratterizzate rispettivamente da: *a*) presenza di vocabolari controllati che hanno una stessa struttura e lingua (in questo caso è più sem-

plice stabilire una relazione di equivalenza fra termini, addirittura di tipo bi-direzionale); *b*) rapporto di non equivalenza fra termini (che differiscono nell'ambito, lingua e struttura); *c*) possibilità di utilizzare uno dei vocabolari con una funzione di riferimento su cui operare delle mappature verso gli altri vocabolari. Spesso nelle applicazioni esistenti questi modelli sono fra loro combinati ed è evidente che i risultati di ricerca su collezioni diversamente indicizzate saranno differenti in termini di precisione e richiamo secondo le situazioni che si presentano.

Ciascuno dei modelli sopra delineati implica che la mappatura possa avvenire a vari livelli di qualità. A parte il caso, piuttosto raro, in cui vi sia esatta equivalenza, per cui a seguito di una ricerca su più collezioni sono recuperati esattamente gli stessi concetti identificati in vocabolari diversi, accade spesso che la funzionalità di cattura dell'informazione ne sia influenzata, con la conseguente perdita di precisione nei risultati della ricerca, se non il mancato recupero dell'informazione stessa.

La differente copertura (quindi la diversa selezione dei concetti inclusi in un thesauro) ed il diverso modo in cui sono rappresentati i concetti nei vari tipi di vocabolari controllati rendono complessa la loro comparazione. Si tratta infatti di gestire la relazione fra varie lingue nel caso di accesso a collezioni multilingue e in tal modo indicizzate, ma oltre a questo caso occorre spesso cercare soluzioni in presenza di termini differenti che connotano un concetto quando ciò è dovuto a diversità culturali, all'appartenenza del concetto stesso a discipline differenti, alla diversa natura e necessità dei destinatari dell'informazione cercata.

Le strategie utilizzate per creare ed accettare le corrispondenze variano secondo i contesti e le finalità

dei sistemi di accesso all'informazione. In mancanza di un esatto equivalente una opzione consiste nella mappatura verso un termine di significato più generale. La corrispondenza di un termine più ampio verso uno più dettagliato è invece generalmente meno soddisfacente, preferendosi una mappatura verso termini simili, nella convinzione che spesso l'opzione migliore è creare una corrispondenza verso un termine che non è né più ampio né più dettagliato, ma che si approssima, nel significato, a quello da equiparare. Nell'area del diritto la ricerca di una equivalenza funzionale che si basa su categorie astratte o genotipi rappresenta una scelta opportuna, che è analizzata in seguito.

Per quanto riguarda i gradi di equivalenza, occorre precisare che l'equivalenza esatta fra concetti espressi dai termini che si pongono in relazione è indubbiamente la situazione ideale e ciò si verifica quando i vocabolari sono stati costruiti con concetti identici. Una equivalenza inesatta può essere espressa, se ciò risulta utile, secondo il livello di sovrapposizione che si può ottenere e il contesto in cui la mappatura viene eseguita. Diversa è l'equivalenza parziale, che si verifica quando un concetto corrispondente al vocabolario di partenza non è presente né parzialmente né in modo inesatto.

Nei casi di vocabolari con diversa struttura sono possibili confronti del tipo uno a molti, come nel caso in cui il vocabolario di destinazione contenga uno o più concetti che insieme corrispondono al concetto di partenza. Quando il livello di specificità differisce fra vocabolari, più concetti nel vocabolario fonte possono essere mappati verso un concetto nel vocabolario di destinazione, corrispondente all'equivalente più vicino ivi contenuto. I thesauri multilingue presentano una casistica simile, ma anche spe-

cifiche particolarità. Le differenze culturali che si riflettono nella lingua hanno un peso notevole nell'operazione di confronto e non è sempre semplice arrivare ad una situazione in cui le varie lingue rappresentate abbiano un identico stato. Anche in questo caso si può accettare un equivalente non esatto in una lingua o, in totale mancanza di un concetto, questa lacuna può essere compensata da un termine preso a prestito dalla lingua originale o da uno coniato per l'occasione, possibilmente corredato da una nota di spiegazione, utile all'utente che esegue la ricerca.

È opinione diffusa che non esiste un'unica soluzione corretta per la realizzazione di una mappatura fra vocabolari controllati, data la diversità di contesti in cui la corrispondenza è applicata. Il principio che una corrispondenza abbastanza simile sia accettabile se i risultati della ricerca sono soddisfacenti si fa sempre più strada, insieme alla consapevolezza della diffusa necessità di accesso a risorse diversificate da parte di utenti che devono poter formulare ricerche nel linguaggio a loro familiare.

Le peculiarità del linguaggio giuridico

Nel settore giuridico l'interoperabilità tra thesauri è un tema molto dibattuto poiché soddisfa due importanti e diverse funzionalità: da una parte il *cross-collection retrieval*, dall'altro il *cross-language retrieval*. Nel primo caso, forte è oggi la necessità di un dialogo costante tra thesauri che indicizzano diverse collezioni affinché l'utente possa, attraverso un'unica interfaccia di ricerca, avere accesso a più fonti informative che utilizzano strumenti di indicizzazione differenti. Nel secondo caso l'interoperabilità tra thesauri in lingue diverse permette all'utente di recupera-

re il materiale informativo anche in lingue differenti da quella utilizzata nell'interrogazione. In entrambi i casi le peculiarità del linguaggio giuridico sono da prendere in considerazione per avviare in modo opportuno e adeguato un lavoro di mappatura di qualità.

In linea generale un termine si compone di due elementi fondamentali: la sua *forma linguistica*, cioè la parola nella sua realizzazione alfabetica e il *contenuto* o *concetto* di cui la forma linguistica è l'espressione.⁸ Nell'ambito del diritto i concetti fondamentali della scienza giuridica costituiscono, insieme alle espressioni linguistiche loro assegnate, la terminologia propria di questa disciplina. Il linguaggio giuridico rappresenta infatti un sottoinsieme di quello comune, configurandosi come un vero e proprio linguaggio specialistico.

Le caratteristiche dei linguaggi settoriali come quello giuridico sono principalmente due: la loro insostituibilità per lo studio di una determinata disciplina e per l'esercizio di un'attività ed in secondo luogo la loro attitudine ad indicare l'appartenenza ad un gruppo di individui uniti da legami di interessi e conoscenze più stretti di quelli che intercorrono normalmente tra i membri di una comunità linguistica. Senza dubbio il linguaggio giuridico è ricco di espressioni formali e sorge la domanda se questo formalismo aiuti o impedisca la comunicazione e se la semplificazione del linguaggio giuridico comporti necessariamente una riduzione della sua precisione. La maggior parte degli operatori del diritto ritiene che un eccesso di formalismo riduca la qualità dell'espressione, sostenendo che "il rispetto della legge non deve confondersi con l'adorazione delle vecchie formule".⁹ Vi è quindi un certo favore verso l'opinione che il modo di esprimersi debba essere il più chiaro e semplice pos-

sibile, privilegiando l'utilizzo del linguaggio naturale piuttosto che di quello tecnico, in modo da avvicinarsi all'interlocutore anche se la terminologia, la struttura, lo stile del diritto e i suoi concetti sono necessariamente complessi e specialistici. Molto dipende anche dal genere di testo trattato, poiché vi è una notevole differenza, ad esempio, fra norme costituzionali e specifiche disposizioni di settore.

Tuttavia, il linguaggio giuridico si caratterizza per essere sempre più tecnico, settoriale e specialistico. Tali caratteristiche sono sottolineate dalla distinzione che viene fatta sul piano dottrinale¹⁰ fra:

- tecnicismi specifici, che identificano termini di significato tendenzialmente univoco, esclusivi di ciascun settore (ad esempio: "anatomismo");
- ridefinizioni, che consistono nell'attribuire a termini della lingua comune un significato diverso da quello con il quale vengono genericamente adoperati (ad esempio: "confusione", intesa come modo di estinzione dell'obbligazione);
- tecnicismi collaterali, definiti come specifiche espressioni stereotipiche, non necessarie, ma preferite per la loro connotazione tecnica (ad esempio: "escusione dei testi").

Naturalmente la terminologia giuridica deve essere sempre considerata in uno specifico contesto, per cui un termine può avere una valenza semantica diversa in relazione all'ambito giuridico nel quale è utilizzato. Quindi anche se da una parte il linguaggio giuridico è caratterizzato dall'utilizzo di termini tecnici e costruiti *ad hoc*, dall'altro adotta dal linguaggio comune termini generici, che sono poi recepiti attraverso un processo di tecnicizzazione che è tipico dello specifico ambito giuridico in cui sono calati ed utilizzati.

Tale tipo di linguaggio coincide, in

alcune circostanze, con quello della politica, basato su giudizi di valore, ricco di ambiguità e orientato alla persuasione. In certi casi la complessità di questa terminologia esclude dalla comunicazione il non esperto. Da ciò deriva la cosiddetta indeterminatezza del linguaggio giuridico, in merito alla quale sono sostenute le due opposte esigenze da soddisfare: quella di eliminarla o di mantenerla. La prima deriva dal dogma della certezza del diritto, che richiede la limitazione dell'indeterminatezza o ambiguità e si ricollega al profilo scientifico del linguaggio giuridico, mentre la seconda ha origine nella necessità di garantire una conveniente flessibilità ed adeguatezza delle norme alla realtà in continua evoluzione.

Casi di ambiguità possono verificarsi quando uno stesso termine possiede più significati distinti, oppure quando i suoi significati sono pertinenti ad aspetti diversi di una medesima realtà o a fenomeni in qualche modo collegati. Un esempio del primo caso si può ritrovare nel confronto tra gli articoli 1081 e 1342 del Codice civile: il primo intende il termine *modulo* come unità di misura dell'acqua corrente, il secondo fa riferimento ai contratti conclusi mediante *moduli*: le possibilità di confusione sono minime, per non dire inesistenti. Più delicato è invece il caso di termini come *condizione*, usato nell'articolo 1354 del Codice civile per indicare la clausola cui le parti subordinano l'efficacia di un contratto e nell'articolo 1359 per indicare l'evento. In questi casi è fondamentale operare un'attività di contestualizzazione, che consiste in un processo di identificazione dell'ambito in cui è inserito il concetto. Ciò permetterà di superare l'ambiguità semantico-lessicale, di evitare risultati imprecisi e irrilevanti e infine di rendere consapevoli i destinatari dell'informazione dei di-

versi contesti in cui si collocano gli istituti giuridici.

Tutto ciò vale anche e soprattutto in un ambiente multilingue in cui i diversi linguaggi giuridici rappresentano, per così dire, i portavoce di specifiche culture giuridiche. In campo giuridico infatti i concetti che ne stanno alla base sono intrinsecamente legati al sistema giuridico di appartenenza; da qui la notevole difficoltà di una perfetta e completa trasposizione linguistica.

È facile intuire ad esempio il forte legame tra la lingua inglese ed il sistema giuridico nato in Gran Bretagna in seguito alla conquista normanna: un sistema di *common law* è stato adottato da tutti i popoli anglofoni ed esso non si è concesso a traduzioni in altre lingue.¹¹ L'inglese e il *common law*¹² rappresentano due aspetti indivisibili della stessa cultura, anche sulla base del fatto che nessun sistema di *common law* si è mai sviluppato in totale autonomia dalla lingua inglese.¹³

Questo stretto legame spesso impedisce di trovare equivalenze adeguate a un discreto numero di termini, come ad esempio *welfare*, *trust*, *joint venture* e *fair*, portatori di un preciso significato per i giuristi di *common law*.

Al contrario, gli ordinamenti di *civil law* sembrano meno legati alla lingua in cui sono espressi. Essi derivano dalle codificazioni napoleoniche redatte in francese ed anche dalla codificazione tedesca, diffondendosi poi in tutta l'Europa continentale.¹⁴

Questo fenomeno dimostra la relazione che il diritto istituisce tra lingua e paese ed evidenzia la capacità che l'uomo ha di "padroneggiare almeno in parte l'evoluzione della propria lingua".¹⁵

Nonostante i notevoli ostacoli in tema di traduzione, la realtà dimostra che anche nel campo del diritto è possibile tradurre e trasportare concetti con ragionevole preci-

sione, raggiungendo risultati accettabili se si segue un approccio che fa riferimento alla cosiddetta equivalenza funzionale tra il testo di origine e quello di destinazione.

L'equivalenza funzionale consiste nel rapporto tra due termini i quali, essendo distinti l'uno dall'altro in ogni parte costituente, possono in qualche modo essere messi in corrispondenza in virtù di un elemento comune: nello studio del diritto comparato si parla di *extensional concepts*. In altre parole, si tratta di un'equivalenza semantica soltanto di base che implica l'uso prevalente di termini appartenenti alla lingua di destinazione in modo da ottenere maggiore familiarità e comprensibilità.¹⁶

Due sono le correnti di pensiero su questo tipo di equivalenza. Alcuni studiosi¹⁷ sostengono che un equivalente funzionale è accettabile solo se il concetto che esprime ha la stessa funzione generale del concetto di partenza.

Altra corrente¹⁸ afferma che l'equivalente funzionale è accettabile solo se il concetto corrispondente presenta lo stesso contenuto giuridico. Per la prima categoria di scienziati quindi, ad esempio, i termini utilizzati per indicare il trasferimento di proprietà mobiliare sono equivalenti in Italia, Francia, Austria, Germania, Inghilterra. Se invece si ritiene necessario che i concetti possiedano lo stesso contenuto giuridico, l'equivalenza non sarà possibile poiché, affinché il trasferimento di proprietà sia valido, ciascun ordinamento sopra menzionato prevede elementi e requisiti diversi.

Nella creazione di sistemi informativi occorrerebbe affidarsi all'approccio dell'equivalenza funzionale e condurre l'analisi giuridica non tanto a livello del diritto positivo, quanto basandosi sulla teoria generale del diritto. Questo tipo di indagine si serve necessariamente di definizioni più generali ed ha co-

me obiettivo lo studio delle somiglianze e delle differenze tra i sistemi in considerazione. Si possono distinguere cioè modelli generali, nozioni “astratte” (genotipi) e singole manifestazioni del diritto positivo di ciascun sistema (fenotipi). Ad esempio, il modello astratto del contratto è l'accordo; le sue singole manifestazioni nel diritto italiano sono costituite in gran parte dai requisiti elencati dagli articoli 1325 e ss. del Codice civile, mentre nel diritto inglese esse derivano dalla lunga serie di sentenze relative a offerta, accettazione, intenzione di vincolarsi.¹⁹

In realtà l'ambiguità tra modello astratto e singole manifestazioni può essere considerata un difetto del linguaggio giuridico, in quanto l'elasticità comporta una mancanza di precisione innegabilmente dannosa per il linguaggio di una disciplina. Questa mancanza di precisione è però, secondo la scienza della traduzione giuridica, uno dei fattori grazie ai quali è possibile la traduzione stessa e la trasposizione dei concetti. Solo grazie alla corrispondenza a livello generale è possibile parlare, nell'ottica comparatistica, di contratto, proprietà, matrimonio e, in definitiva, di diritto.

Uno sguardo ai thesauri giuridici multilingue

In questo contesto così complesso e in particolare nell'ottica di un accesso all'informazione giuridica multilingue sempre più auspicabile, i thesauri, in quanto sistemi di organizzazione della conoscenza costituiti da una raccolta di termini e dall'indicazione delle relazioni semantiche che intercorrono fra loro (gerarchiche, associative e di equivalenza), rappresentano una soluzione per garantire la corrispondenza fra l'interrogazione dell'utente e i documenti opportunamente indicizzati.²⁰ Se si dispone di the-

sauri multilingue l'operazione di traduzione è già avvenuta nella fase della loro preparazione e si può utilizzare così la funzionalità che mette in relazione i termini in ciascuna lingua con quelli corrispondenti in altra lingua.

Per rendere interoperabili determinati sistemi di conoscenza, perché di questo si tratta quando si mettono in relazione vocabolari controllati in una varietà di lingue, diversi sono gli approcci utilizzati, sulla base delle raccomandazioni elaborate a livello internazionale per la creazione di thesauri multilingue e per la loro interoperabilità. Queste riguardano la traduzione di un thesauro multilingue già esistente in una o più lingue, la fusione di diversi thesauri monolingue o la creazione di un thesauro plurilingue *ex novo*.

Forte è l'invito da parte degli esperti del settore ad adottare un vero approccio multilingue nella costruzione di thesauri²¹ che avvenga possibilmente in modo originario per non privilegiare la lingua fonte, garantendo così che la descrizione dei concetti sia ugualmente analitica e dettagliata in tutte le lingue trattate. Infatti un thesauro multilingue è molto di più dell'accorpamento di diversi thesauri monolingue; esso deve necessariamente adottare il principio di equiparazione fra lingue, offrire l'apparato concettuale e terminologico di ciascuna lingua considerata, presentando per ognuna le reciproche relazioni semantiche. Si tratta dunque di un lavoro estremamente complesso che richiede risorse consistenti.

Un esempio significativo di strumento multilingue è offerto da Eurovoc, il thesauro multilingue pluritematico dell'Unione europea, disponibile in 22 lingue ufficiali nonché nella lingua di un paese candidato (croato) e di un paese extra UE (serbo).²² Esso copre tutti i settori di attività dell'Unione e

quindi anche il diritto. Eurovoc rappresenta un ottimo esempio di dizionario concettuale con supporto alla gerarchizzazione di termini inerente all'ordinamento giuridico comunitario.²³ Il thesauro è strutturalmente impostato a livello generico su una classificazione gerarchica a due livelli (settori e microthesauri) e a livello specifico dei descrittori e non descrittori, su relazioni semantiche di diversa tipologia. Fra queste si distinguono la nota operativa o di definizione, la relazione di appartenenza al microthesauro, quella di equivalenza, di gerarchia e di associazione. Tutte le versioni linguistiche del thesauro comprendono quindi: settori, microthesauri, descrittori, relazioni gerarchiche reciproche e relazioni associative reciproche e sono rigorosamente equivalenti fra una lingua e l'altra.

Per poter ottenere una corrispondenza completa fra i termini in ogni lingua in cui Eurovoc è tradotto, ad ogni descrittore di una lingua corrisponde obbligatoriamente un descrittore in ciascuna delle altre lingue, mentre non ci sono corrispondenze fra le varie lingue per quanto riguarda i non descrittori. Il thesauro europeo è senz'altro un valido strumento per effettuare una ricerca giuridica multilingue. Tuttavia l'approccio seguito nella creazione e gestione della terminologia giuridica presenta alcune limitazioni. In particolare, ogni termine può essere tradotto in un singolo termine nelle altre lingue: si ha così una corrispondenza univoca. Ciò è accettabile in Eurovoc, in quanto esso contiene concetti appartenenti all'ordinamento europeo, ma non appare adeguato per la mappatura fra termini dell'ordinamento europeo verso termini di ordinamenti nazionali, i quali possono essere anche molto diversi, essere mappati verso più termini o non avere addirittura equivalenti. Eurovoc è comunque uno stru-

mento documentario di qualità che vanta un notevole utilizzo da parte del Parlamento europeo, dell'Ufficio delle pubblicazioni delle Comunità europee, dei parlamenti nazionali e regionali in Europa, di amministrazioni nazionali e di vari centri di documentazione, ma anche dei cittadini europei sempre più interessati ad avvicinarsi alla terminologia multilingue.

Un altro strumento da menzionare per il settore giuridico è il thesaurus giuridico trilingue svizzero del Tribunale federale e dell'ex Tribunale federale delle assicurazioni (attualmente: Corti di diritto sociale del Tribunale federale),²⁴ consultabile in tedesco, francese ed italiano ed elaborato in collaborazione con biblioteche giuridiche svizzere. Destinato all'indicizzazione delle leggi, della giurisprudenza e della dottrina, la sua terminologia è costituita essenzialmente da termini giuridici, ma vi sono contenuti anche termini propri di altre discipline (per esempio geografia e medicina), nella misura in cui si rivelano utili per la ricerca giuridica. È seguito uno schema di classificazione sistematico: i termini usati per l'indicizzazione con funzione di descrittori sono organizzati in gerarchie strutturate principalmente sul modello della raccolta sistematica del diritto federale. Da menzionare è infine il thesaurus di GLIN (Global Legal Information Network)²⁵ utilizzato dalla Biblioteca giuridica del Congresso di Washington che da sempre svolge una importante funzione di raccolta di materiale sul diritto nazionale e internazionale al servizio dei membri del Congresso americano, della Suprema corte degli Stati Uniti, della comunità di professionisti e accademici e del grande pubblico sia americano che internazionale.

Il sistema in rete GLIN²⁶ consta di un database pubblico che contiene leggi, disposizioni, decisioni giudiziarie ed altre fonti giuridiche

complementari provenienti dalle Americhe, Africa, Asia ed Europa, messe a disposizione da agenzie governative ed organizzazioni internazionali.²⁷ I membri di GLIN forniscono al database i testi completi ufficiali di documenti pubblicati in lingua originale.

Il servizio si caratterizza per la fornitura di informazioni autorevoli sulla legislazione, presenti nella lingua di origine in formato testo standard pdf. I documenti in lingua originale sono ricercabili in arabo, cinese, inglese, francese, tedesco, italiano, coreano, lituano, portoghese, romeno, russo, spagnolo e ucraino. La ricerca su parole chiave e voci di soggetto è espressa in una delle lingue di interrogazione consentite puntando ai termini contenuti nel thesaurus GLIN. La sua costruzione, a cui lavora un gruppo di esperti giuristi provenienti da varie giurisdizioni, segue un approccio che rispetta la varietà dei diversi sistemi giuridici. Per questa ragione nei casi in cui non è possibile trovare un esatto equivalente di termini o frasi che esprimono concetti propri di un determinato ordinamento, i termini figurano nella lingua originale.²⁸ GLIN rappresenta forse l'unica consistente soluzione disponibile per il trattamento dell'informazione giuridica multilingue, adottando una funzionalità di tipo *cross-language* che consente la ricerca in una lingua con recupero di documenti in lingue diverse.

Un primo studio sulla questione: l'applicazione THALEN - THesauri ALIGNment ENvironment

A compendio di questa breve analisi sulla problematica dei thesauri giuridici e sulla loro interoperabilità, viene di seguito brevemente presentato uno studio di fattibilità realizzato nell'ambito del progetto EUROVOC Studies – LOT2: Interope-

rability Feasibility Study, promosso dall'Ufficio delle Pubblicazioni dell'Unione europea (OPOCE TENDER n. 10118) che ha avuto come oggetto l'allineamento di thesauri per i servizi di *e-government* all'interno dell'Unione.²⁹

Il caso di studio ha riguardato la mappatura di cinque thesauri rilevanti per l'Unione europea:

- Eurovoc;
- ECLAS;³⁰
- Thesaurus dell'UNESCO;³¹
- European Training Thesaurus (ETT);³²
- General Multilingual Environmental Thesaurus (GEMET).³³

Eurovoc, come sopra illustrato, è il principale thesaurus europeo che copre 21 settori, con particolare attenzione agli aspetti di interesse specifico e comune per i paesi dell'UE. ECLAS, il thesaurus della biblioteca centrale della Commissione europea, copre 19 domini ed è particolarmente orientato al lavoro della Commissione, ai programmi comunitari e all'accesso alla loro documentazione. ECLAS è stato costruito per indicizzare documentazione sull'integrazione europea e gli obiettivi politici delle Comunità europee, monografie ed altri documenti sugli Stati membri, i loro sistemi giuridici e le loro condizioni socio-economiche, nonché rapporti tecnici necessari per i diversi servizi della Commissione.

GEMET è stato sviluppato come strumento di recupero, indicizzazione e controllo per l'European Topic Centre on Catalogue of Data Sources (ETC/CDS) e l'European Environment Agency (EEA) di Copenhagen. Le sue aree tematiche sono l'ambiente naturale e antropico, le misure di politica ambientale e gli effetti sull'ambiente.

Il Thesaurus dell'UNESCO è un vocabolario controllato sviluppato dall'organizzazione delle Nazioni Unite per l'educazione, la scienza e la cultura ed include termini di soggetto per i settori della conoscenza relativi a istruzione, scien-

za, cultura, scienze sociali e umane, informazione e comunicazione, politica, diritto ed economia.

Infine, l'European Training Thesaurus (ETT) è un thesaurus di 2.500 termini (1.550 descrittori, 950 non descrittori) redatto dall'European Centre for the Development of Vocational Training (Cedefop) che descrive il settore della formazione professionale in Europa identificando, tramite lemmi corredati di note, i soggetti, gli oggetti ed i processi della formazione professionale. ETT si articola in quattro aree: qualità e sistemi di formazione, apprendimento, qualifiche e certificazioni di competenze, orientamento professionale, ambiente di lavoro. I primi tre thesauri sono multidisciplinari, mentre gli ultimi due possono essere considerati thesauri specialistici.

L'obiettivo dello studio, affidato all'Istituto di teoria e tecniche dell'informazione giuridica del Consiglio nazionale delle ricerche (ITTIG-CNR),³⁴ è la valutazione dell'interoperabilità del thesaurus Eurovoc verso gli altri thesauri. La scelta di Eurovoc come referente è stata dettata dal fatto che l'Ufficio delle pubblicazioni si prefigge di offrire ai propri utenti la possibilità, partendo dal thesaurus europeo, di estendere le ricerche ad altre banche dati che sono indicizzate tramite strumenti diversi: si tratta infatti di realizzare una funzionalità del tipo *cross-collection retrieval*. A livello tecnologico la soluzione adottata tiene conto delle recenti direttive e studi sull'utilizzo di standard aperti, provenienti dalla comunità del W3C e della necessità di proporre soluzioni orientate al web semantico. In particolare, per descrivere i concetti di ciascun thesaurus, le loro relazioni, nonché le relazioni di mappatura fra concetti di thesauri diversi è stato adottato lo standard SKOS (Simple Knowledge Organisation System).³⁵ Quest'ultimo fornisce infatti un me-

todo standard per rappresentare i sistemi di organizzazione della conoscenza utilizzando una infrastruttura uniforme per la descrizione di risorse in rete e dei loro reciproci legami, denominata Resource Description Framework (RDF) che permette quindi di rappresentare strutture organizzate di concetti come thesauri, schemi di classificazione, liste di soggetto, tassonomie, ontologie. Codificare queste informazioni in RDF garantisce l'interoperabilità tra applicazioni. Organizzazioni concettuali rappresentate nei formati standard del semantic web con RDF consentono, in modo trasparente, l'annotazione di ulteriori dataset rispetto ai relativi descrittori secondo il modello Linked Open Data del W3C che si occupa di estendere il web tradizionale pubblicando dataset RDF liberi e aperti e mettendo in relazione tra loro dati provenienti da diverse fonti.

L'ITTIG inoltre ha sperimentato metodologie di mappatura automatica di supporto al lavoro manuale effettuato da esperti. In particolare, la metodologia proposta per la mappatura fra i thesauri consiste nell'applicazione di metodi di tipo lessicale (similarità fra stringhe), di tipo semantico (calcolo della similarità semantica fra rappresentazioni vettoriali dei concetti) di tipo strutturale (calcolo della similarità strutturale fra rappresentazioni a grafo dei concetti e delle loro relazioni).

La correttezza di tali metodologie automatiche di mappatura è stata testata confrontando il risultato della mappatura automatica con un insieme di mappature stabilite da esperti giuristi (*gold standard*). Tale base di conoscenza, frutto esclusivo di un lavoro intellettuale, rappresenta l'insieme ideale di mappature considerate corrette, ed è stata costruita sul settore del diritto del lavoro, in continua evoluzione e comprensivo di diversi a-

spetti trattati nei singoli thesauri in esame. Il *gold standard*, così come le strategie di mappatura da applicare, ha coinvolto i soli descrittori, utilizzando l'inglese come lingua di riferimento (*pivot*). Le relazioni di mappatura sono state descritte utilizzando SKOS, limitatamente alle relazioni di *broad-match*, *exact-match*, e *narrow-match*. L'attività di costruzione del *gold standard* è stata portata avanti da esperti giuristi dell'ITTIG, utilizzando un'interfaccia specifica per rendere più facile il loro lavoro. Per l'applicazione pratica sono state valutate alcune soluzioni come Protégé PROMPT,³⁶ Vocabulary Integration Environment (VINE)³⁷ e AIDA.³⁸ In un primo momento la soluzione AIDA Thesaurus Browser³⁹ è sembrata la più promettente. Si tratta infatti di un browser SKOS che supporta la navigazione in parallelo di due thesauri, la ricerca per concetto, la mappatura trasversale, nonché il cambiamento e la rimozione delle mappature del Mapping SKOS Vocabulary. Questo strumento è stato sviluppato nel contesto del progetto VL-e "Virtual Labs for e-Science".⁴⁰

Tuttavia per mantenere il controllo locale dei dati, necessario per i tempi e gli scopi della sperimentazione, nonché per la mancanza di alcune funzioni fondamentali per la creazione del *gold standard*, si è deciso di realizzare una specifica applicazione, denominata THesauri ALignment ENvironment (THALEN) che, attraverso un database relazionale, fornisce funzionalità per stabilire le relazioni tra termini. Il software permette di selezionare i thesauri (visualizzandoli contemporaneamente), di effettuare la ricerca per campi o per scorrimento, di stabilire le mappature (selezione dei termini e delle relazioni), di procedere alla esportazione in SKOS e alla fusione di thesauri. È possibile estendere tale applicazione ad uno sviluppo su altre

piattaforme creando una versione web e l'integrazione di essa con strumenti automatici di mappatura. Per quanto riguarda l'implementazione su altri domini non si può prescindere dalle specificità linguistiche e concettuali delle diverse discipline.

Conclusioni

L'interoperabilità tra thesauri è una funzionalità auspicabile in uno scenario come quello attuale in cui la domanda di accesso ad una pluralità di collezioni diversificate è sempre più forte. Di fatto la convergenza di vocabolari controllati consente di utilizzare un thesaurus particolare e di mapparlo con thesauri in altre lingue, con più vocabolari specializzati o con versioni diverse di un thesaurus per recuperare documenti di più collezioni coerentemente con l'interrogazione originale. Di conseguenza le funzionalità di *cross-collection retrieval* e di *cross-language retrieval* sono strettamente legate alla possibilità di rendere i thesauri fra loro interoperabili.

Negli ultimi anni l'Unione europea si è fatta promotrice fra le organizzazioni internazionali e sovranazionali di politiche di trasparenza e di apertura verso forme avanzate di diffusione dell'informazione. In particolare sono state sviluppate strategie e strumenti che facilitano l'accesso alle diverse fonti giuridiche in un ambiente distribuito, con l'obiettivo di promuovere la partecipazione attiva dei cittadini alle attività della società civile. Ma vi sono ancora limitazioni in questa direzione ed oggi molti thesauri giuridici che sono definiti multilingue, utilizzati per l'accesso per argomento, sono di fatto thesauri monolingue associati a dizionari bilingue o monolingue.⁴¹

Senza dubbio i thesauri multilingue sono strumenti complessi, che

non solo richiedono un lavoro oneroso di sviluppo e manutenzione, ma comportano competenze specifiche per l'utilizzo sia in fase di indicizzazione dei documenti che di ricerca. Questa difficoltà è soprattutto dovuta alle problematiche ben note insite nella traduzione, in particolare alla complessità nel far corrispondere concetti di una lingua con quelli espressi in un'altra quando sono trattati termini di un vocabolario controllato necessariamente concisi e potenzialmente avulsi dal contesto.⁴²

Resta il fatto che il recupero dell'informazione su più fonti richiede che ogni termine usato in fase di ricerca sia confrontato con i termini utilizzati nei diversi vocabolari, possibilmente attribuendo un peso a ciascun termine per stabilire la rilevanza dei risultati.

Dal momento che il recupero dell'informazione giuridica è strettamente condizionato dalla specificità dei vari ordinamenti e quindi dei concetti che ne stanno alla base, non è tanto la diversità di lingue con cui tali concetti vengono espressi in un contesto globalizzato che deve essere gestita, quanto la peculiarità degli istituti che sono il patrimonio storico e culturale dei vari ordinamenti. L'esame comparato tra concetti giuridici appartenenti ai vari ordinamenti e parallelamente lo studio sulla traduzione intesa come ricerca di equivalenti funzionali sono attività essenziali per raggiungere un soddisfacente livello di mediazione tra identità giuridiche diverse, assicurando così una comunicazione interculturale e allo stesso tempo valorizzando le diversità intese come punto di forza e di unione.

La sfida non è quella di operare una scelta su una determinata lingua di comunicazione o su un thesaurus unico di termini giuridici, ma di trovare un metodo per far coesistere e convivere armoniosamente le diversità culturali e linguistiche.

Alla luce delle riflessioni e delle esperienze sopra esposte si può affermare che non esiste un criterio unico per stabilire la corrispondenza fra concetti. Fermo restando che la precisione in tema di linguaggio giuridico è una prerogativa irrinunciabile per le conseguenze che inevitabilmente inesattezza e approssimazione comportano in ogni azione in cui è coinvolto il diritto, si può affermare che nei sistemi di recupero dell'informazione giuridica possono essere accettate, per la comparazione fra concetti, soluzioni in qualche modo meno rigide rispetto a quanto accade per la traduzione di testi giuridici, la cui inesatta formulazione in altra lingua può avere implicazioni assai gravi. Minore rigore non comporta l'abbandono del principio di chiarezza e precisione che sempre deve guidare ogni operazione di raffronto fra concetti giuridici appartenenti ad ordinamenti diversi, ma piuttosto significa l'adozione di un approccio flessibile. Ciò che si deve assicurare nei sistemi di ricerca giuridica è la funzione di comunicare la conoscenza attraverso la fornitura di documenti effettivamente rilevanti, operando anche dei compromessi. Questo significa ad esempio potersi fermare, nella comparazione, ad un livello più generale nell'espressione di un concetto rispetto alla specificità dell'originale ed anche ricorrere in modo sistematico a tecniche di disambiguazione dei termini presentando più risultati e interagendo con l'utente per la scelta del contesto.

L'approccio dell'equivalenza funzionale, da intendersi come svolgimento della stessa funzione generale del concetto di partenza, appare come il metodo più efficace e versatile per presentare all'utente documenti rilevanti in diverse lingue e variamente indicizzati.

Al fine di disporre di una base su cui costruire strumenti che per-

mettano ai cittadini di recuperare ed utilizzare le risorse informative giuridiche messe a disposizione da fonti di natura istituzionale e commerciale, sembra necessario avviare un lavoro congiunto che riunisca le diverse competenze di giuristi, traduttori, linguisti, giuristi linguisti, specialisti dell'informazione e studiosi delle nuove tecnologie e intraprendere studi di diritto comparato sempre più estesi sulle varie aree del diritto. In particolare, attraverso il lavoro di elaborazione di corrispondenze linguistico-concettuali tramite la cooperazione con organismi istituzionali di paesi attivi nel campo dell'indicizzazione e documentazione del diritto, è possibile avviare la sperimentazione di sistemi che garantiscano una migliore comprensione degli istituti giuridici propri dei singoli paesi.

Note

¹ ISO 2788-1986: Guide to establishment and development of monolingual thesauri; ISO 5964-1985: Guide to establishment and development of multilingual thesauri; ANSI/NISO Z.39.19-1993 (R2003): Guidelines for the construction, format and management of monolingual thesauri; BS 8723: Structured Vocabularies for Information Retrieval. London: British Standard Institution, 2007.

² ALBERT SCHRÖDER, *Cross concordances of classifications and thesauri*, WP12 of the CARMEN Project, <<http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en>>.

³ Renardus: Academic Subject Gateway Service Europe, <<http://www.uko.ln.ac.uk/metadata/renardus/>>.

⁴ PATRICE LANDRY, *Providing multilingual subject access through linking of subject heading languages: the MACS approach*, <<http://www.cacaoproject.eu/fileadmin/media/AT4DL/paper-09.pdf>>.

⁵ SUSANNAH WAKE – DENNIS NICHOLSON, *HILT – High Level Thesaurus Project: building consensus for interoperable*

subject access across communities, "D-Lib Magazine", 7 (2001), 9, <<http://www.dlib.org/dlib/september01/wake/09wake.html>>.

⁶ Unified Medical Language System – UMLS, <http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/>.

⁷ JANE KRAUSE, *Virtual libraries, library content analysis, metadata and the remaining heterogeneity*, in: *Proceedings of the the 3rd International Conference of Asian Digital Library (ICADL)*, Seoul, Corea, 2000.

⁸ PETER SANDRINI, *La terminologia giuridica: difficoltà di traduzione e elementi per una metodologia specifica*, in: *Tradurre italiano e tedesco. II: lessici settoriali a confronto*, Padova, Unipress, 1996, p. 57-73; FERDINAND DE SAUSSURE, *Corso di linguistica generale = Cours de linguistique générale*, Roma-Bari, Laterza, 2005.

⁹ GÉRARD SNOW – JACQUES VANDERLINDEN, *Français juridique et science du droit*, Bruxelles, Bruylant, 1995, p. 46.

¹⁰ BICE MORTARA GARAVELLI, *Le parole e la giustizia*, Torino, Einaudi, 2001, p. 4-18.

¹¹ ANDREA ORTOLANI, *Le lingue del diritto: nuove prospettive in tema di traduzione e interpretazione del diritto plurilingue*, "Rivista critica del diritto privato", 2003, 1, p. 203-216.

¹² FRANCESCO DE FRANCHIS, *Dizionario giuridico*, Vol. I: Inglese-Italiano, Milano, Giuffrè, 1984.

¹³ Cfr. GEORGE FLETCHER, *Comparative law as a subversive discipline*, "The American Journal of Comparative Law", 46 (1998), 4, p. 683-700.

¹⁴ Gli ordinamenti di *civil law* sono stati presi come modello, oltre che nei sistemi che utilizzano una lingua neolatina come l'Italia ed i paesi dell'America Latina, anche in paesi di tradizione giuridica e linguistica molto diversi tra loro, come ad esempio Grecia, Turchia, Russia e Giappone.

¹⁵ RODOLFO SACCO, *Riflessioni di un giurista sulla lingua uniforme (lingua del diritto uniforme e il diritto al servizio di una lingua uniforme)*, "Rivista di diritto civile", 1996, 1, p. 63.

¹⁶ MARTIN WESTON, *Theoretical and practical approaches to translation: the linguistic issues*, Part I, in: An English Reader's Guide to the French Le-

gal System, New York, Berg, 1990, p. 9-42.

¹⁷ DOROTY KENNY, *Equivalence*, in: *Routledge Encyclopedia of translation Studies*, London, Routledge, 1998, p. 77-80.

¹⁸ SUSAN ŠARČEVIĆ, *Bilingual and multilingual legal dictionaries: new standards for the future*, "Meta", 36 (1991), 4, p. 615-626.

¹⁹ Ciò è in linea con il modello Functional Requirements for Bibliographic Records (FRBR), <http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf>, con riferimento in particolare ai livelli di espressione e manifestazione, trovando anche una corrispondenza con i tipi di relazioni semantiche previsti in Functional Requirements for Subject Authority Data (FRSAD), <<http://nkos.slis.kent.edu/FRSAR/report090623.pdf>>.

²⁰ FATIHA SADAT – MASATOSHI YOSHIKAWA – SHUNSUKE UEMURA, *Exploiting thesauri and hierarchical categories in cross-language information retrieval*, "Lecture Notes in Computer Science", 2448 (2006), p. 139-146.

²¹ MICHÈLE HUDON, *Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts*, "Knowledge Organisation", 24 (1997), 2, p. 84-91; DAGOBERT SOERTEL, *Multilingual thesauri in cross-language text and speech retrieval*, in: Working notes of AAAI Symposium on Cross-Language Text and Speech Retrieval, 24-26 marzo 1997, p. 164-170.

²² Oltre alle versioni in queste lingue, sono disponibili traduzioni effettuate a cura dei parlamenti nazionali di diversi paesi come Albania, Russia e Ucraina.

²³ MARCO MARTIN, *Sviluppo di uno strumento per la gestione di dizionari giuridici basati sulle ontologie*, Torino, Università degli studi, 2005, <<http://www.notmart.org/files/tesi%20-%20Ontologie%20Giuridiche.pdf>>.

²⁴ <<http://www.bger.ch/it/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-jurivoc-home.htm>>.

²⁵ <<http://www.glin.gov/subjectTermIndex.action>>.

²⁶ <<http://www.glin.gov/search.action>>.

²⁷ JANICE S. HYDE, *Building the Global Legal Information Network (GLIN)*,

“Legal Reference Services Quarterly”, 19 (2001), 3-4, p. 157-173.

²⁸ Esempi sono il termine spagnolo “amparo” e “kafala”, quest’ultimo appartenente a giurisdizioni delle comunità arabe. Ambedue i termini richiamano il concetto di tutela, ma con precisi connotati propri dei rispettivi sistemi.

²⁹ ENRICO FRANCESCONI – SEBASTIANO FARO – ELISABETTA MARINAI, *A framework for semantic mapping between thesauri*, in: *Proceedings of the 2nd International Conference on theory and practice of electronic governance (ICE-GOV)*, Cairo, 1-4 dicembre 2008, a cura di T. Janowski, T.A. Pardo, p. 251-257.

³⁰ <<http://ec.europa.eu/eclas/>>.

³¹ <<http://databases.unesco.org/thesaurus/>>.

³² <<http://libserver.cedefop.europa.eu/ett/en/>>.

³³ <<http://www.eionet.europa.eu/gemet/>>.

³⁴ L’ITTIG studia e applica le tecnologie dell’informazione e della comuni-

cazione nei settori della documentazione del diritto e del linguaggio giuridico, della tecnica legislativa, della decisione giuridica, della formazione dei giuristi. Studia inoltre le tematiche connesse al diritto e alla politica dell’informazione pubblica e dell’informatica. Per ulteriori informazioni: <<http://www.ittig.cnr.it/>>.

³⁵ <<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>>.

³⁶ NATALYA NOY – MARK MUSEN, *The PROMPT Suite: interactive tools for ontology merging and mapping*, “International Journal of Human-Computer Studies”, 59 (2003), 6, p. 983-1024.

³⁷ LUIS BERMUDEZ, *Vocabulary Integration Environment: VINE*, “Avances en Sistemas e Informática”, 4 (2007), 2, p. 53-59.

³⁸ <<http://adaptivedisclosure.org/aida/docs/storage/>>.

³⁹ La scelta di AIDA si basa essenzialmente su due ragioni. La prima riguarda l’interfaccia basata su Web e l’architettura distribuita client-server, ideale per un lavoro in collaborazio-

ne. La seconda fa riferimento al fatto che essa è in grado di gestire thesauri in formato SKOS. L’applicazione è basata su Open Office Calc, per garantire la disponibilità dell’interfaccia su diverse piattaforme (Windows, Linux, etc.), mentre un apposito foglio di stile XSL è stato utilizzato per trasformare il formato XML RDF prodotto in SKOS Mapping.

⁴⁰ <<http://www.vl-e.nl/>>.

⁴¹ KERSTIN JORNA – SYLVIE DAVIES, *Multilingual thesauri for the modern world: no ideal solution?*, “Journal of Documentation”, 57, (2001), 2, p. 284-95.

⁴² OLIVER STREITER, *A semantic description language for multilingual NLP*, 1998, <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.3573&rep=rep1&type=pdf>>; HERVÉ DÉJEAN – ERIC GAUSSIER – FATIA SADAT, *Bilingual lexicon extraction: using and enriching multilingual thesauri*, in: *Proceedings of Terminology Knowledge Extraction (TKE)*, Nancy, August 2002, <<http://muchmore.dfki.de/pubs/dejean.pdf>>.

Abstract

Nowadays the opportunity to access heterogeneous information resources is enormously increased due to the availability, in the global network, of large collections from various institutional, commercial, public and private sources. Hence the need to effectively deal with the issues arising from the use of various indexing tools, namely controlled vocabularies (thesauri, classification schemes, subject strings, taxonomies, authority lists, ontologies) which may considerably differ in their language, aim and level of abstraction and detail. In particular, this paper focuses on the interoperability among thesauri in the legal domain, investigating the possibility of creating mappings among terms and as more accurate matches as possible among the concepts expressed by the terms themselves. Nowadays this is a much-debated topic in the area of jurilinguistics, which seeks to define methods for improving the quality of legal texts, while taking into account the complexity and shades of meaning of each system as well as the unique character of each language. It is not so much the diversity of languages – in which those concepts are expressed – rather than the comparison of regulatory schemes forming the historical and cultural heritage of the various legal orders that has to be managed. Surely the interoperability between thesauri is a possible operative solution, in so far as it meets two important and different functions: cross-collection retrieval and cross-language retrieval. In both cases the peculiarities of legal language are to be seriously taken into account to reach high quality mapping. Comparative studies concerning concepts belonging to different legal orders and translation studies oriented to find “functional equivalents” are essential measures to achieve a satisfactory level of mediation between different legal identities, thus ensuring intercultural communication. After addressing the general issue of interoperability among thesauri and then focusing on legal terminology and semantics of the law, one of the first feasibility studies carried out in this domain is presented. The study, as part of the EUROVOC Studies – LOT2, is sponsored by the Office of Publications of the European Union. It aims at aligning thesauri used in e-government services within the EU. New methods of automatic mapping have also been tested to support the legal experts’ intellectual mapping activity.