



UNIVERSITÀ DEGLI STUDI DI SALERNO

**Dipartimento di Scienze Politiche, Sociali e della Comunicazione**

Corso di Dottorato in Scienze della Comunicazione, Sociologia, Teorie e Storia  
delle Istituzioni, Ricerca Educativa, Corporeità Didattiche, Tecnologie e  
Inclusione, XXIX ciclo

Tesi di dottorato

## **Information Technology per IATE, la banca dati terminologica multilingue dell'Unione Europea**

Candidato: Francesco Rossi

matr. 8887400011

Tutor:  
Prof.ssa  
*Daniela Vellutino*

Coordinatore:  
Prof.  
*Annibale Elia*

**Anno Accademico 2016/2017**



# Sommario

Abstract .....	VI
INTRODUZIONE .....	1
CAPITOLO I – QUADRO TEORICO .....	6
1.1 Wüster e le origini della terminologia .....	6
1.2 Terminologia: status autonomo o dipendente da altre discipline? .....	9
1.3 Approcci teorici - La socioterminologia .....	13
1.4 Approcci teorici - L’approccio (socio)cognitivo .....	16
1.5 Approcci teorici - “Frame-Based Terminology” .....	19
1.6 Approcci teorici - l’approccio comunicativo .....	22
1.7 La terminologia e la traduzione .....	27
1.8 La terminologia e l’informatica .....	31
CAPITOLO II – LA TERMINOLOGIA NEL PARLAMENTO EUROPEO E L’ATTIVITÀ DI TERMINOLOGY COORDINATION UNIT .....	38
2.1 La Direzione Generale per la Traduzione - DGTRAD .....	39
2.2 TermCoord – Terminology Coordination Unit e i tool di terminologia .....	46
2.2.1 Cenni storici .....	47
2.2.2 Il ruolo di TermCoord nella Direzione Generale per la Traduzione .....	50
2.2.3 Terminology Macro e TermSafe .....	52
2.2.4 DocHound e GlossaryLinks .....	57
2.2.5 European Advanced Multilingual Information System - <i>Euramis</i> e <i>Quest</i> .....	60
2.2.6 TermFolder .....	63
2.2.7 Progetti di terminologia dei tirocinanti .....	64
2.2.8 La comunicazione di TermCoord .....	66
2.2.9 EurTerm: una piattaforma collaborativa per traduttori, terminologi, esperti di dominio e comunicatori .....	67
CAPITOLO III – IATE, INTER-ACTIVE TERMINOLOGY FOR EUROPE .....	70
3.1 Cenni storici .....	71
3.2 IATE – principi generali .....	74
3.3 La struttura delle schede terminologiche di IATE .....	77
3.3.1 Language-Independent Level .....	79
3.3.2 Language Level .....	83
3.3.3 Term Level .....	86
3.4 Fonti .....	90
3.4.1. Tipi di fonti .....	91
3.5 Validazione e Consolidamento .....	93

3.6 Sviluppi futuri .....	95
3.6.1 IATE e API .....	95
3.6.2 IATE 2 .....	98
CAPITOLO IV – PROGETTI DI RICERCA PER L’INTEROPERABILITÀ SEMANTICA DI IATE .....	100
4.1 Il web semantico e i Linked Data .....	100
4.2 L’interoperabilità semantica .....	103
4.3 Il linguaggio di marcatura TermBase eXchange - TBX .....	107
4.4 Lo standard di riferimento dell’interoperabilità semantica: gli scopi di ISO/TC 37/ SC 3108	
4.5 Progetti di ricerca per l’interoperabilità semantica di IATE .....	111
4.5.1 Il progetto LIDER - <i>Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe</i> .....	111
4.5.2 Il progetto BabelNet .....	114
4.5.3 Il progetto EuroTermBank .....	116
CAPITOLO V - L’INFORMATICA E LA TERMINOLOGIA PER IATE: DALLE TERMBASE VERSO L’INTEROPERABILITÀ SEMANTICA .....	120
5.1 SDL Studio, una nuova sfida per IATE – il progetto “Terminology in Studio” .....	123
5.1.1 Terminology in Studio, fase 1 – IATE Advanced Export Interface .....	127
5.1.2 Terminology in Studio, fase 2 - Term Recognition Module .....	135
5.1.3 Terminology in Studio – manual termbases .....	139
5.2 Accordo di collaborazione tra il DSPSC e Terminology Coordination Unit del Parlamento europeo .....	143
5.3 URI ed ELI in IATE, un primo passo per l’interoperabilità semantica .....	149
5.4. I termini di MWM Lexicon in IATE .....	152
5.5 Da una scheda IATE di qualità ad una proposta per l’interoperabilità semantica .....	166
CONCLUSIONI .....	172
APPENDICE - ANALISI DI SCENARIO DELLE TECNOLOGIE DELLA GESTIONE ED ESTRAZIONE DELLA TERMINOLOGIA .....	178
A.1 La gestione della terminologia e il rapporto con la traduzione .....	178
A.2 L’estrazione terminologica .....	183
A.3 Schede analitiche di estrattori terminologici .....	185
Bibliografia .....	198
Documenti Unionali .....	204
Sitografia .....	207
Estrattori Terminologici Online .....	214
Ringraziamenti .....	216



## Abstract

*L'Information Technology* è motrice e matrice di cambiamento e innovazione. Con essa, discipline e scienze di ogni dominio evolvono contaminandosi e, tra queste, vi è la terminologia.

InterActive Terminology for Europe (IATE) è la banca dati terminologica, interistituzionale e multilingue dell'Unione Europea realizzata con l'obiettivo di offrire un'infrastruttura centralizzata per tutte le risorse terminologiche dell'UE, migliorandone la disponibilità e la standardizzazione.

Obiettivo di questa ricerca è investigare le modalità per migliorare le risorse linguistico-terminologiche di IATE.

In questo elaborato si presenteranno gli sviluppi delle tecnologie per l'informazione (IT) orientati alla gestione della terminologia della banca dati IATE prodotta dalle istituzioni europee. Ci si concentrerà sui progetti realizzati nella Direzione Generale per la Traduzione del Parlamento europeo mirati al miglioramento della terminologia unionale tramite l'uso delle *termbase* e sull'accordo di cooperazione tra il Dipartimento di Scienze Politiche, Sociali e della Comunicazione dell'Università di Salerno e Terminology Coordination Unit del Parlamento europeo, per rendere la terminologia del database uno strumento utile per l'interoperabilità semantica, ovvero l'abilità di un sistema informatico di elaborare e scambiare informazioni mantenendone il significato preciso e a prescindere dal modo e dal mezzo in cui viene rappresentato.

Nel primo capitolo si presenteranno alcuni dei maggiori approcci teorici della letteratura scientifica in materia, mostrandone le origini, le diverse prospettive, metodologie e finalità. Si illustrerà inoltre il rapporto della terminologia con la traduzione e l'informatica, con una particolare attenzione all'utilità e allo sviluppo di *termbase*, ovvero i database, spesso multilingui, caratterizzati da entrate contenenti informazioni sui concetti di uno o più domini e sui termini che li designano.

Nel secondo capitolo si esporrà l'attività dell'Unità di Coordinamento della Terminologia - *Terminology Coordination Unit* - (TermCoord) del Parlamento europeo, che opera all'interno della Direzione Generale per la Traduzione (DGTRAD). Saranno descritti gli strumenti impiegati dai traduttori e dai terminologi del Parlamento europeo per la gestione della terminologia.

Nel terzo capitolo si analizzerà la banca dati terminologica multilingue delle istituzioni e degli organismi dell'Unione europea IATE. Dopo alcuni cenni storici, si illustreranno i principi generali, la struttura, le raccomandazioni e le buone pratiche per la creazione delle entrate terminologiche e per la loro gestione. Si concluderà il capitolo con gli sviluppi futuri per migliorare questo enorme database terminologico, che ad oggi conta 8,5 milioni di termini nelle 24 lingue ufficiali dell'UE nella versione accessibile al pubblico e 8,6 milioni di termini in 125 lingue in quella interna (che comprende termini in altre

101 lingue non ufficiali dell'UE) ad uso dei traduttori e terminologi delle istituzioni dell'Unione europea.

Nel quarto capitolo si presenteranno le iniziative e i programmi promossi dall'Unione Europea per l'interoperabilità semantica e si descriveranno i progetti di ricerca dell'industria e dell'accademia mirati all'interoperabilità semantica di IATE.

Nel quinto capitolo si mostrerà come gli strumenti dell'informatica e la terminologia possono contribuire all'interoperabilità semantica di IATE. Si analizzeranno i più recenti sviluppi nella cornice delle istituzioni europee, con particolare attenzione al Parlamento europeo, relativi all'utilizzo della terminologia di IATE negli strumenti di traduzione. Si descriveranno gli sviluppi del progetto "Terminology in Studio", mirato al miglioramento della coerenza terminologica e della qualità dei documenti unionali tramite l'impiego di termbase nel software di traduzione SDL Studio.

Si presenteranno inoltre le attività del progetto nato dall'accordo di collaborazione tra il Dipartimento di Scienze Politiche, Sociali e della Comunicazione dell'Università degli Studi di Salerno con Terminology Coordination Unit del Parlamento europeo, finalizzato alla redazione di schede terminologiche che contengono elementi a sostegno dell'interoperabilità semantica di IATE.

Nell'appendice si proporranno schede analitiche che presentano le attuali tecnologie per l'estrazione terminologica, considerando il rapporto tra la gestione della terminologia e la traduzione.





# INTRODUZIONE

Nel XVIII secolo, durante la simultanea espansione della conoscenza e dello sviluppo della tecnologia, la terminologia fu vista come uno strumento per superare alcune delle difficoltà associate a questi molteplici cambiamenti<sup>1</sup>.

La necessità di attribuire un termine ad un concetto scientifico si sviluppò a partire dall'internazionalizzazione delle scienze e dal bisogno degli scienziati di disporre di regole e di nomenclature specifiche e condivise nei vari campi di ricerca. Oggi, in un contesto in cui le discipline raggiungono altissimi livelli di specializzazione, il bisogno di una terminologia condivisa e della sua gestione accurata rappresenta un requisito fondamentale.

In ciò, un ruolo cruciale è giocato dall'informatica, che ha rivoluzionato il campo della terminologia e ha permesso di eseguire azioni precedentemente impensabili. Basti pensare, ad esempio, che nell'Unione Europea, negli anni '80, il lavoro terminologico veniva eseguito a mano e archiviato in faldoni il cui contenuto veniva filtrato e inserito in grossi glossari cartacei che venivano poi messi a disposizione di ogni traduttore in scatole grigie presenti su ogni scrivania.

Nell'era dei computer, gli sviluppi in campo informatico hanno portato, in una prima fase, alla realizzazione dei primi database istituzionali e, in una seconda fase, alla loro

---

<sup>1</sup> Rey A., (1995), *"Essays on Terminology"*.

convergenza in un unico database interistituzionale, che nel 2004 prese il nome di InterActive Terminology for Europe, meglio conosciuto con la sua abbreviazione: IATE.

IATE è il più grande database terminologico dell'Unione Europea: contiene oggi più di 8,6 milioni di termini nelle 24 lingue ufficiali dell'UE.

Oggi molti servizi pubblici sono online e la digitalizzazione delle risorse non è certamente una sorpresa. Molti di questi servizi hanno connessioni tra di loro che potrebbero essere rese più efficienti se si rimuovessero le barriere tra servizi, sistemi informatici e dati.

In altre parole, non basta più che i servizi pubblici siano digitali, ma devono essere anche interoperabili. L'interoperabilità ha l'obiettivo di:

*«permettere la presentazione e il trattamento delle informazioni in modo uniforme nei diversi sistemi gestionali, indipendentemente dalla tecnologia, dall'applicazione o dalla piattaforma utilizzate. La piena interoperabilità comprende la capacità di interoperare in termini di contenuto (semantico), formato (sintassi) e trasmissione. L'interoperabilità semantica vuol dire che il significato preciso dell'informazione scambiata è mantenuto e compreso correttamente e senza ambiguità, a prescindere dal modo in cui viene rappresentato fisicamente o trasmesso»<sup>2</sup>.*

---

<sup>2</sup> Proposta di direttiva del Parlamento europeo e del Consiglio relativa alla fatturazione elettronica negli appalti pubblici, COM (2013) 449 final.

L'interoperabilità semantica va inquadrata nell'European Interoperability Framework (EIF)<sup>3</sup>, che fornisce linee guida sull'allestimento di servizi digitali pubblici interoperabili.

Il programma sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA) (2010-2015)<sup>4</sup> e il nuovo programma ISA<sup>2</sup> (2016-2020)<sup>5</sup> sono i principali strumenti attraverso cui sono stati attuati la strategia europea per l'interoperabilità e il quadro europeo di interoperabilità vigenti.

La terminologia può essere uno strumento per l'interoperabilità semantica nella misura in cui essa possa essere utilizzata negli schemi di metadatozione sviluppati dai programmi europei ISA e ISA<sup>2</sup>. Applicando ciò al contesto di IATE, come sostiene Vellutino (2016),

*«i termini usati dagli applicativi per l'e-government nelle diverse lingue naturali potrebbero essere indicizzati, metadati e collegati alle schede terminologiche IATE che ne descrivono il significato e l'uso nei vari contesti istituzionali; le schede terminologiche di IATE potrebbero a loro volta diventare un hub per collegare i termini ai documenti istituzionali in cui sono presenti. Le schede terminologiche IATE devono essere, però, di qualità: costantemente aggiornate, complete delle*

---

<sup>3</sup> European Interoperability Framework (EIF) o Quadro europeo di interoperabilità (QEI), è parte della Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni (COM (2017)134).

<sup>4</sup> Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del settembre 2009, sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA).

<sup>5</sup> Decisione (UE) 2015/2240 del Parlamento europeo e del Consiglio, del 25 novembre 2015, che istituisce un programma sulle soluzioni di interoperabilità e quadri comuni per le pubbliche amministrazioni, le imprese e i cittadini europei (programma ISA<sup>2</sup>) come mezzo per modernizzare il settore pubblico.

*varianti dei termini presenti nei diversi documenti istituzionali, collegate alla documentazione unionale»<sup>6</sup>.*

Alla luce di ciò, si presenteranno in questo elaborato le attività relative all'accordo di cooperazione tra il Dipartimento di Scienze Politiche Sociali e della Comunicazione dell'Università degli Studi di Salerno e l'Unità di Coordinamento Terminologico del Parlamento europeo, Direzione Generale della Traduzione.

Il progetto di cooperazione ha l'obiettivo di redigere schede terminologiche di qualità per IATE affinché possano diventare strumenti a supporto dell'interoperabilità semantica.

Si illustrerà, inoltre, lo scenario attuale della terminologia istituzionale europea, con particolare attenzione al contesto del Parlamento europeo e alle attività dell'Unità di Coordinamento Terminologico. Si analizzerà in dettaglio la struttura della banca dati terminologica multilingue IATE e le relative metodologie di inserimento, modifica e aggiornamento dei termini e si mostreranno gli ultimi sviluppi relativi all'uso della terminologia del database negli strumenti di traduzione, che mirano al miglioramento della qualità e della coerenza dei documenti unionali.

Un'attenzione particolare sarà rivolta alle attività del Parlamento europeo riguardo all'impiego di nuovi strumenti tecnologici, quali le termbase, nei processi traduttivi e al loro utilizzo in relazione alla gestione delle risorse terminologiche.

---

<sup>6</sup> Vellutino D., Maslias R., Rossi F., (2016), *“Verso l'interoperabilità semantica di IATE, Studio preliminare per il dominio «Gestione dei rifiuti urbani»”*, in *Terminologie specialistiche e diffusione dei saperi*, a cura di Zanola M. T., Diglio C., Grimaldi C., pp.221-240.



# CAPITOLO I – QUADRO TEORICO

## 1.1 Wüster e le origini della terminologia

La terminologia, come disciplina che indaga sui termini specialistici, non è un campo di studio nuovo, ma si può ritenere che solo negli ultimi decenni sia stata sviluppata in piena considerazione dei suoi principi e metodologie.

Maria Teresa Cabré<sup>7</sup> (1998) osserva che già a partire dal XVIII secolo, la ricerca in campo chimico (ad esempio Lavoisier) e in botanica e zoologia (Linneo) cominciava a rendere emergente la necessità di attribuire dei nomi a concetti scientifici.

Nel secolo successivo, con l'internazionalizzazione delle scienze, per gli esperti si faceva sempre più chiara l'esigenza di disporre di una serie di regole per formulare termini nelle rispettive discipline.

In quegli anni, gli scienziati erano gli artefici, fautori e destinatari della terminologia e fu poi a partire dal 1900 che la terminologia divenne di gran interesse anche per ingegneri e tecnici, poiché la tecnologia si sviluppava rapidamente ed occorreva non solo nominare nuovi concetti e oggetti, ma anche condividerli nelle rispettive comunità.

---

<sup>7</sup> Cabré M.T. (1998), *“Terminology – Theory, methods and applications”*, p. 3.

Si crearono così le condizioni per la nascita della terminologia moderna, di cui in letteratura si riconosce Eugen Wüster (1898 – 1977) come padre fondatore.

Pierre Auger<sup>8</sup> identifica quattro periodi principali dello sviluppo della terminologia moderna:

1. *Le origini* (1930-1960), caratterizzato dalla descrizione dei metodi per la formazione dei termini;
2. *La strutturazione del campo* (1960-1975), in cui avvennero le maggiori innovazioni tecnologiche in campo informatico e di tecniche documentative e i primi approcci mirati alla standardizzazione della terminologia;
3. *Il boom* (1975-1985), in cui alcuni Stati (quali ad esempio URSS e Israele), riconoscevano l'importanza della terminologia e creavano le prime politiche linguistiche e, allo stesso tempo, la diffusione dei *personal computer* apportò una svolta nei metodi di gestione dei dati terminologici;
4. *Il presente* (1985-oggi), in cui l'informatica rappresenta una delle forze motrici principali dietro ai cambiamenti e alle evoluzioni della terminologia. È in questo periodo che emerge il mercato dell'industria linguistica e la cooperazione internazionale si amplia e consolida.

Negli anni '30, nella sua tesi di dottorato, Wüster presentò argomenti per sistematizzare metodologie di terminologia, descrivendone una serie di principi e illustrando i punti principali per la gestione dei processi dei dati terminologici. Egli era concentrato particolarmente sulla metodologia e sugli standard in contrapposizione alla teoria,

---

<sup>8</sup> Auger P., (1988), *“La terminologie au Quebec et dans le monde, de la naissance à la maturité”*.

poiché considerava la terminologia uno strumento da impiegare per eliminare l'ambiguità dalla comunicazione tecnico-scientifica.

Nei trent'anni seguenti alla sua tesi di dottorato, tali metodi di compilazione e standardizzazione vennero applicati e raccolti nell'opera "*the Machine Tool*" (1968), e solo successivamente l'accademico austriaco si concentrò sugli aspetti della teoria dei termini<sup>9</sup>.

Fu quindi a partire dagli studi di Wüster che nacque il primo impianto teorico generale in ambito terminologico, noto come *Teoria Generale della Terminologia (TGT)*.

La TGT indaga sul significato dei concetti, sulla loro distinzione e natura, sulle relazioni tra i concetti e quella tra termini e concetti.

La sfera concettuale è vista come indipendente da quella dei termini e, poiché ai concetti viene attribuito un predominio sui termini, ogni lavoro terminologico comincia da essi.

La TGT, per distinguere chiaramente la terminologia dalla lessicografia, sostiene che i termini vadano assegnati ai concetti e adotta quindi un approccio onomasiologico, a differenza di quello adottato in lessicografia, che segue la direzione opposta<sup>10</sup>.

Negli anni, la *TGT* ha avuto numerosi sostenitori, da cui è stata ulteriormente sviluppata. Cabré (2003) definisce l'insieme di queste attività di sviluppo come "*Extended general theory*"<sup>11</sup>, mentre Myking (2001), ne riassume i punti cardine, ovvero: l'indipendenza dei

---

<sup>9</sup> Nel 1969 Wüster pubblicò "*Die vier Dimensionen der Terminologierarbeit*", in cui presentava gli aspetti del lavoro terminologico mentre il suo approccio complessivo alla teoria fu pubblicato, postumo, nel 1979 in "*Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie*".

<sup>10</sup> Cabré M.T., (1998), *cit.* p. 7.

<sup>11</sup> Cabré M.T., (2003) "*Theories of Terminology. Their description, prescription and explanation, in Terminology*", p. 175.



concetti, l'approccio operativo onomasiologico e il problema/obiettivo della standardizzazione<sup>12</sup>.

La Teoria generale ha avuto altrettanti detrattori, come ad esempio Felber<sup>13</sup> (1984) e Kaguera<sup>14</sup> (2002) che criticano la precedenza dei concetti sui termini e la mancanza di appropriati approfondimenti sulla loro relazione.

## 1.2 Terminologia: status autonomo o dipendente da altre discipline?

Non è una coincidenza che la terminologia si sia sviluppata nel XX secolo grazie agli studi di scienziati e tecnici. Una disciplina e la sua metodologia si sviluppano a partire da un bisogno specifico e dunque, nel caso della terminologia, la crescita accelerata di scienza e tecnologia in tempi recenti ha determinato la creazione di un gran numero di nuovi concetti e campi che necessitano di nomi e accordi condivisi su di essi.

Quando, a partire dagli studi di Wüster negli anni '30, il lavoro terminologico cominciò a prendere consistenza e si intraprese uno studio dei termini più organizzato, si svilupparono simultaneamente diverse scuole di terminologia.

---

<sup>12</sup> Myking J., (2001) *"Against Prescriptivism? The Sociocritical Challenge to Terminology"*, p. 49-64

<sup>13</sup> Felber H., (1984), *"Terminology manual"*.

<sup>14</sup> Kaguera K., (2002), *"The dynamics of terminology"*.

Maria Teresa Cabré (1998) ne identifica tre e le definisce come le “tre scuole classiche della terminologia”, ovvero quella Austriaca, Sovietica e Ceca. I tre poli adottano approcci diversi, non mutuamente esclusivi, alla terminologia:

- la scuola austriaca considera la terminologia come interdisciplinare ma comunque un campo di studi autonomo al servizio delle discipline tecniche e scientifiche;
- la scuola sovietica si concentra sulla filosofia, che è principalmente interessata alla classificazione logica dei concetti e all’organizzazione della conoscenza;
- la scuola ceca si concentra sulla linguistica, come considera la terminologia come una componente del lessico del linguaggio e le lingue speciali come sottosistemi del linguaggio generale<sup>15</sup>.

In generale, va riconosciuta l’ampia varietà di approcci della terminologia e la sua multidimensionalità.

In quest’ottica, è particolarmente opportuna una riflessione di Robert Dubuc (1985):

*«As with any relatively new concept, terminology is subject to change, which depends on the theorists and specialists that practise it. The first reason for change derives from the relationship that terminology has with its preceding disciplines, especially with semantics, lexicology and lexicography. Some people believe that terminology is nothing more than a new perspective adopted by these older disciplines. Terminology is currently seen as an art or practice rather*

---

<sup>15</sup> Cabré M.T., (1998), *cit.* p.7.

*than as a science. Even though it has a well-defined aim, namely to satisfy the expressive needs of its users, its working methods are mainly empirical. Theoretical research and the refinement of the process of recognition, analysis and creation of terms must improve before terminology can be placed among the sciences deriving from linguistics»<sup>16</sup>.*

Secondo lo studioso canadese, la teoria dei termini e la pratica terminologica conducono a posizioni diverse. Si possono identificare due estremi: quello che stabilisce che la terminologia sia una disciplina separata con la propria teoria; quello secondo cui la terminologia debba i suoi assunti teorici ad altre discipline più consolidate.

Quest'ultima posizione è presa da Juan C. Sager (1990) che, nonostante non attribuisca alla terminologia uno status indipendente, riconosce che la pratica terminologica ha fondamenta teoriche:

*«There is no substantial body of literature which could support the proclamation of terminology as a separate discipline and there is no likely to be. Everything of import that can be said about terminology is more appropriately said in the context of linguistics or information science or computational linguistics. We see terminology as a number of practices that have evolved around the creation of terms, their collection and explication and finally their presentation in various printed and electronic media. Practices however well-established, do not constitute a discipline, but there is no denying a long history of methodologies*

---

<sup>16</sup> Dubuc R., (1985) *“Manuel pratique de terminologie”*.

*which themselves require theoretical underpinnings to justify their distinctive nature. Disciplines establish knowledge about things and as such are justified in their own right; methodologies are only means to an end, in the case of terminology, how to do things»<sup>17</sup>.*

Seguendo Sager, si può dunque dire che, sebbene la terminologia manchi di un'epistemologia, è purtuttavia caratterizzata da metodi e principi finalizzati specificamente ai suoi obiettivi, quali la facilitazione della comunicazione tra specialisti e professionisti e l'ottimizzazione della standardizzazione del linguaggio.

Questa "svolta" pragmatica, che ha a che fare con requisiti comunicativi reali, è anche dovuta al cambio di direzione verificato nelle discipline scientifiche e nella società. Si è passati dalla "ricerca della forma corretta" ad un approccio più pragmatico e funzionale, a cui corrisponde anche un graduale abbandono degli approcci prescrittivi.

Si può quindi asserire che la terminologia può essere ben definita in considerazione delle proprie attività pratiche e in relazione ai vari campi e discipline in cui si applica. Puntuale in questo senso la considerazione di Roger Goffin (1985):

*«As an intersectional and multidisciplinary science, terminology is located at the crossroads of a large number of disciplines, but it is not their preserve. [...] Terminology is closely linked to an activity carried out within the field of knowledge and thus is inseparable from its social context and its obvious applications»<sup>18</sup>.*

---

<sup>17</sup> Sager J., (1990) "A practical course in Terminology Processing".

<sup>18</sup> Goffin R., (1985) "La Science Terminologique".

A partire dalla teoria wüsteriana, molteplici sono stati gli approcci che si sono sviluppati recentemente. Evidentemente, non se ne può fornire che una sintesi, proprio in ragione dello status multidimensionale della terminologia.

A partire dagli anni '80, i ricercatori hanno constatato sempre più chiaramente i limiti delle teorie terminologiche classiche e hanno cominciato a privilegiare gli approcci descrittivi che si interessano all'uso dei termini nei differenti contesti comunicativi.

Questi nuovi approcci considerano il contesto del termine come fondamentale, superando il triangolo tradizionale concetto-termine-oggetto.

Si passeranno dunque in rassegna alcuni dei maggiori approcci teorici della terminologia, ovvero i filoni socioterminologico, sociocognitivo, la teoria dei frame e l'approccio comunicativo.

Di essi, si presenteranno alcuni degli assunti che criticano, o talvolta capovolgono, quelli della teoria classica.

### 1.3 Approcci teorici - La socioterminologia

La socioterminologia offre un modo di studiare la normalizzazione del linguaggio e la produzione dei termini di cui gli elementi contestuali e sociali giocano un ruolo prominente.

La socioterminologia si definisce in relazione con la terminologia e la sociolinguistica e sostiene che la lingua e la terminologia debbano essere analizzati e compresi nei contesti appropriati, come le società. Ciò differisce dal modello ideale Wüsteriano che promuove un uso prescrittivo e universale della lingua.

L'approccio classico è dunque non sufficiente e non efficiente nel considerare le strutture sociali e le funzioni terminologiche; la socioterminologia, invece, considera la terminologia come un fenomeno *"effectively at the service of the society"*<sup>19</sup> e ridiscute i fondamenti stessi della teoria classica<sup>20</sup>. In altre parole, la socioterminologia prova a ricondurre lo studio della terminologia verso uno studio dell'uso reale del linguaggio, promuovendo approcci descrittivi.

*«Son objet est l'amélioration des connaissances sur le discours spécialisés, scientifiques et techniques, sur les aménagements de ces discours, sur les circonstances de leur élaboration (le codage et l'encodage) et sur celles de leur saisie (le décodage) et de leur reformulation»*<sup>21</sup>.

Anzitutto, l'approccio descrittivo incorpora lo studio della sinonimia e della polisemia che contrasta gli ideali monosemici delle scuole tradizionali.

---

<sup>19</sup> Rey A., (1998), «Préface», Le MicroRobert, Langue française plus noms propres, chronologie, cartes, Paris, Le Robert.

<sup>20</sup> Auger P., (1999), *"L'implantation des officialismes halieutiques au Québec: essai de terminométrie"* p. 24.

<sup>21</sup> Boulanger J-C., (1995), *"GAUDIN, François (1993) : Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles"* p. 199.

La socioterminologia critica, inoltre, l'esistenza di limiti chiari tra i domini. *“La pureté sectorielle n'existe pas”*<sup>22</sup> e quindi sarebbe preferibile considerare le scienze e le tecnologie come “nodi di conoscenza”<sup>23</sup>.

I terminologi e i linguisti non possono tagliare la conoscenza in parti omogenee che sono ben definite e protette da influenze esogene.

Per Yves Gambier *«un “domaine” est constamment le résultat de la dialectique entre l'intégration (interdisciplinaire) et la parcellisation (hyperspécialisation)»*<sup>24</sup>.

La socioterminologia, inoltre, si allontana dagli assunti esclusivamente sincronici e strutturalisti Wüsteriani: *«if sciences are networks of inherited nodes instead of monolithic blocks, then the diachronic study of the history of conceptualisation and naming should be taken up»*<sup>25</sup>.

La socioterminologia, piuttosto, si fissa come obiettivo lo studio dei termini sia in sincronia che in diacronia, includendo anche l'analisi e la modellizzazione dei significati e dei concetti.

L'approccio socioterminologico, secondo François Gaudin (2005) possiede una “dimensione sociocritica”<sup>26</sup> nella misura in cui collega la produzione del senso dei termini alle condizioni in cui essi appaiono. I termini vanno considerati, secondo l'autore,

---

<sup>22</sup> Ivi, p. 198.

<sup>23</sup> Gambier, Y. (1991), *“Travail et vocabulaires spécialisés: prolégomènes à une socioterminologie”* p.37

<sup>24</sup> *Ibidem*.

<sup>25</sup> Gaudin F., (1995). *“Usage sociaux des termes: théories et terrains”*.

<sup>26</sup> Id. (2005). *« La socioterminologie »*, p. 81.

nell'ottica della diversità del loro uso sociale e quindi devono essere studiati in considerazione delle loro condizioni e dei contesti.

I termini vanno considerati come «*des signes linguistiques, et non comme des étiquettes de concepts*»<sup>27</sup>.

La socioterminologia dunque si pone in forte contrapposizione con la teoria di Wüster che, riprendendo le parole di Gaudin (2005), «*a perdu de sa superbe [...] dans la mesure où cela témoigne que l'horizon de la terminologie ne s'inscrit plus dans le paysage des initiatives généreuses du début du XX<sup>e</sup> siècle*»<sup>28</sup>. Secondo l'autore, le aspettative sulla terminologia erano fin troppo rosee per Wüster, che vedeva la terminologia come una sorta di esperanto, e dunque concepita in maniera deliberatamente artificiale e concertata. Gaudin sostiene piuttosto che «*la terminologie n'est plus tout à fait wüsterienne*»<sup>29</sup>.

#### 1.4 Approcci teorici - L'approccio (socio)cognitivo

Uno degli elementi che caratterizza il termine è il concetto, ovvero la rappresentazione astratta dell'oggetto, che, secondo la teoria classica, viene descritto e definito dalla norma.

---

<sup>27</sup> *Ibidem.*

<sup>28</sup> *Ivi*, p. 80.

<sup>29</sup> *Ibidem.*



Gli approcci descrittivi recenti capovolgono il paradigma tradizionale e conducono a pensare a dei concetti collegati non solo ai termini, ma agli utilizzatori.

L'approccio cognitivo condivide molti assunti con la socioterminologia. Rita Temmerman, ad esempio, «*definisce un modello "sociocognitivo" che incorpora gli elementi della teoria dei prototipi e prende in considerazione le dimensioni sociale e diacronica dei termini*»<sup>30</sup>.

Il concetto diventa dunque centrale per questo tipo di approccio, che si concentra sulla creazione di mappe di concetti collegate tra loro e che creano a loro volta altri sistemi di concetti<sup>31</sup>.

L'approccio sociocognitivo sposta l'attenzione dai concetti alle unità di comprensione. Secondo tale assunto, numerosi concetti sono prototipici e quindi non possono essere delineati utilizzando una serie di condizioni restrittive, così come definito dalla Teoria Generale della Terminologia.

Il termine "unità di comprensione" è stato introdotto originariamente da Rita Temmerman<sup>32</sup> ed è spiegato come «*the understanding of a referent studied in a given subject-field. The unit of understanding is communicated via specialised discourse in which it can be expressed in several ways*»<sup>33</sup>. Temmerman propone le unità di comprensione per sopperire all'inadeguatezza della teoria classica riguardo alla struttura concettuale dei domini specialistici.

---

<sup>30</sup> L'Homme M.C. (2004), "La Terminologie: principes et techniques", p.24, (traduzione dell'autore).

<sup>31</sup> Sager J., (1990), "A practical course in Terminology processing".

<sup>32</sup> Temmerman R., (2000), "Towards new ways of terminology description: the sociocognitive approach".

<sup>33</sup> Kerremans K., Temmerman R., De Baer P., (2008), "Construing domain knowledge via terminological understanding", p. 180.

Le unità di comprensione, dunque, demarcano il passaggio dal significato alla comprensione, un assunto che sottolinea che esistono diversi gradi di comprensione che dipendono dall'utente, dal contesto e dai continui cambiamenti a cui la comprensione della realtà è soggetta.

*«The constant development of units of understanding can be explained as the result of several simultaneously active factors: a) the urge for more and better understanding; b) the interaction between different language users; c) prototype structure in the understanding of categories which can be seen simultaneously as the result of and as one of the causes of meaning evolution. Cognitive models play an important role in the development of new ideas»<sup>34</sup>.*

Le unità di comprensione sono caratterizzate da una struttura prototipica e quindi non possono essere definite come chiaramente delineabili. Del resto, *«one 'ideal' definition of units of understanding – whereby meaning is defined on the basis of unique and sufficient properties – is often neither possible nor desirable<sup>35</sup>»*. In quanto prototipiche, le unità di comprensione sono viste come categorie:

*«categories are all the units of understanding which are impossible to describe according to the principles of traditional Terminology. There is prototype structure, both within and between categories (intracategorially and intercategoryally)»<sup>36</sup>.*

---

<sup>34</sup> *Ibidem.*

<sup>35</sup> Temmerman R., (1997), "Questioning the univocity ideal. The difference between sociocognitive terminology and traditional terminology", p. 51- 90.

<sup>36</sup> Ead., (2000), *cit.* p. 73.

Conseguentemente, poiché le unità di comprensione possono essere strutturate in categorie, un'ulteriore differenziazione del significato è ottenuta sulla base del valore che viene ad esse attribuito. Queste unità aiutano a creare un *template* di comprensione includendo altri fattori, come la prospettiva dalla quale l'unità viene compresa<sup>37</sup>.

Un altro principio dell'approccio sociocognitivo, in contrapposizione alla teoria classica, è la critica all'univocità.

Secondo la teoria tradizionale della terminologia, un concetto è designato da uno e un solo termine e un termine deve riferirsi ad un solo concetto. Temmerman sostiene che l'univocità è piuttosto irrealistica e che la polisemia e la sinonimia siano piuttosto fenomeni molto più tipici<sup>38</sup> e che quindi debbano essere prese in piena considerazione nello studio dei termini.

## 1.5 Approcci teorici - "Frame-Based Terminology"

La terminologia "frame-based" (FBT), è un approccio cognitivo alla terminologia «*based on frame-like representations in the form of conceptual templates underlying the knowledge encoded in specialized texts* (Faber 2011, 21; 2012; Faber et al. 2007, 42)».

Secondo la FBT, per "frame" si intende:

---

<sup>37</sup> Ead., (2000), *ivi*, cap. 3

<sup>38</sup> Ead. (1997), *cit.*

*«a schematisation of experience (a knowledge structure), which is represented at the conceptual level and held in long-term memory and which relates elements and entities associated with a particular culturally embedded scene, situation or event from human experience»<sup>39</sup>.*

Pamela Faber (2015), inoltre, distingue i frame dai template:

*«a frame is a representation that integrates various ways of combining semantic generalizations about one category or a group of categories, whereas a “template” is the representational pattern for individual members of the same category»<sup>40</sup>.*

I frame, dunque, sono dispositivi atti alla strutturazione cognitiva basata sull'esperienza e rappresentazioni ad ampia scala.

Nei testi scientifici e tecnici, le unità di conoscenza specialistica attivano frame semantici specifici di un determinato dominio e che sono in consonanza con la conoscenza di partenza dell'utente. Questi frame sono il contesto in cui la FBT specifica gli aspetti semantici, sintattici e pragmatici delle unità linguistica specialistiche. Essi forniscono dunque la conoscenza di base e la motivazione dell'esistenza delle parole in una lingua e come tali parole sono usate nel discorso.

*«Frames have the advantage of making explicit both the potential semantic and syntactic behaviour of specialized language units. This necessarily includes a description of conceptual relations as well as a term's combinatorial potential»<sup>41</sup>.*

---

<sup>39</sup> Evans V., (2007), "A Glossary of Cognitive Linguistics", p. 85.

<sup>40</sup> Faber P., (2015) "Frames as a framework for terminology", in "Handbook of terminology", a cura di Kocaert H., J., e Steurs F.

<sup>41</sup> Faber P., (2009) "The cognitive shift in terminology and specialized translation.", p. 107-134.

In altri termini, per comprendere realmente il significato delle parole in una lingua, è necessario avere comprensione dei frame semantici o delle strutture concettuali che sottostanno al loro uso. Ovviamente questo principio si applica anche alle unità di linguaggio specialistiche.

Inoltre, la FBT, usando le parole di Faber, si concentra su:

*«(1) conceptual organization; (2) the multidimensional nature of terminological units; and (3) the extraction of semantic and syntactic information through the use of multilingual corpora. In frame-based terminology, conceptual networks are based on an underlying domain event, which generates templates for the actions and processes that take place in the specialized field as well as the entities that participate in them»<sup>42</sup>.*

Un altro aspetto importante della terminologia frame-based è che essa prevede anche il ruolo delle immagini nella rappresentazione dei concetti specialistici. Essa esprime come le descrizioni linguistiche e grafiche di entità specifiche siano collegate e ciò sottolinea ancora più chiaramente la natura multidimensionale dei concetti così come le relazioni entro un dominio<sup>43</sup>. Tradizionalmente, le immagini sono classificate in accordo con la loro morfologia, ad esempio in fotografie, disegni, video, diagrammi ecc. Tuttavia, è più utile categorizzare le immagini in termini delle loro funzioni salienti o in termini delle loro relazioni con le entità del mondo reale che rappresentano.

---

<sup>42</sup> *Ibidem*.

<sup>43</sup> Faber P., et al. (2007), "Linking Images and Words: the description of specialized concepts".

## 1.6 Approcci teorici - l'approccio comunicativo

L'approccio comunicativo di Maria Teresa Cabré parte da due assunti fondamentali: il primo è che la terminologia è «*a set of needs, a set of practices to resolve these needs and a unified field of knowledge*»<sup>44</sup>. Riguardo tale assunto, crea una lista di bisogni che sono associati con le attività effettuate nei diversi campi:

«*Firstly, terminology presupposes a need for all the activities related to the representation and transfer of specialised knowledge such as technical translation, the teaching of languages for specific purposes, technical writing, the teaching of special subjects, documentation, special language engineering, language planning, technical standardisation etc*»<sup>45</sup>.

Enunciando ciò, Cabré sottolinea che lo sviluppo di una teoria terminologica è dettata dall'emergenza di un bisogno. Utilizzando le parole della ricercatrice, si può dire che: «*subject matter and methodology develop when there is a need, and are pursued to the extent that they are the result of clear social needs*»<sup>46</sup>. Conseguentemente, è necessario che si sviluppino “prodotti” orientati alla soluzione di tali bisogni e quindi disporre degli elementi che consentano di riconoscere la terminologia come un “dominio unificato di conoscenza”.

---

<sup>44</sup> Cabré M.T., (2003), *cit.*, p. 182.

<sup>45</sup> Ead., (2003), *ibidem*.

<sup>46</sup> Ead., (1998), *cit.*, p. 2.

È necessario, secondo Cabré, gettare le basi per una teoria terminologica perché è ciò che la rende una disciplina separata. La terminologia, per la ricercatrice è:

*«a discipline, and as such it is an organised set of basic essentials about an object of knowledge. And a theory of terminology must describe this object and provide a sufficiently broad methodological framework which includes the practises intended for the satisfaction of diverse needs»<sup>47</sup>.*

Il secondo assunto è che gli elementi della terminologia sono le cosiddette “unità terminologiche”. Tali unità sono multidimensionali, ovvero sono allo stesso tempo unità di conoscenza, unità di linguaggio e unità di comunicazione.

La loro descrizione avviene quindi secondo tre diverse prospettive: linguistica (il termine), cognitiva (il concetto) e socio-comunicativa (contesto).

Secondo la prospettiva linguistica i termini sono unità lessicali del linguaggio che attivano un valore specializzato quando usate in determinati contesti. Tale valore risulta in un significato preciso, riconosciuto e stabilizzato entro una comunità di esperti.

Dal punto di vista cognitivo, i termini costituiscono unità concettuali che rappresentano nodi di conoscenza che sono necessari nella costruzione della struttura di un dominio e che vedono nelle unità lessicali il loro equivalente linguistico.

L’approccio comunicativo vede i termini come unità di discorso utili all’interazione e alla condivisione della conoscenza all’interno di un determinato contesto pragmatico/comunicativo. Questa tripla composizione delle unità terminologiche non

---

<sup>47</sup> Ead., (2003), *ivi*, p. 183.

mostra, però, differenze con altre unità del linguaggio quali le parole, quindi è necessario specificare le caratteristiche delle unità terminologiche e le condizioni sotto le quali esse sono tali. Inoltre, è necessario fare una distinzione anche con altre unità che esprimono conoscenza specifica, come ad esempio le unità morfologiche e fraseologiche.

La specificità delle unità terminologiche si trova nel fatto che adempiono a condizioni restrittive in ognuno dei suoi livelli costituenti, ovvero cognitivo, grammaticale e pragmatico.

Dal punto di vista cognitivo, esse dipendono dal contesto tematico, occupano una posizione precisa in una struttura concettuale e il loro significato è esplicitamente fisso, stabilito e riconosciuto da una comunità di esperti.

Dal punto di vista linguistico, le unità terminologiche devono rispettare le condizioni di essere unità lessicali, avere una struttura lessicale e sintattica, appartenere a categorie semantiche più ampie, quali entità, eventi, proprietà o relazioni.

Dal punto di vista comunicativo, le condizioni sono di comparire in un discorso specialistico, essere acquisite tramite processi di apprendimento e quindi gestite da specialisti nei loro campi<sup>48</sup>.

Questo assunto conduce, secondo l'autrice, ad un dilemma teorico, ovvero: «*how can we reconcile the multidimensionality of the object and the subject field with the condition of unity of a field of knowledge of terminology?*»<sup>49</sup>.

---

<sup>48</sup> Per la lista completa delle condizioni cognitive, linguistiche e comunicative si rimanda a Cabré (2003), *cit.* p. 183-185.

<sup>49</sup> Ead. (2003), *Ivi*, p. 185.



In altre parole, come formulare una teoria in cui i vari filoni terminologici si combinino?

Per rispondere a tali quesiti Cabré introduce un modello che chiama “*theory of doors*”.

Il modello rappresenta un accesso plurale, ma non simultaneo, all’oggetto, ovvero le unità terminologiche, a partire dal concetto, dal termine o dal contesto.

In questo senso, ribadendo la multidimensionalità dell’oggetto, è possibile avere una descrizione del dato reale e in tutta la sua complessità (Cabré 2003, 2005).

La teoria delle porte è spiegata con un’immagine di una casa:

*«let us assume a house with several entrance doors. We can enter any one of its rooms through a different door, but the choice of the door conditions the way to the inside of the house. The internal arrangement of rooms is not altered, what does change is the way one chooses to get there»<sup>50</sup>.*

Dunque, se si accetta la natura multidimensionale delle unità terminologiche, le tre dimensioni, ovvero linguistica, cognitiva e comunicativa, devono essere sempre impiegate nell’accesso all’oggetto.

*«The conceptual strand of a unit (the concept and its relations) may be the door to the description and explanation of terminological units, without thereby rejecting their multidimensionality. Equally, its linguistic strand is another door to description. Logically, keeping in mind that, even though we analyse them as linguistic units, they do not lose their cognitive and social nature. Finally, if we approach terminology via the door of communication we are faced with different*

---

<sup>50</sup> Ead. (2003), *Ivi*, pp. 195-196.

*communicative situations in which linguistic units share the expressive space with those of other systems of communication»<sup>51</sup>.*

Cabré, nel suo lavoro più che ventennale, riconosce la coerenza interna della teoria tradizionale ma la definisce come insoddisfacente nella descrizione dei dati reali e della loro complessità.

La sua ricerca sostiene che una teoria multidimensionale è l'unico sistema in grado di comprendere le diverse posizioni teoriche:

*«I have, over the past few years, thought about how a multidimensional theory of terminology could be made broad enough to encompass all the different existing theoretical positions. And so, from my background in linguistics, I was determined to delineate this broad theoretical framework and within it to develop an approach to the description of terminology on the basis of a theory of natural language which describes and explains terminological units»<sup>52</sup>.*

La studiosa catalana sostiene, infine, che la cornice entro la quale si studiano le unità terminologiche è quella della comunicazione specialistica, caratterizzato da condizioni esterne quali l'emittente, il ricevente e il medium, da condizioni di trattamento delle informazioni quali la determinazione, fissazione e validazione del concetto da una comunità di esperti e dalle condizioni contestuali e da quelle che ne restringono e determinano la funzione e gli obiettivi.

---

<sup>51</sup> Ead. (2003), *Ivi*, p. 185.

<sup>52</sup> Ead. (2003), *Ivi*, p.187.

Cabré, anticipando che una teoria terminologica sia ancora in fase di elaborazione e allestimento, sostiene che essa debba adeguatamente identificare le unità terminologiche e le relazioni che esse hanno con le altre unità di conoscenza specialistica con cui si relazionano:

*«Within the general framework of specialised communication this theory must integrate basic justifications, principles and conditions which adequately describe terminological units, their characteristics and properties, the relations between their intrinsic components, their functions, the relations they establish among themselves and with other units of specialised knowledge and the processes they follow. There is no doubt that from the contributions to this cooperative effort a more appropriate theoretical model will emerge»<sup>53</sup>.*

## 1.7 La terminologia e la traduzione

La terminologia e la traduzione presentano numerose similitudini e punti di incontro ma sono caratterizzate da obiettivi distinti.

Per Cabré (2010) la traduzione è l'insieme dei processi traduttivi mirati all'espressione, in una determinata lingua, di una struttura semantico-pragmatica prodotta originariamente in un'altra lingua. Mentre la terminologia è, invece, concentrata sulle

---

<sup>53</sup> Ead. (2003), *Ivi*, p. 194.

forme lessicali e sui nodi di contenuto che rappresentano conoscenza con l'obiettivo di raccogliere termini specialistici per la creazione di risorse terminologiche, quali glossari, dizionari, vocabolari e database<sup>54</sup>.

La terminologia e la traduzione hanno in comune una lunga tradizione da materie applicative, in contrasto con la recente ricerca di affermazione come discipline autonome e, conseguentemente, sono entrambe caratterizzate da interdisciplinarietà con altre scienze.

Il rapporto tra terminologia e traduzione va analizzato sia dalla prospettiva della traduzione che da quella della terminologia. Dal punto di vista della traduzione, la terminologia è concepita come uno strumento per la traduzione.

Le risorse terminologiche forniscono ai traduttori informazioni necessarie per la risoluzione di dubbi, ovvero, per trovare un equivalente nella lingua di destinazione, per imparare il significato di un termine nella lingua di origine o selezionare l'opzione migliore tra termini diversi. A questa funzione puramente strumentale va però aggiunto che la terminologia serve ai traduttori anche come un mezzo per acquisire conoscenza in un dominio specifico.

I termini costituiscono strutture di conoscenza e dunque conoscere la terminologia di un dominio implica acquisirne e accrescerne la conoscenza. In questo senso, la terminologia acquisisce *«una funzione metacognitiva che aiuta i traduttori ad organizzare la propria conoscenza di una materia e fornisce loro le unità lessicali*

---

<sup>54</sup> Cabré M.T., (2010) "Terminology and Translation" in Handbook of Translation Studies, a cura di Y. Gambier e L. Van Doorslaer, John Benjamins, pp. 356-365.

*(termini) per esprimere adeguatamente la unità di conoscenza specialistica del campo»<sup>55</sup>.*

L'orientamento secondo il quale la terminologia è di supporto ed integrata alla traduzione è altamente sviluppato negli Stati a legislazione bilingue, come ad esempio il Quebec e il Belgio (la zona della Vallonia) e forma le basi delle attività condotte da organismi e istituzioni internazionali, quali la FAO, le Nazioni Unite, l'Unesco e l'Unione Europea.

Tale approccio rappresenta la principale motivazione per la creazione di banche dati terminologiche istituzionali, quali Termium Plus<sup>56</sup> del governo canadese e InterActive Terminology for Europe (IATE), gestito dall'Unione Europea.

IATE è il database interistituzionale, multilingue e interattivo che contiene la terminologia prodotta dall'Unione Europea.

In quest'elaborato, un'ampia sezione sarà dedicata a IATE, per cui è importante illustrare subito le caratteristiche e le finalità della terminologia in un contesto interistituzionale come quello dell'Unione Europea.

Uno degli obiettivi dell'UE è di preservare il multilinguismo integrale, che esprime *«la capacità dei cittadini di esprimersi in ogni lingua ufficiale dell'Unione e contribuisce ai valori europei della democrazia, dell'uguaglianza, della trasparenza e della*

---

<sup>55</sup> Ead. (2010), *Ivi*, p. 358. Traduzione dell'autore.

<sup>56</sup> Termium Plus è la banca dati terminologica del governo canadese. Per maggiori informazioni consultare [http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&i=1&index=alt&codom2nd\\_wet=1](http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&i=1&index=alt&codom2nd_wet=1).

*competitività»<sup>57</sup>. Per questo motivo, è stata realizzata un'imponente struttura traduttiva che garantisca l'impegno delle istituzioni verso il multilinguismo.*

Nel Parlamento europeo, ad esempio, la struttura demandata è la Direzione Generale per la Traduzione (DGTRAD), che ha come finalità la traduzione dei documenti in tutte le lingue ufficiali dell'UE. Relativamente alla terminologia, DGTRAD ha come obiettivo "lo sviluppo di strumenti IT appropriati e database terminologici per sostenere i traduttori e integrarli nei loro processi lavorativi"<sup>58</sup>.

A livello UE, la terminologia, dunque, è considerata come uno strumento subordinato alla traduzione. Tuttavia, proprio grazie al ruolo centrale delle istituzioni europee, alla terminologia possono essere attribuite anche altre funzioni oltre a quella di sostegno all'attività traduttiva.

Le istituzioni europee determinano politiche che impattano l'Unione a livello sociale, politico, economico e conseguentemente anche a livello linguistico. La terminologia, in questo senso, è il veicolo ideale per la rappresentazione della conoscenza e una sua appropriata gestione, mirata all'armonizzazione e alla standardizzazione dei contenuti, apporterebbe benefici ben superiori a quelli di mero sostegno alla traduzione.

---

<sup>57</sup> Multilinguismo al Parlamento, il valore aggiunto dell'Europa, URL: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM%20PRESS+20071017FCS11816+0+DOC+XML+V0//IT>.

<sup>58</sup> Directorate-General. Directorate-General for Translation. URL: <http://www.europarl.europa.eu/the-secretary-general/en/organisation/directorate-general-for-translation>. Traduzione dell'autore.

## 1.8 La terminologia e l'informatica

Il progresso dell'informatica ha determinato un grande cambiamento nei processi di gestione del linguaggio.

La relazione con la terminologia si è evoluta gradualmente, in proporzione con il grado di complessità e stato di avanzamento dei sistemi di trattamento automatico del linguaggio. Si possono distinguere diversi stadi di tale relazione: uno stadio iniziale, in cui i dati linguistici venivano usati soltanto a livello formale e senza l'intervento di manipolazione o analisi (un esempio possono essere il cosiddetto *spellchecker* e *grammar checker* in Word, ovvero delle funzioni che permettono rispettivamente l'identificazione automatica di errori di battitura e di errori grammaticali); uno stadio intermedio, in cui si cominciano a sviluppare i primi database, dizionari elettronici, software di supporto alla traduzione<sup>59</sup>; uno stadio finale, quello attuale, in cui gli strumenti informatici sono dei sistemi esperti, che possono assistere, e in taluni casi sostituire, la componente umana. Un esempio di questo stadio sono i software di identificazione ed estrazione automatica dei termini, *machine translation* e, ovviamente, le ontologie e i più recenti sviluppi nel campo del Web Semantico<sup>60</sup>.

Particolarmente significativa, in questo contesto, una riflessione di Pierre Auger (1989) che auspicava che:

---

<sup>59</sup> Comunemente denominati CAT tool (Computer Assisted Translation tools).

<sup>60</sup> Per Web Semantico si intende l'evoluzione del World Wide Web, un nuovo ambiente dove, associando dati ed informazioni ad altri dati (che prendono il nome di metadati) se ne migliora la specificazione semantica, permettendone l'interrogazione da parte di macchine e umani (Berners-Lee, 2009).

*«in a futuristic scenario, terminologists will have access to huge data (or knowledge) banks; from these banks, they will download items belonging to their corpora, they will go through these items without having to previously tackle the text manually, they will automatically establish their working terminologies, contrasting the incoming terms with the semantic descriptors that will later be used in the automatic writing of definitions; they will classify, choose, merge and edit data bases reducing their intervention to a minimum. Their work station, equipped with advanced and intelligent office tools, will allow them to control, by themselves and throughout the whole process, the elaboration of their product and carry it out under the best conditions»<sup>61</sup>.*

Una buona parte di quanto auspicato da Auger si è effettivamente verificato e senza dubbio l'informatica ha cambiato sia la metodologia del lavoro terminologico che la pratica. Grazie ad essa, è possibile oggi identificare strutture che sarebbe impossibile scorgere senza l'impiego di strumenti tecnologici, quali la consultazione di un database terminologico come IATE, le operazioni di estrazione automatica della terminologia o l'impiego di una termbase in un software di traduzione.

Una termbase è un glossario che ha la finalità di fungere da supporto terminologico al traduttore per le funzioni di recupero, consultazione e riutilizzo dei dati terminologici. Le termbase ricoprono un ruolo fondamentale nel contesto della traduzione e rappresentano il punto di intersezione tra terminologia, traduzione e informatica. Ad

---

<sup>61</sup> Auger P., (1989) "La terminotique et les industries de la langue".



oggi rappresentano lo standard di riferimento per l'archiviazione di informazioni terminologiche e per il loro impiego nei software di traduzione assistita.

Alan K. Melby (2012) la definisce come:

*«a computer database consisting primarily of information about domain-specific concepts and the terms that designate them. Specialised translation deals with domains of knowledge, and every domain is organised through concepts that are linked to objects or ideas relevant to that domain. Termbases may be monolingual, bilingual, or multilingual»<sup>62</sup>.*

È però doverosa un'ulteriore distinzione, ovvero tra termbank e termbase.

Puntuale la definizione di Lynne Bowker (2015) che descrive le prime come *«strictly controlled multilingual resources»* e le seconde come *«smaller and more personalized terminological resources»*.

Le termbank sono essenzialmente grandi collezioni di schede terminologiche che contengono informazioni sui termini e concetti che rappresentano. Solitamente sono destinate all'uso di un ampio spettro di utenti.

Juan C. Sager (1990), ne identifica cinque categorie:

*«(1) subject specialists, who often consult such resources for reassurance, but who sometimes need to ascertain the meaning of an unknown term; (2) information scientists, who perform tasks such as the indexing of specialized documents; (3) language planners, who are charged with developing or maintaining a language and*

---

<sup>62</sup> Melby A. K., (2012), *“Terminology in the age of multilingual corpora”*. Sulle termbase si rimanda a Wright & Budin (2001), Melby (2012) e Bowker (2015).

*have a particular interest in standardization; (4) teachers and students, whether studying language or a specialized field and (5) professional communication mediators, such as journalists, technical writers or specialized translators»<sup>63</sup>.*

Ad ogni modo, sono numerose le distinzioni tra le due basi di dati.

Le termbank sono solitamente create e gestite da terminologi specializzati, che seguono linee guida predefinite e metodi condivisi, spesso orientati alla standardizzazione.

Le termbase sono invece create da traduttori e raccolgono entrate prodotte da un unico individuo, condivise limitatamente o comunque prive da ogni obiettivo di pubblicazione. Sono, inoltre, solitamente create ad hoc e concepite con un chiaro orientamento alla produzione.

Le termbank sono spesso caratterizzate da un ampio spettro di domini e contengono materiale le cui fonti sono altamente controllate. Le entrate stesse, che contengono termini in forma canonica, sono dettagliate e possono offrire numerosi campi per ulteriori informazioni.

Le termbase sono usualmente limitate ad un dominio (o pochi), possono contenere termini con fonti moderatamente controllate (visto l'orientamento alla produttività) e il livello di dettaglio è minimalista e concentrato sugli equivalenti.

I termini possono non essere in forma canonica ed essere anche integrati da elementi fraseologici o altri elementi linguistici di uso ricorrente.

Un'ulteriore distinzione è nell'approccio di costruzione: nelle termbank è solitamente onomasiologico, mentre per le termbase si possono anche adottare soluzioni sia onomasiologiche che semasiologiche. Infine, le termbank sono limitatamente integrate

---

<sup>63</sup> Sager J., (1990), *cit.*, p. 197.

in altri strumenti, mentre le termbase trovano la loro finalità principale nell'integrazione ad altri *tool*, come i software di traduzione assistita.

Un esempio concreto di termbank è il database interistituzionale IATE, alimentato da traduttori, terminologi e dalle Unità di Coordinamento Terminologico delle istituzioni dell'Unione Europea.

Un recente sviluppo di IATE<sup>64</sup> ha offerto la possibilità di estrarre risorse terminologiche dal database e archivarle in termbase da utilizzare direttamente nei software di traduzione assistita utilizzati dai traduttori dell'Unione Europea.

L'analisi sull'uso delle termbase e di IATE sarà una delle parti fondamentali di questo elaborato, per cui per ora ci si soffermerà sul fatto che la distinzione tra termbank e termbase non segue due estremi contrapposti e rigidi ma va vista piuttosto come un continuum.

Effettuare la distinzione tra termbank e termbase sottolinea che analizzare il rapporto tra terminologia e tecnologia senza considerare l'influenza del contesto traduttivo sarebbe un chiaro errore.

Oggi, nell'era dell'*overload* informativo, i traduttori riconoscono un'importanza crescente nei sistemi informatici dedicati alla gestione della terminologia<sup>65</sup>. Numerosi sono gli autori (come Bowker e Pearson 2002, 20; Homme 2004, 119; Pavel e Nolet 2001, xx) che sostengono che oggi è virtualmente impensabile effettuare ricerche terminologiche senza l'ausilio degli strumenti informatici. Bowker (2015) nota che, se il

---

<sup>64</sup> Ci si riferisce ad una funzione accessibile solo dalla versione ad uso esclusivamente interno del database, non a quella pubblica.

<sup>65</sup> Allard M. G. P. (2012) "*Managing Terminology for Translation Using Translation Environment Tools: Towards a Definition of Best Practice*".

lavoro dei terminologi è stato senz'altro influenzato dall'uso di *tool* computerizzati, l'impatto del lavoro terminologico effettuato dai traduttori è ancora più pronunciato, perché gli strumenti di gestione della terminologia possono ora essere integrati direttamente nei software traduttivi.

Con una considerazione di più ampio respiro, si può dunque dire che la terminologia, tramite l'ausilio della tecnologia, ha influenzato profondamente (ed è stata influenzata da) le discipline ad essa prossime.

Attraverso il suo approccio multidisciplinare, si sostiene che essa sia senza dubbio la disciplina che si presta più favorevolmente alle specifiche della gestione della conoscenza.

Nei tempi recenti della rivoluzione digitale, sono stati sviluppati strategie e metodi per identificare, raccogliere, conservare e rendere accessibile la conoscenza. In questo senso, una realtà sempre più concreta è il Web Semantico. Con l'interpretazione del contenuto dei documenti che il Web Semantico consente, saranno possibili ricerche molto più evolute delle attuali, basate sulla presenza nel documento di parole chiave, e altre operazioni specialistiche come la costruzione di reti di relazioni e connessioni tra documenti secondo logiche più elaborate del semplice collegamento ipertestuale.

Molte di queste tecnologie possono, e devono, essere utilizzate per migliorare la gestione delle risorse terminologiche e il loro (ri)utilizzo, non soltanto nel contesto della traduzione, ma in ogni sistema esperto di gestione della conoscenza. Viceversa, le risorse terminologiche sono lo strumento ideale per arricchire le nuove tecnologie in cui il trattamento del linguaggio naturale ha un ruolo centrale.



## CAPITOLO II – LA TERMINOLOGIA NEL PARLAMENTO EUROPEO E L'ATTIVITÀ DI TERMINOLOGY COORDINATION UNIT

Il Parlamento europeo è l'Istituzione di espressione democratica e di controllo dell'Unione europea e rappresenta gli oltre 500 milioni di cittadini dei 28 Stati che ne fanno parte.

Il Parlamento tiene le sue sedute plenarie<sup>66</sup> a Strasburgo e a Bruxelles, dove si riuniscono anche le commissioni parlamentari<sup>67</sup>, mentre il Segretariato generale ha sede a Lussemburgo.

Il Segretariato, sotto la leadership politica del Presidente del Parlamento europeo, l'Ufficio di Presidenza<sup>68</sup> e la Conferenza dei Presidenti del Parlamento europeo<sup>69</sup>, è responsabile per l'organizzazione delle plenarie e delle altre riunioni, del coordinamento

---

<sup>66</sup> la seduta plenaria rappresenta il punto d'arrivo del lavoro legislativo effettuato in seno alle commissioni parlamentari e ai gruppi politici. La seduta plenaria rappresenta altresì la sede in cui i rappresentanti dei cittadini dell'Unione europea – i deputati europei – partecipano al processo decisionale comunitario e fanno valere i propri punti di vista presso la Commissione e il Consiglio.

<sup>67</sup> I deputati sono suddivisi in 20 commissioni specializzate. Esse danno seguito alle proposte legislative attraverso l'approvazione di relazioni, presentano emendamenti da sottoporre all'Aula e nominano i membri per negoziare la legislazione dell'UE con il Consiglio. Per un elenco completo delle commissioni permanenti, si consulti <http://www.europarl.europa.eu/committees/it/parliamentary-committees.html>

<sup>68</sup> L'Ufficio di presidenza del Parlamento viene eletto assieme al presidente del Parlamento e rimane in carica per due anni e mezzo. È composto dal presidente, dai 14 vicepresidenti e da 5 questori eletti dall'assemblea. L'Ufficio di presidenza si occupa del bilancio del Parlamento e di varie questioni di tipo amministrativo e organizzativo.

<sup>69</sup> Conferenza dei presidenti delle commissioni del Parlamento europeo.

delle attività legislative e dell'assistenza tecnica agli organi parlamentari. Il Segretariato coordina inoltre l'attività delle dodici Direzioni Generali<sup>70</sup> che lo compongono e deve, come parte fondamentale del suo mandato, garantire il multilinguismo integrale<sup>71</sup> per tutte le sedute plenarie e le altre riunioni.

Non ci si soffermerà, in questa sede, sull'analisi dettagliata della vasta complessità dell'attività europarlamentare, ci si focalizzerà sull'attività della Direzione Generale per la Traduzione, DGTRAD, e nello specifico sulle mansioni, progetti e risorse collegate alla gestione della terminologia in seno all'istituzione europea.

## 2.1 La Direzione Generale per la Traduzione - DGTRAD

La Direzione Generale per la Traduzione, anche conosciuta con l'acronimo "DGTRAD", garantisce che i documenti del Parlamento siano disponibili in tutte le lingue ufficiali<sup>72</sup>

---

<sup>70</sup> Ogni Direzione Generale è responsabile per un macro-dominio dell'attività parlamentare, che sono: Presidenza, Politiche Interne, Politiche Esterne, Servizi di Ricerca Parlamentare, Comunicazione, Personale, Infrastrutture e Logistica, Traduzione, Interpretazione e Conferenze, Finanza, Innovazione e Supporto Tecnologico, Sicurezza. Per un elenco esaustivo delle Direzioni Generali, consultare <http://www.europarl.europa.eu/the-secretary-general/en/organisation.html>.

<sup>71</sup> Il multilinguismo integrale si riferisce alla capacità dei cittadini dell'Unione Europea di esprimersi in ogni lingua ufficiale dell'Unione. Il regolamento del Parlamento europeo, articolo 158, elenca quattro disposizioni generali per lo svolgimento delle sedute: 1) tutti i documenti del Parlamento sono redatti nelle lingue ufficiali; 2) tutti i deputati hanno il diritto di esprimersi in Parlamento nella lingua ufficiale di loro scelta [...]; 3) durante le riunioni di commissione e delegazione è assicurata l'interpretazione da e verso le lingue ufficiali [...]; 4) Durante le riunioni di commissione o di delegazione al di fuori dei luoghi abituali di lavoro è assicurata l'interpretazione da e verso le lingue dei membri che hanno confermato la propria presenza alla riunione. Per il testo integrale riferirsi a <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+RULES-EP+20160715+RULE-158+DOC+XML+V0//IT&navigationBar=YES>, <http://www.europarl.europa.eu/aboutparliament/it/20150201PVL00013/Multilinguismo>, [http://ec.europa.eu/education/policy/multilingualism\\_it](http://ec.europa.eu/education/policy/multilingualism_it).

<sup>72</sup> 24 sono le lingue ufficiali dell'Unione Europea per i 28 stati membri: Bulgaro, Ceco, Croato, Danese, Estone, Finlandese, Francese, Greco, Inglese, Irlandese (Gaelico), Italiano, Lettone, Lituano, Maltese, Olandese, Polacco, Portoghese, Rumeno, Slovacco, Sloveno, Spagnolo, Svedese, Tedesco, Ungherese.

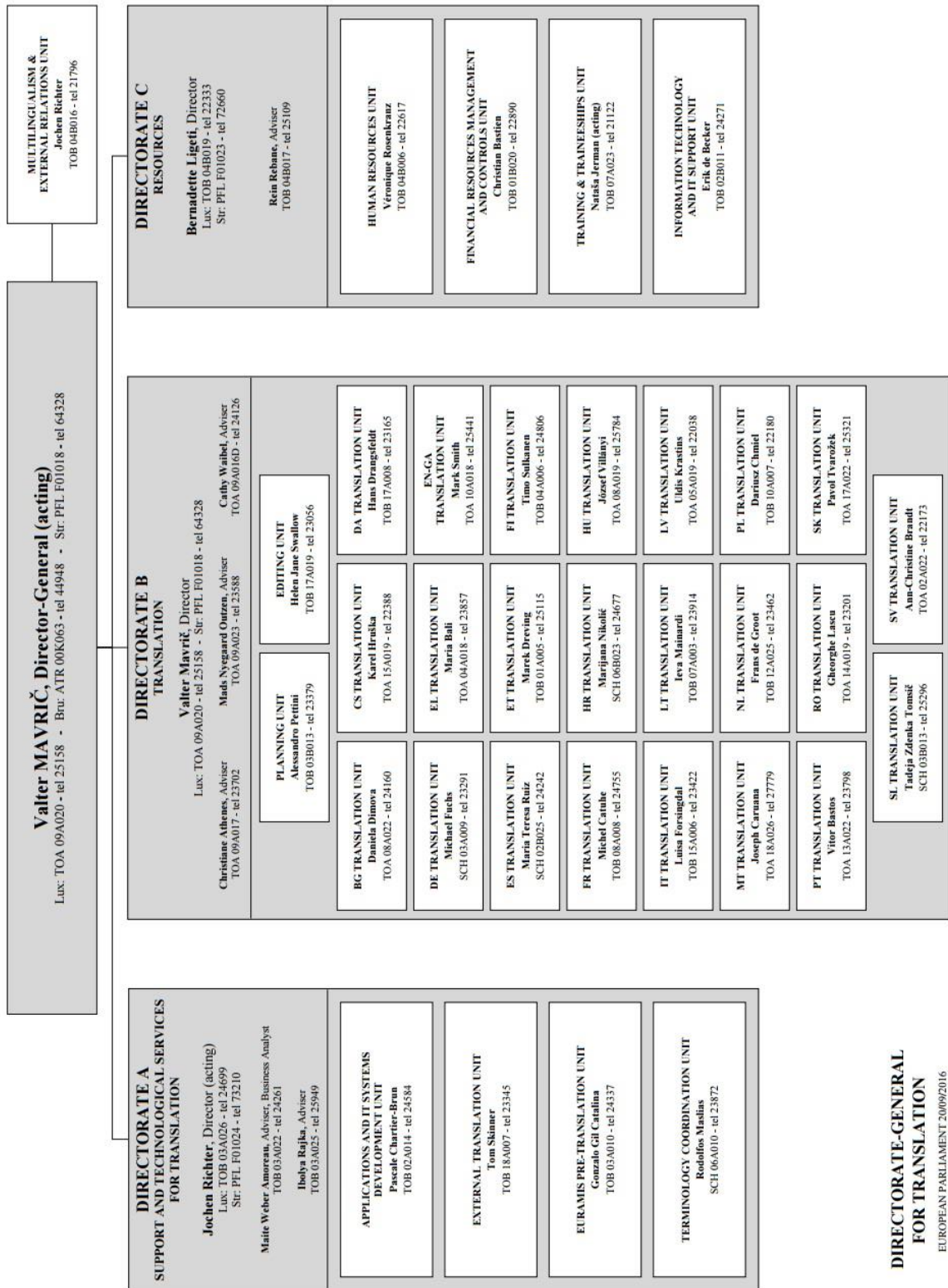
dei 28 Stati membri dell'Unione Europea, facilita la trasparenza, lo scambio e l'accesso ai documenti e, più generalmente, gioca un ruolo fondamentale nella preservazione della diversità linguistica e culturale dell'Unione.

Il suo mandato prevede la traduzione dei documenti da e verso le 24 lingue ufficiali, raggiungendo 552 combinazioni linguistiche e offrendo a tutti i cittadini dell'Unione l'accesso ai documenti e l'opportunità di comunicare con le istituzioni nella loro lingua madre.

Il servizio di traduzione deve garantire la qualità e l'efficienza anche per lo sviluppo di risorse, quali strumenti informatici, database terminologici e altre piattaforme traduttive, che supportino l'attività di traduzione. DGTRAD inoltre si occupa della gestione delle traduzioni eseguite al di fuori del Parlamento e della loro revisione e dei tirocini formativi per traduttori.

DGTRAD si divide in tre direzioni, nominate "A - Support and technological services for translation", "B - Translation" e "C - Resources", denominate nel cosiddetto *eurojargon* "A", "B" e "C".





**DIRECTORATE-GENERAL  
FOR TRANSLATION**  
 EUROPEAN PARLIAMENT 2009/2016

Fig. 2.1 Organigramma della Direzione Generale della Traduzione<sup>73</sup>

Le direzioni fanno capo ad un direttore generale. Ogni direzione è composta da un numero variabile di Unità, ovvero un'entità amministrativa posta sotto la responsabilità di un capo Unità. È quindi un gruppo di professionisti del Parlamento europeo che si occupa di una determinata attività.

Al direttore generale fa riferimento diretto un'Unità che non rientra in alcuna delle Direzioni, ovvero *Multilingualism & External Relations Unit*, il cui obiettivo è di gestire le attività legate all'implementazione di risorse mirate al multilinguismo e alle relazioni esterne di DGTRAD.

La Direzione B è senz'altro riconosciuta come la spina dorsale della Direzione Generale, in quanto è composta dalle 23 Unità Linguistiche, che si occupano della traduzione dei documenti, e dalle Unità *Planning* e *Editing*. L'Unità *Planning* è responsabile per il coordinamento della traduzione di tutti i documenti del Parlamento. Si occupa del monitoraggio delle attività delle commissioni parlamentari, informando le Unità linguistiche circa le priorità e le previsioni di traduzione<sup>74</sup>, ovvero fornisce anticipazioni sul carico di lavoro in relazione alle attività parlamentari.

L'Unità *Planning* si occupa dell'ammissibilità dei documenti, del conteggio pagine e della negoziazione delle scadenze, tenendo conto delle priorità e urgenze delle commissioni e della capacità delle Unità linguistiche. Ruolo fondamentale di quest'Unità è inoltre la creazione di *Feuille de Route (FdR)*, ovvero un documento, che contiene tutte le

---

<sup>74</sup> Letteralmente "translation forecasts".

informazioni relative ad un documento da tradurre, come ad esempio l'ente committente, il tipo di traduttore a cui il documento viene assegnato, note tecniche, suggerimenti e riferimenti ad altri documenti in ausilio alla traduzione.

L'*Editing Unit* si occupa dell'adattamento del documento sulla base di una forma e di una struttura specifica per tipologia di documento, la cosiddetta *mise en forme* nei diversi tipi di testo istituzionale nella lingua di partenza, ovvero in lingua inglese.

L'Unità si occupa inoltre della verifica di documenti quali le iniziative, i progetti di relazione, le relazioni finali e le opinioni e offre supporto alle Unità linguistiche durante il lavoro di verifica di testi specifici.

Le 23 Unità linguistiche si occupano ognuna delle 24 lingue dell'Unione Europea e della traduzione dei documenti istituzionali. Le Unità linguistiche traducono principalmente da una lingua di origine verso la propria, ma raggiungono anche altre combinazioni traduttive in relazione a determinate necessità dell'attività parlamentare.

Le Unità Linguistiche sono supportate dalle cosiddette "Unità orizzontali", dislocate nelle direzioni C e B, che sono orientate ai servizi come ad esempio formazione, informatica, sviluppo.

La Direzione C è composta da quattro Unità. *Human Resources, Financial Resources Management and Controls Unit, Training & Traineeships Unit, Information Technology and IT Support Unit.*

L'Unità *Human Resources* si occupa della gestione amministrativa del personale di DG TRAD. È responsabile dell'assunzione del nuovo personale, della pianificazione delle

risorse collegate allo staff ed offre anche uno sportello di consulenza per lo sviluppo della carriera. L'Unità coopera con la Direzione Generale del Personale, DG PERS, per facilitare i processi relativi alle esigenze dei concorsi pubblici<sup>75</sup>. *Human Resources* si occupa anche del telelavoro, ovvero la possibilità di lavorare dal proprio domicilio, degli scambi di staff a livello interistituzionale e dell'attrezzatura degli uffici.

*Financial Resources Management and Controls Unit*, più comunemente nota come *Finance*, ha il ruolo di assistere all'esecuzione del budget finanziario della Direzione Generale, integrandosi al lavoro del direttore generale con rapporti, dati statistici e analisi di necessità e obiettivi per le strategie di DGTRAD.

*Training & Traineeships Unit*, o anche *T&T*, si divide in due servizi, ovvero *Training* e *Traineeship*. *Training* gestisce la formazione per lo staff della Direzione Generale e definisce strategie e politiche di formazione. Si occupa inoltre di formazione per l'uso di sistemi e software di supporto alla traduzione, per i quali organizza seminari specifici, corsi di formazione linguistici per traduttori ed eventi tematici. *Traineeships* garantisce l'assunzione e la gestione dei profili dei tirocinanti traduttori che trascorrono un periodo di fino a sei mesi in una delle Unità di DGTRAD. Questo settore organizza, per i quattro turni annuali di tirocinio, workshop di formazione, missioni a Bruxelles e Strasburgo e in generale controlla il budget allocato per i tirocini. *Information Technology and IT Support Unit*, o anche *ITS*, è un'Unità chiave in seno alla Direzione C e per tutta la Direzione Generale. Offre supporto informatico, tramite un servizio apposito denominato "ServiceDesk", sui software di traduzione e affini; gestisce l'attrezzatura tecnologica di

---

<sup>75</sup> I concorsi pubblici che selezionano gli ufficiali dell'Unione Europea vengono gestiti da EPSO, l'Ufficio Europeo per la Selezione del Personale.

ogni ufficio e della sala di riunione, fornendo un primo punto di contatto per l'assistenza ad essa relativa; si occupa della gestione di progetti e del loro coordinamento tecnologico, fornendo ambienti di sviluppo, test e pre-produzione. ITS offre inoltre soluzioni per l'archiviazione dei file in cooperazione con altre Unità di DG ITEC (Directorate General for Innovation and Technology Support), la gestione dei diritti di accesso alle risorse informatiche e del loro rilascio e installazione.

La Direzione A è composta da *Applications & IT Systems Development*, *External Translation Unit*, *Euramis Pre-Translation Unit*, e *Terminology Coordination Unit*. *Applications & IT Systems Development*, o DAS, è l'Unità che si occupa di sviluppo di software in integrazione all'attività traduttiva con l'obiettivo di offrire servizi di alta qualità alle Unità traduttive, migliorandone il contesto tecnologico. L'Unità DAS opera in stretta collaborazione con l'Unità ITS per il coordinamento tecnologico dei progetti, dei test e degli aggiornamenti e nuovi strumenti.

*External Translation Unit*, o ETU, è responsabile dell'esternalizzazione delle traduzioni non eseguite dal Parlamento. Le traduzioni sono mandate all'Unità ETU dall'Unità Planning o dai capi Unità delle Unità linguistiche, in base a valutazioni di capacità, urgenza e importanza del lavoro. L'Unità ETU fa da volano tra le richieste di traduzioni esterne che vengono poi inoltrate ai servizi competenti e gestisce i contratti con le agenzie esterne di traduzione, effettuando processi di controllo e monitoraggio dei livelli di qualità.

*Euramis Pre-Translation Unit*, o più comunemente *PreTrad*, ha avuto, negli anni recenti, un'evoluzione che ne ha cambiato profondamente gli obiettivi e l'identità. Inizialmente,

era dedicata all’inserimento di dati nel database *Legis-Process* che gestisce i testi legali consolidati<sup>76</sup> del Parlamento. Oggi gestisce le memorie traduttive disponibili in *Euramis*<sup>77</sup>, stabilendone ruoli e attributi, coordinando le politiche di gestione dei documenti ed effettuando manutenzione agli archivi *Euramis*. Si occupa inoltre di *Safe Working Protocols*, manuali tecnici che dettagliano il *workflow*<sup>78</sup> legato ad uno specifico documento da tradurre. PreTrad fornisce servizi di verifica e di validazione, se non in alcuni casi di correzione, degli originali ricevuti dalle Unità linguistiche.

## 2.2 TermCoord – Terminology Coordination Unit e i tool di terminologia

*Terminology Coordination Unit*, o anche TermCoord, è l’Unità che si occupa della gestione della terminologia del Parlamento europeo.

I suoi campi principali sono il coordinamento e la gestione delle schede terminologiche prodotte dai traduttori del Parlamento europeo attraverso il database interattivo terminologico interistituzionale IATE; supporto alla produzione e alla gestione della terminologia per le Unità linguistiche e la preparazione di terminologia proattiva<sup>79</sup>.

---

<sup>76</sup> Per “consolidamento” si intende l’integrazione in un atto legislativo delle sue successive modifiche e rettifiche.

<sup>77</sup> Euramis è un database che ospita memorie traduttive condivise da tutte le istituzioni. Fu creato nel 2006 ed ospita 12.424.803 segmenti (ovvero frasi, paragrafi, titoli abbinati in due lingue). Fonte dati: “Pre Trad Statistical Report 2015” (consultato il 18 Dicembre 2016).

<sup>78</sup> “workflow” è un anglicismo traducibile con “flusso di lavoro”, ovvero la descrizione di una serie ripetibile di operazioni, pianificate e condivise *a priori*, finalizzate alla realizzazione di un’attività, quale la traduzione, in questo caso.

<sup>79</sup> Per “terminologia proattiva” si intende la creazione di liste di termini (denominate TermFolder), estratti da documenti o procedure che contengono terminologia particolarmente complessa. Le TermFolder

Gestisce strumenti di gestione terminologica e implementa progetti terminologici in seno alle Unità. Una attività di particolare importanza è il ruolo della comunicazione e della cooperazione interistituzionale.

### 2.2.1 Cenni storici

Fino alla fine degli anni '80, il lavoro terminologico era realizzato da traduttori che disponevano di un sistema di schede, organizzate in ordine alfabetico, raccolte in scatole grigie presenti su ogni scrivania e sistematizzate dalla Divisione di Terminologia in glossari cartacei. Tali glossari<sup>80</sup> venivano poi stampati e distribuiti a tutti i traduttori. Nel caso di documenti importanti e difficili, venivano organizzate consulte, principalmente tra Parlamento, Consiglio e Commissione, in presenza dei responsabili delle bozze, amministratori e traduttori di ogni lingua, in modo da produrre collezioni di terminologia relativa ad una determinata procedura.

Il vero primo passo della terminologia verso l'automazione si ebbe con la creazione di un sistema interattivo per la terminologia tra divisioni linguistiche, scritto con *dBase*, un software per la gestione di database, e creato da Rodolfo Maslias, oggi Capo Unità di TermCoord.

Quando negli anni '80 il panorama informatico mondiale cominciava ad esibire i primi software di traduzione e gestione terminologica, il Parlamento decise di accorpate la

---

vengono preparate da TermCoord e messe a disposizione dei traduttori del Parlamento prima (e quindi "proattivamente") che i testi arrivino in DGTRAD per la traduzione.

<sup>80</sup> Si tratta di circa di 100 glossari multilingui su diversi domini.

divisione di terminologia ai servizi tecnologici. L'obiettivo della fusione era di convertire tutti i glossari stampati in un vero e proprio database. Tale database fu nominato *Euterpe* e nacque in concomitanza con il database della Commissione Europea, *Eurodicautom*. Dalla fusione di questi database nacque, agli inizi degli anni 2000, IATE, InterActive Terminology for Europe, un database che doveva fungere da strumento unico per l'archiviazione delle risorse terminologiche.

Nel 2008, nuovi Stati Membri erano entrati a far parte dell'Unione Europea<sup>81</sup> e quindi l'esigenza di un servizio centralizzato per il coordinamento del lavoro terminologico crebbe in proporzione. Il nuovo *Terminology Coordination Service* fu creato nell'ottobre 2008. Il servizio era gestito da un piccolo team composto da un capo-servizio, un amministratore a tempo pieno e un amministratore a tempo parziale. Diverse mansioni erano – e sono – svolte anche dai tirocinanti, studenti specializzati in linguistica computazionale e/o comunicazione provenienti da diverse Università e Istituti di Ricerca dei Paesi UE, per un periodo di solito non superiore a sei mesi, attraverso procedure di selezione centralizzate gestite dal Parlamento europeo<sup>82</sup>. Dal 2008, TermCoord ha realizzato numerosi progetti; oltre all'attività collegata a IATE, ha partecipato, nel 2011 alla creazione di un portale interistituzionale di terminologia (denominato EurTerm), ha promosso la realizzazione di terminologia proattiva ed è stata attiva in numerose

---

<sup>81</sup> Si tratta del cosiddetto "allargamento ad Est", del 2004-2007, in cui 10 nuovi Membri (Cipro, Rep. Ceca, Estonia, Ungheria, Lettonia, Lituania, Malta, Polonia, Slovacchia e Slovenia) entrarono a far parte dell'UE nel 2004 e due (Bulgaria e Romania) si aggiunsero nel 2007.

<sup>82</sup> Il Parlamento offre diverse tipologie di borse per un tirocinio. DGTRAD accoglie solitamente i vincitori di borsa per tirocini di formazione alla traduzione, ma è anche possibile usufruire di un altro tipo di borsa destinata a titolari di diplomi universitari, i "tirocini Schuman".

Per maggiori informazioni si rimanda alla pagina dedicata ai tirocini del PE <http://www.europarl.europa.eu/atyourservice/it/20150201PVL00047/Tirocini> (data consultazione: 4 Dicembre 2016).



cooperazioni interistituzionali, con università, accademie, organi e associazioni linguistiche e terminologiche europee e non-europee, e nondimeno una capillare quanto vasta attività di comunicazione. TermCoord inoltre offre formazioni per traduttori, terminologi e tirocinanti ed è responsabile per numerosi strumenti di gestione terminologica.

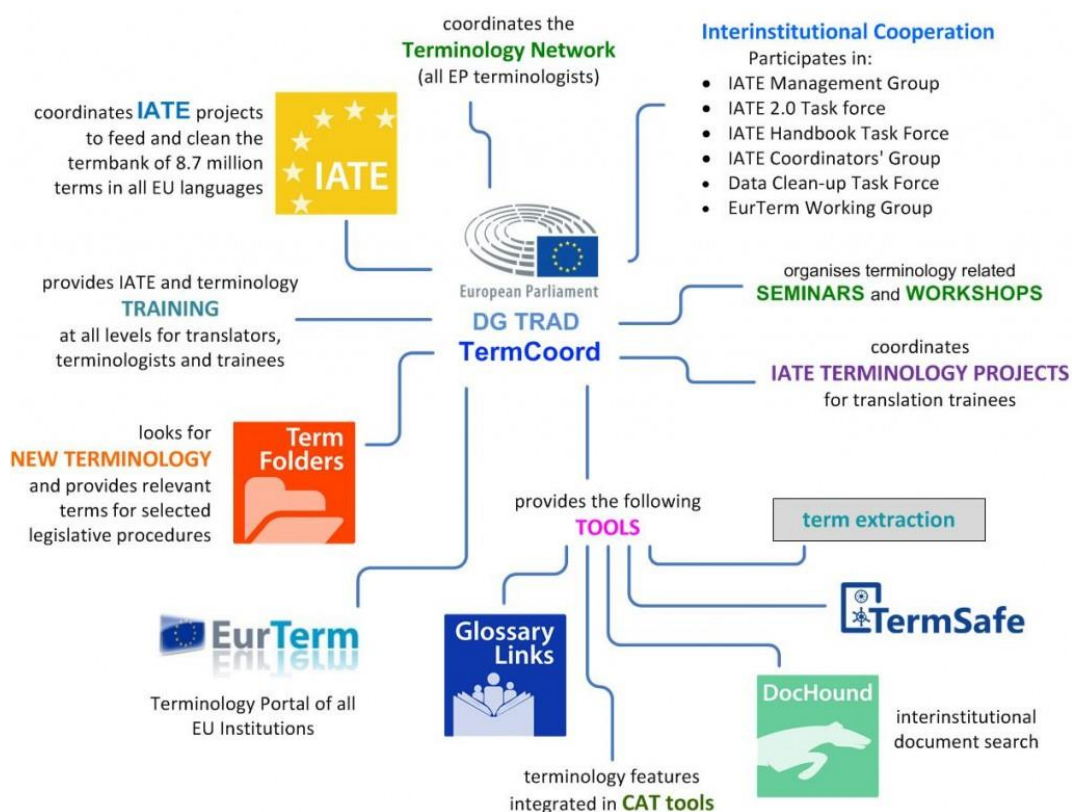


Fig. 2.2 Infografica sull'attività di TermCoord<sup>83</sup>

<sup>83</sup> Fonte immagine

<https://epnet.extranet.ep.europa.eu/http://tradportal.ep.parl.union.eu/terminology/what-we-do/>  
Consultata il 15 dicembre 2016.

## 2.2.2 Il ruolo di TermCoord nella Direzione Generale per la Traduzione

La Direzione Generale per la Traduzione del Parlamento europeo gioca, come abbiamo visto, un ruolo fondamentale nel multilinguismo. Ogni traduzione è legalmente vincolante e l'uso erraneo delle parole può causare diversi tipi di problemi che senz'altro inficerebbero l'uniformità dei concetti legali che esse veicolano. È centrale, quindi, l'importanza della terminologia in seno alla traduzione e il ruolo di TermCoord a sostegno delle altre Unità linguistiche.

Le Unità linguistiche sono composte principalmente da traduttori e assistenti di traduzione. Alcuni traduttori però, rivestono anche il ruolo di terminologo. Nel Parlamento europeo non esistono terminologi a tempo pieno<sup>84</sup>, quindi quei traduttori che hanno competenze terminologiche vengono indicati come terminologi<sup>85</sup> di riferimento per la loro Unità. La produzione della terminologia viene fatta o direttamente in IATE o tramite altri strumenti.

Nello specifico, il *workflow* della creazione di nuova terminologia in IATE nella Direzione Generale per la Traduzione può essere rappresentato nella figura sottostante.

---

<sup>84</sup> Non è così in tutte le istituzioni con dipartimenti di traduzione. La Commissione, ad esempio, prevede la figura di *full-time terminologist* all'interno di ogni dipartimento linguistico, con mansioni specifiche relative alla terminologia e quindi esente dal lavoro di traduzione.

<sup>85</sup> Solitamente sono almeno due per Unità.

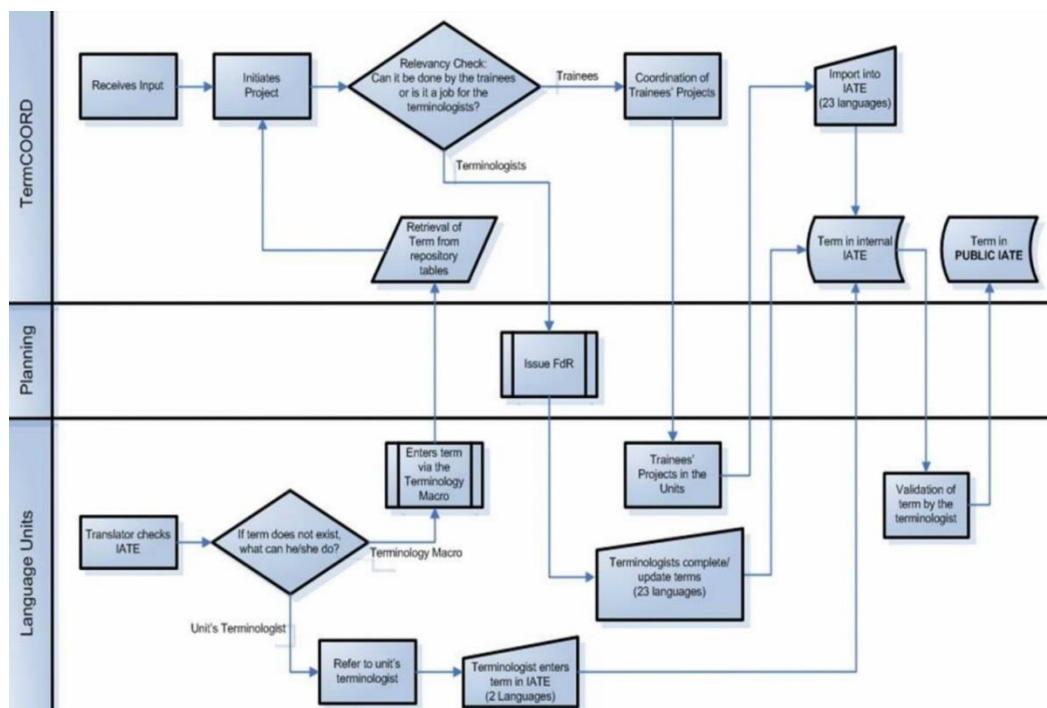


Fig. 2.3 *workflow* della creazione di nuova terminologia in IATE<sup>86</sup>

Il *workflow* si distribuisce su tre unità operative: TermCoord, Planning e le Unità linguistiche.

Osservando lo schema si può dire che la gestione della terminologia avviene attraverso il terminologo delle Unità linguistiche, che inserisce un nuovo termine direttamente in IATE o tramite TermCoord, che attiva progetti di terminologia appositi. In quest'ultimo caso, TermCoord riceve informazioni circa l'eventualità di inserire un nuovo termine o una nuova collezione di termini.

<sup>86</sup> Fonte <https://epnet.extranet.ep.europa.eu/http://tradportal.ep.parl.union.eu/terminology/what-we-do/terminology-in-the-ep-presentations-08122015/> (data consultazione: 15 Dicembre 2016).

TermCoord crea quindi un progetto su cui viene effettuato un controllo di rilevanza. In base ad essa, l'Unità sceglie di affidare il progetto ad un terminologo o ad un tirocinante. Nel caso in cui venga assegnato ad un tirocinante, a quest'ultimo viene richiesto di compiere ricerche terminologiche su un gruppo di termini che gli vengono assegnati. A conclusione del lavoro di ricerca terminologica, i termini candidati vengono validati da un terminologo e successivamente importati in IATE, dove saranno disponibile sia per la versione interna che per quella esterna del database. Se invece il progetto viene assegnato ai terminologi, interviene l'Unità *Planning*, che crea una *Feuille de Route* per quel determinato progetto. Ugualmente a quanto eseguito da un tirocinante, una volta che i terminologi completano la loro ricerca, i termini candidati vengono importati in IATE. In integrazione a questo sistema, la gestione di terminologia avviene tramite segnalazione del traduttore di un'Unità linguistica; se il termine non è presente in IATE, il traduttore può riferirsi al terminologo dell'Unità, che inserisce il termine, solitamente in due lingue, ovvero nella lingua del documento e nella lingua madre del traduttore, direttamente in IATE. In alternativa può usare la Terminology Macro oppure la TermSafe, che creano liste di termini da inserire in IATE, raccolte e gestite da TermCoord, che provvede all'inserimento nel database terminologico.

### 2.2.3 Terminology Macro e TermSafe

Il Parlamento ha utilizzato fino al 1 Aprile 2016 il software di traduzione Translation WorkBench<sup>87</sup>, prodotto da SDL Trados<sup>88</sup>, per poi passare esclusivamente ai software SDL

---

<sup>87</sup> Nel 2014 il Parlamento ha adottato il software commerciale di traduzione assistita SDL Studio, ma sta anche sviluppando un software proprietario, a partire dal 2014, denominato CAT4Trad.

<sup>88</sup> Trados è l'azienda produttrice di Translation Workbench. Nel 2005 fu acquistata dal SDL International, diventando quindi SDL Trados.

Studio, sviluppato dalla stessa azienda, e Cat4Trad<sup>89</sup>, sviluppato internamente dal Parlamento, che erano stati introdotti in DGTRAD già a partire dal 2014. Translation WorkBench, o TWB, era direttamente integrato in Microsoft Word e, dunque, l'attività di traduzione avveniva nell'applicazione. Unità quali DAS (*Applications & IT Systems Development*) e ITS (*Information Technology and Support Unit*) avevano sviluppato altri software e strumenti in ausilio ai traduttori, tra cui un *ribbon*<sup>90</sup> di Word, denominato *DGTRAD Ribbon*<sup>91</sup>.

Il *ribbon* raccoglie molti strumenti utili al lavoro di traduzione, tra cui, Terminology Macro, visibile nell'immagine sottostante, che permette, da un lato, di avere un accesso immediato, e direttamente da Word, a IATE

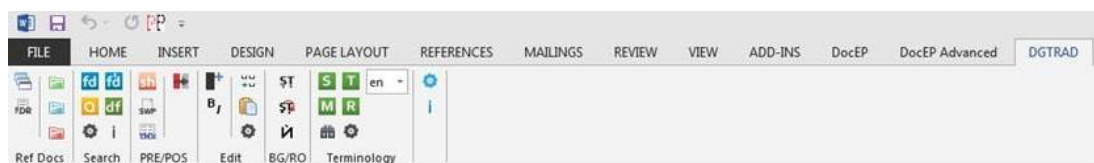


Fig. 2.4 DGTRAD Ribbon<sup>92</sup>

<sup>89</sup> CAT4Trad sta per "CAT, Computer Assisted Translation (Tool) 4 (DG) TRAD".

<sup>90</sup> Un *ribbon* è una interfaccia grafica composta da una striscia nella parte superiore di una finestra che raccoglie tutte le funzioni che un programma può compiere.

<sup>91</sup> Dopo la completa migrazione da TWB a SDL Studio, il ribbon non è più uno strumento centrale per il lavoro del traduttore ma è tuttavia ancora ampiamente impiegato.

<sup>92</sup> Fonte: screenshot di un'installazione *Word* in un computer del PE (data consultazione: 3 Novembre 2016).

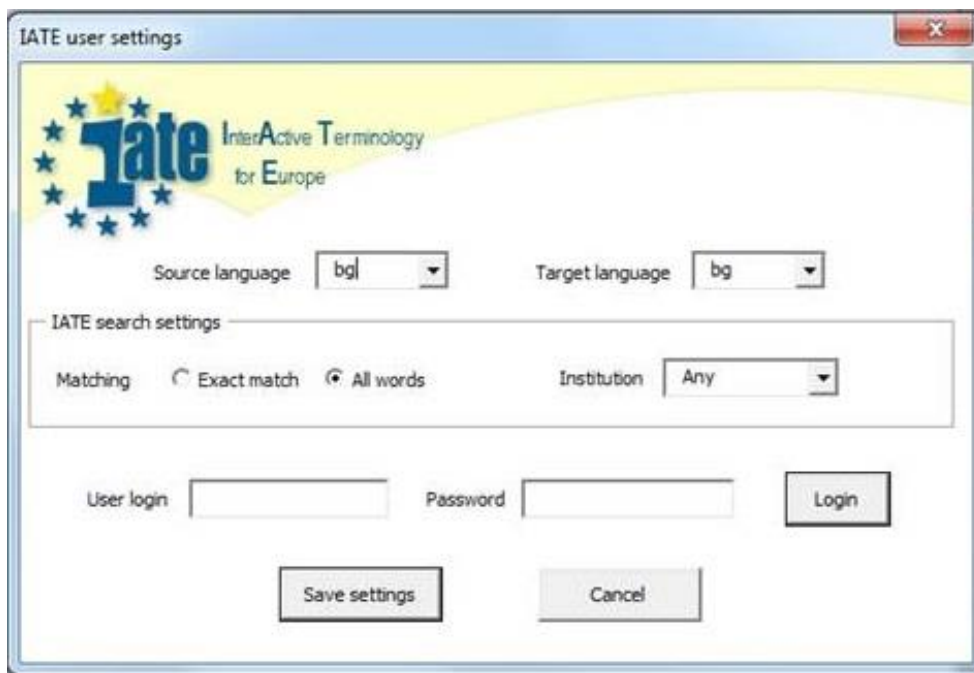


Fig. 2.5 L'interfaccia di IATE accessibile tramite Terminology Macro

e dall'altro di raccogliere terminologia individuata durante il lavoro di traduzione. Selezionando il termine nella lingua di origine e in quella di destinazione, la macro crea un nuovo documento contenente una tabella con in colonna i campi delle schede terminologiche di IATE. La tabella, successivamente compilata dalle Unità linguistiche, viene mandata a TermCoord per un coordinamento centrale della scheda terminologica e per l'inserimento in IATE. Qui sotto, un esempio della tabella autogenerata.

Source language	Source term	Target term	Author	Target file name	Date/time	Terme reference	Definition	Definition reference	Context	Context reference
en	cohesion policies	политике ди cohesion	Rossi Francesco	Document1						

Fig. 2.6 Tabella di termini generata dalla Terminology Macro<sup>93</sup>

<sup>93</sup> Fonte: screenshot di un'installazione *Word* in un computer del PE (data consultazione: 3 Novembre 2016).

Con il passaggio da TWB a SDL Studio e Cat4Trad, anche il *workflow* di gestione di terminologia si è spostato da Word all'interfaccia di Studio e la Terminology Macro, sebbene tuttora attiva e disponibile nelle installazioni dei traduttori di DGTRAD, non è più utilizzata.

TermCoord ha quindi optato per una soluzione alternativa, in attesa di un *workflow* che consideri i nuovi software. L'alternativa è *TermSafe*, un'interfaccia web disponibile sul portale internodi DGTRAD, che esegue pressappoco le stesse funzioni della Terminology Macro, con la differenza che *TermSafe* non crea un documento con tabelle in Word ma invia un'email direttamente ad una casella di posta funzionale di TermCoord.

Il flusso di lavoro si sposta quindi da Word ad Outlook e necessita dello stesso intervento da parte del Coordinamento centrale per l'inserimento dei termini in IATE.

## TermSafe

---



[About](#)

[Term Repository](#) (Read Only)

### Mandatory Fields

---

Your username

Source Term

en ▼

Target Term

bg ▼

FdR Number and/or Domain

Add additional information below or

---

Fig. 2.7 Interfaccia di TermSafe<sup>94</sup>

---

<sup>94</sup> Fonte

<https://epnet.extranet.ep.europa.eu/http://tradportal.ep.parl.union.eu/terminology/termsafe/#EN>  
(data consultazione: 3 Novembre 2016).



Come mostrato nella figura soprastante, *TermSafe* ha un'interfaccia piuttosto intuitiva.

Si compone di due parti, una con campi obbligatori e l'altra con campi opzionali.

Il traduttore/terminologo deve necessariamente scrivere il termine nella lingua di origine e di destinazione e indicare il numero di *Feuille de Route* e/o il dominio in cui esso è stato ritrovato. In aggiunta, può integrare quest'informazione con altri campi opzionali, quali *Term Reference*, *Definition*, *Definition Reference*, *Context*, *Context Reference* e *Remarks*, che, con eccezione per *Remarks*, sono alcuni dei campi che compongono le schede terminologiche di IATE.

#### 2.2.4 DocHound e GlossaryLinks

TermCoord gestisce anche altri due strumenti di ausilio al traduttore: *DocHound* e *GlossaryLinks*. *DocHound*<sup>95</sup> è uno strumento di ricerca documentaria interistituzionale, con una versione pubblica e una versione accessibile attraverso la *Intranet* del Parlamento.

---

<sup>95</sup> Consultabile pubblicamente <http://termcoord.eu/dochound/>.



## DOC Hound

<p><b>BASIC DOCUMENTS</b></p> <ul style="list-style-type: none"> <li>• Treaties (Europa)</li> <li>• Treaties (EUR-Lex)</li> <li>• Official Journal Search</li> <li>• EUR-Lex Simple Search</li> <li>• Budget of the EU</li> <li>• National legislation (N-Lex)</li> <li>• Financial regulation</li> <li>• International agreements</li> </ul>	<p><b>LEGISLATIVE DRAFTING</b></p> <ul style="list-style-type: none"> <li>• Interinstitutional style guide</li> <li>• Manual of precedents for acts established within the Council of the EU</li> <li>• Joint practical guide</li> <li>• Legislative Drafting – A Commission Manual</li> </ul>	<p><b>OVERVIEW OF PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• OEIL – Legislative Observatory of the EP</li> <li>• PreLex database – monitoring of the decisionmaking process between institutions</li> <li>• Council – Ordinary legislative procedure database</li> </ul>	<p><b>PRESS &amp; LIBRARY</b></p> <ul style="list-style-type: none"> <li>• EP Press Archives</li> <li>• RAPID Press release archive</li> <li>• Council press release database</li> <li>• Press services of other EU Institutions</li> <li>• EP Library</li> <li>• EU News</li> </ul>
<p><b>COUNCIL</b></p> <ul style="list-style-type: none"> <li>• SEARCH IN COUNCIL DOCUMENTS</li> <li>• Legislative acts under 'codecision'</li> <li>• Agreements and conventions database</li> <li>• Council meetings</li> <li>• Minutes</li> </ul>	<p><b>PARLIAMENT</b></p> <ul style="list-style-type: none"> <li>• SEARCH IN EP DOCUMENTS</li> <li>• Search in plenary sittings documents</li> <li>• Search in committees' documents</li> <li>• Ordinary legislative procedure</li> <li>• Parliamentary questions</li> <li>• Conference of Presidents – Bureau – Quaestors</li> <li>• Rules of procedure</li> <li>• Compendium of rules</li> <li>• EP Establishment Plan</li> <li>• Joint parliamentary assemblies</li> <li>• RDM Document models</li> <li>• Staff Regulations</li> <li>• Delegations and related documents</li> <li>• Studies according to policy areas</li> <li>• Rare documents on Epades</li> </ul>		<p><b>COMMISSION</b></p> <ul style="list-style-type: none"> <li>• SEARCH IN COMMISSION DOCUMENTS</li> <li>• EUR-Lex</li> <li>• Green papers</li> <li>• White papers</li> <li>• Comitology register</li> <li>• Annual reports on the EU's activities</li> </ul>
<p><b>COURT OF JUSTICE</b></p> <ul style="list-style-type: none"> <li>• SIMPLE SEARCH IN CASE-LAW</li> <li>• Advanced search in case-law</li> <li>• Search in EU case-law via EUR-Lex</li> <li>• Annual reports</li> <li>• National case-law concerning EU</li> </ul>	<p><b>COURT OF AUDITORS</b></p> <ul style="list-style-type: none"> <li>• SEARCH IN ECA DOCUMENTS</li> <li>• Annual reports</li> <li>• Special reports</li> <li>• Specific annual reports</li> <li>• Opinions</li> <li>• Annual activity reports</li> </ul>	<p><b>OTHER INSTITUTIONS &amp; BODIES</b></p> <ul style="list-style-type: none"> <li>• European Council</li> <li>• European Central Bank</li> <li>• Economic and Social Committee</li> <li>• Committee of the Regions</li> <li>• European Investment Bank</li> <li>• European External Action Service</li> <li>• European Ombudsman</li> <li>• European Data Protection Supervisor</li> <li>• Publications Office</li> </ul>	

Fig. 2.8 DocHound<sup>96</sup>

Nella versione pubblica, non tutti i documenti sono accessibili in quanto alcune istituzioni hanno inserito i link ai propri documenti caricati sui propri portali di riferimento, ma rimane tuttavia un utile ausilio per la ricerca di documenti.

<sup>96</sup> Fonte:

<https://epnet.extranet.ep.europa.eu/http://tradportal.ep.parl.union.eu/terminology/terminology-material/dochound/> (consultata il 3 novembre 2016).

*GlossaryLinks* è un database di glossari dotato di una funzione di ricerca<sup>97</sup> composto da circa 2500 risorse. I glossari sono controllati, aggiornati e integrati da TermCoord che li ricerca solitamente sul web e li cataloga in base al dominio. È anche disponibile in una versione pubblica<sup>98</sup>.



Fig. 2.9 *GlossaryLinks*<sup>99</sup>

Nella sezione “Category” sono elencati i 25 domini, mentre nel campo “Keyword” è possibile ricercare un glossario in base alla parola chiave.

Il *tool* offre anche la possibilità di cercare glossari partendo dalla lingua. *GlossaryLinks* colleziona glossari anche in lingue non ufficiali dell’Unione Europea, per un totale di circa 110 lingue.

<sup>97</sup> La ricerca viene effettuata solo nel nome del glossario e non nel suo contenuto. TermCoord è attualmente attiva per rendere “ricercabili” i glossari ed espandere quindi il potenziale del *tool*.

<sup>98</sup> Disponibile presso <http://termcoord.eu/glossarylinks/>.

<sup>99</sup> Fonte immagine <http://termcoord.eu/glossarylinks/> (consultata il 3 Novembre 2016).

## 2.2.5 European Advanced Multilingual Information System - *Euramis* e *Quest*

*European Advanced Multilingual Information System - Euramis* è un sistema di gestione di memorie traduttive<sup>100</sup> condivise da tutte le Istituzioni Europee, lanciato dalla Commissione Europea nel 1995, implementato in DGTRAD nel 2006 e gestito dall'Unità Pre-Trad.

L'obiettivo di *Euramis* è di ridurre la ripetitività nel lavoro traduttivo e ottenere coerenza linguistica e metodologica. Il Parlamento produce un'ampia varietà di documenti con una struttura testuale standard. *Template*, intestazioni, titoli e tutte le altre componenti strutturali di un documento rimangono invariate oppure sono leggermente modificate. Affidare questo tipo di lavoro al traduttore o all'assistente di traduzione rappresenterebbe rischi: la possibilità di commettere errori, la mancata uniformità dei documenti e l'eccessivo carico di lavoro.

*Euramis* è un sistema *client-server* che fornisce una varietà di servizi nel campo del *Natural Language Processing*. *Euramis* offre un sistema di archiviazione centrale di memorie traduttive condivisibili, analisi di massa dei dati linguistici e la possibilità di integrazione con altre applicazioni e servizi. Al livello utente, *Euramis* può essere interrogato tramite un'interfaccia web che comunica con il server e dà il risultato nella sequenza appropriata. Le applicazioni più comuni sono:

---

<sup>100</sup> Una memoria traduttiva è un database utilizzato da software di traduzione assistita al computer. Per una disamina sulle memorie traduttive si rimanda ad Austermuhl (2001), Bowker (2002) e (2005).

- ricerca documentaria, che avviene in EUR-Lex<sup>101</sup>, o ricerca di concordanze, che riguarda i vari segmenti delle memorie traduttive;
- ricerca di memorie traduttive con la possibilità di effettuare download ed editing;
- modifica delle memorie traduttive.

*Quest* è un *metasearch tool*<sup>102</sup> che centralizza, semplifica ed accelera la ricerca terminologica e documentaria tramite un'interfaccia unica.

Per le istituzioni europee, *Quest* rappresenta un formidabile strumento di supporto di lavoro. I traduttori e le traduttrici del Parlamento europeo dispongono di un link diretto a *Quest* nel software di traduzione SDL Studio. L'uso più comune è di selezionare un determinato termine e lanciare l'applicazione *Quest* che lo ricercherà nelle risorse utili al lavoro traduttivo ovvero, tra le altre, IATE, Eur-Lex ed Euramis.

Nella figura che segue, un esempio di ricerca del termine "politiche di coesione" che mostra risultati di *Euramis*.

La sezione principale mostra tutti i segmenti di memorie traduttive che contengono quel termine. Ogni risultato mostra dati e metadati aggiuntivi quali la memoria traduttiva di provenienza, il numero di documento, il codice del richiedente, l'anno, il tipo di documento, la lingua d'origine e Obs. (ovvero il numero di procedura e *Fueille de Route*).

---

<sup>101</sup> Eur-Lex è un sito web dell'Unione Europea che offre la consultazione online gratuita di tutti i testi di legge dell'UE.

<sup>102</sup> Per "metasearch tool" si intende uno strumento che permette di effettuare ricerche da fonti diverse e accessibili da un'unica interfaccia.

The screenshot shows the Quest Metasearch interface with the search term 'pubbliche di coesione'. The search parameters are: Source: Italian, Target(s): Italian, Options: Exact string, All words. The results table shows 30 hits, with the first few rows displaying document details like 'EESC-COR Doc. No. 00368' and 'EP Committees Doc. No. 1104418'. The interface also includes navigation options like 'LIVE', 'Eur-Lex', and 'Eur-Lex'.

Fig. 2.10 Quest con Euramis<sup>103</sup>

La Figura 2.11 mostra invece la stessa query ma con i risultati di Eur-Lex

The screenshot shows the Quest Metasearch interface with the search term 'pubbliche di coesione'. The results are displayed in a list format, showing document titles, dates, and authors. The first result is 'COMUNICAZIONE DELLA COMMISSIONE AL PARLAMENTO EUROPEO, AL CONSIGLIO E ALLA CORTE DEI CONTI TUTELA DEL BILANCIO DELL'UE FINO ALLA FINE DEL 2015' dated 18/07/2016. The interface also includes navigation options like 'LIVE', 'Eur-Lex', and 'Eur-Lex'.

Fig. 2.11 Quest con Eur-Lex

<sup>103</sup> Fonte Figure 10, 11 e 12: <https://webgate.ec.testa.eu/questmetasearch/> (data consultazione: 27 novembre 2016).

Infine, la figura sottostante mostra il termine “politiche di coesione” ricercato in IATE.

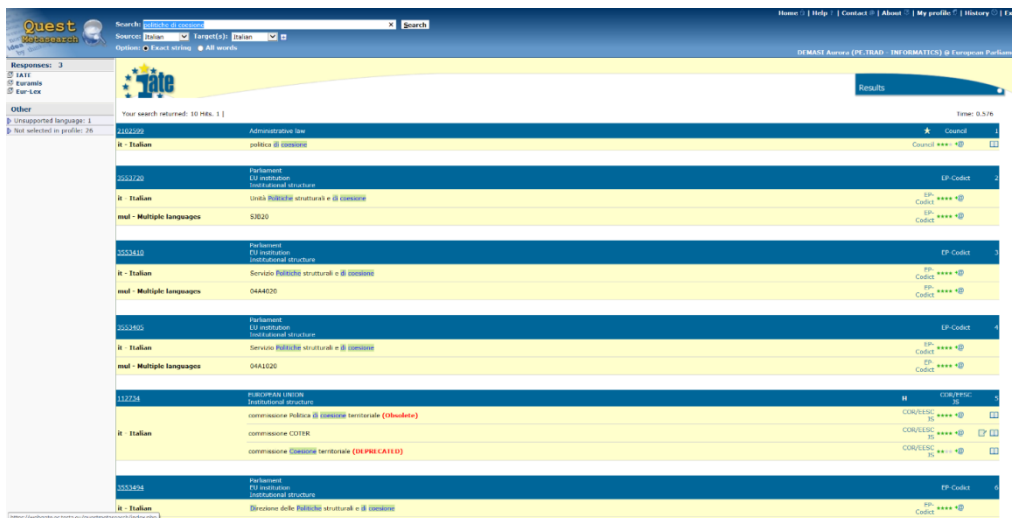


Fig. 2.12 Quest con IATE

I tre risultati sono le funzioni principali che i traduttori dell’Unione Europea usano ogni giorno per il lavoro terminologico. Tuttavia un vero e proprio *workflow* per la gestione della terminologia è ancora in fase di sviluppo.

## 2.2.6 TermFolder

Un’altra importante attività di TermCoord è la creazione delle cosiddette *TermFolder*, ovvero cartelle di terminologia create per le procedure legislative o gli altri tipi di documenti istituzionali che presentano problematiche linguistiche e terminologiche particolarmente complesse.

Le *TermFolder* contengono la cosiddetta “terminologia proattiva”, ovvero liste di termini che vengono estratte da testi legislativi contenenti terminologia complessa e prima che arrivino in seno alle Unità linguistiche di DGTRAD.

Il *workflow* prevede che la Direzione B nomini, tra i traduttori, un coordinatore, denominato *file coordinator*, e informi TermCoord. Il coordinatore raccoglie eventuali informazioni rilevanti su uno spazio di comunicazione ospitato sulla *Intranet* del Parlamento. Basandosi su queste informazioni, TermCoord decide se è necessaria la creazione di una *TermFolder* e procede con l'estrazione di termini candidati dai documenti disponibili. I concetti vengono quindi ricercati, definiti e documentati in IATE in inglese. TermCoord crea una pagina web sul proprio sito interno, dove include un glossario in inglese con una lista delle entrate di IATE aggiornate, relative definizioni e altre informazioni che non possono essere documentate nel database terminologico, oltre ad eventuali materiali aggiuntivi trovati durante la ricerca. TermCoord conclude il proprio lavoro con la pubblicazione di comunicazioni sulla creazione della *TermFolder*.

Le schede terminologiche sono poi completate nelle altre lingue ufficiali da parte dei terminologi del Parlamento e quindi disponibili per la consultazione da parte di tutti i traduttori a cui saranno assegnati quei documenti problematici.

### 2.2.7 Progetti di terminologia dei tirocinanti

Il Parlamento, insieme alle altre istituzioni, è dotato di una grande macchina organizzativa di tirocini formativi per studenti europei.



Nel caso di DGTRAD, ogni tre mesi il Parlamento riceve un gruppo di stagisti traduttori, che vengono assegnati alle varie Unità linguistiche o orizzontali. Una parte obbligatoria del loro tirocinio prevede la gestione di un progetto di terminologia, che permette loro di sviluppare competenze terminologiche e di utilizzo di IATE.

I progetti vengono realizzati grazie alla stretta collaborazione tra TermCoord e le Unità linguistiche. L'obiettivo del tirocinio è dare la possibilità ai giovani traduttori di avere un'esperienza nel campo del lavoro terminologico, ma anche di alimentare ed aggiornare IATE, ad esempio completando alcune schede terminologiche assenti in alcune lingue o aggiungendo nuovi termini in un determinato dominio.

La selezione dei termini può quindi avvenire sia per i termini già esistenti in IATE, ma mancanti in alcune lingue, o con campi vuoti o obsoleti, oppure registrando nuovi termini su domini che vengono accordati tra TermCoord e Unità linguistiche. Più precisamente, TermCoord è referente per progetti tematici su vari domini, come ad esempio *acquacultura, ornitologia, disabilità, finanza, diritti umani, IT (Information Technology)* ecc.; oppure sono le stesse Unità linguistiche a proporre dei progetti, su temi di rilevanza oppure per un aggiornamento delle entrate a rischio obsolescenza o con bassa affidabilità.

TermCoord identifica liste di termini organizzate in progetti terminologici tematici e li assegna ai tirocinanti di un'Unità linguistica. Questi cominciano così il lavoro di ricerca dei termini, che avviene sotto la supervisione del terminologo della propria Unità linguistica.

Durante tutta la durata del progetto, TermCoord fornisce supporto e offre formazioni in cui vengono date indicazioni su come effettuare le ricerche e/o come modificare le schede di IATE.

### 2.2.8 La comunicazione di TermCoord

Terminology Coordination Unit è l'unica Unità della Direzione Generale a comunicare direttamente con il cittadino.

Il canale principale di comunicazione con il "mondo esterno" è [www.termcoord.eu](http://www.termcoord.eu), un sito in cui il team di comunicazione dell'Unità raccoglie articoli, interviste, video, curiosità, ma anche strumenti software ed e-books sulla terminologia, traduzione e interpretazione.

Il sito nacque grazie all'iniziativa di Maria Pia Montoro, tirocinante presso TermCoord nel 2011, nonché alla sensibilità e all'attenzione sulla comunicazione di Rodolfo Maslias, capo Unità di TermCoord. [termcoord.eu](http://termcoord.eu) era inizialmente un blog che è velocemente diventato, in seguito ai numerosi e sempre vivaci contributi dei successivi team di comunicazione, un punto di riferimento per linguisti, terminologi e in generale amanti delle lingue.

Ad oggi, 750mila sono le visualizzazioni uniche e più di 1000 visite uniche al giorno.

### 2.2.9 EurTerm: una piattaforma collaborativa per traduttori, terminologi, esperti di dominio e comunicatori

L'accordo sulla terminologia specialistica è da sempre frutto di un processo di collaborazione all'interno di una comunità di esperti. In passato il confronto avveniva e attraverso riunioni e consultazioni con i rappresentanti delle divisioni traduttive delle varie istituzioni, principalmente del Parlamento, Consiglio e Commissione nei casi di terminologie particolarmente difficili. Oggi, con l'aumento della mole di documenti da tradurre e il conseguente incremento delle terminologie specialistiche è emersa l'esigenza di sviluppare una piattaforma per la collaborazione tra traduttori, terminologi, esperti di dominio e comunicatori.

A rispondere a questa esigenza è EurTerm, il Portale di Terminologia interistituzionale delle Istituzioni Europee, uno spazio collaborativo per la comunicazione e la condivisione delle risorse terminologiche.

I contenuti di EurTerm sono gestiti dai servizi di terminologia delle istituzioni mentre il coordinamento e l'amministrazione è affidato allo IATE Management Group (IMG), l'organo di governo interistituzionale che orienta le strategie di IATE.

L'obiettivo della piattaforma EurTerm è di fungere da punto di accesso unico alle risorse terminologiche, strumenti e database dell'UE, agevolare la comunicazione interistituzionale, condividere informazioni terminologica e la conoscenza linguistica e ospitare spazi di discussione per gruppi di interesse, per progetti, di terminologia o per sviluppi tecnologici, per far evolvere IATE. Principali utilizzatori di EurTerm sono i

terminologi delle istituzioni, i giuristi linguisti, i traduttori, gli interpreti, gli amministratori e redattori di documenti unionali.

EurTerm offre un ampio ventaglio di risorse linguistiche e traduttive, anche interattive, in ausilio al lavoro terminologico. Dal sito è possibile accedere ad una *wiki* che ogni comunità linguistica può organizzare e gestire autonomamente, allestendo aree di discussione e sezioni di archiviazione. Per ogni lingua si può creare uno spazio di discussione a proposito della terminologia di una determinata procedura e raccogliere i pareri dei linguisti e degli esperti di dominio.

Il portale è popolato da una nutrita sezione di materiali sulla terminologia dell'Unione Europea quali glossari istituzionali, presentazioni, *framework* terminologici e meeting; un'area è dedicata allo *IATE Management Group*, ai suoi obiettivi e componenti, a IATE e alla sua gestione, con archivi, documentazioni e statistiche, ma anche ai suoi processi operativi e ai relativi gruppi di lavoro. Un'ampia pagina è dedicata ai progetti terminologici, in cui ogni istituzione adotta uno o più domini e ne gestisce i termini.

EurTerm mostra, inoltre, risorse per la formazione, quali materiali e calendari, ma anche risorse esterne, quali database, siti di terminologia, *search tools*, estrattori terminologici, *widget*, estensioni per browser, e-books, riviste e articoli. Una sezione è dedicata agli aspetti tecnici dell'integrazione di IATE in SDL Studio.

La piattaforma ha un'area in cui è possibile trovare i terminologi di ogni istituzione e poterli contattare facilmente, a cui si aggiunge un'altra in cui si può ricercare un esperto in base al dominio di conoscenza. Il portale include anche un blog dove vengono postati articoli su eventi, attività ed altre notizie relative alle terminologie specialistiche. Nella

sezione 'Calendario' è possibile avere informazioni aggiornate sui workshop, le conferenze, le giornate di formazione ed altri eventi –anche dal mondo accademico - che riguardano la traduzione e la terminologia.

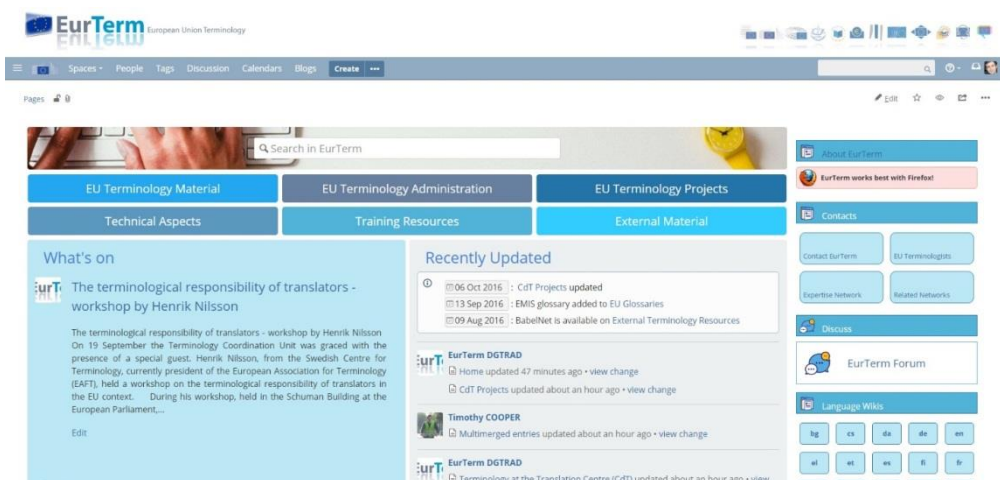


Fig. 2.13. Home page del portale Euterm<sup>104</sup>

EurTerm, se ben utilizzata, potrà essere una piattaforma di condivisione della conoscenza terminologica interistituzionale, offrendo risorse e soluzioni per gli addetti ai lavori delle istituzioni, aprendosi anche ai contributi dell'industria linguistica e dell'accademia. In questo senso, EurTerm rappresenta l'unico fenomeno di (social) network terminologico, un ambiente florido dove proporre, discutere e armonizzare la terminologia interistituzionale e le sue variazioni intra e interlinguistiche.

<sup>104</sup> Fonte

<https://webgate.ec.europa.eu/fpfis/wikis/display/iatetp/Home;jsessionid=C2D3D6E2098C370C7855234BD27F7678?ticket=ST-2693775-UlqrJFJtKoOEJXIYidnzYACZpMFOobe4cvZhC4AIKFQ14ZCoVMteEGrG2j7sPS8tA41eFLRBdoJ8ZmJ77NmjLK-zGb7phj4Ee1KpAUIOeOwOu-WF6d9OY7S0apDNMbZow2SYeD736zcsviStPPnHAI9zkO> (data consultazione 15 settembre 2016).

## CAPITOLO III – IATE, INTER-ACTIVE TERMINOLOGY FOR EUROPE

IATE, acronimo di Inter-Active Terminology for Europe, è il database interistituzionale di terminologia dell'Unione Europea. È alimentato da dieci istituzioni: Commissione Europea, Parlamento europeo, Consiglio, Corte di Giustizia, Corte dei Conti, Comitato e Sociale, Comitato delle Regioni, Banca Centrale Europea, Banca di Investimenti Europea, Centro di Traduzione degli Organismi dell'Unione Europea e contiene ad oggi circa 1,4 milioni di entrate e 8,6 milioni di termini.

Il progetto IATE nasce con l'obiettivo di disporre di un'infrastruttura comune, *web-based*, per tutte le risorse terminologiche dell'Unione Europea, aumentandone la condivisione, la facilità di accesso e la standardizzazione. Il database è gestito dallo IATE Management Group, un organo di governo composto da rappresentanti delle divisioni terminologiche delle istituzioni partner.

In questo capitolo si parlerà della storia di IATE, delle caratteristiche, struttura, modalità di gestione delle fonti e inserimento dei dati terminologici e degli sviluppi futuri.

### 3.1 Cenni storici

Prima di IATE, ogni istituzione aveva un proprio database terminologico: Eurodicautom, alla Commissione, TIS al Consiglio, Euterpe al Parlamento, Euroterms al Centro di Traduzione e CDCTERM alla corte di Conti.

Negli anni '90, tuttavia, l'esigenza di un unico database che fungesse da singolo punto di accesso alla terminologia e ne migliorasse la gestione e la standardizzazione si fece sempre più chiara.

Il Centro di Traduzione lavorò inizialmente su un database decentralizzato il cui nome iniziale era *Inter-Agency Terminology Exchange*, o *IATE*<sup>105</sup>. Il suo obiettivo era gestire la terminologia degli organismi e delle agenzie europee, ma fu poi deciso che tutte le istituzioni dovessero far parte del progetto. Fu quindi eseguito uno studio di fattibilità<sup>106</sup> le cui raccomandazioni furono di unire tutta la terminologia esistente in un unico database, con un'interfaccia *user-friendly* e regole di gestione comuni.

Nell'autunno del 1999 il Centro di Traduzione pubblicò un bando di gara per lo sviluppo di un nuovo database terminologico che migliorasse l'integrazione delle risorse terminologiche e garantisse una futura cooperazione tra le istituzioni e gli organi dell'Unione Europea.

Il contratto fu assegnato all'industria greca "Quality & Reliability" e all'istituto di ricerca danese "Center for Sprogteknologi". In un documento redatto dal *Telematics Between*

---

<sup>105</sup> Il titolo fu poi cambiato in "Interactive Terminology for Europe" per riflettere la natura interistituzionale del progetto.

<sup>106</sup> Assegnato alla compagnia Atos.

*Administrations Committee* (TAC) si presentavano le nuove caratteristiche, i benefici e i componenti tecnici del nuovo database, il cui *Global Implementation Plan* fu presentato da Alastair Macphail, che al tempo rivestiva il ruolo di Capo della Standardizzazione presso il Centro di Traduzione degli Organismi dell'Unione Europea. La prima fase di sviluppo durò circa tre anni e mezzo, che videro un susseguirsi di diverse task force, con diversi obiettivi.

La responsabilità generale del progetto era stata affidata all'*Interinstitutional Committee for Translation* (ICT)<sup>107</sup> mentre i finanziamenti erano forniti dal programma IDA, *Interchange of Data between Administrations*.

Altri gruppi, come l'*Expert Group for EU Terminology database* (EGEUT), composto da rappresentanti delle varie istituzioni, furono create con finalità di contributo tecnico e di monitoraggio del progetto.

Il primo prototipo fu consegnato a marzo 2001. Esso usava un'architettura software, "Oracle Forms", che fu velocemente considerata non accessibile per gli utenti e soprattutto non adatta alla gestione di dati linguistici.

L'interfaccia fu quindi ricostruita usando un'architettura HTML. Alla prima fase pilota seguì una seconda, rilasciata a giugno 2002, in cui l'interfaccia fu ulteriormente rivista basandosi sul feedback di un pool di esperti e traduttori.

Tra dicembre 2000 e gennaio 2001 e nel bimestre marzo/giugno 2002 i dati terminologici a disposizione delle varie istituzioni furono convertiti e caricati su IATE.

---

<sup>107</sup> Il Comitato, fondato nel 1995 e composto da membri di provenienza interistituzionale, si occupava di ottenere economie di scala in relazione alla traduzione.



Questa fu senz'altro una delle maggiori sfide di IATE poiché i dati erano tutti strutturati in maniera eterogenea nei diversi database preesistenti.

La struttura che venne adottata come riferimento per IATE fu quella di Eurodicautom della Commissione, sia perché era ritenuta la migliore in termini di accessibilità sia perché il database era il più completo rispetto agli omologhi delle altre istituzioni.

La conversione dei dati terminologici e la migrazione verso un contenitore unico ha, probabilmente tuttora, un impatto sulla ridondanza delle entrate.

Molti furono gli sforzi di pulitura ma di fatto il fenomeno dei duplicati gioca tuttora un ruolo importante nei progetti attuali.

La versione finale di IATE fu consegnata a dicembre 2002 e migrata presso il Data Centre della Commissione, dove è tuttora ospitato. Altri test furono condotti nei primi mesi del 2003, quando il contratto tra IDA e gli assegnisti venne al termine<sup>108</sup>.

Il 2004 fu l'anno dell'allargamento a Est dell'Unione Europea<sup>109</sup>, in cui dieci nuovi Stati membri entrarono ufficialmente nell'Unione. Nel 2007 si aggiunsero anche Bulgaria e Romania e lo stesso anno l'Irlandese diventò lingua ufficiale per l'Irlanda che già dal 1973 era Stato membro. Nel 2013 anche la Croazia divenne un membro ufficiale e le lingue ufficiali salirono a 24.

I nuovi ingressi hanno avuto ovviamente un chiaro impatto anche su IATE che ha visto il proprio bacino espandersi esponenzialmente.

---

<sup>108</sup> Il costo totale fu di 1,41 milioni di euro tra il 1999 e il 2003.

<sup>109</sup> Si rimanda a <http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=URISERV%3Ae50017>.

IATE fu lanciato ufficialmente nel 2004 come risorsa interna dell'Unione Europea. Conteneva 8,4 milioni<sup>110</sup> di termini in 127 lingue<sup>111</sup>. IATE diventò rapidamente uno strumento utilizzato quotidianamente dai traduttori delle istituzioni europee, che effettuano oggi 18 milioni di *query* all'anno, aggiungendo 120.803 termini e modificandone 310.333<sup>112</sup>.

Un'importante svolta per IATE si ebbe a giugno 2007 quando fu lanciata la versione pubblica di IATE. Rendere IATE pubblico fu una grande sfida in termini di supporto e aggiornamenti ma riscosse anche un impressionante numero di query da parte degli utenti esterni: nel 2016 sono state 36 milioni le ricerche effettuate sul database con una media di 100mila *queries* effettuate ogni giorno<sup>113</sup>.

### 3.2 IATE – principi generali

IATE è un database dinamico creato, principalmente, per supportare la creazione di *draft* di testi, anche legali, dell'Unione Europea. Come database terminologico, la sua funzione specifica è di fornire dati rilevanti, affidabili, controllati e facilmente accessibili che rappresentino un valore aggiunto distinto in paragone con altre risorse di

---

<sup>110</sup> Oggi sono circa 8,6 milioni. L'incremento, che può apparire non vistoso, è dettato soprattutto dalle numerose operazioni di consolidamento, mirate appunto alla riduzione della ridondanza e duplicati. I dati sono estratti da "IATE indicators", documento ufficiale per le statistiche del database pubblicato su base annuale. I dati sono aggiornati all'anno 2016.

<sup>111</sup> La versione interna di IATE contiene anche entrate di lingue non ufficiali dell'Unione Europea, come ad esempio il cinese, l'afrikaans, lo yiddish e il latino.

<sup>112</sup> I dati sono relativi all'anno 2016 (data di rilevamento: 13 gennaio 2017).

<sup>113</sup> Per ulteriori informazioni sulla storia di IATE, consultare F. Rossi, (2014) su <http://termcoord.eu/2014/05/10-years-iate-timeline/>.

informazione lessicale, in particolare le memorie traduttive e gli archivi di documenti elettronici.

Il bisogno crescente di definizione terminologica precisa concernente le attività dell'UE è sempre più complessa e diversificata e la necessità di una legislazione in linguaggio preciso e accurato presuppone che questo tipo di dati garantisca qualità di alto livello. Per questa ragione, IATE deve essere alimentato e gestito con rigore poiché eventuali discrepanze tra le diverse versioni linguistiche non possono essere sempre rilevate automaticamente, ma spesso solo tramite l'intervento dell'utente.

in *"IATE HandBook"*<sup>114</sup>, si legge che:

*« IATE's usefulness for multilingual drafting, translation and interpretation in the EU sphere must always be the prime concern [...]. It allows terminologists to make a greater contribution to the EU policy of encouraging multilingualism and making EU legislation more transparent to the citizen [...]. The strength and uniqueness of IATE lie in its multilingualism It is very important to promote the development of entries towards multilingualism, by consolidating and merging any monolingual or bilingual entries for a concept and by adding as many languages as possible to new entries»*<sup>115</sup>.

Il multilinguismo e l'utilità per la traduzione, *drafting* e interpretazione sono quindi i principi generali su cui si basa IATE.

---

<sup>114</sup> IATE HandBook è un documento interistituzionale che definisce i principi e le regole generali sulla gestione delle entrate e del lavoro terminologico e contiene guide e istruzioni pratiche per l'utilizzo del database.

<sup>115</sup> IATE HandBook, p. 4 (data consultazione: 27 ottobre 2016).

Come si legge in IATE Best Practice<sup>116</sup>, «*the information fed into IATE must have an added value over and above what can be found in documentary databases or on the Internet*»<sup>117</sup> e deve essere il risultato di un processo terminologico mirato all'accrescimento dell'affidabilità della soluzione proposta.

«*The terminologist must ensure that the term entered is relevant, in other words, that it genuinely corresponds to a past, present or potential drafting, translating or interpreting problem in a EU area. He or she must also ensure, as far as possible, that the term corresponds to a real concept used in the relevant context and not simply to a one-off name or occurrence*»<sup>118</sup>.

Prima di creare un'entrata, l'utente deve assicurarsi che il concetto non sia già presente e deve anche escludere l'eventualità che quel concetto possa essere contenuto nel database in lingue usate più di frequente, come il francese e l'inglese.

Per quanto riguarda i criteri di input dei termini, IATE Best Practice rimanda alla "conciseness of the term", ovvero:

«*the term must be as concise as possible, i.e. it should be the smallest possible indivisible unit to designate a precise concept. Complex expressions combining several concepts must be broken down into their constituent parts and a separate entry created for each concept*»<sup>119</sup>.

---

<sup>116</sup> Anche *IATE Best Practice* è un documento interistituzionale che descrive i fondamenti del lavoro terminologico dell'UE, un aspetto che ha in comune con il documento *IATE Handbook*. Tuttavia se ne differenzia perché non contiene guide e istruzioni dettagliate per l'utilizzo del database.

<sup>117</sup> IATE Best Practice, *cit.* [https://iate.cdt.europa.eu/iatenew/help/best\\_practice.html](https://iate.cdt.europa.eu/iatenew/help/best_practice.html) (data consultazione 28 ottobre 2016)

<sup>118</sup> *Ibidem.*

<sup>119</sup> *Ibidem.*

Tale documento considera anche l'eventualità di inserire parole di uso comune. L'aggiunta di lessico comune deve essere limitato a quei casi in cui sia strettamente necessario o in quelli in cui la presenza in IATE crei un valore aggiunto in paragone con quello dei dizionari. I neologismi, inoltre, sono fenomeni lessicali con cui i traduttori hanno a che fare ogni giorno e la cui gestione non è una facile attività.

Il terminologo, si cita in IATE Best Practice,

*«must be able to help the user by proposing well-founded solutions with due consideration for natural processes for forming new terms and the socio-cultural factors which determine the acceptance of neologisms in a language community. If a term is a neologism it should be indicated in a note»<sup>120</sup>.*

### 3.3 La struttura delle schede terminologiche di IATE

Le entrate di IATE sono divise in tre livelli:

1. Language-Independent Level (LIL). Contiene i metadati e le informazioni sul concetto (come i domini, ad esempio) e si applica a tutti i dati ad esso sottostante;
2. Language Level (LL). Riguarda il concetto, scritto in una determinata lingua, e si applica a tutti i termini di quella lingua. La definizione del concetto deve essere simile in tutte le lingue e applicabile a tutti i termini della stessa scheda. La definizione, quindi, va idealmente inserita nel LIL, ma per permettere la creazione di una definizione in ogni lingua, le definizioni dei concetti di IATE appaiono in questo livello;

---

<sup>120</sup> *Ibidem.*

3. Term Level (TL). È il livello di un particolare termine in una data lingua.

The image shows a screenshot of an IATE terminological card with three main sections highlighted by red boxes and numbered 1, 2, and 3 with red arrows.

- Section 1:** The top section, labeled 'Domains: 1016004 - European Union' and '24 - FINANCE'. It includes 'Problem language: en' and a list of 'Cross references' with narrower terms like 933855, 771576, 3537723, 890098, 890097, and 1162442. A large red '1' is placed to the right of this section.
- Section 2:** The middle section, labeled 'Council'. It contains a 'Definition' in Italian: 'denominazione comune per cinque fondi: il Fondo europeo di sviluppo regionale (FESR), il Fondo sociale europeo (FSE), il Fondo di coesione (FC), il Fondo europeo agricolo per lo sviluppo rurale (FEASR) e il Fondo europeo per gli affari marittimi e la pesca (FEAMP)'. It also includes a 'Reference' to a Council communication and a 'Note' indicating an update by BARBAIL on 21/10/2013. A red arrow points to this section with a large red '2'.
- Section 3:** The bottom section, labeled 'COM'. It shows a 'Term Group: 1' with 'Term: Fondi strutturali e d'investimento europei' and 'Reliability:3'. It includes a 'Reference' to a Commission communication, a 'Context' explaining the term's use, and another 'Reference' to a COM(2012) 496 document. Below this, it shows another 'Term Group: 1' with 'Short Form: Fondi ES1' and 'Reliability:2', and a third 'Term Group: 1' with 'Short Form: fondi SIE' and 'Reliability:3'. A red arrow points to this section with a large red '3'.

Fig. 3.1 una scheda terminologica di IATE<sup>121</sup>

<sup>121</sup> Fonte immagine <https://iate.cdt.europa.eu/iatenew/login.do?method=logout> (data consultazione, 27 Ottobre 2016).

### 3.3.1 Language-Independent Level

Questo livello contiene metadati e informazioni sul concetto e sul dominio o domini.

Il dominio è «il campo della conoscenza a cui il concetto appartiene»<sup>122</sup>, esso deve identificare il linguaggio e il contesto di cui il concetto fa parte. Solitamente in IATE un'entrata ha almeno un dominio ma è possibile che anche sceglierne di più, laddove necessario. I domini di IATE sono rappresentati dai codici a due, quattro e sette caratteri che corrispondono ai domini di EuroVoc<sup>123</sup>.

Gli utenti di IATE sono incoraggiati a consultare EuroVoc, verificare se il termine vi è già contenuto ed eventualmente paragonarlo con gli altri termini ad esso potenzialmente attinenti<sup>124</sup>.

L'interfaccia<sup>125</sup> per l'inserimento di un termine in IATE è mostrata nella figura sottostante.

---

<sup>122</sup> IATE Handbook, *cit.* p. 10 (traduzione dell'autore).

<sup>123</sup> EuroVoc è un thesaurus multilingue e pluridisciplinare che comprende la terminologia dei settori d'attività dell'Unione europea. Contiene termini in 23 lingue dell'UE (bulgaro, ceco, croato, danese, estone, finlandese, francese, greco, inglese, italiano, lettone, lituano, maltese, olandese, polacco, portoghese, rumeno, slovacco, sloveno, spagnolo, svedese, tedesco e ungherese), oltre che in tre lingue di paesi candidati all'adesione: македонски (mk), shqip (sq) e српски (sr).

EuroVoc è curato dall'Ufficio delle pubblicazioni, che è passato ad una gestione del thesaurus basata sull'ontologia e alle tecnologie del web semantico, in linea con le raccomandazioni del W3C e con gli ultimi sviluppi negli standard di classificazione.

<sup>124</sup> In alcune entrate di IATE è possibile trovare la dicitura "Domain code not specified, Domain 00". Questo significa che l'entrata non ha dominio.

<sup>125</sup> Si tratta dell'interfaccia, avanzata, di IATE interno. Gli utenti possono scegliere anche un'interfaccia "basic" ma è fortemente raccomandato di non usarlo in quanto carente di alcuni campi.

Fig. 3.2 Interfaccia di creazione di un'entrata<sup>126</sup>

I **domini** sono scelti tramite l'apposito pulsante "Add". Si potranno scegliere tre livelli di domini, a cascata, come mostrato nell'esempio sotto.

Fig. 3.3 Menu di scelta dei domini

I livelli vanno quindi da un livello più generale (livello 1), a uno più specifico (livello 3).

Le **note del dominio** vanno compilate in inglese e danno maggiori informazioni sul contesto in cui il concetto viene usato. "**Origins**" è un campo da compilare se un

<sup>126</sup> Fonte immagine <https://iate.cdt.europa.eu/iatenew/login.do?method=logout> (data consultazione, 27 Ottobre 2016).



concetto è specifico di un certo Stato membro o specifico dell'UE. Le note del campo "Origins" seguono le stesse regole delle note di dominio.

"Problem Language (PL)" è la lingua cardine da cui tutte le altre devono dipendere, funge quindi da àncora e le definizioni delle altre lingue devono essere quanto più simili ad essa. Solitamente Problem Language è la lingua del testo fonte in cui il termine compare. Nel caso di termini relativi ad animali e piante, PL è il latino, per tutti gli altri concetti è solitamente l'Inglese. Inserire sempre Problem Language è utile soprattutto in termini di affidabilità del termine perché il valore di un termine in un'altra lingua è correlato al concetto espresso il PL. Un esempio è illustrato in figura 3.4<sup>127</sup>.

Domains:	24 - FINANCE 2416 - Financial institutions and credit	Note:	
Origins:	CY - CYPRUS	Note:	
Problem language:	el		
<a href="#">Printer Friendly</a>   <a href="#">Admin Info</a>   <a href="#">History</a>   <a href="#">Collections</a>			
<a href="#">cs</a>   <a href="#">el</a>   <a href="#">en</a>   <a href="#">fi</a>   <a href="#">fr</a>   <a href="#">ga</a>   <a href="#">mt</a>   <a href="#">pl</a>			
<b>en</b> Council			
Definition:	financial regulatory agency of the Republic of Cyprus		
Reference:	COUNCIL-EN		
Note:	Not to be confused with the equivalent body in Greece, the Hellenic Capital Market Commission [ATE:914244]		
<a href="#">Admin Info</a>   <a href="#">History</a>			
Term Group: 1	Council	Reliability:3	
Term:	<b>Cyprus Securities and Exchange Commission</b>		
Reference:	CySEC website, <a href="http://www.cysec.gov.cy/defa...">www.cysec.gov.cy/defa...</a> [14.03.2012]		
<a href="#">Admin Info</a>   <a href="#">History</a>			
Term Group: 1	Council	Reliability:3	
Term:	<b>CySEC</b>		
Reference:	CySEC website, <a href="http://www.cysec.gov.cy/defa...">www.cysec.gov.cy/defa...</a> [14.03.2012]		
<a href="#">Admin Info</a>   <a href="#">History</a>			

Fig. 3.4 Dettaglio di una scheda terminologica con Problem Language<sup>128</sup>

Un'entrata viene marcata come "**Primary Entry**" quando vanta di un livello particolarmente alto di qualità o va preferita tra altri possibili duplicati. In generale, sono

<sup>127</sup> In IATE 2, la prossima versione di IATE, di cui si parlerà nel par. 3.6.2., "Problem Language" verrà rinominata in "Anchor Language".

<sup>128</sup> Fonte immagine: IATE Handbook.

Primary Entries le entrate più affidabili e che compaiono al primo posto della lista di risultati, marcate con l'icona di una stellina. Sono solitamente i coordinatori terminologici al livello centrale che marcano le entrate come Primary Entries aggiungendo la stella al Language-Independent Level durante i lavori di consolidazione. La stella va conferita solo quando i criteri di qualità minima sono rispettati e non soltanto quando una scheda è affidabile. Solitamente, al livello centrale, i terminologi passano in rassegna tutte le entrate che coprono un determinato concetto e ne esaminano la qualità. Si tratta di un lavoro complesso poiché vanno ricercate tutte le varianti di spelling di tutti i termini che possono essere utilizzati per rappresentare quel concetto, come ad esempio i sinonimi. Il terminologo considera anche le informazioni contenute nei domini e negli altri campi, controlla gli equivalenti nelle lingue di competenza e infine seleziona un set di termini candidati usando alcuni criteri di importanza decrescente. Anzitutto si valuta la coerenza generale, la qualità, il numero di lingue, presenza di una Problem Language ecc. Se una entrata rispetta i criteri, sarà definita come Primary, in caso contrario si crea una nuova entrata. Non è possibile avere diverse entrate primarie per lo stesso concetto, ma è possibile averne per lo stesso termine se denota concetti diversi.

Il campo **Management Field** è usato solo dai membri dei team di coordinamento terminologico per raccogliere informazioni sulla gestione dell'entrata, quali il progetto di cui faceva parte e altre informazioni rilevanti per il coordinamento.

Il campo **Historical Field** viene usato per marcare un'entrata come storica, quindi per indicare che l'oggetto non è più in uso, come ad esempio un organismo che non esiste

più. Le entrate vengono marcate con una grossa “H” e non dovrebbero essere confuse con “obsolete”, che è una valutazione attribuita ad un singolo termine che non viene più utilizzato.

Un importante sviluppo recente è il campo “**Cross-References**” per cui un’entrata può riferirsi ad un’altra entrata di IATE. Selezionando “Cross Reference” si accede al menu di dettaglio del Language Independent Level; qui va inserito il codice identificativo di una o delle altre entrate di IATE con cui si vuole creare la relazione e si può scegliere se l’altra entrata è “broader”, “narrower”, “related” o “antonym”.

Di recente, lo IATE Management Group (IMG) ha voluto ampliare le *cross-references* in IATE introducendo, per la prima volta nella storia di IATE, le relazioni ontologiche.

Per ora il progetto è ancora in fase embrionale ma oltre alle relazioni esistenti si vorrebbero introdurre, ad esempio, “*partOf*” e “*hasPart*”, tipiche espressioni di alcuni linguaggi del Web Semantico.

Il campo “**Collections**” è usato per creare subset di dati in IATE che riguardano specifici progetti, sottodomini o lingue e per semplificarne la gestione. Il campo “**Graphics**”, infine, viene usato per inserire un’immagine, fotografia e disegno, quindi in generale un contenuto grafico che aiuti alla comprensione dell’entrata. Se ne possono aggiungere per un massimo di tre e ogni immagine deve essere libera da copyright e diritto d’autore.

### 3.3.2 Language Level

Come opzione predefinita, la lingua indicata nel profilo del terminologo che crea una nuova entrata è anche la prima lingua dell’entrata stessa. In alternativa, il database

estrae e seleziona automaticamente la Problem Language. Nella figura seguente l'interfaccia della creazione dell'entrata al Language Level.

The screenshot displays a web-based interface for creating a new entry in a Language Level. At the top, a yellow header bar contains the text 'Entry Number 3551837'. Below this, a blue bar shows 'Domains: 5621 - Cultivation of agricultural land' and a 'Note:' field. A navigation bar includes 'Printer Friendly', 'Admin Info', and 'History' buttons. The main content area is titled 'NEW' and features a 'Language' dropdown menu set to 'en-English'. The 'Definition' field contains '[No Title]'. Below this are 'Reference' and 'Note' fields, each with a small blue arrow icon. A 'Confidentiality' dropdown menu is set to 'Public'. At the bottom of the form, there are buttons for 'Admin Info', 'Related Material', and 'Graphics', followed by 'Save' and 'Exit' buttons. A '<< Previous entry' button is located below the form. The interface is set against a light yellow background.

Fig. 3.5 Interfaccia entrata Language Level<sup>129</sup>

Probabilmente tra le attività più complesse della redazione di una scheda terminologica è la definizione. Il campo “**Definition**” necessita di contenere informazioni chiare e concise, in modo che l’utente possa prontamente capire ciò a cui l’entrata si riferisce esattamente. Ricordando che dal Language Level deriva il Term Level, la definizione deve essere valida per tutti i termini inclusi sotto di essa e in generale va applicato il principio secondo cui in IATE ogni concetto deve corrispondere ad una singola entrata. La definizione può essere anche una traduzione della definizione fornita da *Problem Language*, soprattutto se validata da esperti.

<sup>129</sup> Fonte immagine: IATE Handbook.

Nel documento IATE HandBook si distingue:

- Definizione per comprensione: basata su un gruppo di caratteristiche che costituiscono un concetto;
- Definizione per estensione: basata sull'enumerazione esaustiva degli oggetti a cui il concetto si riferisce o degli specifici concetti al livello successivo di astrazione.

Inoltre, ci sono casi in cui le definizioni non coprono tutti gli aspetti di un concetto, come ad esempio nei concetti legali, in cui è spesso difficoltoso esprimere dei significati specifici, spesso legati a particolari ordinamenti legali.

In questi casi il terminologo può utilizzare il campo "Note", spiegando eventuali gradi di significato e aggiungendo ulteriori informazioni laddove rilevante.

Le definizioni formali presenti in legislazioni che si applicano solo a quel determinato strumento dovrebbero essere usate solo se sono sufficientemente ampie e se coprono un contesto più generale e non uno specifico atto.

Il campo "**Definition Reference**" esprime le fonti delle definizioni. La fonte deve essere, generalmente, credibile. Wikipedia è ritenuta, a questo livello, una fonte accettabile, a condizione che venga effettuata una ricerca più ampia preventivamente. Nel caso in cui la definizione venga scritta dal terminologo, si deve inserire come *reference* il nome del suo dipartimento/unità linguistica e istituzione. Se invece usa altre fonti, deve citarle seguendo lo schema "SiglaIstituzione-SiglaLingua, based on: [riferimento alla fonte originale]".

### 3.3.3 Term Level

Il livello del termine può comprendere uno o più termini per lingua e tutti devono riferirsi allo stesso concetto ma devono essere inseriti separatamente. Ciò viene denominato, in IATE Handbook, il “principio dell’autonomia del termine”. Non è corretto, ad esempio, creare un’entrata che come termine abbia “InterActive Terminology for Europe (IATE)”, ovvero sia il nome completo del database che il suo acronimo che, invece, andrebbero separati in due schede distinte.

In IATE ogni termine viene indicato con un codice identificativo univoco, denominato “**Term ID**”, un numero seriale assegnato automaticamente ad ogni livello del termine in una particolare lingua. L’ID non può essere cambiato e non può essere riutilizzato dopo l’eventuale cancellazione del termine. Viene impiegato per raccogliere i termini in gruppi e per agevolare eventuali operazioni di *merging*<sup>130</sup>.

I “**Term Group**” sono gruppi di termini morfologicamente correlati tra di loro e dunque devono avere lo stesso numero in questo campo<sup>131</sup>. Un esempio può essere quello usato poco sopra: a “IATE” ed “InterActive Terminology for Europe” sarà assegnato lo stesso numero e saranno quindi indicati come parte dello stesso gruppo.

IATE prevede inoltre una delle tipologie specifiche di termini, raggruppate sotto il campo “**Term Type**”. La lista comprende:

---

<sup>130</sup> Il *merging* è la fusione/unione di due o più entrate terminologiche. Quest’operazione viene effettuata per ridurre il numero di duplicati.

<sup>131</sup> Il numero dei *Term Group* non è il *Term ID*.

- **Term:** una parola o una polirematica che designano un concetto definito in una determinata lingua;
- **Abbrev:** abbreviazioni, che includono acronimi, contrazioni, troncamenti ecc.;
- **Phrase:** le unità fraseologiche non sono strettamente dei termini ma, poiché compaiono ripetutamente in testi da tradurre e necessitano di una traduzione standardizzata, sono anch'esse contenute in IATE. Le frasi che illustrano l'uso di un termine devono essere inserite nel campo "Contesto".
- **Formula:** formula chimiche, matematiche, ISO, espressioni scientifiche;
- **Short Form:** ad esempio, il nome comune di un accordo o il nome non ufficiale, ma abbreviato, di uno Stato, come ad esempio "Czechia" per "Czech Republic".

La fraseologia inoltre, è spesso difficilmente inquadrabile nei campi standard di IATE e forzarla in una cornice complessa come quello del database interistituzionale ne mina potenzialmente la già precaria integrità. D'altra parte, va ricordato che IATE è visto, nelle istituzioni, principalmente come strumento di supporto alla traduzione e quindi la presenza di elementi non puramente terminologici viene tollerata. Inoltre le varie istituzioni partner, nonostante la presenza di documenti, task force e buone pratiche condivise, sono piuttosto indipendenti per la gestione della terminologia e quindi il fenomeno è difficilmente controllabile.

Ad ogni modo, ai termini va attribuita una valutazione che in IATE corrisponde al campo "**Evaluation**". I metadati utilizzati, come illustrato in IATE Handbook, sono:

- **Preferred:** il miglior termine da utilizzare per la traduzione. Può essere preferito perché è intrinsecamente migliore degli altri termini candidati o perché è stato preferito agli altri per assicurare maggior coerenza terminologica;

- **Admitted:** si usa per indicare un termine corretto per il quale esistono sinonimi migliori. Usare questo tag in maniera esplicita è piuttosto raro e viene fatto solo in casi eccezionali;
- **Deprecated:** un termine che è usato ampiamente e quindi potenzialmente frequente nei documenti ma che non dovrebbe essere usato né negli originali né nelle traduzioni perché non è corretto o adatto ai testi EU;
- **Obsolete:** è un termine che veniva usato precedentemente per denotare un concetto ma non è più in uso.

Nel campo “*Term*” va aggiunto un singolo termine. Un termine può avere differenti forme ma deve riferirsi ad un solo concetto. In IATE HandBook sono stabilite alcune regole lessicografiche:

- usare la forma canonica per le entrate di sostantivi e aggettivi con eccezioni per quei termini abitualmente usati al plurale;
- non usare articoli o maiuscole;
- usare le forme flesse corrette in accordo con le regole della lingua impiegata.

Inoltre, si raccomanda di non creare entrate per i titoli di legislazione secondaria che sono già disponibili in Eur-Lex a meno che questi non abbiano ricevuto una variazione di denominazione, come ad esempio un nome breve che differisce significativamente dal titolo ufficiale. IATE inoltre offre un sistema di rilevazione dei duplicati che viene attivato se un termine corrisponde ad un altro termine che esiste già nel database, anche se riferito ad un concetto diverso. Il terminologo riceve una notifica e sceglie se continuare con la creazione della scheda potenzialmente duplicata o meno.



Il campo “**Reliability**” indica l’affidabilità di un termine, la sua corrispondenza con il concetto definito nella *Problem Language* e l’affidabilità della fonte. Esso non va confuso con la validazione. Si distinguono, in IATE HandBook, quattro livelli di affidabilità con rispettivi codici:

- **Codice 0, termine in attesa di cancellazione:** assegnato ad un termine da cancellare, ad esempio durante un progetto di consolidamento;
- **Codice 1, affidabilità non verificata:** viene assegnato ai termini inseriti da utenti non nativi. Il codice cambia quando un utente con diritti di modifica conferma l’entrata (solitamente è un’attività affidata ad un terminologo);
- **Codice 2, affidabilità minima:** assegnata automaticamente alle entrate create dai nativi di una lingua. I terminologi aumentano solitamente questo valore a “3” o a “4” a meno che non notino mancanze, come ad esempio dubbi sulle fonti, fonti inesistenti, ecc.;
- **Codice 3, affidabile:** assegnato manualmente, questo valore indica che un termine deve rispettare almeno una tra le seguenti proprietà:
  - deve avere una fonte affidabile;
  - deve essere stato validato dai terminologi di un determinato organismo;
  - deve essere la designazione comune di un concetto in uno specifico dominio.

Il codice, inoltre, si riferisce generalmente a termini per cui è stata fatta approfondita ricerca terminologica;

- **Codice 4: molto affidabile:** questo valore viene attribuito ai termini che sono ampiamente accettati dagli esperti e piuttosto diffusi e confermati da un fonte autorevole.

Il campo “**Term reference**” stabilisce l’affidabilità di un termine e quindi è ampiamente preferibile che contenga fonti credibili e ufficiali.

“**Context**”, infine, è il campo che indica una piccola citazione di una fonte attendibile che illustra l’uso del termine all’interno del contesto testuale. In IATE esistono diversi livelli di accordi interistituzionali in merito, ma in generale si stabilisce che il contesto può contenere forme flesse ed è particolarmente importante quando manca una definizione. “**Context Reference**” indica la fonte del contesto. Come per gli altri campi simili, deve essere credibile e affidabile.

### 3.4 Fonti

Un aspetto fondamentale delle entrate terminologiche è l’uso delle fonti. Per IATE, ogni termine, definizione e contesto devono avere un riferimento ad una fonte. Se essa è un documento, deve contenere il termine, la sua definizione e il contesto. I riferimenti alle fonti devono essere comprensibili e precisi e l’informazione in esse contenuta deve essere credibile<sup>132</sup>.

Per ogni tipo di riferimento, le informazioni chiave che essi devono contenere sono:

---

<sup>132</sup> Wikipedia, ad esempio, può essere usata solo per la *definition reference*.

- Titolo: Il titolo principale dell'oggetto che si sta citando. Risulterebbe difficile, ovviamente, lavorare su una fonte di cui non si identifica un elemento che ne descriva l'entità e la provenienza. In IATE HandBook si invita quindi ad usare "il senso comune" e indicare sempre l'informazione che identifichi la fonte;
- Autore: laddove ammissibile, l'autore deve essere esplicitamente espresso. In caso in cui l'autore non sia specificato, come ad esempio in alcuni siti web, va usata la dicitura "*corporate author*".
- Data di pubblicazione e di consultazione: l'indicazione dell'anno è solitamente sufficiente, ma si richiede maggior precisione per fonti quali riviste e giornali.

#### 3.4.1. Tipi di fonti

**Documenti unionali.** Le versioni linguistiche della legislazione dell'UE sono ritenute come originali autentici, quindi fungono da fonte primaria per la terminologia specialistica istituzionale. In generale, se il concetto è stato creato dalle istituzioni dell'UE, va fatto il riferimento al testo legale. Tuttavia, i documenti UE non sono sempre una fonte attendibile di terminologia di concetti che esistono anche al di fuori del contesto europeo. In questi casi la fonte dovrebbe essere, idealmente, ugualmente attendibile, come ad esempio i termini prodotti da organismi interazionali come Food and Agriculture Organization of the United Nations (FAO), Federal Aviation Administration (FAA), International Maritime

Organization (IMO) ecc. Per i documenti interni, è necessario esplicitare l'istituzione, il numero di documento e il suo titolo.

**Documenti di Eur-Lex.** I documenti di Eur-Lex adottano pressappoco lo stesso protocollo dei documenti unionali. La differenza è la presenza, dopo il titolo del documento e la data di consultazione, del numero Celex. Il numero Celex identifica in maniera univoca ciascun documento in EUR-Lex, indipendentemente dalla versione linguistica. È composto nel modo seguente: "SaaaaT(T)nnn(n)", dove "S" indica il settore, "aaaa" l'anno, "T(T)" il tipo di documento e "nnn(n)" il numero del documento<sup>133</sup>. I documenti di Eur-Lex posseggono un identificativo unico (URI) con il codice Celex come metadato. Il codice Celex segue solitamente il titolo di un documento e la data di consultazione non è richiesta in quanto si tratta di un link permanente. Un esempio può essere "Decision No 633/2009/EC granting a Community guarantee to the European Investment Bank against losses under loans and loan guarantees for projects outside the Community, CELEX:32009D0633/EN".

**Trattati, convenzioni, accordi:** Per le loro denominazioni si usa solitamente la forma concisa seguita dal numero dell'articolo. Un esempio può essere

---

<sup>133</sup> Il numero CELEX è l'identificatore univoco di ciascun documento in EUR-Lex, indipendentemente dalla versione linguistica. Ulteriori informazioni sul numero Celex, consultare <http://eur-lex.europa.eu/content/help/faq/celex-number.html>.

“Treaty on European Union, Article 3”. È raccomandata l’aggiunta di un URL laddove il testo sia consultabile.

**Pagine e siti web.** Se si consulta una risorsa web, è necessario indicare il titolo, l’autore, l’URL e la data di consultazione, dando anche indicazioni su eventuali sottopagine/sottosezioni consultate.

**Fonti umane.** Si dividono in esterne ed interne. Se, ad esempio, un terminologo propone un termine o una definizione, sulla base di fonti autorevoli, si può usare la formula “based on” seguita da un identificativo dell’istituzione e dipartimento linguistico di cui quella fonte, in questo caso il terminologo stesso, fa parte. Per le fonti esterne si adotta la stessa formula, aggiungendo anche la funzione ricoperta dall’esperto ed eventuali altre informazioni rilevanti.

### 3.5 Validazione e Consolidamento

Per “validazione” si intende quell’operazione effettuata su un’entrata per verificare che essa rispetti le regole di IATE. Viene eseguita dai terminologi di ogni lingua di ciascuna istituzione e sono responsabili per la validazione delle entrate create dal loro team. IATE offre diversi filtri per agevolare il processo, permettendo di scegliere le entrate anche in base ai campi. Esiste anche una funzione “Multivaldation” con cui poter verificare le entrate multiple. L’aspetto più importante della validazione è che solo le entrate

validate sono visibili nella versione pubblica di IATE, dunque è prioritario che i terminologi delle istituzioni europee controllino regolarmente i dati non verificati.

Il consolidamento è invece il processo che viene attivato laddove vengano identificati problemi con entrate, ad esempio le voci che corrispondono a più di un concetto, una definizione troppo superficiale, i duplicati ecc. In questi casi vengono creati i cosiddetti “progetti di consolidamento”<sup>134</sup> che vanno quindi visti come degli interventi mirati al miglioramento di molteplici aspetti di IATE. Solitamente, se le entrate, o le lingue, sono una o due, l’intervento viene effettuato direttamente dal terminologo dell’istituzione responsabile, se invece si coinvolgono più di due lingue e due entrate, è compito del servizio di coordinamento terminologico delle varie istituzioni di attivare il progetto. I progetti di consolidamento si concentrano principalmente sull’armonizzazione e sulla coerenza delle risorse terminologiche e sono spesso finalizzati o alla scelta di una *primary entry* o ad un’operazione di *merging*. Quest’ultima è un’operazione effettuata sulle entrate duplicate.

Al livello *Language-Independent Level* e al *Language Level*, vi sono tre possibilità:

- Si mantiene solo l’informazione dell’entrata primaria;
- Si sovrascrive l’informazione dell’entrata secondaria su quella della primaria;
- Si allega l’informazione dell’entrata secondaria a quella primaria, senza ulteriori cancellazioni.

---

<sup>134</sup> Terminologia di IATE HandBook, *cit.* p 33.

A livello *term level* si può o mantenere o cancellare l'entrata primaria, mentre per la secondaria si può:

- Creare un nuovo *term level* aggiunto alla *primary entry*;
- “concatenare” il *term level* ad uno preesistente della primaria;
- Non copiare il *term level*.

Spesso i progetti di consolidamento riguardano numerose entrate di competenza di istituzioni diverse. In quel caso, vengono creati progetti interistituzionali coinvolgendo i team di coordinamento terminologico di ogni istituzione interessata ed allestendo gruppi di lavoro ad hoc.

## 3.6 Sviluppi futuri

IATE è costantemente aggiornato, non solo dal punto di vista contenutistico, ma anche da quello strutturale e tecnologico. Sono molti i progetti di sviluppo di cui è oggetto e in questa sezione si illustreranno quelli più rilevanti.

### 3.6.1 IATE e API

Nell'attuale scenario tecnologico del web, un ruolo importante è giocato dalle API (Application programming interface) ovvero un'interfaccia, o una serie di strumenti utili per la realizzazione di software e applicazioni.

Un'API permette di sviluppare, facilmente, un programma basato su un altro programma, le cui componenti strutturali di base sono già fornite. Applicando ciò a IATE,

le API permetterebbero di sviluppare software che si interfaccino con il database e che ne implementino, in specifiche sezioni, alcune funzioni.

In questo momento, la pubblicazione delle API è al vaglio dello IATE Management Group e un eventuale sviluppo in questo senso non è stato ancora deciso. Tuttavia, esistono già delle piccole iniziative realizzate grazie alle competenze e alla lungimiranza di TermCoord.

L'unità ha difatti sviluppato un'applicazione per smartphone che permette la consultazione di IATE direttamente dal proprio dispositivo mobile. Creata da un valido tirocinante dell'Unità, è attualmente disponibile per la piattaforma Android<sup>135</sup>.

Se sviluppate ufficialmente a livello centrale e con la dovuta cautela in termini di proprietà intellettuale e pubblicabilità, le API rappresenterebbero per IATE una notevole accelerazione di sviluppo, sia "dal basso", ovvero grazie alla partecipazione di sviluppatori ed entusiasti di terminologia da tutto il mondo per la creazione di applicativi e strumenti linguistici, sia "dall'alto", tramite i team degli sviluppatori delle istituzioni<sup>136</sup>.

A livello istituzionale, una prima, importante, innovazione potrebbe verificarsi nel contesto dei CAT tool impiegati dai traduttori dell'Unione.

Il Parlamento europeo, negli ultimi anni, sta sviluppando un CAT tool proprietario denominato CaT4Trad. Lo strumento, basato sul Web, avrebbe la finalità di rimpiazzare

---

<sup>135</sup> Il download è disponibile presso

<https://play.google.com/store/apps/details?id=termcoord.termcoord.termcoordsearch>.

<sup>136</sup> Il team di sviluppo di IATE è parte del Translation Centre (CdT) ma la sua attività è governata da IATE Management Group (IMG).



i software esterni (quali SDL Studio) e diventare l'unica piattaforma di riferimento per la traduzione.

La necessità di sviluppare un tool di traduzione proprietario nasce dall'ampia varietà delle tipologie dei documenti e dagli alti standard di qualità che DGTRAD deve garantire, che portano spesso a incompatibilità, errori, bug ed altri problemi causati dall'uso di strumenti commerciali.

Lo sviluppo di CaT4Trad è tuttora in corso e solo una parte dei documenti prodotti dal Parlamento viene attualmente gestita in questa piattaforma<sup>137</sup>.

Ad ogni modo, un uso efficace delle API di IATE permetterebbe di creare una sezione in CaT4Trad per la consultazione del database direttamente nel software.

Le API potrebbero altresì permettere lo sviluppo di un'interfaccia che raccolga input per modifiche, aggiornamenti e creazione di nuove entrate da convalidare successivamente in IATE e tramite il lavoro di esperti e terminologi.

La letteratura di riferimento delle API vuole che siano concepite sempre meno come un *web service* basato su SOAP (Simple Object Access Protocol) e *Service-oriented Architecture* (SOA) verso architetture basate su risorse (ROA = Resource-oriented architecture), che sono parte del trend del Web Semantico, che promuove tecnologie basate sull'ingegnerizzazione delle ontologie.

L'implementazione di API rappresenterebbe quindi, per IATE, un piccolo passo verso le tecnologie del Semantic Web.

---

<sup>137</sup> Attualmente si possono tradurre in CAT4Trad solo alcuni tipi di emendamenti.

### 3.6.2 IATE 2

Fin dalla sua messa in produzione nel 2004 IATE è stato protagonista di numerosi sviluppi e cambiamenti attraverso un numero incalcolabile di progetti di consolidamento e raffinazione del contenuto terminologico. IATE non ha mai subito, tuttavia, un restyling completo, una ristrutturazione integrale che tenga conto delle necessità attuali dei terminologi, di una gestione avanzata dei dati, dell'aumento del numero delle entrate e della compatibilità con gli standard e le tecnologie più moderne.

A rispondere a quest'esigenza è IATE 2, un progetto coordinato dallo IATE Management Group mirato alla pubblicazione di una nuova versione di IATE più adeguato all'attuale contesto tecnologico. Gli sviluppi di IATE 2 possono essere distinti in generali e specifici, ovvero relativi ai livelli delle entrate.

Quelli generali riguardano il miglioramento delle fonti, che saranno trattate differenzialmente e catalogate in sezioni separate, in modo da permetterne il riutilizzo e la gestione avanzata. I numeri Celex saranno anch'essi immagazzinati separatamente e in corrispondenza delle fonti e i link saranno automaticamente verificati. Simile operazione sarà effettuata per le URL, allegate alle fonti e con un sistema di verifica di *broken links*<sup>138</sup>. Si miglioreranno inoltre i campi di gestione e i *lookup terms*.

A livello delle entrate si vorrebbe introdurre un algoritmo che calcoli automaticamente la qualità dell'entrata basata sull'analisi della presenza di nuovi elementi di marcatura,

---

<sup>138</sup> Per "broken link" si intende un link ad una pagina web che non è più accessibile a causa di vari motivi, tra cui cambio dell'URL, rimozione della pagina, cancellazione o indisponibilità temporanea del sito, ecc.

quali ad esempio l'opzione per la cancellazione. Il campo *cross reference* sarà ampliato con relazioni ontologiche di base mentre il campo *Problem language* sarà rinominato in *Anchor language*. Una priorità dello IMG è di sviluppare un *Quick new entry creation system* che permetta di creare nuove entrate più velocemente in modo da contribuire all'arricchimento di IATE 2 in considerazione dell'aumentata velocità di creazione di nuovi termini e conseguentemente del loro inserimento.

I componenti delle varie task-force per lo sviluppo di IATE propongono miglioramenti quali l'identificazione dei duplicati, la formattazione dei testi ed un miglior collegamento con la classificazione di EuroVoc ed Eur-Lex. Sono stati proposti anche accorgimenti sull'interfaccia web per la quale sono state già create delle bozze.

Ad ogni modo, un prototipo di IATE che contenga solo le funzioni di base è previsto per la seconda metà del 2017, mentre le funzioni più avanzate potrebbero essere implementate nel terzo o quarto quadrimestre dello stesso anno. Eventuali sviluppi mirati a *web service* per partner pubblici e/o specifici e l'integrazione con altri strumenti sono previsti per la fine del 2017.

# CAPITOLO IV – PROGETTI DI RICERCA PER L'INTEROPERABILITÀ SEMANTICA DI IATE

## 4.1 Il web semantico e i Linked Data

Il World Wide Web ha totalmente cambiato il modo in cui condividiamo la conoscenza, abbattendo le barriere geografiche e fornendo la possibilità di pubblicare documenti e di accedervi in ambito globale. I collegamenti ipertestuali permettono agli utenti di traversare i documenti grazie all'impiego di Web browser, mentre i motori di ricerca li indicizzano ed analizzano la struttura dei link per ricavare un *ranking* di rilevanza rispetto ad una determinata *query* di ricerca. Tali funzionalità hanno rappresentato la chiave di volta per una crescita esponenziale del Web.

Tuttavia, gli stessi principi che hanno permesso al Web dei documenti di prosperare non sono adatti per rappresentare con accuratezza le relazioni tra gli elementi del mondo. In altre parole, nel web convenzionale, le modalità di rappresentazione della relazione tra documenti non risultano essere sufficienti se applicate a quella tra entità. Si esige quindi un nuovo tipo di approccio, in cui sussistano dati strutturati e non strutturati che sostengano forme di interconnessione più proficue e semanticamente più espressive. Tale evoluzione va sotto il nome di Web Semantico, ovvero l'evoluzione del World Wide Web e un nuovo ambiente dove, associando dati ed informazioni ad altri dati (che

prendono il nome di metadati) se ne migliora la specificazione semantica, permettendone l'interrogazione da parte di macchine e umani (Berners-Lee, 2009).

Nel contesto del web semantico è stata introdotta una nuova modalità di pubblicazione di dati strutturati, che permette che questi siano collegati fra loro e quindi utilizzabili attraverso interrogazioni semantiche. Questa nuova modalità prende il nome di Linked Data<sup>139</sup> con cui, tramite l'utilizzo di tecnologie e standard web e stringhe di identificazione univoche come gli URI (Uniform Resource Identifier), si rappresentano informazioni che possano essere lette e comprese da computer, rendendo così possibile collegare e utilizzare dati provenienti da diverse sorgenti (Heat e Bizer, 2011).

A formalizzare queste condizioni fu Berners-Lee (2006) con i suoi quattro principi dei Linked Data:

1. *«Use URIs as names for things;*
2. *Use HTTP URIs so that people can look up those names;*
3. *When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL);*
4. *Include links to other URIs. so that they can discover more things»<sup>140</sup>.*

Con i quattro principi, Berners-Lee voleva promuovere l'aumento dei dati linkati, che fino ad allora non lo erano, e in «quantità sorprendenti»<sup>141</sup>.

---

<sup>139</sup> <http://linkeddata.org/>.

<sup>140</sup> Berners-Lee, T., (2006), *“Design Issues”*, W3C in <https://www.w3.org/DesignIssues/LinkedData.html>

<sup>141</sup> *Ibidem*. Berners-Lee non ha mai imposto, *de facto*, i quattro principi, definendoli più che come regole, «*expectations of behavior*». È utile ricordare, inoltre, la celebre presentazione di Berners-Lee in occasione del TED Talk (2009), in cui incoraggiava il pubblico alla pubblicazione di dati non adulterati: «OK, we have to ask for raw data now. And I'm going to ask you to practice that, OK? Can you say “raw”?» [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web/transcript?language=en#t-640180](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web/transcript?language=en#t-640180).

Signore (2002), definisce RDF come «*lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e che consente l'interoperabilità tra applicazioni che si scambiano sul Web informazioni machine-understandable*»<sup>142</sup>. SPARQL è definito, nei documenti del W3C, come «*a query language and protocol for RDF*»<sup>143</sup> e quindi «*the query language for the Semantic Web*»<sup>144</sup> (Herman, 2008).

Una categoria specifica di Linked Data è quella dei Linked Open Data<sup>145</sup>, ovvero dati pubblicati secondo i principi dei contenuti aperti, meglio conosciuti semplicemente come *open*, che vengono così definiti quando vengono pubblicati sotto libera licenza e dietro autorizzazione per il riuso, la copia e la modifica da parte di terzi, a condizione che questi ne citino la fonte (Wiley, 1998). Il progetto che sostiene il principio di Linked Open Data, patrocinato dal W3C<sup>146</sup>, ha appunto l'obiettivo di estendere il Web tramite la pubblicazione di *dataset open* e impostando collegamenti tra di essi. Una celebre rappresentazione dei Linked Open Data è il Linked Open Data cloud<sup>147</sup>, un diagramma che mostra i collegamenti tra i diversi *dataset* che lo compongono.

In questo contesto si inserisce un ulteriore movimento concentrato sulla condivisione e il riutilizzo di risorse linguistiche in accordo con i principi dei Linked Data, ovvero quello

---

<sup>142</sup> Signore O., (2002), «*RDF per la rappresentazione della conoscenza*», p. 3. Per la definizione, le specifiche e la sintassi complete di RDF si rimanda a Lassila O., e Swick R. R. (1997), <https://www.w3.org/TR/WD-rdf-syntax-971002/>.

<sup>143</sup> Grant Clark K., et al., (2008), «*SPARQL Protocol for RDF*».

<sup>144</sup> Per specifiche tecniche e sintassi di SPARQL si rimanda a Prud'hommeaux E., e Seaborn A., (2008) <https://www.w3.org/TR/rdf-sparql-query/>.

<sup>145</sup> Il cui portale di riferimento è reperibile presso <https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.

<sup>146</sup> World Wide Web Consortium.

<sup>147</sup> Linking Open Data cloud diagram (2017), di Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch e Richard Cyganiak. URL: <http://lod-cloud.net/>.

che fa capo ai Linguistic Linked Open Data (LLOD)<sup>148</sup>. Ideato dall'Open Linguistics Working Group (OWLIG) della Open Knowledge Foundation, il movimento ha l'obiettivo di sviluppare un *subcloud* popolato di risorse linguistiche. Chiarcos (et al., 2001), definiscono il *cloud* LLOD come un sistema dove

*«linguistic resources (lexical semantic resources, corpora, metadata repositories) are not only provided in an interoperable way (using RDF), but also freely accessible (under an open license) and linked with each other (so that applications can combine information from different knowledge sources)»<sup>149</sup>.*

I LLOD sono impiegati ampiamente per collegare risorse eterogenee come ad esempio WordNet e Wikipedia, per tendere alla creazione di forme di standardizzazione di informazioni di risorse linguistiche e soprattutto per rispondere alla sfida dell'interoperabilità semantica.

## 4.2 L'interoperabilità semantica

*«L'obiettivo dell'interoperabilità è di permettere la presentazione e il trattamento delle informazioni in modo uniforme nei diversi sistemi gestionali, indipendentemente dalla tecnologia, dall'applicazione o dalla piattaforma utilizzate. La piena interoperabilità comprende la capacità di interoperare in termini di contenuto (semantico), formato (sintassi) e trasmissione. L'interoperabilità*

---

<sup>148</sup> Il sito di riferimento è disponibile in <http://linguistic-lod.org/>.

<sup>149</sup> Chiarcos C., et al., (2001) "Towards a Linguistic Linked Open Data cloud: The Open Linguistics Working Group", in TAL (Traitement automatique des langues), vol. 52, 3, pp. 245-275.

*semantica vuol dire che il significato preciso dell'informazione scambiata è mantenuto e compreso correttamente e senza ambiguità, a prescindere dal modo in cui viene rappresentato fisicamente o trasmesso»<sup>150</sup>.*

L'interoperabilità semantica è quindi un requisito per la rappresentazione della conoscenza ed è ottenuta attraverso l'impiego di dati su altri dati (metadati) e linkando ogni elemento tramite un vocabolario controllato.

La capacità delle macchine di inferire e interpretare i dati è l'associazione del vocabolario controllato ad un'ontologia, ovvero una rappresentazione formale di un determinato dominio espressa tramite logiche descrittive<sup>151</sup>.

L'importanza dell'interoperabilità semantica è stata più volte misurata anche in termini di stime di costi (solitamente in termini di perdita di efficienza) da alcuni studi<sup>152</sup> che dimostrano perdite di ingenti somme di denaro a causa della carenza di interoperabilità semantica. Tra i campi che più beneficiano dell'interoperabilità semantica figura senza dubbio quello dell'eGovernment, ovvero l'uso di sistemi informatici e altre tecnologie di telecomunicazione volti a migliorare l'efficacia e l'efficienza del servizio offerto nel settore pubblico.

L'Unione Europea, nel giugno 2007, ha lanciato l'iniziativa SEMIC (Semantic Interoperability Centre Europe), finalizzata ad attuare i principi della strategia European

---

<sup>150</sup> COM(2013) 449 final 2013/0213, Proposta di Direttiva del Parlamento europeo e del Consiglio relativa alla fatturazione elettronica negli appalti pubblici.

<sup>151</sup> Non è obiettivo di questo lavoro di tesi soffermarsi sull'inquadramento teorico delle ontologie, per cui si rimanda, per una lettura più approfondita, a Oberle *et al.* (2009), Natalya F. & McGuinness D., (2001) e Gruber T.R. (1993).

<sup>152</sup> Ad esempio quello condotto da Walker J. et al. per Health Affairs (2005).



Interoperability Framework (EIF)<sup>153</sup>, che fornisce linee guida su come allestire servizi digitali pubblici interoperabili. Il programma sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA) (2010-2015)<sup>154</sup> e il suo successore, il programma ISA<sup>2</sup> (2016-2020)<sup>155</sup> sono i principali strumenti attraverso cui sono stati attuati la strategia europea per l'interoperabilità e il quadro europeo di interoperabilità vigenti.

Oggi le attività messe in atto per l'iniziativa SEMIC si sono evolute nella piattaforma collaborativa creata dalla Commissione europea Joinup<sup>156</sup>, che rappresenta lo strumento operativo del programma ISA<sup>2</sup>, che per il quinquennio della nuova programmazione Gennaio 2016-Dicembre 2020<sup>157</sup> ha quattro macro-obiettivi:

*«(a) develop, maintain and promote a holistic approach to interoperability in the Union in order to eliminate fragmentation in the interoperability landscape in the Union; (b) facilitate efficient and effective electronic cross-border or cross-sector interaction between European public administrations on the one hand, and between European public administrations and businesses and citizens on the other, and to contribute to the development of a more effective, simplified and user-friendly e-administration at the national, regional and local levels of public*

---

<sup>153</sup> EIF o Quadro europeo di interoperabilità (QEI), è parte della Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni (COM (2017)134).

<sup>154</sup> Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del settembre 2009, sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA).

<sup>155</sup> Decisione (UE) 2015/2240 del Parlamento europeo e del Consiglio, del 25 novembre 2015, che istituisce un programma sulle soluzioni di interoperabilità e quadri comuni per le pubbliche amministrazioni, le imprese e i cittadini europei (programma ISA<sup>2</sup>) come mezzo per modernizzare il settore pubblico.

<sup>156</sup> <https://joinup.ec.europa.eu/>

<sup>157</sup> Su ISA<sup>2</sup> e relative attività si consulti la Decisione (EU) 2015/2240 succitata e il work programme [http://ec.europa.eu/isa/library/documents/isa2-work-programme-2016-summary\\_en.pdf](http://ec.europa.eu/isa/library/documents/isa2-work-programme-2016-summary_en.pdf).

*administration; (c) identify, create and operate interoperability solutions supporting the implementation of Union policies and activities; (d) facilitate the re-use of interoperability solutions by European public administrations»<sup>158</sup>.*

Per la realizzazione del programma ISA<sup>2</sup> è centrale lo sviluppo di *dataset* in formato Linked Data per l'offerta di servizi di eGovernment, passando da un'interoperabilità tecnica (ad esempio tramite l'implementazione di standard quali TCP/IP) ad un'interoperabilità semantica (con l'impiego di RDF, OWL ecc.).

Per lo scambio dei dati tra i sistemi informativi il programma ISA<sup>2</sup> ha sviluppato sei diversi vocabolari *Core Vocabularies*<sup>159</sup> - *Core Person Registered organisation, Core Location, Core Public Service, Core Criterion and Core Evidence* e *Core Public Organisation*.

I Core Vocabularies sono modelli semplificati, riusabili ed estensibili che rappresentano le caratteristiche fondamentali di un'entità, permettono l'integrazione di dati provenienti da fonti diverse, possono essere impiegati come fondamento di un formato di esportazione di dati e usati come base per disegnare modelli di dati di nuovi sistemi.

---

<sup>158</sup> Decisione 2015/2240 *cit.*

<sup>159</sup> Le pagine informative dei vocabolari sono disponibili presso le seguenti URL: Core Person Registered organisation [https://joinup.ec.europa.eu/asset/core\\_person/description](https://joinup.ec.europa.eu/asset/core_person/description); Core Location [https://joinup.ec.europa.eu/asset/core\\_location/description](https://joinup.ec.europa.eu/asset/core_location/description); [https://joinup.ec.europa.eu/asset/core\\_business/description](https://joinup.ec.europa.eu/asset/core_business/description); Core Public Service [https://joinup.ec.europa.eu/asset/core\\_public\\_service/description](https://joinup.ec.europa.eu/asset/core_public_service/description); Core Criterion and Core Evidence [https://joinup.ec.europa.eu/asset/criterion\\_evidence\\_cv/description](https://joinup.ec.europa.eu/asset/criterion_evidence_cv/description); Core Public Organisation <https://joinup.ec.europa.eu/asset/cpov/description>.

Nell'ambito del programma ISA<sup>2</sup> è stato sviluppato il Data Catalogue Vocabulary - Application Profile for data portals in Europe (DCAT-AP<sup>160</sup>), che presenta un linguaggio comune per descrivere i *dataset* del settore pubblico e che incoraggia le ricerche *cross-dataset*, e l'Asset Description Metadata Schema (ADMS), una specifica per descrivere soluzioni di interoperabilità.

### 4.3 Il linguaggio di marcatura TermBase eXchange - TBX

TBX (TermBase eXchange) è un linguaggio di marcatura fondato su XML, (eXtensible Markup Language), un metalinguaggio per la definizione di linguaggi di marcatura, ovvero un linguaggio che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo. TBX è concepito per lo scambio di dati terminologici, grazie al quale una gran varietà di linguaggi di markup terminologici (TML) possono essere definiti.

Il TBX è inoltre uno standard internazionale pubblicato come ISO 30042:2008 e facilita lo scambio di dati terminologici tra utenti e macchine. Citando la sintesi di ISO, TBX:

*«is modular in order to support the varying types of terminological data, or data-categories, that are included in different terminological databases (termbases). TBX includes two modules: a core structure, and a formalism for identifying a set of data-categories and their constraints, both expressed in XML [...]. To maximize*

---

<sup>160</sup> Una descrizione formale di DCAP-AP è disponibile presso [https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description).

*interoperability of the actual terminological data, TBX also provides a default set of data-categories that are commonly used in terminological databases»<sup>161</sup>.*

Alcuni dei maggiori database terminologici, tra cui IATE, possono essere esportati in TBX in modo da poterne permettere il (ri)utilizzo da parte di altre applicazioni come ad esempio gli strumenti di traduzione assistita dal computer (CAT tool) tramite l'impiego di *termbase*.

Dal sito di IATE<sup>162</sup> è possibile scaricare una versione completa e aggiornata del database terminologico interistituzionale in formato TBX. Il file, di circa 2,1 GB, è disponibile in formato compresso in modo da permetterne un download più agevole. Il sito offre, inoltre, uno strumento di estrazione denominato IATEExtract che ha appunto la funzione di estrarre sezioni del file di IATE in base a filtri quali le lingue o i domini ed esportarle in formato TBX.

#### 4.4 Lo standard di riferimento dell'interoperabilità semantica: gli scopi di ISO/TC 37/ SC 3

ISO/TC 37 è un *Technical Committee* (TC) dell'*International Organization for Standardization* (ISO), un comitato tecnico internazionale che gestisce gli standard e i documenti circa la metodologia e i principi per la terminologia e le risorse linguistiche. Nel 2001, le attività del TC sono state riformate<sup>163</sup> e il titolo fu cambiato in "*Terminology*

---

<sup>161</sup> ISO 30042:2008 "*Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)*", disponibile presso [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=45797](http://www.iso.org/iso/catalogue_detail.htm?csnumber=45797).

<sup>162</sup> Download IATE, Unione Europea (2017), <http://iate.europa.eu/tbxPageDownload.do>.

<sup>163</sup> La struttura operativa di ISO/TC 37 fu riformata già a partire dall'assemblea plenaria del 1989, in cui si ristrutturarono anche le attività, competenze e responsabilità dei *Sub-Committee* (SC). Come conseguenza,

*and other language and content resources*” il cui scopo è la standardizzazione dei principi, del metodo e delle applicazioni della terminologia e delle lingue, delle risorse nei contesti della comunicazione multilingue e della diversità culturale<sup>164</sup>.

Il TC, grazie al contributo di esperti provenienti dall'industria e dall'accademia, opera per definire principi e metodi della terminologia, layout degli standard terminologici, terminografia e lessicografia computerizzata, ingegneria del linguaggio, *content management*, traduzione, interpretazione e relative tecnologie.

Il comitato tecnico è formato da cinque sottocomitati (SC), ognuno con obiettivi, scopi e campi di competenza specifici. Quello più rilevante per le finalità di questo lavoro è SC 3, “Terminology Management Systems and Content Interoperability”.

ISO/TC 37/SC 3 fu fondato nel 1985 per gestire gli standard dell'ISO/TC 37 sulla terminografia computerizzata e sulle relative applicazioni al computer. Oggi, il suo scopo è la standardizzazione dei principi e dei requisiti per l'interoperabilità semantica, la terminologia, i sistemi di gestione dei contenuti e degli strumenti della gestione della conoscenza.

---

si introdussero nuovi metodi di standardizzazione nel campo della terminologia e le ISO pubblicate fino ad allora furono sottoposte a revisione o abbandonate. Cfr. ISO/TC 37 (2004). 50 Years ISO/TC 37 “*Terminology and other language resources*” - a history of 65 years of standardization of terminological principles and methods. (ISO/TC 37 N 499 ISO/TC 37 AG N 117), pp.1-5, consultato presso [http://www.infoterm.info/pdf/activities/Standing\\_document\\_02\\_50\\_years\\_ISO\\_TC\\_37.pdf](http://www.infoterm.info/pdf/activities/Standing_document_02_50_years_ISO_TC_37.pdf) (data ultima consultazione: 22 Dicembre 2016).

<sup>164</sup> Per le specifiche complete si rimanda a [http://www.iso.org/iso/iso\\_technical\\_committee.html%3Fcommid%3D48104](http://www.iso.org/iso/iso_technical_committee.html%3Fcommid%3D48104).

L'obiettivo di ISO/TC 37/SC 3 è di sviluppare standard per l'interoperabilità semantica, incluse specifiche di terminologia tra cui *data modelling*, *mark-up* e sistemi di valutazione della terminologia<sup>165</sup>.

*« Terminology is often embedded in or combined with other kinds of content (mostly specialized texts). In order to make content development less expensive (because of its labour-intensiveness), we need new methods of content creation (and the respective workflow management) [...]. In principle all Content items/units should be prepared and maintained in such a way that they fulfil the requirements of*

- *single-sourcing → uninhibited re-usability*
- *resource-sharing → (net-based distributed) cooperative content development*
- *universal accessibility → incl. access by persons with special needs»<sup>166</sup>.*

La terminologia acquisisce una nuova dimensione, ricoprendo un ruolo fondamentale nella modellazione di un nuovo sistema per gestire i dati terminologici e le risorse linguistiche.

---

<sup>165</sup> Le ISO sotto la responsabilità di ISO/TC 37/SC 3 sono ISO 16642:2003: "Computer Applications in terminology – Terminological markup Framework"; ISO 22274:2013 "Systems to manage terminology, knowledge and content - Concept-related aspects for developing and internationalizing classification systems", ISO 26162:2012 "Systems to manage terminology, knowledge and content -- Design, implementation and maintenance of terminology management systems"; ISO 30042:2008 "Systems to manage terminology, knowledge and content – Termbase eXchange (TBX)". Per un'analisi più dettagliata si rimanda a [http://www.iso.org/iso/standards\\_development/technical\\_committees/list\\_of\\_iso\\_technical\\_committees/iso\\_technical\\_committee.htm?commid=48136](http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=48136).

<sup>166</sup> Galinski (2005), "Semantic Interoperability and Language Resources, the role of ISO/TC 37 standards for Global Semantic Interoperability", p. 11.

## 4.5 Progetti di ricerca per l'interoperabilità semantica di IATE

Lo scenario tecnologico appena illustrato è utile per meglio comprendere alcuni dei più attuali progetti di ricerca mirati all'interoperabilità semantica di IATE. In questa sezione si passeranno in rassegna i progetti che mostrano come un database terminologico come IATE possa essere trasformato in una risorsa del Web Semantico e reso interoperabile con altre risorse.

### 4.5.1 Il progetto LIDER - *Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe*

Nell'ambito del progetto europeo "*Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe*" - LIDER<sup>167</sup>, finanziato con fondi del 7<sup>th</sup> *Framework Programme for Research and Technological Development* (FP7)<sup>168</sup>, si è formato un gruppo di ricerca interdisciplinare (Cimiano *et al.*, 2015), che ha sviluppato un approccio per pubblicare risorse terminologiche usando i principi dei Linked Data<sup>169</sup>. Ha mostrato come rappresentare risorse terminologiche come IATE ed il

---

<sup>167</sup> L'obiettivo del progetto è di creare le basi per la creazione di un Linguistic Linked Data cloud che supporti l'analisi del contenuto di risorse multilingui non strutturate. Il progetto, della durata di 24 mesi, è iniziato il 1 Ottobre 2013 e terminato il 31 Ottobre 2015. Per un'introduzione si rimanda a <http://www.lider-project.eu/>, mentre per una panoramica più completa a <http://www.lider-project.eu/sites/default/files/D4.11-v3.0.pdf>.

<sup>168</sup> Il FP7 è stato un programma, attivo dal 2007 al 2013, di finanziamenti principalmente in forma di borse di ricerca, con l'obiettivo di finanziare sviluppi tecnologici che abbiano un "valore aggiunto europeo" e che stimolino il superamento delle barriere nazionali. V. "*FP7 in Brief, How to get involved in the EU 7th Framework Programme for Research*".

<sup>169</sup> Cimiano P., *et al.*, "*Linked Terminology: Applying Linked Data Principles to Terminological Resources*".

glossario European Migration Network glossary (EMN)<sup>170</sup> in RDF e come collegarle tra loro e con altre risorse.

Gli studiosi hanno applicato i quattro principi fondamentali dei Linked Data ai *dataset* terminologici mentre il lessico è basato sul modello *lemon*<sup>171</sup>.

Per convertire IATE ed EMN in RDF, e quindi per fare in modo che le due risorse potessero supportare il collegamento con altri dataset, sono state effettuate due operazioni di conversione. IATE, che è già disponibile in TBX<sup>172</sup>, è stato convertito in RDF usando il convertitore TBX2RDF<sup>173</sup> e ogni concetto terminologico è stato trasformato in skos:Concept. EMN, che è disponibile solo in formato HTML, è stato trasformato direttamente in RDF<sup>174</sup>.

I *dataset* sono stati infine collegati tra loro e con altre risorse linguistiche esterne a quelle prodotte dalle istituzioni e dagli organismi dell'Unione europea, quali BabelNet<sup>175</sup> e

---

<sup>170</sup> Il glossario EMN è stato sviluppato dall'“European Migration Network” (EMN), costituito dalla Commissione Europea e dai National Contact Point (ovvero la struttura che offre guide, informazioni e assistenze sulla partecipazione ad Horizon 2020). I NCP partecipanti sono stati tutti quelli degli Stati Membri dell'UE, più la Norvegia. Per maggiori informazioni su EMN si rimanda a “*Asylum and Migration Glossary 3.0, a tool for better comparability produced by the European Migration Network*”, Ottobre 2014, [https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/european\\_migration\\_network/docs/emn-glossary-en-version.pdf](https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/european_migration_network/docs/emn-glossary-en-version.pdf) mentre il glossario interattivo è disponibile presso [https://ec.europa.eu/home-affairs/what-we-do/networks/european\\_migration\\_network/glossary/index\\_a\\_en](https://ec.europa.eu/home-affairs/what-we-do/networks/european_migration_network/glossary/index_a_en) (consultato il 22 Dicembre 2016).

<sup>171</sup> Lemon sta per Lexical Model for Ontologies ed è stato concepito per rappresentare le informazioni lessicali in combinazione con ontologie. Cfr. McCrae J., et al., (2011), “*Linking lexical resources and ontologies on the semantic web with lemon. In The Semantic Web: Research and Applications*”.

<sup>172</sup> Presso la pagina dedicata <http://iate.europa.eu/tbxPageDownload.do>.

<sup>173</sup> Il convertitore è disponibile presso <http://tbx2rdf.lider-project.eu>.

<sup>174</sup> La versione Linked Data di IATE è disponibile su <http://tbx2rdf.lider-project.eu/data/iate> mentre quella di EMN è disponibile su <http://data.lider-project.eu/emn>.

<sup>175</sup> BabelNet è sia un dizionario enciclopedico multilingue che un network semantico che connette concetti in un network di relazioni semantiche composto da 14 milioni di entrate. Si veda il paragrafo successivo e <http://babelnet.org>.



MASC<sup>176</sup>. Il risultato è stato la creazione di 3028 link tra EMN e IATE, di cui 2038 sono *best match*<sup>177</sup>, e di 700 link tra IATE e BabelNet tramite EMN come pivot.

Con questo lavoro, il team di ricerca ha presentato una nuova modalità di pubblicare e linkare terminologie usando i principi dei Linked Data.

Tra i principali vantaggi di pubblicare terminologie in RDF vi è la possibilità che supporti i collegamenti tra risorse diverse tramite l'impiego dello stesso linguaggio. Specificamente su IATE, sebbene il database non offra un *dump*<sup>178</sup> direttamente in RDF e quindi convertito da TBX, può essere inserito nel cloud dei Linguistic Linked Open Data. Secondo gli autori, questo tipo di approccio può trovare delle applicazioni pratiche.

Un esempio potrebbe essere il caso di Tilde Terminology<sup>179</sup>, un servizio online dalle varie funzionalità: è un motore di ricerca terminologico, estrattore di terminologia e piattaforma di condivisione e *storage* di terminologia. Il servizio si interfaccia inoltre con i principali CAT tool ed è compatibile coi sistemi di Machine Translation.

Un altro lavoro concentrato su IATE è quello condotto da un gruppo del centro di Eccellenza dell'Università di Bielefeld (Siemoneit *et al.*, 2015), che ha pubblicato risorse linguistiche in formato Linked Data nel cloud Linguistic Linked Data. Vantaggio fondamentale per gli autori è la creazione automatica di collegamenti tra elementi del

---

<sup>176</sup> Manually Annotated Subcorpus, un corpus di testi e trascrizioni recuperati principalmente da *Open American National Corpus* (OANC), un corpus di testi in Inglese Americano composto da 22 milioni di parole, scritte e parlate.

<sup>177</sup> Il numero totale dei link tra EMN e IATE è stato di 3028 perché si considerano tutti gli abbinamenti possibili. 2038 sono invece i "best match", ovvero i migliori abbinamenti possibili tra due concetti nei due database. Cfr. Cimiano *et al.*, (2015).

<sup>178</sup> Ovvero una conversione di una risorsa in un determinato formato già disponibile per il download.

<sup>179</sup> Per l'utilizzo e maggiori informazioni si rimanda a <https://term.tilde.com/>.

*Cloud*, essenziale ai fini dell'accrescimento della qualità delle relazioni attraverso cui è possibile l'integrazione e l'esplorazione di dati attraverso i vari *dataset*.

Il gruppo esamina il collegamento automatico di quattro diversi *dataset*, di cui due sono risorse terminologiche, EMN e IATE, e due sono di tipo lessico-concettuali, MASC e BabelNet. IATE e EMN sono state rappresentate in RDF usando il modello *lemon*, mentre per MASC è stato impiegato NIF (acronimo di NLP Interchange Format) (Hellmann *et al.* 2013), un formato basato su RDF/OWL mirato all'interoperabilità tra NLP, risorse linguistiche e annotazioni<sup>180</sup>. Le quattro risorse sono state convertite in Linked Data.

Nel caso di IATE è stato utilizzato un convertitore TBX-RDF similmente a quanto usato nel precedente approccio di ricerca. Il progetto ha prodotto 3082 link tra EMN e IATE, di cui 2038 sono corrispondenze perfette. È notevole la precisione di collegamento tra i due repertori terminologici, che si attesta sul 94%, dato dovuto soprattutto alla somiglianza di contenuto tra le due risorse, in cui ad ogni termine viene associato un unico senso, a differenza di quanto accade, ad esempio, in BabelNet. Anche questo lavoro mostra che è possibile descrivere una metodologia per collegare entità linguistiche anche piuttosto diverse tra loro tramite l'impiego di Linked Data.

#### 4.5.2 Il progetto BabelNet

BabelNet è una rete semantica ma anche un dizionario enciclopedico multilingue ed è sviluppato dal Laboratorio di Linguistica Computazionale del Dipartimento di

---

<sup>180</sup> Hellmann S., *et al.*, (2013), "Integrating NLP using linked data".

Informatica dell'Università "La Sapienza", guidato da Roberto Navigli<sup>181</sup>. È stato creato collegando, automaticamente, Wikipedia e WordNet<sup>182</sup> ed è composto da più di 13 milioni di entrate, chiamate *synsets*<sup>183</sup>, che rappresentano ognuna un significato e contengono tutti i sinonimi che esprimono un determinato significato in un ventaglio di lingue (fino a 271).

Anche per questo progetto, è stata scaricata ed impiegata la versione in TBX di IATE.

Collegare IATE e BabelNet è stata possibile attraverso tre passi:

1. Il recupero dei concetti che presentano gli stessi lemmi in inglese in entrambe le risorse;
2. Il paragone tra ogni concetto estratto da IATE con ogni corrispondente concetto di BabelNet per quel lemma;
3. Il collegamento da un concetto di IATE ad un concetto BabelNet in modo da massimizzare la sovrapposizione dei lemmi e delle traduzioni (Navigli, 2016).

Per fare ciò è stato impiegato Babelfy<sup>184</sup>, un algoritmo per la disambiguazione basato sull'*Entity Linking*<sup>185</sup> e la *Word Sense Disambiguation*<sup>186</sup>. Sono state raccolte tutte le traduzioni e i sinonimi per generare un testo multilingue come input per Babelfy.

---

<sup>181</sup> BabelNet è consultabile presso <http://babelnet.org/about>.

<sup>182</sup> WordNet è un database lessicale per la lingua inglese che ha lo scopo di organizzare e descrivere i concetti espressi dai vocaboli. Per maggiori informazioni, si rimanda a <https://wordnet.princeton.edu/>

<sup>183</sup> Un *synset* è un insieme di sinonimi che possono essere descritti da un'unica definizione perché esprimono uno stesso senso.

<sup>184</sup> Per l'utilizzo, si rimanda a <http://babelfy.org/>.

<sup>185</sup> L'*Entity Linking* è l'operazione di determinare l'identità di un'entità in un testo. L'obiettivo di EL è di scoprire menzioni di entità entro un testo e linkarle all'entrata più adatta in una base di conoscenza (Moro *et al.*, 2014).

<sup>186</sup> Word Sense Disambiguation (WSD), "è l'abilità di identificare il significato di parole in un contesto in maniera computazionale" (Navigli, 2009). Sul WSD si rimanda ancora a Navigli (2010), e a Ide N., & Véronis J., (1998).

Le conclusioni del lavoro sono che la strategia migliore per collegare IATE e BabeNet è quella di disambiguare le definizioni di IATE e di BabeNet associate con un lemma; successivamente, si paragonano i *synset*<sup>187</sup> con ogni definizione; per ogni coppia “concetto IATE – *synset* BabeNet” si seleziona la coppia di definizioni più adatta in base ad un grado di similarità predefinita<sup>188</sup>.

I vantaggi di collegare BabeNet a IATE, secondo l'autore, sarebbero nell'arricchimento del database terminologico in numero di lingue e di BabeNet in termini di concetti specialistici, inoltre IATE diventerebbe, grazie a BabeNet, un network semantico e beneficerebbe della quantità di informazioni aggiuntive contenute in BabeNet. Sarebbe infine possibile identificare concetti duplicati.

#### 4.5.3 Il progetto EuroTermBank

EuroTermBank<sup>189</sup> è un portale di ricerca terminologica online e gratuito, una banca terminologica centralizzata che collega altre banche terminologiche tra loro. L'obiettivo è di facilitare il miglioramento dell'informazione del settore pubblico dei nuovi Stati membri dell'Unione Europea (Estonia, Ungheria, Lituania, Lettonia, Polonia) tramite

---

<sup>187</sup> Navigli descrive il gruppo di *synset* ottenuti come “*bag-of-synsets*” (2016).

<sup>188</sup> Un esempio è di inserire in Babelfy la definizione di “albero” (del dominio della botanica) di IATE e BabeNet. I *synset* risultanti, dopo la disambiguazione con Babelfy, avranno un maggior grado di similarità poiché saranno maggiori i sensi condivisi. Il contrario, invece, risulterà se una delle due risorse disambiguerà in Babelfy la definizione di “albero” come grafo. A nessun senso condiviso dopo la disambiguazione, non corrisponderà alcuna similarità tra definizioni. (Cfr. Navigli, 2016).

<sup>189</sup> EuroTermBank è un progetto co-finanziato dal programma eContent della Commissione Europea (eContent è stato attivo dal 2001 al 2004 ed è ora concluso) e condotto da un consorzio composto da Tilde (coordinatore), Centre for Language Technology at University of Copenhagen, Information Processing Centre, Polonia, Institute for Information Management at Cologne University of Applied Science, Germania, Institute of Lithuanian Language, Lituania, MorphoLogic, Ungheria, Terminology Commission of the Latvian Academy of Sciences, Lettonia, Università di Tartu, Estonia. Per una consultazione della banca terminologica si rimanda a <http://www.eurotermbank.com/about.aspx>, per una più completa documentazione, a Rirdance S., Vasiljevs A., (2006).

l'armonizzazione, collezione e disseminazione delle risorse terminologiche pubbliche, ritenute «risorse terminologiche disperse» (Vasiljevs *et al.*, 2005), da cui però partire per ottenere interoperabilità.

*«EuroTermBank is a new type of standards-based multilingual terminology gateway that provides unified access to centralized and distributed multilingual terminology resources, applying existing and emerging standards and concepts to ensure interoperability on several levels across diverse terminology data. Its federation approach enables consolidation of centrally stored resources with resources residing in diverse established terminology databases»<sup>190</sup>.*

La consolidazione, dunque, va vista come quell'attività atta alla riduzione della dispersione di risorse terminologiche multilingui. Il lavoro terminologico è organizzato in maniera diversa nei diversi paesi e nelle diverse istituzioni ed è determinato dagli obiettivi degli attori coinvolti. Il consorzio divide il lavoro di terminologia in tre livelli: locale, nazionale e internazionale (Henriksen *et al.*, 2005). A livello locale, il lavoro terminologico è utile, solitamente agli scopi di una particolare organizzazione, come ad esempio un'agenzia di traduzione, un'azienda ecc. Il livello locale si limita solitamente ad uno o pochi domini e il lavoro terminologico è spesso limitato nello scopo. Conseguentemente la consolidazione terminologica non è un'attività particolarmente sviluppata. A livello nazionale, il lavoro terminologico è solitamente bilingue o monolingue: a questo livello operano le istituzioni coinvolte nella standardizzazione dei termini e nello sviluppo e mantenimento dei sistemi di gestione terminologica. A livello

---

<sup>190</sup> Vasilievs A., *et al.*, (2008) "EuroTermBank: Towards Greater Interoperability of Dispersed Multilingual Terminology Data", pp. 213-220.

internazionale, il lavoro terminologico riguarda la consolidazione e l'armonizzazione di termini conosciuti a livello nazionale e locale e richiede la coordinazione e la gestione della terminologia in un'infrastruttura complessa. Un buon esempio in termini di consolidazione e armonizzazione della terminologia su un livello sovranazionale è ovviamente IATE.

EuroTermBank offre lo scambio di dati terminologici stabilendo relazioni tra banche dati, allineando metodologie e standard e sviluppando meccanismi di scambio. Il progetto propone il formato unificato TeDIF (Terminology Documentation Interchange Format) (Betz e Schmitz, 1999) per l'annotazione delle risorse terminologiche, che offre schemi descrittivi per descrivere dati bibliografici e collezioni di termini. TeDIF è anche in formato compatibile con XML<sup>191</sup>. Per favorire lo scambio di dati, invece, è utilizzato il linguaggio di marcatura TBX che, come visto, è anch'esso un formato compatibile con XML ed è utile per importare ed esportare dati da e in EuroTermBank. Il TBX assicura inoltre consistenza e coerenza dei dati e favorisce senz'altro l'interoperabilità all'interno del database così come al suo esterno.

Il consorzio introduce il concetto di "federazione di banche terminologiche", riprendendo la definizione di Galinski (2007):

*«Federation in general is a fairly recent concept in linking portals and data repositories that goes beyond the establishment of pointers or links and reaches out to the level of semantic interoperability of data and data structures.*

---

<sup>191</sup> È un subset di SGML, Standard Generalized markup Language.

*Especially terminology and other kinds of structured content can be made to enable interoperability in the form of network(s) of federated databases»<sup>192</sup>.*

Gli autori risolvono tali requisiti tramite meccanismi di interconnessioni tra EuroTermBank e database esterni, che possono essere interrogati tramite un'interfaccia web comune. Connettendo la risorsa esterna tramite tale modalità, si assicura un collegamento *machine-machine* e ciò viene eseguito tramite messaggi XML seguendo gli standard SOAP (Simple Object Access Protocol). Purtroppo EuroTermBank, nonostante la vocazione ambiziosa, non ha applicato, nei suoi sviluppi futuri, i principi dei Linked Data e verso i modelli di rappresentazione più diffusi nel LLOD cloud, quali RDF. Il progetto rimane purtuttavia un pioniere nel contesto della centralizzazione e armonizzazione della gestione delle risorse terminologiche.

---

<sup>192</sup> Galinski C., (2007) *“New ideas on how to support terminology standardization projects”*.

## CAPITOLO V - L'INFORMATICA E LA TERMINOLOGIA PER IATE: DALLE TERMBASE VERSO L'INTEROPERABILITÀ SEMANTICA

Le istituzioni dell'Unione Europea sono costantemente attive per migliorare la qualità e l'affidabilità delle risorse terminologiche da esse prodotte. Particolarmente puntuale, in questo senso, l'intervento di Rodolfo Maslias, capo di Terminology Coordination Unit del Parlamento europeo, a proposito del ruolo della terminologia nella legislazione dell'Unione europea e sulle sfide della gestione di IATE:

*«The availability of reliable terminology resources is a key factor for achieving high quality in all fields of multilingual communication. Terminology needs to be connected with all IT tools in the changing world of translation»<sup>193</sup>.*

In un contesto storico caratterizzato dall'alta complessità dei documenti unionali, è emergente l'esigenza delle istituzioni di migliorare le proprie risorse terminologiche favorendone l'integrazione con gli strumenti di traduzione.

Lo *IATE Management Group* (IMG), organo di governo di IATE composto da esponenti dei dipartimenti terminologici di ogni istituzione europea, è da tempo attivo in questa direzione e ha promosso una serie di sviluppi per il database terminologico, mirati ad un

---

<sup>193</sup> Da "in termo qualitas", webinar allestito da International Association of Professional Translators and Interpreters", 17 giugno 2017.



impiego diretto della termbank nei software di traduzione e alla standardizzazione della gestione della terminologia.

Nel Parlamento europeo è in atto un progetto denominato “Terminology in Studio”, con l’obiettivo di migliorare la qualità e la coerenza della terminologia istituzionale, impiegando i prodotti dei recenti sviluppi di IATE nel CAT tool<sup>194</sup> SDL Studio e creando quindi un nuovo *workflow* delle risorse terminologiche nei processi traduttivi.

Allo stesso tempo, il Dipartimento di Scienze Politiche Sociali e della Comunicazione dell’Università degli Studi di Salerno ha sottoscritto un accordo di collaborazione<sup>195</sup> con la Terminology Coordination Unit del Parlamento europeo, finalizzato alla redazione di schede terminologiche per lo sviluppo della banca dati multilingue IATE.

Presupposto di questo progetto è che il database terminologico, che contiene il repertorio delle terminologie specialistiche dell’Unione europea, possa essere impiegato per un nuovo utilizzo, ovvero non solo come una risorsa traduttiva e di divulgazione dei termini, ma anche come strumento per l’interoperabilità semantica.

Come sostiene Daniela Vellutino (2016):

*«L’Unione Europea è una fonte primaria di diritto per gli Stati Membri e attraverso le norme, le attività amministrative e comunicative delle sue istituzioni*

---

<sup>194</sup> Computer Assisted Translation tool.

<sup>195</sup> Responsabile scientifico del gruppo di ricerca è Daniela Vellutino, al gruppo di ricerca partecipano studenti di laurea magistrale e dottorato, tra cui Francesco Rossi, responsabile per i rapporti con l’Unità di Coordinamento Terminologico del Parlamento europeo.

*diffonde terminologia specialistica dei domini delle sue molteplici sfere d'azione»<sup>196</sup>.*

In altre parole, l'Unione europea è produttrice di terminologia specialistica del lessico istituzionale:

*«il complesso di termini di differenti domini semantici il cui significato è oggetto di definizioni presenti nei testi legali delle istituzioni europee [...]. Sono termini istituzionali anche quelli il cui significato è definito in testi legali prodotti dalle negoziazioni tra i rappresentanti degli Stati membri. [...] La terminologia istituzionale di IATE, se controllata e convalidata, può dunque essere impiegata come strumento per gli schemi di metadattazione e nei vocabolari controllati per la rappresentazione di servizi pubblici, in linea con l'Azione "Improving semantic interoperability in European eGovernment systems", promossa da ISA<sup>2</sup>, Interoperability solutions for public administrations, business and citizens<sup>197</sup>, che offre soluzioni per la gestione dei metadati istituzionali e il loro riuso»<sup>198</sup>.*

---

<sup>196</sup> Vellutino D., Maslias R., Rossi F., (2016), "Verso l'interoperabilità semantica di IATE, Studio preliminare per il dominio «Gestione dei rifiuti urbani»", in *Terminologie specialistiche e diffusione dei saperi*, a cura di Zanola M. T., Diglio C., Grimaldi C., pp.221-240.

<sup>197</sup> Interoperability Solutions for European Public Administration (ISA) è il programma approvato dalla Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del 16 settembre 2009, sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA) (GU L 260 del 3.10.2009, pag. 20). Oggi è stato sostituito da ISA<sup>2</sup>, la sua evoluzione per il quinquennio in atto. Per maggiori informazioni sull'Azione "Improving semantic interoperability in European eGovernment systems" si rimanda a [https://ec.europa.eu/isa2/actions/improving-semantic-interoperability-european-egovernment-systems\\_en](https://ec.europa.eu/isa2/actions/improving-semantic-interoperability-european-egovernment-systems_en).

<sup>198</sup> Vellutino D., Maslias R., Rossi F., (2016), *cit.* p. 225-227.

## 5.1 SDL Studio, una nuova sfida per IATE – il progetto “Terminology in Studio”

Il Parlamento ha utilizzato fino al 1 Aprile 2016 il software di traduzione *Translation WorkBench* (TWB), prodotto da SDL Trados<sup>199</sup>, integrato nel software Microsoft Word e, dunque, l’attività di traduzione avveniva nella cornice dello stesso programma di videoscrittura.

Dal 2014, però, in seguito ad una gara d’appalto interistituzionale, il Parlamento ha adottato il software di traduzione SDL Studio, ultimo prodotto commerciale di SDL ed “erede” di TWB.

La migrazione alla nuova piattaforma ha cambiato radicalmente il *workflow* di gestione della terminologia del Parlamento. Si è già presentato<sup>200</sup> che, in un ambiente “Word”, le funzioni terminologiche venivano esplicitate tramite la funzione “Terminology macro” di *DGTRAD ribbon* che permette, da un lato, di avere un accesso immediato al database a IATE e, dall’altro, di raccogliere terminologia individuata durante il lavoro di traduzione.

Con il passaggio a SDL Studio, tali funzioni non sono più utilizzabili ed è dunque emersa l’esigenza di sviluppare nuovi strumenti che permettano di avvalersi delle risorse di IATE in SDL Studio, perfezionando dunque la qualità dei documenti tradotti attraverso una migliore coerenza terminologica.

---

<sup>199</sup> Trados è l’azienda produttrice di Translation Workbench. Nel 2005 fu acquistata dal SDL International, diventando quindi SDL Trados.

<sup>200</sup> Si consulti il par. 2.2.3.

SDL Studio offre tra le funzioni native la possibilità di gestire le risorse terminologiche in una termbase. È opportuno riprendere il contributo di Alan Melby (2012), che la definisce come:

*«a computer database consisting primarily of information about domain-specific concepts and the terms that designate them. Specialised translation deals with domains of knowledge, and every domain is organised through concepts that are linked to objects or ideas relevant to that domain. Termbases may be monolingual, bilingual, or multilingual»<sup>201</sup>.*

Parafrasando la definizione del ricercatore americano, una termbase è un glossario con la finalità di fornire informazioni terminologiche al traduttore per le funzioni di recupero, consultazione e riutilizzo dei dati terminologici.

Nella cornice interistituzionale, una termbase è un database, solitamente bilingue o multilingue, che contiene risorse terminologiche estratte da IATE da impiegare nella traduzione dei documenti unionali.

Gli indubbi vantaggi dell'utilizzo di una termbase per un traduttore sono la consultazione, direttamente nel software, dei dati terminologici e il riconoscimento e il suggerimento automatico di un termine, con la possibilità di inserirlo direttamente nel testo da tradurre.

---

<sup>201</sup> Melby (2012), *cit.* Su una letteratura sulle termbase si rimanda a Wright & Budin (2001), Melby (2012) e Bowker (2015).

Tuttavia, dal momento in cui SDL Studio è stato adottato, le termbase non sono state subito inserite nel *workflow* di traduzione e, dunque, non impiegate fin dal principio dell'adozione di SDL Studio.

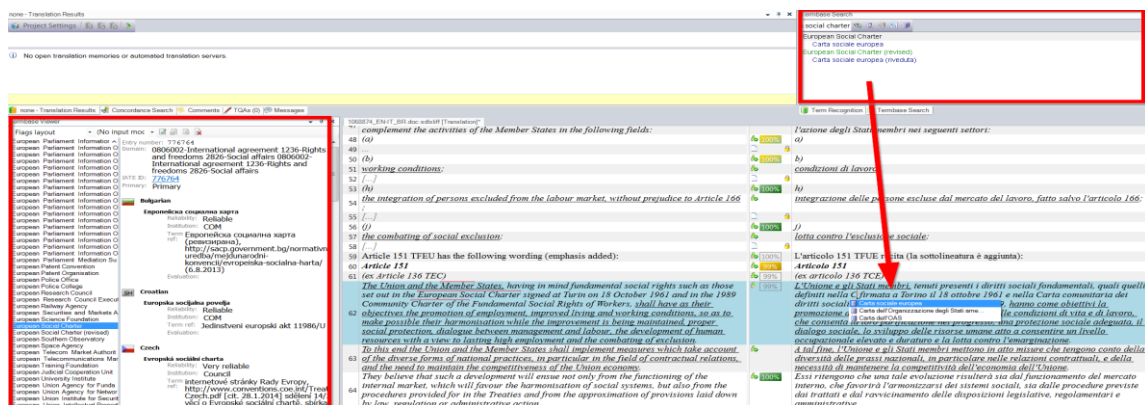


Fig. 5.1 Le funzioni principali delle termbase in SDL Studio

In figura, un esempio delle funzioni delle termbase in Studio, che ritrae un documento in fase di traduzione.

Nell'area centrale è visibile il documento nella lingua originale, in questo caso in inglese, già diviso in segmenti per facilitare il lavoro di traduzione, mentre nell'area destra è visibile il documento nella lingua di destinazione, in questo caso l'italiano.

In alto a destra è visibile il riquadro della termbase. Tale riquadro mostra i termini in essa contenuti che possono essere impiegati per la traduzione. Più precisamente, il software identifica che un termine contenuto nella termbase è presente nel documento di origine (evidenziato da una sovralineatura in rosso) e ne propone l'equivalente nella lingua di destinazione. La freccia indica, inoltre, la funzione di autosuggerimento, ovvero la funzione secondo la quale il termine, o i termini, vengono

proposti in un menu a tendina tramite cui è possibile inserirli nel testo senza dover digitarne tutte le lettere.

Evidenziato a sinistra, invece, il menu *termbase viewer* in cui è possibile consultare tutte le entrate della termbase ed avere accesso a maggiori dettagli dei campi delle schede IATE.

Lo IATE Management Group quindi ha creato una Task Force per l'integrazione di IATE e Studio con il risultato della creazione di una funzione di IATE che permette di estrarne sezioni convertite in basi terminologiche compatibili con il software Studio.

A partire dall'estate 2015, il Parlamento europeo ha allestito il progetto "*Terminology in Studio*", il cui gruppo di lavoro è composto da membri di TermCoord, ITS (Information Technology Support Unit) e dai terminologi delle Unità linguistiche.

Come anticipato poco sopra, gli obiettivi del progetto sono l'analisi delle potenzialità dell'uso delle termbase in Studio, la loro implementazione nel *workflow* di traduzione, il perfezionamento della qualità dei documenti e il miglioramento della gestione della terminologia tramite un nuovo *workflow*.

Naturalmente, l'impiego delle termbase nei processi traduttivi implica un miglioramento della standardizzazione della terminologia e conseguentemente ne limita la variabilità poiché questa, idealmente, sarebbe già validata prima di essere utilizzata dal traduttore.

Le risorse contenute nelle termbase devono, dunque, essere di qualità, pre-approvate e, soprattutto, aggiornate in base alle continue evoluzioni della terminologia istituzionale. Come sostiene Daniela Vellutino (2016),

«i significati dei termini sono definiti nei documenti di programmazione e di attuazione della strategia e, soprattutto, nei Regolamenti che hanno una validità giuridicamente definita, per cui anche i termini di questo micro dominio semantico hanno una precisa data di scadenza. Pertanto la terminologia specialistica del lessico istituzionale è variabile, ma ha significati stabili perché definiti nei testi legali che, in alcuni casi, così come attestano la nascita di un termine ne attestano anche l'obsolescenza e/o la risemantizzazione»<sup>202</sup>.

Francesco Rossi ricopre, per il Working Group, la funzione di coordinatore tecnologico. Si occupa di supporto avanzato sull'uso, la creazione e la gestione di termbase, fornisce soluzione per allestire ambienti informatici per il loro utilizzo, effettua valutazioni di rischio e impatto sulle infrastrutture tecnologiche e contribuisce allo sviluppo del nuovo *workflow* per il loro impiego nei processi traduttivi.

### 5.1.1 Terminology in Studio, fase 1 – IATE Advanced Export Interface

Nel 2015, in IATE è stata implementata la funzione *IATE Advanced Export Interface* che permette di estrarre dati da IATE e convertirli in formato Excel, XML, o in SDLTB, formato proprietario di SDL Studio per le termbase.

La funzione *Export* è senza dubbio uno degli sviluppi più importanti di IATE in quanto gli output esportati, specialmente le termbase, possono essere utilizzati direttamente nei CAT tool impiegati nel Parlamento e nelle altre istituzioni nel lavoro traduttivo. Questo

---

<sup>202</sup> Vellutino D., Maslias R., Rossi F., (2016), *cit.* p. 225-226.

sviluppo ha permesso per la prima volta l'utilizzo delle risorse terminologiche senza l'ausilio di altri strumenti esterni al software di traduzione.

IATE permette tre tipi di *Export*:

1. Creazione di un glossario informativo. IATE produce un file Excel che permette di avere una visione completa di un determinato *dataset* su cui intervenire successivamente. Questo formato mostra tutti i dati delle istituzioni e non è utilizzabile per un lavoro off-line;
2. Creazione di un *export* per il lavoro offline. Questo output è un file Excel per un lavoro offline mirato ad implementare alcune modifiche in IATE e quindi non effettuandole direttamente nell'interfaccia del database. Il file con i cambiamenti potrà poi essere reimportato successivamente. Gli utenti non possono accedere ai dati delle altre istituzioni;
3. Creazione di un file SDL XML (con formato output XML o SDLTB<sup>203</sup>). Questo output è il più avanzato e complesso e permette la creazione di termbase per SDL Studio.

Questa terza funzione produce due possibili output:

- **Un file XML**, formato compatibile con SDL Multiterm, software proprietario di SDL per la gestione delle risorse terminologiche, in cui è possibile effettuare operazioni di personalizzazione ed editing e reimportare il file in un *template* SDLTB per Studio;

---

<sup>203</sup> L'estensione SDLTB indica un file termbase (TB) nel formato proprietario SDL. È l'estensione standard delle termbase utilizzabili nel software SDL Studio.



- **Un file SDLTB**, formato proprietario di SDL che permette l'utilizzo di funzioni di terminologia direttamente in Studio tramite termbase.

Il "download" dei file è asincrono, ovvero avviene prima una selezione dei campi rilevanti e successivamente la creazione del file, scaricabile da un'altra interfaccia del sito.

Nella figura sottostante viene mostrata l'interfaccia di *Export* di IATE.

The screenshot shows the IATE Export interface. At the top left is the IATE logo. The user is logged in as 'FROSSI' with the role 'EP Administrator'. The interface has three radio buttons for output types: 'Creation of an information glossary', 'Creation of an export for off-line work', and 'Creation of SDL XML file' (which is selected). Below this, there is a link to download a standard SDLTB template. A dropdown menu for 'Please select export format' has 'XML' and 'SDLTB' (selected) options. There is a text input for 'A name of the project (optional)' and a 'Shared' checkbox. The 'Basic criteria' section includes a text area for 'Entry IDs (separated by commas)' and a 'Languages' section with 'Source language' and 'Target language' dropdowns. The 'Source language' is set to 'Not selected', and the 'Target language' dropdown shows options: 'Not selected', 'it - Italian', 'bg - Bulgarian', and 'cs - Czech'. At the bottom, there are radio buttons for 'Primary entries' (Included, Only, Excluded) and 'Pre-Iate terms' (Included, Only, Excluded). An 'Extend criteria Show' link is also present. On the right side, there are two 'Export' and 'Reset' buttons.

Fig. 5.2 Interfaccia estrazione di glossari e file SDL XML<sup>204</sup>

In alto si seleziona una delle tre tipologie di output e, nel caso dei file SDL, si può ulteriormente scegliere tra file in formato XML e SDLTB.

<sup>204</sup> Fonte immagine <https://iate.cdt.europa.eu/iatenew/login.do?method=logout> (data consultazione, 3 marzo 2017).

Il manuale IATE Handbook raccomanda di attribuire un nome al file che si intende creare in modo da riconoscere facilmente il proprio *export* nell'interfaccia di download.

I criteri di esportazione sono piuttosto vari e dettagliati, ma non capillari come quelli degli output in XML. Anzitutto, è obbligatorio selezionare sempre una lingua fonte e almeno una lingua di destinazione. Si può poi scegliere se inserire o escludere le *Primary entries* e i termini *Pre-IATE*<sup>205</sup>. Si può inoltre creare una termbase specifica, direttamente inserendo gli ID delle schede.

Ulteriori opzioni di filtraggio sono illustrate nella Figura 5.3.

---

<sup>205</sup> I termini pre-IATE sono entrate, spesso glossari, aggiunti in IATE tramite la Pre-IATE Macro, «*that is used to import batches of entries (e.g. glossaries) into IATE en masse, making potentially useful information available quickly and effortlessly*», cit. IATE Handbook, p.32. I terminologi incorporano tali dati in IATE in tempi successivi, seguendo processi di validazione. Per questo motivo i dati pre-IATE non sono disponibili per IATE pubblico ed escludibili anche dalle funzioni di ricerca di IATE interno.

The screenshot displays the IATE search interface with the following sections and options:

- Languages:** A table with columns for 'Source language' (set to 'Not selected') and 'Target language' (showing a list including 'gd - Gaelic (Scots)', 'ny - Chichewa; Nyanja', 'fj - Fijian', and 'oc - Occitan (post-1500)- Provenca').
- Primary entries:** Radio buttons for 'Included' (selected), 'Only', and 'Excluded'.
- Pre-Iate terms:** Radio buttons for 'Included' (selected), 'Only', and 'Excluded'.
- Extend criteria:** A link to 'Hide' the extended criteria section.
- Institutions:** A list of institution codes: 44 - EP-Codict, 45 - MAOC-N-CdT, 46 - Eurojust-CdT, 47 - EIGE-CdT, 48 - EDA-CdT. To the right are radio buttons for 'LIL level', 'LL level', and 'TL level' (selected).
- Domain code:** A search box with a 'Find!' button. Results include: 'Domain code not specified(00)', 'Politics(04)', 'Political framework(0406)', 'Political ideology(0406001)', and 'Political institution(0406002)'. Below this is an empty list box.
- Collection name:** A search box with a 'Find!' button and an empty list box below it.
- Reliability equal or greater than:** A dropdown menu set to '0 - downgraded prior to deletion'.
- Actions:** A link labeled 'Delete selected items' is located below the collection name search area.

Fig. 5.3 Opzioni di filtraggio delle termbase<sup>206</sup>

È possibile filtrare la ricerca in base all'istituzione, ai livelli delle entrate, al codice dominio, al nome delle collezioni e al grado di affidabilità. Al termine delle operazioni di selezione, il file è pronto per il download nell'apposita sezione.

Tramite questa piattaforma, sono state prodotte termbase contenenti terminologia normativa che sono state utilizzate dal Working Group di Terminology in Studio come primo elemento per la valutazione di stabilità e risoluzione di errori.

<sup>206</sup> Fonte immagine <https://iate.cdt.europa.eu/iatenew/login.do?method=logout> (data consultazione, 3 marzo 2017)

Le termbase normative contengono risorse terminologiche ad alta affidabilità (4 stelle in IATE), inclusi nomi di trattati, organismi e istituzioni.

Nella fattispecie, è stata creata un'unica termbase contenente tutte le 552 possibili combinazioni linguistiche delle 24 lingue ufficiali e, quindi, ugualmente accessibile<sup>207</sup> da tutti i traduttori.

Inoltre, sono anche state create termbase contenenti l'intera collezione di terminologia in una data combinazione linguistica (ad esempio EN-IT, oppure EN-DE, chiamate "*full IATE retrievals*") e altre termbase con contenuto filtrato in base al dominio.

A dicembre 2015 è stato avviato un primo progetto pilota per l'uso di tali risorse.

L'uso delle termbase normative è stato accolto piuttosto positivamente dalle Unità traduttive, mentre le opinioni sulle *full IATE retrievals* non hanno accolto lo stesso tasso di soddisfazione. Il motivo è da ricercare nella qualità delle risorse terminologiche contenute in IATE.

Tradizionalmente le lingue ufficiali si distinguono in "pre-2004" e "post-2004", dove il 2004 è l'anno del maggior allargamento dell'Unione Europea<sup>208</sup>, e corrisponde anche all'anno in cui il database IATE è stato impiegato con piena funzionalità dalle istituzioni

---

<sup>207</sup> SDL prevede che una termbase sia accessibile da un unico utente. Accessi multipli e contemporanei possono rallentarla o bloccarne l'utilizzo. Tuttavia, replicare una termbase di dimensioni piuttosto notevoli (le normative e le *full IATE retrievals* sono grandi fino a 500MB) un numero "n" di volte avrebbe avuto un impatto peggiore sulla capacità di *storing* del Parlamento, per cui la condivisione è stata ritenuta comunque fattibile, anche se ha richiesto un certo livello di interventi manuali sulla struttura del server e del database.

<sup>208</sup> Il cosiddetto "allargamento UE a Est" è stato il maggiore ampliamento dell'Unione, in cui 10 nuovi Paesi sono diventati, simultaneamente, membri dell'UE. I paesi sono Cipro, Repubblica Ceca, Estonia, Ungheria, Lettonia, Lituania, Malta, Polonia, Slovacchia, Slovenia. Successivamente, nel 2007, anche Bulgaria e Romania aderirono al progetto europeo.

europee. Si ricorda che IATE è il risultato della convergenza di diversi database preesistenti che, naturalmente, non contenevano informazioni terminologiche nelle lingue dei paesi del post-2004. Dunque, le risorse terminologiche nelle lingue “pre-2004” contengono potenzialmente molto più rumore, duplicati e informazioni obsolete e non aggiornate a causa dell’unione dei diversi database, mentre le lingue “post-2004” risultano molto più pulite e aggiornate poiché il relativo lavoro terminologico è stato fatto direttamente in IATE e quindi l’influenza della sovrapposizione delle risorse da database precedenti è stata nulla.

Per le lingue “post-2004”, inoltre, le schede di IATE sono state create anche sulla base degli insegnamenti appresi dalla convergenza. Per *“Terminology in Studio”*, ciò comporta che le risorse terminologiche delle lingue “vecchie” siano meno soddisfacenti di quelle “nuove” e, conseguentemente, le *full IATE retrievals* delle lingue “pre-2004” contengono minor qualità di quelle delle lingue “post-2004”.

Ad ogni modo, a livello tecnico, per alcune settimane di test non sono stati riportati problemi di performance del server o dei database.

Tuttavia questa prima fase pilota ha subito una battuta d’arresto perché alcuni progetti di traduzione rallentavano vistosamente e bloccavano gradualmente il software e il computer stesso.

Le analisi hanno riportato che tale comportamento è causato da un bug, riconosciuto come “LTB-247” da SDL, che si verifica con la presenza di termini più lunghi di 100 caratteri in una termbase.

Nel modulo di terminologia di Studio, se viene rilevato un termine maggiore di 100 caratteri, la funzione di riconoscimento attiva la ricerca dell'equivalente nella lingua di destinazione, ma il processo entra in *loop*. Ciò provoca uno sforzo di RAM del computer e successivamente un rallentamento e/o un blocco del software o del PC stesso.

IATE è un database terminologico complesso: contiene termini costituiti da sequenze lessicali complesse e anche da fraseologie ed è possibile che un termine possa essere piuttosto lungo e superare i 100 caratteri. Di conseguenza, poiché alcune delle termbase scaricate da IATE causano potenziali rallentamenti e interruzioni, è stata decisione unanime dello Studio Steering Committee<sup>209</sup>, di sospendere l'utilizzo delle termbase estratte da IATE.

L'errore è stato successivamente identificato anche da parte delle altre istituzioni. ITS (*Information Technology Support Unit*) del Parlamento ha proposto una prima pulitura delle termbase, rimuovendo tutti i termini con più di 100 caratteri, ma rimandando la responsabilità di una pulitura sistematizzata allo *IATE Management Group* e all'azienda produttrice del software. SDL ha dichiarato a dicembre 2016 che il bug è stato risolto grazie ad un aggiornamento dell'ultima versione di Studio, ovvero SDL Studio 2017. Un eventuale aggiornamento dell'installazione del software nel Parlamento non è però, al momento in cui si scrive, stato programmato.

---

<sup>209</sup> Studio Steering Committee è l'organo di gestione delle strategie e delle decisioni per il progetto di Studio nel Parlamento.

### 5.1.2 Terminology in Studio, fase 2 - Term Recognition Module

Data la complessità di IATE e delle risorse terminologiche da esso contenute, un problema come il bug “LTB-247” sarebbe un fenomeno piuttosto frequente, nonostante sia stato riprodotto solo nel 5% dei casi durante la prima fase di test. Cionondimeno, considerato l’impatto sull’intero ambiente informatico, il progetto è stato congelato per alcune settimane. Una soluzione, però, è stata trovata dalla Commissione, che ha proposto un nuovo sistema di filtraggio dei contenuti delle termbase che elimina alla fonte, ovvero già in fase di creazione della termbase in IATE, quegli elementi che possono provocare il bug.

Tale sistema di filtraggio è stato implementato in un nuovo modulo in IATE, denominato “Term Recognition Module”. A differenza del precedente, questo modulo non crea output terminologici sulla base di campi scelti dall’utente, ma effettua estrazione terminologica<sup>210</sup> da un dato documento, ovvero ne riconosce automaticamente i termini, ne associa quelli nelle lingue di destinazione e restituisce una termbase contenente le combinazioni linguistiche desiderate.

La termbase è poi salvata sul server di IATE, dove rimane per 48 ore, da cui è possibile il download per un utilizzo in SDL Studio. Il file è inoltre disponibile anche nei formati XML o CSV.

---

<sup>210</sup> L’estrazione terminologica è un’operazione che consiste nell’identificazione di termini a partire da un dato corpus, sulla base dei quali si possono creare risorse terminologiche quali, ad esempio, le termbase. Gli algoritmi di TRM sono calibrati specificamente sulle risorse di IATE, rendendolo, in questo senso, un *tool* unico. Tuttavia, il contesto degli strumenti tecnologici applicati alla traduzione offre molti strumenti di estrazione (semi)automatica. Una panoramica su tali strumenti, categorizzati per schede analitiche, è disponibile nell’appendice di questo elaborato.

Fig. 5.4 Term Recognition Module

In figura 5.4 è possibile vedere l'interfaccia, molto intuitiva, del Term Recognition Module<sup>211</sup> (TRM). Nel campo in alto l'utente inserisce il nome della termbase, che si può inoltre decidere se condividere con altri utenti o meno. Tramite la funzione "Scegli file" si può scegliere il documento da cui estrarre la terminologia e tramite il menu "Target Language" si scelgono le lingue di destinazione. Si possono inoltre caricare i più comuni tipi di file editabili, ovvero file Word, Excel, PowerPoint, html, xml, csv ecc.

Sono esclusi dalle termbase:

- i termini con affidabilità 0, 1 o 2;
- i termini marcati come "Deprecated" o "Obsoleti";
- i termini più lunghi di 10 parole;
- termini "MUL", in latino e Lookup;

<sup>211</sup> Fonte immagine: <https://iate.cdt.europa.eu/iatenew/login.do?method=logout>. Data ultima consultazione: 13 Maggio 2017.



Sono inoltre applicate liste di *stopwords*<sup>212</sup> multilingui e una lista specificamente per la lingua Inglese.

Tali liste di “non termini” vengono aggiornate dai componenti del *Working Group* e dalle altre istituzioni. Un esempio di parola presente in una *stopword list* è “ITS”, contenuta in IATE perché è l’acronimo di “Information Technology Support”, ma è anche l’aggettivo “its” in Inglese, estremamente ricorrente nei documenti unionali.

Il *Working Group di Terminology* in Studio ha avviato una fase di test tra il 30 gennaio e il 29 marzo 2017. Partecipanti a questa nuova fase di prova sono stati i terminologi di ogni Unità traduttiva, supportati da *TermCoord* e da *ITS*.

355 sono state le termbase scaricate durante la fase di test, per un totale di 6,2 GB e una media di 17 MB.

Le lingue di origine più comune sono state Inglese, Francese e Tedesco. Sono state create 43 termbase aventi come lingua di destinazione il tedesco, 42 in danese, 30 in olandese, 25 in gaelico, 21 in romeno, 18 in francese, seguite dalle altre lingue con una media di 10 termbase. In italiano è stata creata una sola termbase.

La creazione di termbase *ad hoc* per documenti è senza dubbio un’utile funzione per ridurre il numero di dati irrilevanti o non necessari e per disporre solo delle informazioni relative al dominio/documento che si sta traducendo.

---

<sup>212</sup> Una *stopword list* è una lista di termini che vengono esclusi dai risultati degli algoritmi di estrazione. Solitamente si tratta di falsi termini o di altri elementi linguistici la cui presenza comporterebbe un eccessivo rumore della termbase.

L'affidabilità di tali risultati, ancora una volta, dipende però da IATE. Anche in questo caso il livello di soddisfazione della qualità dei risultati contenuti nelle termbase è stato estremamente variabile in base alla lingua.

Nelle lingue “pre-2004” le termbase contenevano molto rumore, dati ridondanti e talvolta poco affidabili. Nelle lingue “post-2004” le termbase invece contenevano risultati aggiornati e di maggior qualità.

Due i casi estremi e particolarmente emblematici: le termbase del Francese si sono rivelate essere poco utili o addirittura inutili, mentre quelle in Gaelico si sono rivelate una risorsa preziosa, riscuotendo l'entusiasmo dei suoi utilizzatori<sup>213</sup>.

Il motivo è da rintracciare ancora una volta nel fatto che i termini francesi in IATE sono il risultato di una convergenza di database precedenti, spesso non aggiornati e quindi di bassa qualità e affidabilità. Il Gaelico, che invece è una lingua ufficiale dell'Unione dal 2007, è stato gestito in maniera più efficiente e contiene, in termini assoluti, molti meno termini<sup>214</sup>.

Ad ogni modo, i risultati positivi della prima fase di test del TRM e l'assenza di errori bloccanti hanno finalmente permesso l'utilizzo delle termbase di IATE nella cornice di SDL Studio.

Ad oggi, 979 sono state le document-based termbase impiegate dalle Unità traduttive nel periodo tra fine gennaio e giugno 2017. Le termbase erano caratterizzate

---

<sup>213</sup> Fonte: estrazione da moduli di feedback della fase di test.

<sup>214</sup> In IATE sono presenti 67.497 termini in Gaelico e 1.297.538 termini in Francese. I dati sono aggiornati al quarto quadrimestre del 2016. Fonte: “IATE indicators”, documento interno che raccoglie le statistiche del database, tra cui il numero di termini per lingua, di duplicati, di modifiche e aggiornamenti delle schede ecc.

principalmente da coppie di lingue aventi l'inglese o il francese come lingua di origine. Per quanto riguarda le lingue di destinazione, le cinque più comuni sono state il tedesco con 232 termbase, l'olandese con 182, lo slovacco con 87, il romeno con 75 e il maltese con 58. Nei 5 mesi, solo 3 termbase sono state create per l'italiano.

Un primo beneficio dell'utilizzo di tali termbase è senz'altro la possibilità di accedere alle risorse del database direttamente dall'applicazione e conseguentemente di risparmiare il tempo della ricerca terminologica.

È fondamentale notare che l'impiego di termbase, in maniera sistematica, rappresenti un primo passo verso la standardizzazione e l'aumento della qualità e della coerenza terminologica.

Le risorse terminologiche devono ovviamente essere validate, aggiornate e controllate e molti sono i potenziali miglioramenti in questo senso, sia per le termbase che per IATE in sé. In quest'ottica l'implementazione delle termbase nel contesto della traduzione di una grande istituzione europea, a nostro parere, è strumentale al miglioramento della qualità della terminologia e della sua gestione.

### 5.1.3 Terminology in Studio – manual termbases

Sulla scia dei risultati positivi ottenuti con le termbase generate dal TRM, *Terminology Coordination Unit* ha proposto un nuovo, ulteriore impiego delle termbase in Studio.

Tale utilizzo è stato chiamato “manual termbases”, che rappresenta la possibilità di avere termbase “personalizzabili”, ovvero che possano essere create in risposta ad una necessità specifica di un traduttore, che ne gestisce il contenuto in maniera autonoma.

Il contenuto di tali termbase avrebbe sia provenienza da IATE che da liste e glossari create manualmente. In quest’ultimo caso i traduttori potrebbero disporre anche di fraseologie ed elementi testuali ricorrenti.

Tra le Unità di traduzione del Parlamento c’è un interesse crescente per quest’opportunità. Alcune hanno già avviato indipendentemente una fase di prova di tali termbase, popolandole delle espressioni frequenti usate nei documenti delle sessioni plenarie, oppure nei regolamenti “Financial Regulation” e “European Globalisation Adjustment Fund regulation”.

Queste termbase sono ideali per assicurare consistenza terminologica nelle traduzioni e conterranno solo elementi selezionati da documenti rilevanti e specifici per una determinata procedura, come ad esempio i termini e le espressioni di una proposta della Commissione o quelli contenuti in una TermFolder prodotta da TermCoord.

L’aspetto innovativo di tali termbase è la possibilità che contengano, talvolta, più elementi di quelle estratte da IATE, poiché il database interistituzionale non è sempre esaustivo in tutti i domini e in tutte le lingue. Inoltre, le *manual termbase* potrebbero essere un ottimo elemento complementare alle *document-based termbase*, specialmente per le lingue “pre-2004”.

Al fine di implementare queste nuove funzioni, è necessario dotare ogni Unità traduttiva di alcuni strumenti, ovvero *Synchroterm 2014* e *Glossary Converter*<sup>215</sup>.

Il primo, già in dotazione in DGTRAD, è tra i migliori software di estrazione terminologica nel panorama commerciale, mentre il secondo è uno strumento atto alla conversione di file in termbase compatibili con SDL Studio.

*Glossary Converter* è uno strumento che permette di convertire file in formato glossario, come ad esempio un semplice file Excel contenente due colonne con due lingue diverse, in un file SDLTB direttamente utilizzabile nel CAT tool.

Tramite questi due strumenti, la creazione di termbase sarebbe più facile e veloce.

Tuttavia, l'integrazione delle termbase manuali nei processi traduttivi cela una complessità maggiore rispetto a quelle viste nelle precedenti fasi.

Per le *normative termbase* e *full-IATE retrievals* è stato sufficiente creare un unico punto di accesso per ogni utente.

Le *document-based termbases* estratte con il *Term Recognition Module* sono, per natura, direttamente collegate al documento di origine e, dunque, una volta completata la traduzione, vengono rimosse automaticamente<sup>216</sup>.

Le *manual termbase*, invece, sarebbero molteplici, multilingui, divise per dominio e/o procedure e verrebbero usate sul medio/lungo termine e su diversi documenti e progetti. Si presenta quindi la necessità di preparare una nuova infrastruttura

---

<sup>215</sup> Il software è accessibile presso <http://appstore.sdl.com/app/glossary-converter/195/>.

<sup>216</sup> Più precisamente, alcuni algoritmi di manutenzione intervengono sui server di DGTRAD e rimuovono ogni progetto già finalizzato con SDL Studio, inclusi ogni file ad essi connessi, tra cui le termbase.

informatica che tenga in considerazione una manutenzione semi-automatica dei file e la loro gestione, autonoma, da parte di terminologi e traduttori.

È necessaria inoltre la creazione di un *workflow* per l'utilizzo di tali termbase, con il quale si considerino:

- il ruolo e le attività eseguite dai terminologi, ovvero la definizione delle fasi di estrazione dei termini, la creazione delle liste e la conseguente creazione delle termbase e con quali strumenti;
- l'allestimento di server specifici per immagazzinare le termbase e renderle operabili con i CAT Tool;
- la manutenzione dei server, includendo la valutazione di rischio e gestione dello spazio;
- l'utilizzo effettivo delle termbase da parte dei traduttori, in piena considerazione dell'esistente *workflow* traduttivo.

Al momento in cui si scrive, si sta lavorando alla creazione di tale nuovo *workflow*, che prenda in considerazione tutte le tipologie di termbase impiegabili nei processi traduttivi e le loro diverse finalità. Ad ogni modo, l'impiego di *manual termbases* rappresenta chiaramente un altro passo verso l'accrescimento della qualità dei documenti tradotti.

IATE, per il momento, non può essere onnicomprensivo e immediatamente aggiornato nella completa totalità delle sue risorse.

Tali termbase introdurranno per la prima volta nel *workflow* di traduzione l'estrazione sistematica e semi-automatica di terminologia dai documenti di origine e quindi

contribuiranno a sopperire alle carenze di IATE e all'impossibilità di essere aggiornato tempestivamente.

La terminologia in esse contenute può essere successivamente inserita in IATE (escludendo le espressioni e le fraseologie), migliorandone i tempi di aggiornamento e qualità.

## 5.2 Accordo di collaborazione tra il DSPSC<sup>217</sup> e Terminology Coordination Unit del Parlamento europeo

L'accordo di collaborazione tra il DSPSC e Terminology Coordination Unit nasce dall'assunto che IATE possa essere uno strumento per l'interoperabilità semantica oltre che risorsa traduttiva. Le schede che saranno sviluppate nella cornice di questo progetto conterranno dunque informazioni linguistico-terminologiche utili per l'interoperabilità semantica.

Come sostiene Daniela Vellutino (2016):

*«Le informazioni linguistico-terminologiche possono essere un valido strumento a supporto delle tecnologie del web semantico al servizio della modernizzazione delle amministrazioni pubbliche e dei rapporti interistituzionali e commerciali all'interno dell'Unione europea. Ad oggi, però, non sono ancora sapientemente utilizzate come*

---

<sup>217</sup> Dipartimento di Scienze Politiche Sociali e della Comunicazione dell'Università degli Studi di Salerno.

*metadati delle soluzioni per l'interoperabilità tra i sistemi informativi delle istituzioni ed organismi dell'Unione europea e delle amministrazioni pubbliche degli Stati membri»<sup>218</sup>.*

Si è già visto che:

*«l'obiettivo dell'interoperabilità è di permettere la presentazione e il trattamento delle informazioni in modo uniforme nei diversi sistemi gestionali, indipendentemente dalla tecnologia, dall'applicazione o dalla piattaforma utilizzate. La piena interoperabilità comprende la capacità di interoperare in termini di contenuto (semantico), formato (sintassi) e trasmissione. L'interoperabilità semantica vuol dire che il significato preciso dell'informazione scambiata è mantenuto e compreso correttamente e senza ambiguità, a prescindere dal modo in cui viene rappresentato fisicamente o trasmesso»<sup>219</sup>.*

In questo senso, il database terminologico IATE, che contiene il repertorio delle terminologie specialistiche dell'Unione europea, può essere impiegato per un nuovo utilizzo, ovvero non solo come risorsa traduttiva e di divulgazione dei termini, ma anche come sostiene ancora Daniela Vellutino (2016):

*«uno strumento a due vie per le soluzioni per l'interoperabilità dei sistemi informativi delle PA locali, nazionali e dell'Unione. I termini usati dagli applicativi per l'e-government nelle diverse lingue naturali potrebbero essere indicizzati, metadati e collegati alle schede terminologiche IATE che ne descrivono il significato e l'uso nei vari contesti istituzionali; le schede terminologiche di IATE*

---

<sup>218</sup> Vellutino D., Maslias R., Rossi F., (2016), *cit.* p. 221.

<sup>219</sup> Proposta di Direttiva del Parlamento europeo e del Consiglio relativa alla fatturazione elettronica negli appalti pubblici, COM(2013) 449 final 2013/0213.



*potrebbero a loro volta diventare un hub per collegare i termini ai documenti istituzionali in cui sono presenti. Le schede terminologiche IATE devono essere, però, di qualità: costantemente aggiornate, complete delle varianti dei termini presenti nei diversi documenti istituzionali, collegate alla documentazione unionale»<sup>220</sup>.*

Da queste riflessioni nasce lo studio per l'accordo di collaborazione su un dominio ristretto che riguarda un servizio pubblico: la "Gestione dei rifiuti urbani".

Il Dipartimento di Scienze Politiche, Sociali e della Comunicazione dell'Università degli Studi di Salerno ha sottoscritto un accordo di collaborazione<sup>221</sup> con Terminology Coordination Unit del Parlamento europeo, finalizzato alla redazione di schede terminologiche per la banca dati multilingue IATE.

L'obiettivo del progetto è sviluppare risorse linguistico-terminologiche nel dominio della "Gestione dei rifiuti urbani" da raccogliere nel "lessico MWM", una risorsa terminologica bilingue nella forma di *Linguistic Linked Open Data* condivisa e usata nel Web Semantico.

Sotto la direzione di Daniela Vellutino il gruppo di ricerca sta dunque sviluppando schede terminologiche per IATE che possano avere una specifica stringa nello schema di codifica della localizzazione informatica, *URI (Uniform Resource Identifier)*, che le identifichi come risorse aperte e riusabili in Rete.

---

<sup>220</sup> Vellutino D., Maslias R., Rossi F. (2016), *cit.* p.222.

<sup>221</sup> Responsabile scientifico del gruppo di ricerca è Daniela Vellutino, al gruppo di ricerca studenti di laurea magistrale e dottorato, tra cui Francesco Rossi, responsabile per i rapporti con l'Unità di Coordinamento Terminologico del Parlamento europeo.

Gli *URIs* delle schede terminologiche IATE possono essere annotati negli schemi di metadattazione del *Core Public Service Vocabulary*, ovvero un vocabolario controllato già sviluppato per questo servizio pubblico<sup>222</sup> nell'ambito del programma ISA<sup>2</sup>.

A descrivere le fasi del progetto, l'approccio teorico e gli sviluppi futuri è il documento "Unisa project: Long Description" (Vellutino, 2017), che spiega che:

*«Nelle prime fasi dello studio, i termini sono stati estratti da "CPI-GRU", un corpus che raccoglie informazioni in Italiano sul dominio "Gestione Rifiuti Urbani", sviluppato su "CPI", modello di classificazione dei testi per l'informazione e la comunicazione del settore pubblico (Vellutino 2012, 2016), e su "WMC", un glossario Inglese/Italiano/Inglese creato per essere impiegato come risorsa di traduzione (Formisano, 2015). Il corpus CPI-GRU include testi legali e amministrativi sulla gestione dei rifiuti redatti dall'Unione Europea e in Italia e anche testi per l'informazione e la comunicazione con il cittadino. Il corpus WMC è stato creato per essere rappresentativo della varietà dell'Inglese scritto nell'ambito di questo dominio. Include 147 testi raccolti da riviste accademiche e professionali, create da nativi e non nativi, tra il 2008 e il 2011»<sup>223</sup>.*

Il corpus CPI-GRU è composto da testi, in lingua italiana, di varie fonti istituzionali (unionali, nazionali, regionali) che hanno differenti finalità pragmatico-comunicative (ad esempio: testi legali, pubblicitari, amministrativi ecc.). Come illustra Vellutino (2015):

---

<sup>222</sup> Vassilios Peristeras, "Semantic Standards: Preventing Waste in the Information Industry", IEEE Intelligent Systems, vol.28, no. 4, pp. 72-75, July-Aug. 2013.

<sup>223</sup> Vellutino D., (2017), "UniSA project: Long Description", Accordo di collaborazione tra DSPSC e Terminology Coordination Unit del Parlamento europeo.

*«i documenti istituzionali sono stati selezionati seguendo il modello di classificazione dei testi per la comunicazione pubblica e l'informazione istituzionale – Modello CPI -, che individua i tipi di testo in relazione alle finalità della legge 150/00, che disciplina le attività di informazione e comunicazione della pubblica amministrazioni e individua le loro strutture operative»<sup>224</sup>.*

L'approccio in cui si inquadra il corpus si basa sulla *«traduzione intralinguistica come riformulazione, dunque, come attività d'interpretare i segni verbali per mezzo di altri segni della stessa lingua»<sup>225</sup>*. Lo studio ha quindi un orientamento variazionista che considera cruciale l'identificazione dei diversi termini, impiegati nei vari contesti pragmatico-comunicativi, che esprimono però lo stesso concetto.

Il corpus WMC, come spiega Vellutino (2017), è:

*«concepito per utenti con livelli diversi di competenza nel campo, con l'obiettivo di spiegare la terminologia della gestione dei rifiuti con particolare riferimento alle tecnologie della gestione dei rifiuti (quali ad esempio i trattamenti biologici) atte ad uno smaltimento appropriato. Il corpus è uno strumento utile per processi di traduzione e un supporto agli esperti delle materie ambientali e ingegneristiche»<sup>226</sup>.*

Dai corpora WMC e CPI-GRU sono stati estratti i termini per creare la risorsa "MWM Lexicon", composta da 791 termini. È stata verificata in IATE la presenza di tali termini

---

<sup>224</sup> Vellutino D., (2015), *“Risorse linguistiche e open data per la comunicazione pubblica della gestione dei rifiuti urbani”*, in Vellutino D., Zanola M.T., *Comunicare in Europa. Lessici istituzionali e terminologie specialistiche*, Educatt.

<sup>225</sup> Ead. (2015), *ibidem*.

<sup>226</sup> Vellutino D., (2017), *cit.*

con l'obiettivo di controllare lo stato di aggiornamento dei riferimenti ai documenti unionali e italiani delle schede del database.

Il progetto si svilupperà successivamente in due fasi:

1. la verifica dei campi relativi ai riferimenti alle risorse documentarie, in collaborazione con esperti di dominio;
2. lo sviluppo di schede non presenti in IATE, da presentare al Dipartimento Italiano di Traduzione della Commissione, che presentano campi utili per l'interoperabilità semantica.

Come illustra Vellutino (2017):

*«collaboreremo con esperti di dominio da ARPAT (Agenzia regionale per la protezione ambientale della Toscana) e l'associazione di promozione sociale e innovazione "CittadiniReattivi", un progetto di civic journalism e sito di crowdmapping su ambiente salute e legalità, che indaga sui siti inquinati in Italia e mappa le buone pratiche dei cittadini. Questi esperti verificheranno i campi seguenti nelle schede di IATE: "Term Reference", "Definition", "Definition Reference", "Context", "Context Reference".*

*Svilupperemo poi schede di IATE di termini non presenti nel database e li sottoporremo al Dipartimento Italiano di Traduzione della Commissione. Inseriremo, inoltre, nelle nostre schede, due nuovi campi da inserire nel template delle schede di IATE: "Reference to the Institutional Source" e "Time Validation". Questi campi incoraggiano l'interoperabilità di IATE perché connettono le schede alle risorse documentarie dell'UE e degli Stati Membri che trasformano le*

*direttive europee in legge nazionale. Questi campi saranno inclusi nelle schede consegnate a TermCoord sui termini non presenti in IATE e che conterranno termini con link alle risorse documentarie nazionali»<sup>227</sup>.*

### 5.3 URI ed ELI in IATE, un primo passo per l'interoperabilità semantica

Al fine di far diventare la banca dati terminologica multilingue IATE centrale nel “Web dei dati” in quanto strumento per il collegamento tra *linked data* delle risorse documentarie di *Eur-Lex* e le risorse terminologiche multilingui in essa contenute, il gruppo di ricerca sta sviluppando schede terminologiche in cui è annotato il riferimento ai testi legali dell'UE.

Come sostiene Daniela Vellutino questo è

*«già un primo facile passo verso l'interoperabilità semantica di IATE perché i testi legali in Eur-Lex hanno già una specifica URI, che fa da “collegamento” (linked) ad altri dati (data). Inoltre, molti atti legislativi e amministrativi UE hanno anche un altro tipo di stringa di identificazione denominata European Legislation Identifier (Identificatore della Legislazione Europea) ELI, che permette di associare ad ogni provvedimento legislativo ed amministrativo una URI in maniera univoca»<sup>228</sup>.*

---

<sup>227</sup> Ead. (2017), *ibidem*.

<sup>228</sup> Vellutino D., Maslias R., Rossi F. (2016), *cit.* p. 226-227.

ELI è un sistema comune di identificazione della legislazione con relativi metadati e si applica alla legislazione dell'Unione europea riportata nella Gazzetta ufficiale e nel portale EUR-Lex, gestiti dall'Ufficio delle Pubblicazioni. ELI rende disponibile online la legislazione in un formato standardizzato in modo che sia accessibile, interscambiabile e riutilizzabile ovunque e comprende specifiche tecniche quali: identificatori web (URI) per le informazioni giuridiche, metadati che specificano come descrivere le informazioni giuridiche e un linguaggio specifico per scambiare legislazioni in formato a lettura automatica.

La questione dell'identificazione univoca delle risorse è un punto centrale nella rappresentazione della conoscenza nel contesto del web semantico ed ha ovviamente un ruolo importante nell'identificazione della legislazione. In un documento del 2012 il Consiglio sostiene che:

*«per l'identificazione della legislazione dovrebbe essere usato un identificatore unico che possa essere riconosciuto, letto e compreso sia dalle persone che dai computer e che sia compatibile con gli standard tecnologici esistenti. ELI propone un set di elementi di metadati per descrivere la legislazione nel rispetto di un'ontologia raccomandata. [...] Avvalendosi dell'architettura emergente del web semantico, che consente il trattamento diretto delle informazioni sia da parte delle persone che dei computer, ELI permetterebbe un maggiore e più*

*rapido scambio di dati rendendo possibile uno scambio di informazioni automatico ed efficiente»<sup>229</sup>.*

L'Ufficio di Pubblicazioni ha poi implementa l'ELI in tre fasi: nel 2015 ha attribuito gli URI<sup>230</sup> a tutti gli atti della Gazzetta Ufficiale, serie L, ovvero per i testi di Direttive, Regolamenti, Decisioni, accordi internazionali, opinioni e raccomandazioni, Corrigenda, Trattati e legislazione consolidata; nella seconda fase, i metadati<sup>231</sup> di ELI sono stati applicati a tali atti giuridici; nella fase finale, completata a marzo 2016, gli URI HTTP e i metadati di ELI a lettura ottica sono stati pubblicati mediante il portale EUR-Lex<sup>232</sup>.

ELI utilizza gli URI HTTP per identificare specificamente tutte le informazioni giuridiche online pubblicate ufficialmente in tutta Europa. Tali URI sono descritti in maniera formale mediante *template* URI leggibili elettronicamente, che utilizzano componenti che comprendono dati semantici sia dal punto di vista giuridico che da quello dell'utente finale.

*«Un URI strutturato può già identificare atti utilizzando una serie di componenti predefiniti ma l'attribuzione di metadati supplementari stabiliti nel contesto di una sintassi condivisa porrà le basi per promuovere lo scambio ed accrescere l'interoperabilità tra i diversi sistemi di informatica giuridica [...]. Gli Stati membri*

---

<sup>229</sup> Conclusioni del Consiglio che invitano all'introduzione dell'identificatore della legislazione europea (ELI) del 26 ottobre 2012 (2012/C 325/02).

<sup>230</sup> Per una descrizione formale degli URI consultare [http://publications.europa.eu/mdr/eli/documentation/uri\\_templates.html](http://publications.europa.eu/mdr/eli/documentation/uri_templates.html).

<sup>231</sup> una descrizione formale dei metadati e dell'ontologia di ELI è consultabile presso [http://publications.europa.eu/mdr/eli/documentation/ELI\\_Ontology-v1.0.htm](http://publications.europa.eu/mdr/eli/documentation/ELI_Ontology-v1.0.htm).

<sup>232</sup> v. "Notizie sull'ELI, l'Ufficio delle Pubblicazioni dell'UE migliora la pubblicazione online della legislazione", link: [http://eur-lex.europa.eu/eli-register/news\\_item\\_7.html?locale=it](http://eur-lex.europa.eu/eli-register/news_item_7.html?locale=it) (data consultazione 20 giugno 2017).

*sono liberi di utilizzare il loro proprio sistema di metadati ma sono incoraggiati a seguire e utilizzare gli standard di metadati di ELI con tabelle di autorità condivise ma estensibili»<sup>233</sup>.*

Come sostiene Daniela Vellutino (2016), adattando ciò al contesto di IATE, è possibile sviluppare un sistema di metadattazione specifico per la rappresentazione delle risorse terminologiche in combinazione con gli standard di metadati di ELI. In conformità alla raccomandazione del W3C “RDFa in XHTML: Syntax and Processing”<sup>234</sup>, il sistema permetterebbe di rispondere alle esigenze di metadattazione specifiche di IATE e ne accrescerebbe l’interoperabilità, favorendo il riutilizzo delle informazioni in maniera *open e linked*.

#### 5.4. I termini di MWM Lexicon in IATE

Nel dominio “Environment/Ambiente<sup>235</sup>” (codice 52) di IATE sono contenuti, al momento in cui si scrive, 20.900 termini, di cui 3145 del Consiglio, 5671 della

---

<sup>233</sup> Conclusioni del Consiglio del 26 ottobre 2012 (2012 /C 325/02). *Cit.*

<sup>234</sup> Si rimanda al documento completo presso <https://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>.

<sup>235</sup> Si utilizzerà, per questo paragrafo, sia il nome del dominio inglese che quello italiano, con la struttura “EN/IT”.



Commissione, 5360 del Centro di Traduzione, 8 di COR/EESC<sup>236</sup>, 22 di EFSA<sup>237</sup>, 982 di Com Swiss Data<sup>238</sup> e 8 di ECHA<sup>239</sup>.

Questo dominio ha tre sottodomini, o, utilizzando la terminologia di EuroVoc<sup>240</sup>, il tesaurus sul quale si basano i domini del database interistituzionale, ha tre microtesauri:

- “Environmental policy/Politica dell’ambiente” (codice 5206), che contiene 250 termini, di cui 124 del Consiglio, 83 della Commissione, 28 del Parlamento europeo, 8 della Banca di Investimenti Europea, 4 del Centro di Traduzione, 2 di COR/EESC, 1 di ECHA;
- “Natural environment/Ambiente naturale” (codice 5211) che contiene 216 termini, di cui 48 del Consiglio, 136 della Commissione, 27 del Parlamento europeo, 2 di COR/EESC, 1 di EFSA e 1 di Com Swiss Data;
- “Deterioration of the environment/Degrado ambientale” (codice 5216) che contiene 83 termini, di cui 29 del Consiglio, 48 della Commissione, 2 del Parlamento europeo, 2 del Centro di Traduzione, 1 di ECHA, 1 di EFSA.

---

<sup>236</sup> COR/EESC sta per “Committee of Regions/European Economic and Social Committee”. In IATE le due istituzioni sono classificate sotto la stessa etichettatura.

<sup>237</sup> European Food Safety Authority (Autorità europea per la sicurezza alimentare).

<sup>238</sup> “Com Swiss Data” non è un’istituzione o un’agenzia in senso stretto ma l’etichetta data alla collezione di termini creata dalla Sezione di terminologia della Cancelleria federale svizzera, che gestisce TERMDAT, la banca dati terminologica dell’Amministrazione federale svizzera, e caricata in IATE tramite un accordo con la Commissione.

<sup>239</sup> European CHemicals Agency (Agenzia europea delle sostanze chimiche).

<sup>240</sup> EuroVoc è un thesaurus multilingue e pluridisciplinare che comprende la terminologia dei settori d’attività dell’Unione europea. Contiene termini in 23 lingue dell’UE (bulgaro, ceco, croato, danese, estone, finlandese, francese, greco, inglese, italiano, lettone, lituano, maltese, olandese, polacco, portoghese, rumeno, slovacco, sloveno, spagnolo, svedese, tedesco e ungherese), oltre che in tre lingue di paesi candidati all’adesione: македонски (mk), shqip (sq) e српски (sr).

EuroVoc è curato dall’Ufficio delle pubblicazioni, che è passato ad una gestione del thesaurus basata sull’ontologia e alle tecnologie del web semantico, in linea con le raccomandazioni del W3C e con gli ultimi sviluppi negli standard di classificazione.

“Environmental policy/Politica dell’ambiente” ha sei microtesauri:

- “Environmental policy/Politica ambientale” (codice 5206001) che contiene 175 termini, di cui 49 del Consiglio, 51 della Commissione, 14 del Parlamento europeo, 3 del Centro di Traduzione, 58 di COR/EESC;
- “Environmental protection/Protezione dell’ambiente” (codice 5206002), che contiene 164 termini, di cui 77 del Consiglio, 63 della Commissione, 17 del Parlamento europeo, 2 del Centro di Traduzione, 3 di COR/EESC, 2 di EFSA;
- “Pollution control measures/Lotta contro l’inquinamento” (codice 5206003), che contiene 208 termini, di cui 54 del Consiglio, 134 della Commissione, 14 del Parlamento europeo, 1 della Banca di Investimenti Europea, 3 del Centro di Traduzione, 4 di COR/EESC;
- “Waste management/Gestione dei rifiuti” (codice 5206004) che contiene 151 termini, di cui 16 del Consiglio, 99 della Commissione, 29 del Parlamento europeo, 3 del Centro di Traduzione, 4 di COR/EESC;
- “Water management/Politica in materia di acque” (codice 5206005) che contiene 33 termini, di cui 5 del Consiglio, 23 della Commissione, 1 del Parlamento europeo, 1 del Centro di Traduzione, 1 di COR/EESC;
- “Climate change policy/Politica in materia di cambiamenti climatici” (codice 5206006). che contiene 98 termini, di cui 55 del Consiglio, 35 della Commissione, 3 del Parlamento europeo, 4 di COR/EESC, 1 di EFSA.

Il dominio “Natural environment/Ambiente naturale” (codice 5211) ha cinque microtesauri:

- “Climate/Clima” (codice 5211001), che contiene 203 termini, di cui 49 del Consiglio, 144 della Commissione, 4 del Parlamento, 1 del Centro di Traduzione, 1 di COR/EESC, 4 Com Swiss Data;
- “Geophysical environment/Ambiente geofisico” (codice 5211002), che contiene 70 termini, di cui 7 del Consiglio, 61 della Commissione e 2 del Centro di Traduzione;
- “Natural resources/Risorse naturali” (codice 5211003), che contiene 741 termini, di cui 652 del Consiglio, 61 della Commissione, 14 del Parlamento europeo, 12 di COR/EESC; 1 di ECHA e 1 di EFSA;
- “Physical environment/Ambiente fisico” (codice 5211004), che contiene 103 termini, di cui 15 del Consiglio, 79 della Commissione, 5 del Parlamento europeo, 1 del Centro di Traduzione, 2 di COR/EESC, 1 di EFSA;
- “Wildlife/Vita selvatica” (codice 5211005) che contiene 11041 termini, di cui 245 del Consiglio, 3660 della Commissione, 6733 del Parlamento europeo, 7 del Centro di Traduzione, 395 di Com Swiss Data e 1 di EFSA;

Il dominio “Deterioration of the environment/Degrado ambientale” (codice 5216), ha quattro microtesauri:

- “Degradation of the environment/degradazione dell’ambiente” (codice 5216001), che contiene 88 termini, di cui 16 del Consiglio, 59 della Commissione, 7 del Parlamento europeo, 3 del Centro di Traduzione, 3 di COR/EESC;
- “Nuisance/inquinamento” (codice 5216002), che contiene 72 termini, di cui 10 del Consiglio, 56 della Commissione, 4 del Parlamento europeo, 2 di COR/EESC;

- “Pollution/nocività” (codice 5216003) che contiene 211 termini, di cui 50 del Consiglio, 131 della Commissione, 19 del Parlamento europeo, 5 del Centro di Traduzione, 4 di COR/EESC, 2 di EFSA;
- “Waste/rifiuti” (codice 5216004) che contiene 210 termini, di cui 15 del Consiglio, 185 della Commissione, 5 del Parlamento europeo, 4 del Centro di Traduzione, 1 di COR/EESC.

La tabella riassuntiva che segue mostra il numero di termini italiani divisi per dominio in base alla fonte di provenienza. Successivamente, una rappresentazione grafica della gerarchia del dominio “Ambiente”.

Tab. 5.1 – Numero di termini italiani di IATE divisi per dominio e fonte di provenienza<sup>241</sup>

Istituzione Dominio	COU	COM	EP	EIB	CdT	COR / EESC	EFSA	COM Swiss Data	ECHA	Tot. termini per dominio
<b>Ambiente</b>	3145	5671	5704	0	5360	8	22	982	8	20900
<i>Environmental policy</i>	124	83	28	8	4	2	0	0	1	250
Environmental policy (5206001)	49	51	14	0	3	58	0	0	0	175
Environmental protection	77	63	17	0	2	3	2	0	0	164
Pollution control measures	54	134	14	1	2	3	0	0	0	208
Waste management	16	99	29	0	3	4	0	0	0	151
Water management	5	23	1	0	1	3	0	0	0	33
Climate change policy	55	35	3	0	0	4	1	0	0	98
<i>Natural environment</i>	48	136	27	0	0	2	2	1	0	216
Climate	49	144	4	0	1	1	0	4	0	203
Geophysical environment	7	61	0	0	2	0	0	0	0	70
Natural resources	652	61	14	0	0	12	0	0	1	740
Physical environment	15	79	5	0	1	2	1	0	0	103
Wildlife	245	3660	6733	0	7	0	1	395	0	11041
<i>Deterioration of the environment</i>	29	48	2	0	2	0	1	0	1	83
Degradation of the environment	16	59	7	0	3	3	0	0	0	88
Nuisance	10	56	4	0	0	2	0	0	0	72
Pollution	50	131	19	0	5	4	2	0	0	211
Waste	15	185	5	0	4	1	0	0	0	210
<b>Tot. Termini per istituzione</b>	<b>4661</b>	<b>10779</b>	<b>12630</b>	<b>9</b>	<b>5400</b>	<b>112</b>	<b>32</b>	<b>1382</b>	<b>11</b>	<b>35016</b>

<sup>241</sup> Nella colonna “dominio” il macrodominio “Ambiente” è marcato in grassetto, i suoi tre sottodomini sono marcati in corsivo e i sottodomini di terzo livello hanno formattazione semplice. Si noti che tra “Ambiente” e i suoi sottodomini e tra questi e i sottodomini di terzo livello non sussiste una relazione di inclusione vera e propria a livello terminologico. Ad esempio “Natural environment” ha 216 termini mentre il suo sottodominio “Wildlife” ne ha 11041.

Nella colonna delle istituzioni gli acronimi e le abbreviazioni per esteso sono: COU (Council), COM (Commission), EP (European Parliament), EIB (European Investment Bank), CdT (Centre de Traduction), COR/EESC (Committee of Regions/European Economic and Social Committee), EFSA (European Food Safety Authority), COM Swiss Data (v. nota 238), ECHA (European Chemicals Agency). L’ordine delle colonne delle fonti segue il sistema di identificazione delle istituzioni di IATE: Council (1), Commission (2), European Parliament (3) ecc.

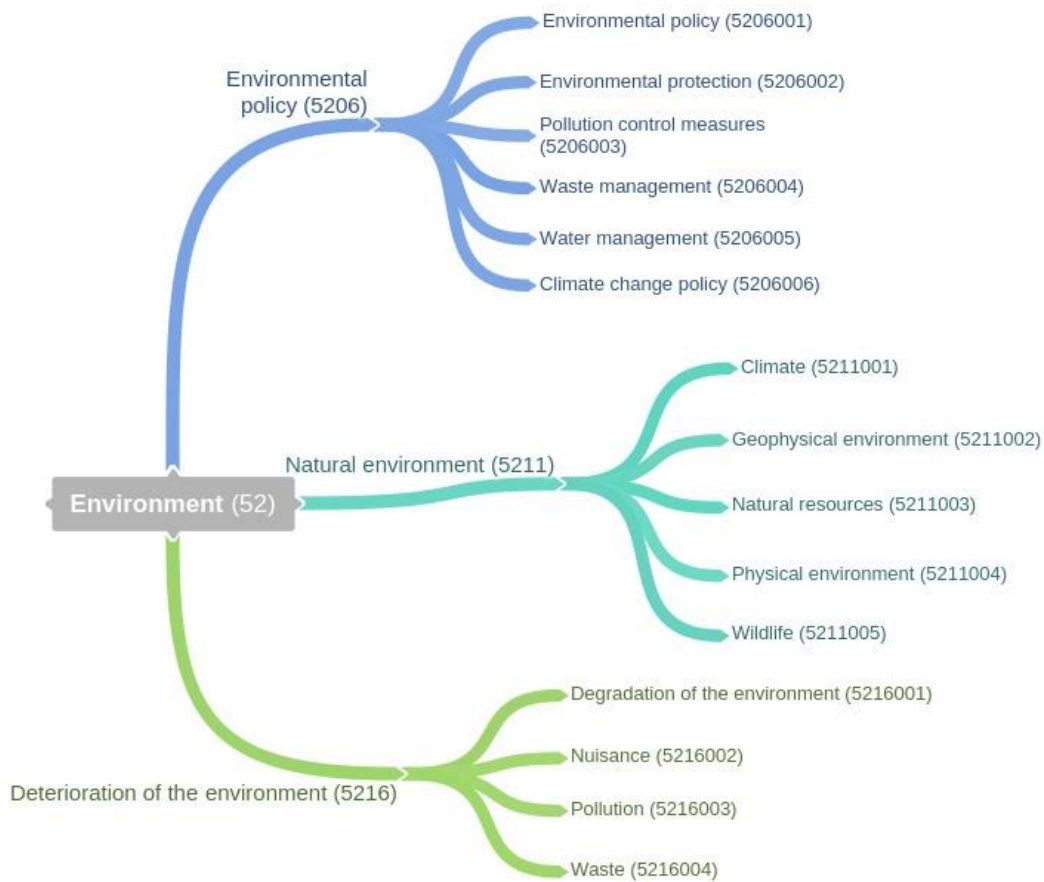


Fig. 5.5 Rappresentazione grafica del dominio “Environment/Ambiente” di IATE<sup>242</sup>

Per le finalità dello studio, il dominio “Natural environment/Ambiente naturale” è poco attinente, mentre “Environmental policy/Politica dell’ambiente” e “Deterioration of the environment/Degrado ambientale” sono più rilevanti. Questi domini, contengono rispettivamente i microtesauri “Waste management” e “Waste”.

<sup>242</sup> L’immagine è stata creata con il tool online coggle. <https://coggle.it/>.

Ciò porta a fare una preliminare considerazione sull'organizzazione e sulla granularità dei domini. Come visto, IATE si basa sui domini di EuroVoc, che è stato concepito per rispondere ai bisogni di sistemi documentari generali inerenti alle attività dell'Unione europea, ma non è utile per l'indicizzazione e la ricerca di documenti specializzati. Conseguentemente, tale strutturazione non risulta perfetta per un database di terminologia istituzionale specialistica perché contiene, inevitabilmente, imprecisioni e approssimazioni già al livello indipendente dalla lingua.

“Waste/Rifiuti” è microtesauro di “Deterioration of the environment/Degrado ambientale” mentre “Waste management/Gestione dei rifiuti” è microtesauro di “Environmental policy/Politica dell'ambiente”, mentre, invece, ci si aspetterebbe che tra loro ci fosse una relazione gerarchica e che fossero contenuti nello stesso dominio anziché essere separati e appartenenti a due microtesauri diversi.

Un altro caso piuttosto particolare è il dominio “Environmental policy” (codice 5206), che ha sei microtesauri, di cui uno omonimo, con l'unica differenza del codice di terzo livello a 7 cifre, ovvero “Environmental policy” (5206001). È singolare, dunque, che esistano due domini aventi lo stesso nome e che siano anche in relazione gerarchica.

Nella versione in lingua italiana di EuroVoc, però, si nota una differenza sintattica: il dominio di livello superiore “environmental policy” (5206) in italiano è tradotto “politica dell'ambiente”, mentre il microtesauro di livello inferiore “environmental policy” (5206001) è tradotto “politica ambientale”. La traduzione con l'uso del nesso di specificazione o quella con l'aggettivizzazione, tuttavia, non riduce l'ambiguità semantica.

Similmente, “Deterioration of the environment” ha come microtesauro “Degradation of the environment”. In EuroVoc, in italiano sono tradotti con “Degrado ambientale” e “degradazione dell’ambiente”. Alla ricerca di “degrado dell’ambiente” in IATE, si ricevono come risultato due schede, di cui una, redatta dalla Commissione, propone anche la variante “deterioramento dell’ambiente”<sup>243</sup>. Ciò sottolinea che, oltre alla poca chiarezza della definizione dei domini, in IATE è visibile poca precisione nella definizione dei termini che costituiscono la propria struttura di domini.

Ad ogni modo, nello studio condotto da Daniela Vellutino (2015) sono state ricercate in IATE 377 polirematiche e 102 parole semplici dal corpus “CPI-GRU”. Per le polirematiche, 210 sono risultate assenti e 167 già contenute in IATE. Di queste 167, però, non tutte sono rilevanti per il dominio individuato.

Nel corpus, ad esempio, è presente “tariffa del servizio”, che però in IATE è presente solo nel dominio “Comunicazione” e che si riferisce quindi a un concetto diverso.

Negli altri casi, ovvero quelli in cui i termini contenuti in IATE sono relativi al dominio di indagine, questi presentano principalmente “Environment/Ambiente” come dominio. Più precisamente, sono 112 i termini marcati con tale dominio. In IATE, è possibile creare schede relative ad uno stesso termine ma marcate con domini diversi. È il caso di quei termini che rappresentano diversi concetti e per i quali è quindi necessario creare schede separate, come ad esempio il termine “capacità nominale” che in IATE ha entrate sia nel dominio della metallurgia, dell’industria e dell’ambiente.

---

<sup>243</sup> Link scheda <http://iate.europa.eu/FindTermsByLilId.do?lilId=1407755&langId=it>.



Ci sono altri casi, invece, in cui un termine rappresenta un solo concetto ma è comunque presente in più di una scheda. Si prenda, come esempio, il termine “dispositivo di depurazione”.

AMBIENTE [EP]		<a href="#">Voce completa</a>
IT	<a href="#">dispositivo di depurazione</a>	★★★★ →@
DA	rensningsanordning	★★★★ →@
DE	Reinigungseinrichtung	★★★★ →@
	Reinigungsvorrichtung	★★★★ →@
EL	σύστημα καθαρισμού	★★★★ →@
EN	purification device	★★★★ →@
ES	dipositivo de depuración	★★★★ →@
	dispositivo de purificación	★★★★ →@
FR	système d'épuration	★★★★ →@
NL	rookgasreinigingsinstallatie	★★★★ →@

Ricerca, AMBIENTE [Council]		<a href="#">Voce completa</a>
IT	<a href="#">dispositivo di depurazione</a>	★★★★ →@
DE	Reinigungsvorrichtung	★★★★
EN	purification device	★★★★
FR	dispositif d'épuration	★★★★
PT	dispositivo de depuração	★★★★

Fig. 5.6 Risultato di ricerca per “dispositivo di depurazione”<sup>244</sup>

In Figura 5.6 si vede che il motore di ricerca di IATE propone due risultati rilevanti. Uno è prodotto dal Parlamento europeo (EP) e l'altro è prodotto dal Consiglio (Council).

Nel primo caso, la scheda prodotta dal Parlamento è stata marcata con il dominio “Ambiente”, mentre la scheda prodotta dal Consiglio è stata marcata sia con “Ricerca” che con “Ambiente”. Ciò mostra diverse particolarità: anzitutto è possibile scegliere, per una stessa scheda, più di un dominio (ed è il caso della scheda del Consiglio); inoltre, è

<sup>244</sup> Fonte immagine: <http://iate.europa.eu/SearchByQuery.do>, data ultima consultazione: 25 maggio 2017.

visibile che si tratta di un caso di duplicato, ovvero dello stesso concetto contenuto in due schede separate che sono state marcate incoerentemente con due domini diversi.

Per le 167 polirematiche presenti in IATE, solo 20 sono marcate con dominio “Waste/Rifiuti” e 16 con dominio “Waste management/Gestione dei rifiuti”. Molti altri termini, tuttavia, potrebbero essere parte di questi due domini, come ad esempio “residuo di lavorazione” (con dominio “Ambiente” e “Agricoltura, Silvicultura e Pesca”) o, in un caso decisamente più lampante, “rifiuto inerte”, marcato solo con il dominio “Ambiente”. Anche in questo caso la granularità dei domini è determinata da diversi fattori. Si è già visto che la strutturazione dei domini di IATE è fondata su quelli di EuroVoc, non sempre adatti alle finalità della rappresentazione della conoscenza terminologica.

Inoltre, non esistono direttive o buone pratiche rigide riguardo alla scelta dei domini delle schede: *IATE Handbook* dà istruzioni pratiche su come scegliere un dominio dall’interfaccia di creazione della scheda ma non offre disposizioni e regole dettagliate.

Il documento “IATE Best Practice”<sup>245</sup>, invece, suggerisce di scegliere un numero limitato di domini e che questi debbano chiaramente identificare il linguaggio speciale e il contesto in cui il concetto viene impiegato:

- *«The Domain should be chosen to identify clearly the special language and context in which the concept is used. Each entry should have only a limited*

---

<sup>245</sup> *IATE Best Practice* è un documento interistituzionale che definisce i principi e le regole generali sulla gestione delle entrate e del lavoro terminologico, un aspetto che ha in comune con il documento *IATE Handbook*. Tuttavia se ne differenzia perché non contiene guide e istruzioni pratiche per l’utilizzo del database.

*number of domains; the presence of several domains may indicate that the term actually covers more than one concept.*

- *Terminologists adding their language to an existing entry must check whether their term is used in the same context as that indicated by the domain names already included on the entry, and must avoid adding new domain names which would result in one entry covering more than one concept.*
- *In computer-assisted translation systems the use of domain names will be even more important. It is therefore vital for anyone creating entries in IATE to be reasonably familiar with the names used in the database and to add them to their entries as rigorously and precisely as possible»<sup>246</sup>.*

È però evidente, purtroppo, che la scelta del dominio non avviene nella maniera più precisa e adeguata possibile, permettendo dunque la realizzazione di fenomeni come quelli presentati sopra, in cui sarebbe legittimo aspettarsi una marcatura dei domini più accurata. Su questa scia, il resto delle polirematiche analizzate, che non sono marcate né da dominio “Ambiente” né da “Rifiuti” o “Gestione dei rifiuti”, sono rappresentate con domini piuttosto diversi, che variano da “Unione Europea”, a “Agricoltura”, “Tecnologia dei materiali”, “Industria meccanica”, “Chimica” ecc.

In considerazione di tutti i domini e i termini, anche duplicati, rilevanti per questo studio, 97 schede sono state prodotte dalla Commissione, 56 dal Parlamento, 52 dal Centro di Traduzione, 49 dal Consiglio e 8 dal Comitato delle Regioni.

---

<sup>246</sup> IATE Best Practice, *cit.*

Per quanto riguarda le 102 parole semplici ricercate in IATE, 88 sono presenti e 12 non sono presenti in IATE. Come per le polirematiche, l'accuratezza dei domini è piuttosto variabile. Inoltre, trattandosi di parole semplici, si prestano potenzialmente ad un numero maggiore di accezioni concettuali.

Le parole semplici assenti in IATE sono, conseguentemente, in minor numero ma aumentano in proporzione i concetti a cui i termini possono riferirsi e dunque anche i domini.

Il dominio "Ambiente" viene utilizzato 35 volte, mentre "Gestione dei rifiuti" viene impiegato 6 volte e "Rifiuti" solo due volte.

Particolarmente significativi sono i casi dei termini "rifiuto" e "rifiuti". Il termine "rifiuto" compare in 10 schede, marcate rispettivamente con i domini "Chimica", "Comunicazione<sup>247</sup>", "Edilizia e lavori pubblici, Trasporto terrestre, Trasporto<sup>248</sup>", "Scienze mediche", "Composto chimico", "Proprietà intellettuale, Assicurazioni<sup>249</sup>", "Diritto", "Assicurazioni", "Diritto dell'UE, Assicurazioni<sup>250</sup>".

È naturale pensare che il termine "rifiuto" rimandi a concetti diversi, per cui è legittimo aspettarsi un certo numero di schede; eppure, nessuna di queste schede mostra il termine associato al dominio "Ambiente" o i suoi sottodomini, quale ad esempio "Rifiuti".

---

<sup>247</sup> Compiono due schede "rifiuto" con dominio "Comunicazione". È un caso di duplicato "puro".

<sup>248</sup> Per questa scheda, prodotta dalla Commissione, sono stati impiegati questi tre domini.

<sup>249</sup> Anche in questo caso sono stati impiegati due domini per la stessa scheda.

<sup>250</sup> Idem come sopra.

Ricercando il termine “rifiuti” in IATE si ottengono quattro schede, di cui una, creata dalla Commissione, che è marcata con dominio “Industria nucleare” e tre, create da Commissione, Parlamento e Com Swiss Data, che sono marcate con “Ambiente”. Nessuna di queste schede è marcata con dominio “Rifiuti”.

Va notato che la Direttiva n.2008/98/CE<sup>251</sup>, ovvero il documento che contiene la definizione di “rifiuto” e che stabilisce misure volte a proteggere l’ambiente e la salute umana, cita che “si intende per «rifiuto» qualsiasi sostanza od oggetto di cui il detentore si disfi o abbia l’intenzione o l’obbligo di disfarsi”. Nella definizione, quindi, si impiega il termine alla forma singolare, che compare un totale di 34 volte nel documento, mentre la forma plurale è presente 382 volte. È singolare verificare che la definizione di “rifiuto”, così come espresso dalla Direttiva, è espressa in un’unica scheda marcata con dominio “Chimica”.

Un altro esempio è il termine “scoria”. Sono presenti due risultati in IATE, nei domini “Elettronica ed elettrotecnica, Chimica” e “Industria metallurgica”, mentre “scorie” compare più spesso, 5 volte, con termini marcati nei domini “Industria carboniera, Industria del legno, Rifiuti”, “Ambiente”, “Industria metallurgica, Rifiuti”, “Ambiente, Metallurgia e siderurgia” e “dominio non relativo all’oggetto”<sup>252</sup>.

---

<sup>251</sup> Direttiva 2008/98/CE del Parlamento europeo e del Consiglio del 19 novembre 2008 relativa ai rifiuti e che abroga alcune direttive. Fonte: <http://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32008L0098&from=IT>.

<sup>252</sup> “Dominio non relativo all’oggetto” è il valore che viene attribuito ad un termine quando non ne viene indicato un dominio in IATE.

Ad ogni modo, in considerazione di tutti i domini e i termini, anche duplicati, rilevanti per questo studio, 46 schede sono state prodotte dalla Commissione, 22 dal Centro di Traduzione, 12 dal Consiglio e 8 dal Parlamento.

Sia per le polirematiche che per le parole semplici si nota una generale mancanza di definizione e/o di riferimento al contesto, ma ciò è dovuto dalla non obbligatorietà dei campi.

Più in generale, la gestione delle fonti e dei contesti della terminologia non ha vera e propria armonizzazione: si citano fonti istituzionali con collegamenti ipertestuali e CELEX ma anche fonti secondarie e poco dettagliate. Una caratteristica, questa, comune tuttavia all'intero database terminologico IATE e decisamente più frequente in una lingua come l'italiano.

Per quanto riguarda le fonti, sono anch'esse piuttosto varie: molte sono direttive europee, come la 2008/98/CE sopra citata, ma sono anche presenti fonti di altro tipo, quali articoli di workshop, glossari multilingui istituzionali, pubblicazioni, dizionari, esperti di settore e Wikipedia.

## 5.5 Da una scheda IATE di qualità ad una proposta per l'interoperabilità semantica

Le schede di IATE analizzate per questo studio sono spesso poco aggiornate e mostrano fonti di provenienza piuttosto eterogenea e in generale manca spesso il riferimento alla

normativa vigente. Per il dominio specifico dei rifiuti, inoltre, l'allegato alla Decisione della Commissione del 18 dicembre 2014 (2014/955/UE)<sup>253</sup>, stabilisce un "elenco dei rifiuti", più noto come "Catalogo europeo dei rifiuti" (CER), che, come definisce la Decisione della Commissione del 20 dicembre 1993 (94/3/CE)<sup>254</sup> è «*un elenco armonizzato, non esaustivo, di rifiuti [...]. Il catalogo vuole essere una nomenclatura di riferimento con una terminologia comune per tutta la Comunità allo scopo di migliorare tutte le attività connesse alla gestione dei rifiuti*». La Decisione (2014/955/UE) non contiene una definizione esplicita dell'elenco, ma specifica che i rifiuti in esso contenuti «*sono definiti specificatamente mediante il codice a sei cifre per ogni singolo rifiuto e i corrispondenti codici a quattro e a due cifre per i rispettivi capitoli*»<sup>255</sup>.

Ogni rifiuto classificato secondo l'elenco della Decisione deve essere dunque etichettato con un codice a sei cifre. Tali codici si presentano nella forma 'XX YY ZZ', dove la prima coppia si riferisce ad un capitolo fondamentale dell'elenco e le successive due alle sue sottocategorie. Ad esempio, i rifiuti da estrazione di minerali metalliferi vengono identificati con il codice '01 01 01', di cui la prima coppia equivale a "Rifiuti derivanti da prospezione, estrazione da miniera o cava, nonché dal trattamento fisico o chimico di

---

<sup>253</sup> Il testo integrale è consultabile presso <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32014D0955&from=IT>

<sup>254</sup> Il testo integrale è consultabile presso <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:31994D0003&from=IT>. Tale decisione, che menziona per la prima volta il termine "Catalogo europeo dei rifiuti", è stata poi sostituita dalla Decisione (2000/532/CE), disponibile presso <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32000D0532&from=IT>. Quest'ultima, è stata a sua volta modificata dalla Decisione della Commissione del 18 dicembre 2014 (2014/955/UE), v. *supra*, che sopprime gli articoli 2 e 3 della (2000/532/CE) e ne sostituisce l'allegato. Il termine "catalogo europeo dei rifiuti" non viene più menzionato e viene sostituito da "elenco di rifiuti". È convenzione però riferirsi al Catalogo e naturalmente ai codici CER nei documenti specialistici e piattaforme di settore.

<sup>255</sup> Allegato alla Decisione (2014/955/UE), *cit.*

minerali”, la seconda coppia a “rifiuti da estrazione di minerali” e la terza a “rifiuti da estrazione di minerali metalliferi”.

Tale sistema di catalogazione può essere uno strumento utile per la metadattazione delle schede di IATE relative ai rifiuti. Vellutino, Maslias e Rossi (2016), propongono tale sviluppo sulla base di un esempio pratico, ovvero dalla modifica della scheda del termine ‘rifiuto ingombrante’. Partendo dalla scheda del termine attualmente presente in IATE

<b>Domain</b>	ENVIRONMENT	it
<b>Term</b>	<b>rifiuto ingombrante</b>	
<i>Reliability</i>	3 (Reliable)	
<i>Term Ref.</i>	Institution: EEA; Title: GEMET - GEneral Multilingual Environmental Thesaurus; Publication Year: 1999; Publication Month: August; Volume No: 5;	
<i>Date</i>	03/07/2002	

Fig. 5.7 L’attuale scheda terminologica di ‘rifiuto ingombrante’<sup>256</sup>

gli autori ne propongono una versione aggiornata, mostrata in figura 5.8, e spiegano che «il termine ‘rifiuto ingombrante’ è presente in IATE con entrata canonica. Sono assenti, però, i riferimenti alla normativa vigente - Regolamento n. 1357/2014/UE - e il riferimento al Catalogo europeo dei rifiuti (CER) con il relativo codice di catalogazione ‘20 03 07’»<sup>257</sup>.

<sup>256</sup> Fonte immagine: Vellutino D., Maslias R., Rossi F., (2016), *cit.* p. 239. Scheda reperibile presso <http://iate.europa.eu/SearchByQuery.do>, data ultima consultazione: 30 maggio 2017.

<sup>257</sup> Vellutino D., Maslias R., Rossi F., (2016), *cit.* p. 238.



**Domínio** AMBIENTE

**Termine** rifiuto ingombrante

**Definizione** Un rifiuto residuo di grandi dimensioni che non ha trovato collocazione in altre tipologie di raccolta differenziata; in altre parole è ingombrante il rifiuto che residua da tutte le raccolte differenziate, avente dimensioni unitarie tali da non poter essere conferito all'ordinario sistema di raccolta del secco residuo.

**Rif. della definizione** Regolamento per la disciplina dei servizi di gestione dei rifiuti urbani e dei rifiuti assimilati di nettezza urbana, Consorzio Azienda Intercomunale di Bacino Treviso Tre, 10/1 1/08  
[http://www.tvre.it/downloads/\\_52\\_D09901%20Regolamento%20disciplina%20gestione%20rifiuti%2010-11-08%20Rev4.pdf](http://www.tvre.it/downloads/_52_D09901%20Regolamento%20disciplina%20gestione%20rifiuti%2010-11-08%20Rev4.pdf)

**Affidabilità** 3 (Affidabile)

**Riferimento relativo ad un termine** Decisione della Commissione del 18 Dicembre 2014 (2014/955/UE), Codice CER 20 03 07  
<http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32014D0955&qid=1432761359210&from=IT> [28.05.2015]

**Nota** La Decisione 2014/955/UE si riferisce al Regolamento (UE) N. 1357/2014 della Commissione del 18 dicembre 2014

**Data** 28/05/2015

Fonte: IATE ID: 112509

iate diffusion version 1.8.4 © Copyright Disclaimer Download IATE About IATE FAQ Contact us

Fig. 5.8 Una proposta per 'rifiuto ingombrante'<sup>258</sup>

Si osservino le modifiche dei campi della scheda che mostrano una definizione, il suo riferimento in un'istituzione regionale, il riferimento alla Decisione (2014/955/UE), al Regolamento 1357/2014 e il codice CER.

Dunque, come continuano gli autori (2016),

*«per trasformare IATE in una risorsa interoperabile sarebbe necessario collegare le schede terminologiche alle risorse documentarie, quali i testi legali in Eur-Lex e i documenti informativi dell'Ufficio Pubblicazione, già identificate da URIs. A loro volta le schede terminologiche dovrebbero essere identificate da URIs per*

<sup>258</sup> Fonte immagine: Vellutino D., Maslias R., Rossi F., (2016), p. 239. Scheda reperibile presso <http://iate.europa.eu/SearchByQuery.do>, data ultima consultazione: 30 maggio 2017.

*diventare risorse aperte e annotabili negli schemi di metadattazione utili per le soluzioni per l'interoperabilità del programma ISA<sup>2</sup>.*

*IATE potrebbe anche contenere informazioni linguistico-terminologiche, quali varianti lessicali e annotazioni sulla formazione lessicale dei termini. Anche queste informazioni potrebbero essere un valido strumento, non solo per il multilinguismo integrale, ma anche a supporto delle tecnologie del web semantico al servizio della modernizzazione delle amministrazioni pubbliche e dei rapporti interistituzionali e commerciali all'interno dell'Unione europea»<sup>259</sup>.*

---

<sup>259</sup> Vellutino D., Maslias R., Rossi F., (2016) *cit.* p.240.



## CONCLUSIONI

Abbiamo mostrato in questo elaborato una modalità per migliorare le risorse linguistico-terminologiche di IATE. Ciò può essere ottenuto sia “dall’alto”, ovvero dall’interno della cornice interistituzionale, che “dal basso”, ovvero dall’esterno, attraverso la collaborazione tra il DSPSC e Terminology Coordination Unit.

In quest’ultimo caso, abbiamo presentato il progetto relativo all’accordo di collaborazione che il Dipartimento di Scienze Politiche Sociali e della Comunicazione dell’Università degli Studi di Salerno ha sottoscritto con l’Unità di Coordinamento Terminologico del Parlamento europeo, con cui si stanno creando, per il dominio specifico “Municipal Waste Management”, schede terminologiche che contengono alcuni elementi utili a sostegno dell’interoperabilità semantica di IATE.

Abbiamo mostrato, inoltre, i più recenti sviluppi nell’ambito delle istituzioni europee sull’uso della terminologia di IATE negli strumenti di traduzione.

Siamo coinvolti direttamente nell’implementazione di progetti nel Parlamento europeo, come “Terminology in Studio”, che mostrano che un passo verso una maggiore coerenza terminologica e standardizzazione può ottenersi tramite l’impiego di strumenti quali le termbase e, parallelamente, grazie agli sviluppi di nuovi moduli di IATE che permettono un’agevole creazione e il download di termbase tematiche o specifiche per un determinato documento.

Siamo ben consci, però, dei limiti di tali approcci. In quello “dal basso”, ovvero quello della creazione di nuove schede terminologiche nella cornice del progetto tra il DSPSC e TermCoord, è naturalmente impossibile sviluppare risorse utili per l’interoperabilità per l’intero parco terminologico di IATE. Possiamo, tuttavia, lavorare a tali risorse in modo che possano essere organizzate in una collezione di IATE, ovvero una raccolta di risorse terminologiche relative a progetti specifici sotto un unico codice di identificazione. Un esempio è la collezione terminologica di IATE marcata dall’etichetta “COM Swiss Data”, ovvero un insieme di risorse provenienti dal database TERMDAT, gestito dalla Sezione di terminologia della Cancelleria federale svizzera, e inserite in IATE tramite un accordo con la Commissione Europea.

Una strutturazione del genere permetterebbe di inserire in IATE le risorse di qualità prodotte dal progetto e ne consentirebbe una facile delimitazione, riducendo il rischio di “inquinare” con altre risorse meno aggiornate.

La questione della granularità dei domini di IATE rimane purtroppo irrisolta. Come visto, IATE si basa sui domini del thesaurus EuroVoc, concepito in origine per rispondere ai bisogni di sistemi documentari generali inerenti alle attività parlamentare dell’Unione Europea. EuroVoc non risulta sempre adatto per l’indicizzazione e la strutturazione di un database complesso come IATE.

Una possibile risposta a ciò arriverebbe da una nuova proposta di strutturazione dei domini di IATE, che potrebbe essere implementata nell’ambito del progetto di IATE 2, ovvero l’iniziativa promossa dallo IATE Management Group per lo sviluppo di una nuova versione del database, che avrà una nuova interfaccia e molte innovative funzioni. Tale

nuova strutturazione è però tuttora ad uno stato embrionale e potrebbe subire sostanziali modifiche nella sua versione definitiva.

Come esposto nei capitoli III e V, i domini che vengono usati per marcare le entrate di IATE permettono un livello di profondità fino a 3 sottolivelli. In IATE 2 i sottolivelli saranno invece 6. I primi due sottolivelli riguarderanno il dominio (identificato da un codice a 2 cifre) e un sottodominio (identificato da un codice a 4 cifre); ad essi si aggiungono altri quattro sottolivelli nella forma di descrittori.

Un descrittore è una parola chiave contenuta in EuroVoc che aiuta a “indicizzare un termine e a denotare il nome del concetto senza ambiguità”<sup>260</sup>. Un esempio dei nuovi 6 livelli può essere l’albero formato da:

- 1) Politics (codice dominio 04);
  - 2) Parliamentary proceedings (codice dominio 0426);
    - 3) Parliamentary procedure;
    - 4) Parliamentary sitting;
    - 5) Document for discussion at a sitting;
    - 6) Committee report.

I punti 1) e 2) sono rispettivamente dominio e sottodominio, mentre gli altri sono descrittori, collegati tra loro da una relazione “Broader Term” (BT) o “Narrower Term” (NT).

---

<sup>260</sup> Traduzione dell’autore. La definizione di “descrittore” è disponibile presso <http://eurovoc.europa.eu/drupal/?q=node/1187&cl=it>.

Sarà possibile, inoltre, utilizzare anche “non descrittivi”, ovvero i cosiddetti “non preferred terms”, ovvero quei termini che, accostati ad altri termini, ne indicano un uso non preferito.

Tra descrittore e non-descrittivi sussiste una relazione di equivalenza<sup>261</sup> nella forma dell’istruzione “USE” oppure “UF” (Used For).

Nel caso in cui diversi termini rappresentino uno stesso concetto, tale relazione di equivalenza chiarisce quale termine deve essere usato. Ad esempio, si supponga che si possa utilizzare sia “marine pollution” che “disposal of waste at sea” per indicare lo stesso concetto, dove il primo sarà il termine preferito mentre il secondo sarà il termine non preferito.

La relazione tra i due termini sarà che a “marine pollution” saranno associati i metadati “UF disposal of waste at sea” e a “disposal of waste at sea” saranno associati i metadati “USE marine pollution”.

Il limite di tale innovazione è che, sebbene la profondità dei livelli sia raddoppiata, la fonte di descrittivi e domini è comunque EuroVoc, che non sarà sempre in grado di offrire un modello che copra il livello di dettaglio di IATE.

L’aspetto più interessante, invece, è che ad ogni descrittore e non-descrittore corrisponde un URI che li identifica univocamente. Dunque implementare metadati di tale tipo per descrivere il dominio di una scheda di IATE rappresenterà indubbiamente

---

<sup>261</sup> Una relazione di equivalenza è una relazione binaria, simmetrica e transitiva che sussiste tra due oggetti. Sulla relazione di equivalenza tra termini e indici di EuroVoc si rimanda a <http://eurovoc.europa.eu/drupal/?q=node/322&cl=en>.

un ulteriore passo del database nel Web Semantico e un miglioramento in direzione dell'interoperabilità.

Per quanto riguarda l'approccio "dall'alto", va ricordato che sebbene le termbase siano in un formato standard e compatibile con diverse piattaforme, ogni filone di sviluppo è concentrato, per ora, solo su quelle impiegabili con il software di traduzione assistita SDL Studio.

Studio è un programma proprietario, sviluppato dall'azienda SDL Trados, ed è attualmente impiegato nelle istituzioni in quanto vincitore di un bando di gara, ma già da tempo alcune istituzioni, tra cui il Parlamento<sup>262</sup>, stanno lavorando allo sviluppo di software proprietari che potrebbero essere dotati di strumenti di gestione terminologica autonomi.

Ad oggi le attività di "Terminology in Studio" non hanno ancora preso in considerazione le potenzialità dei nuovi software proprietari, i cui team di sviluppo non hanno ancora previsto alcuna implementazione delle risorse terminologiche di IATE.

Una nuova sfida sarà, invece, quella di completare le analisi di rischio e di fattibilità legate agli aspetti tecnici e successivamente di allestire un nuovo *workflow* che prenda in considerazione tutte le tipologie di termbase impiegabili nei processi traduttivi e le loro diverse finalità. L'integrazione dell'uso delle termbase nel *workflow* vigente nella complessa macchina traduttiva della Direzione Generale della Traduzione del

---

<sup>262</sup> Abbiamo visto in 3.6.1 che il Parlamento europeo sta sviluppando il software *in house* CAT4Trad.



Parlamento europeo contribuirà al perfezionamento della qualità dei documenti e al miglioramento della gestione della terminologia istituzionale.

## APPENDICE - ANALISI DI SCENARIO DELLE TECNOLOGIE DELLA GESTIONE ED ESTRAZIONE DELLA TERMINOLOGIA

### A.1 La gestione della terminologia e il rapporto con la traduzione

Wright e Budin (1997) definiscono la gestione della terminologia come «*any deliberate manipulation of terminological information*»<sup>263</sup>, un concetto che include «*the systematic recording of such information and its presentation as dictionaries, terminology collections and terminology databases [...]; the retrieval of such information for the purposes of translation and interpreting, for text production as a function of technical communication, and for information management*»<sup>264</sup>. La gestione della terminologia può essere quindi vista come l'insieme dei processi mirati all'organizzazione e alla sistematizzazione dei termini.

Nel contesto della traduzione, l'uso di sistemi di gestione terminologica è un ingrediente primario per assicurare coerenza e qualità nella produzione di testi, particolarmente se il traduttore è specializzato in uno o più campi.

I traduttori che usano strumenti di traduzione automatica possono scegliere di gestire il proprio contenuto terminologico all'interno dei software utilizzati. È il caso, ad esempio,

---

<sup>263</sup> Wright S. E., Budin G. (1997), "Handbook of Terminology Management", volume 1, pp. 1-2.

<sup>264</sup> *Ibidem*.

di SDL Studio e delle termbase. Tali operazioni possono anche essere effettuate, facilmente, tramite software non specialistici come Word, Excel o Access, ma esiste anche una porzione di traduttori che non utilizza alcuno strumento specifico, limitandosi all'impiego di dizionari e risorse online.

I traduttori, talvolta, reputano il lavoro terminologico non efficace in termini di tempi e piuttosto confusionario poiché l'attenzione è concentrata sul processo traduttivo piuttosto che sul tempo impiegato a collezionare termini e gestire termbase. D'altro canto, i CAT tools, talvolta, risultano non soddisfacenti e non offrono mezzi adeguati per completare processi di natura terminologica. Un'altra categoria di traduttori, inoltre, effettua il lavoro terminologico solo alla fine della traduzione e in alcuni casi basa le proprie correzioni terminologiche esclusivamente sulle memorie traduttive. Questo è un metodo che può essere correttamente applicato soltanto in determinati compiti traduttivi, come quelli in cui le memorie coprono ampiamente l'argomento trattato. Tuttavia tale scenario non è sempre verificabile e sul medio periodo l'impiego di termbase *ad-hoc* può avere effetti decisamente benefici nel caso di traduzioni specifiche.

A ciò va aggiunto l'annoso problema del fattore tempo: il traduttore è in continua "lotta" con le scadenze, spesso serrate, imposte dal committente e talvolta il lavoro terminologico non viene effettuato proprio per mancanza di tempo sufficiente. Durante le sessioni plenarie del Parlamento, ad esempio, sono numerosi i documenti, quali risoluzioni e emendamenti, la cui scadenza è fissata per il giorno in cui il documento viene emesso. In tal caso il traduttore potrebbe non aver il tempo necessario per

effettuare le necessarie ricerche terminologiche. Esistono, però, anche lavori di traduzione che necessitano del lavoro terminologico come parte fondamentale della commessa e quindi, talvolta, la ricerca terminologica è un obbligo. È il caso, ad esempio, delle fasi iniziali di una procedura o dei progetti di iniziativa (INI), documenti di grande rilievo che possono introdurre delle proposte legislative e dunque contenere molta terminologia. All'arrivo di tali documenti, è l'Unità di Coordinamento terminologico a intervenire con la creazione di *termfolder*, ovvero liste di termini estratte dai documenti e che vengono inviate alle Unità ancor prima dell'inizio delle attività traduttive, in modo da fornire loro importanti risorse terminologiche.

SDL, l'azienda che produce Studio e MultiTerm<sup>265</sup>, tra i maggiori applicativi nell'attuale scenario traduttivo mondiale, ha condotto nel 2009 un sondaggio<sup>266</sup> per esplorare le opinioni degli utilizzatori a riguardo della terminologia e del suo utilizzo nei processi traduttivi.

Il sondaggio, condotto su 194 utenti, è stato diretto a professionisti della traduzione, di cui l'82% è un traduttore freelance e il 18% è un traduttore di un'azienda, impiegato in un servizio linguistico o un *project manager*. Le aziende coinvolte sono Citi, Unilever, Disney, GE, Cisco, Philips, Panasonic, Siemens, British Airways e Verizon. Tra le conclusioni più significative, è emerso che il 95% dei traduttori è a conoscenza dell'impatto che ha l'uso dei termini sul brand, il 72% crede che la mancata coerenza della terminologia abbia un'influenza negativo sul branding e il 75% sostiene che possa

---

<sup>265</sup> Software proprietario di SDL atto alla gestione di termbase.

<sup>266</sup> Il sondaggio completo è disponibile presso <http://www.sdl.com/download/terminology-an-endtoend-perspective/71114/>.

avere un impatto negativo anche nella comunicazione interna. Difatti, solo il 13% degli intervistati sostiene che la terminologia è coerente all'interno del proprio business.

Mentre il 95% dei partecipanti ha dichiarato di trascorrere molto tempo ad effettuare ricerche terminologiche e il 77% sostiene che sia importante integrare un sistema di gestione della terminologia negli applicativi di traduzione, soltanto il 29% ha risposto di aver già adottato una soluzione di *terminology management*, il 22% lo prenderebbe in considerazione e il 23% si dichiara incerto sull'effettiva implementazione. Lo strumento più usato per gestire terminologia è Microsoft Excel (42%), seguito, nel 31% dei casi, da un tool specifico. La creazione di terminologia avviene via email nel 6% e viene pubblicata in linee guida nel 6%.

Sull'effettiva percezione di produttività tramite un sistema di gestione di terminologia, il 41% degli intervistati sostiene che creerebbe un aumento in produttività dal 25% al 50%. L'estrazione terminologica, invece, viene fatta manualmente nell'84% dei casi mentre solo il 10% utilizza uno strumento.

Appare quindi chiaro che la terminologia viene riconosciuta come avente un ruolo centrale e debba essere gestita efficientemente.

L'incoerenza terminologica è, inoltre, un fattore di preoccupazione notevole che influenza direttamente la qualità del prodotto e aumenta costi e tempi. Tuttavia, ad una chiara consapevolezza sull'importanza del ruolo della terminologia nei processi traduttivi non corrisponde un'effettiva sensibilità sull'uso fattivo di strumenti di gestione terminologica e sui benefici da essi derivati. Come illustrato, in generale, il lavoro terminologico, visto dal livello del traduttore, è un investimento sul tempo:

nonostante un'attività di gestione terminologica abbia effetti benefici sulla produttività e sui tempi di consegna, i traduttori non lo ritengono un fattore sufficientemente determinante per la realizzazione del lavoro traduttivo. Dunque, anche se poche ore di lavoro terminologico per settimana garantirebbero una coerenza e consistenza dei termini e un beneficio di ritorno in termini di tempo e garanzia di qualità, i traduttori risultano essere ancora piuttosto diffidenti nella cura sistematica della terminologia.

D'altronde, idealmente, il lavoro terminologico dovrebbe essere separato da quello traduttivo ed effettuato da team appositi, in cooperazione con esperti e in continui rapporti con gli attori coinvolti. In tal modo la terminologia sterzerebbe chiaramente verso la standardizzazione e la coerenza, riducendo al minimo l'impatto sul lavoro specifico del traduttore. Tuttavia, anche questa è una prospettiva piuttosto utopica, non realizzabile neanche a livello istituzionale, dove il fattore tempo è un elemento ancora più importante. Difatti, come visto, i traduttori del Parlamento europeo destinano una quantità esigua di ore alla terminologia che è sempre curata da traduttori che non hanno necessariamente una preparazione di natura terminologica. In altre istituzioni, come la Commissione, è prevista la figura del terminologo. A differenza del Parlamento, la Commissione prevede un profilo professionale specifico che lavora a tempo pieno sulla terminologia. Una delle motivazioni di questa differenza è da rintracciare nel ruolo stesso della Commissione Europea nell'Unione Europea, della quale è organo esecutivo e promotrice del processo legislativo. Conseguentemente, in ambito terminologico, molta (nuova) terminologia è contenuta nei documenti prodotti dalla Commissione e quindi naturalmente l'istituzione offre una concentrazione maggiore sulla disciplina.

Difatti, nell'ultimo quadrimestre del 2016 la Commissione ha prodotto circa 837mila entrate in IATE, mentre il Parlamento ne ha prodotte circa 216mila<sup>267</sup>.

## A.2 L'estrazione terminologica

L'estrazione terminologica è un'operazione che consiste nell'identificazione di termini a partire da un dato corpus, sulla base dei quali si possono creare risorse terminologiche quali, ad esempio, le termbase. L'estrazione ha dunque come input un documento, o un corpus, e produce come output una lista di termini candidati organizzati in una determinata struttura. L'estrazione può essere eseguita sia manualmente che tramite l'ausilio di strumenti automatici, con un certo spettro di presenza di ridondanza, di rumore o di silenzio, causata da molteplici fattori.

Generalmente, un'estrazione effettuata tramite strumenti è senza dubbio efficace in termini di gestione dei tempi ma i risultati vanno sempre verificati dall'utente, per cui ad oggi, si può definire l'estrazione terminologica nel migliore dei casi, come semi-automatica. Essa può essere monolingue, bilingue o multilingue.

L'estrazione monolingue è solitamente mirata all'analisi di un testo, o di un corpus, con l'obiettivo di identificare dei termini candidati, mentre quella bilingue o multilingue ha l'obiettivo di analizzare i testi fonte in rapporto ai testi nelle lingue di destinazione e identificarne i termini e i loro equivalenti linguistici. L'estrazione è seguita solitamente

---

<sup>267</sup> Fonte: IATE indicators 2016, documento istituzionale che raccoglie le statistiche sui dati di IATE.

dalla creazione di liste di termini candidati, la loro verifica e la creazione automatica o semi-automatica di repertori terminologici.

Si distinguono (Dunaevsky, 2014) tre principali approcci metodologici: linguistico, statistico e ibrido. L'approccio linguistico è mirato all'identificazione di combinazioni morfologiche o sintattiche. In questo caso, tramite l'impiego di *parsers*<sup>268</sup>, si analizza il testo e se ne filtra il contenuto tramite tecniche di *matching*. Naturalmente, tale approccio dipende dalla lingua e quindi i risultati possono variare significativamente da una lingua all'altra in base alle diverse strutture morfosintattiche. L'approccio statistico è invece basato sull'analisi di frequenza degli elementi lessicali. Il vantaggio è che il metodo può essere più indipendente dalla lingua ma è chiaramente meno efficace nel riconoscimento di determinate strutture. L'approccio linguistico ha lo svantaggio di produrre troppo "rumore", ovvero contenuto non prettamente terminologico che però soddisfa i requisiti dei *parser*, mentre quello statistico produce troppo "silenzio", ovvero potrebbe ignorare termini con frequenza bassa.

Preferibile sarebbe un approccio ibrido. Nonostante la maggioranza degli strumenti prediliga quello statistico, regole e filtri linguistici sono introdotti per identificare i termini candidati che hanno una determinata struttura. Un'ulteriore complicazione è il supporto alle lingue. Per quelle dell'Europa occidentale è più facile trovare buoni strumenti, mentre per quelle dell'Est Europa o dell'Asia l'offerta è piuttosto povera.

---

<sup>268</sup> Un parser è un processo che analizza un flusso continuo di dati. In quest'ambito il parser effettua un'analisi sintattica (o parsificazione) del testo input.



### A.3 Schede analitiche di estrattori terminologici

In questo paragrafo si analizzeranno i maggiori strumenti di estrazione di terminologia.

Non è possibile effettuare una classificazione onnicomprensiva a causa della complessità dei software in termini di piattaforme e tecnologie. Tuttavia, le schede presentano sia campi comuni a tutti i *tool* che sezioni dedicate ad eventuali caratteristiche specifiche.

Ogni scheda si chiude con note e valutazioni su ogni estrattore.

Nome	<b>Synchroterm</b>
Casa produttrice	Terminotix
Versione	Synchroterm 2014
Installazione	Sì (con licenza)
File supportati	Documenti: Microsoft Word, Microsoft Excel, RTF, txt, HTML, pdf Memorie Traduttive: Trados, Déjà-vu, Word Fast, SDLX, MemoQ, SDL Studio, MultiTrans, Fusion
Lingue	Inglese, Francese, Spagnolo, Italiano, Portoghese, Tedesco, Svedese, Russo, Greco, Olandese, Ungherese, Norvegese, Polacco, Turco, Ceco, Danese, Bulgaro, Finlandese, Romeno, Lituano, Slovacco, Sloveno, Arabo, Cinese
Algoritmi di estrazione	Statistici, sintattici, morfologici
Export	Terminotix LogiTerm, HTML, Trados MultiTerm 5.5, Trados MultiTerm iX, Trados WinAlign, Microsoft Excel, PROMT

<p>Entrate Terminologiche e funzionalità</p>	<p>Synchroterm prevede tre tipologie di creazione di entrate terminologiche:</p> <p><b>Batch processing:</b> questa modalità permette la creazione automatica di entrate terminologiche. Il numero di entrate e il numero i errori dipende dalle opzioni selezionate nel tab “Tools &amp; Config”, tra cui, ad esempio, “Minimum Match Value” e “Minimum Occurence”. Il Batch processing permette di generare un ampio numero di entrate ma anche il numero di errori sarà più alto, mentre il numero di entrate ad aggiungere manualmente sarà più basso.</p> <p><b>Entrate manuali:</b> questa procedure viene impiegata per creare entrate, singolarmente, dalle liste di termini estratti proposte dal software. È possibile visualizzare i termini nel proprio contesto ed eventualmente modificarlo o aggiungerlo, cancellare espressioni dalle liste di origine e destinazione e modificare l’entrata stessa.</p> <p><b>Entrate manuali di termini non riconosciuti dal software:</b> è possibile aggiungere termini anche nel caso in cui Synchroterm non li abbia riconosciuti o non li abbia riconosciuti correttamente.</p>
<p>Note e Valutazioni</p>	<p>Synchroterm è un valido strumento di estrazione terminologica, che permette un alto grado di personalizzazione dei progetti. Supporta molte lingue e formati ed è dotato di un’interfaccia essenziale ma intuitiva. L’estrazione è possibile in più documenti simultaneamente e la possibilità di aggiungere e verificare i contesti dei termini rende il lavoro di estrazione più semplice. Offre la possibilità di creare liste di termini da ignorare per le estrazioni successive, creando quindi meccanismi di “apprendimento”.</p>

	Synchroterm estrae spesso non-termini ed espressioni non rilevanti. In altri casi, invece, non riesce ad identificare correttamente dei termini e/o espressioni monorematiche (a meno di non creare eccessivo rumore). Non tutte le lingue ufficiali europee sono supportate.
--	---

Nome	<b>MultiTerm Extract</b>
Casa produttrice	SDL
Versione	2009
Installazione	sì
File supportati	txt,doc, html, htm, tmx, rtf, xml, sgm, sgml, ppt, xls, tmx, tmw, ttx
Lingue	Tutte le lingue Unicode
Algoritmi di estrazione	Statistici
Export	TXT, Multiterm XML, Multiterm Export, sdltb (Multiterm Termbase)
Entrate Terminologiche e funzionalità	La creazione di entrate avviene in maniera piuttosto accurata e il software offre possibilità di filtrare tra quelle validate e non validate. MT Extract propone, automaticamente, una lista di stop-words per progetti di estrazione bilingue e monolingue e un Basic Vocabulary, che può essere utilizzato per migliorare ulteriormente i processi estrattivi futuri
Note e Valutazioni	La qualità dell'estrazione è probabilmente la più accurata tra i tool testati. Molto utile la possibilità di filtrare tra liste di termini validati e non validati. I processi di estrazioni impiegano una considerevole quantità di tempo, ma ciò

	dipende anche dalla lunghezza e complessità dei documenti/file impiegati per l'estrazione. Essendo un'estensione della suite SDL MultiTerm, beneficia dell'uso degli standard SDL e gli output sono compatibili al 100% con SDL Studio.
--	---

Nome	<b>TaaS – Terminology as a Service (Tilde Terminology)</b>
Casa produttrice	Tilde
Versione	-
Installazione	No, applicazione online <sup>269</sup>
File supportati	PDF, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Text (.txt), Rich Text (.rtf), XLIFF, HTML, XML, MIF
Lingue	Bulgaro, Croato, Ceco, Danese, Olandese, Inglese, Estone, Finlandese, Francese, Greco, Ungherese, Irlandese, Italiano, Lettone, Lituano, Maltese, Polacco, Portoghese, Romeno, Russo, Slovacco, Sloveno, Svedese, Turco
Algoritmi di estrazione	Linguistici, statistici e basati su corpora  TaaS si basa su diversi strumenti di estrazione. <ul style="list-style-type: none"> <li>• TWSC (Tilde wrapper system for CollTerm), uno strumento basato sull'analisi linguistica (taggatura del discorso, pattern morfosintattici ecc.) integrati da risorse statistiche (ad esempio l'analisi di frequenza).</li> </ul>

<sup>269</sup> <https://term.tilde.com/>.

	<p>L'utente può scegliere tra un'estrazione basata sulla quantità, bilanciata, o qualità;</p> <ul style="list-style-type: none"> <li>• Kilgray: l'estratteore terminologico Kilgray<sup>270</sup> si basa su analisi statistica indipendente dalla lingua</li> </ul> <p>Si può inoltre normalizzare i termini nelle forme canoniche o da dizionario, mantenere i termini esistenti in modo che nuovi termini estratti vadano ad integrarsi nella collezione e avere la possibilità di visualizzare i termini all'interno nel testo tramite processi di evidenziazione.</p> <p>La piattaforma permette inoltre di selezionare fonti per lookup della traduzione di destinazione, tra cui:</p> <ul style="list-style-type: none"> <li>• TaaS public collections, collezioni pubbliche di altri utenti</li> <li>• Le proprie collezioni</li> <li>• EuroTermBank</li> <li>• IATE</li> <li>• TAUS Data, piattaforma di memorie traduttive</li> <li>• Web Data, Database statistico che immagazzina automaticamente dati estratti dal Web</li> </ul>
Export	TBX, CVS, TSV, Moses
Entrate Terminologiche e funzionalità	La creazione di entrate avviene in maniera piuttosto semplice e intuitiva. È possibile modificare i termini e aggiungere nuovi termini fonte, definizioni, note, uso, tipologia di termine, identificatori, etc. Le entrate possono essere

<sup>270</sup> Kilgray è l'azienda produttrice del CATool MemoQ, <https://www.memoq.com/en/about-us/background>.

	validate con semplicità e il tool offre anche strumenti di clean-up semiautomatici.
Note e Valutazioni	TaaS beneficia dell'essenzialità dei tool online e l'interfaccia è particolarmente accessibile e intuitiva. È tra i pochi a permettere un'ampia scelta di algoritmi di estrazione e supporta tutte le lingue ufficiali dell'UE più il Russo. Fondamentale, inoltre, la componente <i>cloud</i> . L'estrazione, in alcuni casi, può però anche impiegare diverse ore o giorni. TaaS notifica automaticamente l'utente quando il processo è completato.

Nome	<b>MemoQ (TermExtraction)<sup>271</sup></b>
Casa produttrice	Kilgray
Versione	2015 Translator Pro 7.8.159
Installazione	sì
File supportati	Documenti UniCode, Memorie traduttive, LiveDocs corpus documents
Lingue	Tutte le UniCode
Algoritmi di estrazione	statistici
Export	CSV, Multiterm XML, TaaS <sup>272</sup>

<sup>271</sup> MemoQ è un CATool che integra una sezione dedicata all'estrazione terminologica. A differenza di altri strumenti simili, come ad esempio SDL Studio, che offre strumenti separati quali MultiTerm Extract, da acquistare separatamente, MemoQ li offre nel pacchetto di installazione di default. In questa scheda si parla dunque solo della funzione TermExtraction del software e non delle altre funzionalità.

<sup>272</sup> Si può esportare la terminologia sulla piattaforma TaaS attraverso un apposito plugin.

<p>Entrate Terminologiche e funzionalità</p>	<p>L'interfaccia di gestione delle entrate terminologiche simula quella che viene impiegata durante il lavoro traduttivo e conseguentemente anche i comandi sono piuttosto simili. Essendo gli algoritmi puramente statistici, il software propone una lista di termini ordinati per frequenza e occorrenza e assumono di default lo status di "candidate". L'utente, similmente a quando traduce, li può confermare (come farebbe con un segmento tradotto) o scartare. È possibile inoltre visualizzare i termini in dettaglio, ovvero il/i contesto/i in cui occorrono e come vengono immagazzinati nella termbase. MemoQ offre anche stopwords lists in ogni lingua con possibilità di configurarle in base alle proprie necessità.</p>
<p>Note e Valutazioni</p>	<p>MemoQ è tra i software emergenti del panorama dei CAT tool ed un punto di forza è senz'altro la possibilità di avere un estrattore terminologico di default. Le personalizzazioni delle entrate e le possibilità di configurazione di parametri ad-hoc è decisamente ampia. Tuttavia, l'estrazione puramente statistica non permette, in alcuni casi, l'identificazione di terminologia potenzialmente rilevante ma a bassa frequenza, nascosta dagli altri termini con occorrenza maggiore.</p>

<p>Nome</p>	<p><b>AlchemyAPI Keyword extraction</b></p>
<p>Casa produttrice</p>	<p>Alchemy API</p>
<p>Versione</p>	<p>-</p>

Installazione	No (online) <sup>273</sup>
File supportati	Documenti online. File HTML, testo (da copiare nell'interfaccia)
Lingue	Inglese, Tedesco, Francese, Italiano, Portoghese, Russo, Spagnolo e Svedese
Algoritmi di estrazione	Statistici, basati su NLP
Export	Metadati in XML, JSON, RDF, formati Microformats REL-tag
Entrate Terminologiche e funzionalità	I termini sono proposti in una lista e marcati come "entity". Essendo questo tool impiegato principalmente per la sentiment analysis, per ogni entità compariranno anche campi quali "Emotion" e "score". La piattaforma propone inoltre una lista di concetti estratti sulla base dell'estrazione e mostra i relativi link a dbpedia, freebase e yago. I termini possono essere visti nel contesto e la piattaforma riesce ad estrarre, spesso con accuratezza, metadati quali titolo, autore, argomento, relazioni, lingua ecc.
Note e Valutazioni	AlchemyAPI utilizza algoritmi di estrazione basati su processi NLP e statistici, rendendo il risultato altamente efficiente. Non sempre, però, tali criteri sono risultati chiari e spesso inconsistenti.

Nome	<b>TermoStat</b>
Casa produttrice	University of Montreal
Versione	3.0

<sup>273</sup> <http://www.alchemyapi.com/products/alchemylanguage/keyword-extraction>.



Installazione	No (online) <sup>274</sup>
File supportati	Documenti plain text (no Word, no pdf)
Lingue	Inglese, Francese, Spagnolo, Italiano, Portoghese
Algoritmi di estrazione	Statistici contrastivi
Export	.txt
Entrate Terminologiche e funzionalità	Non è possibile creare entrate terminologiche <i>stricto sensu</i> poiché si può effettuare un'esportazione di termini solo in formato .txt. I termini vengono estratti in contrasto con un corpora specializzati e non specializzati. Per ogni lingua viene disposto un corpus di svariati milioni di occorrenze (29 milioni per l'italiano) di varia provenienza (l'assemblea parlamentare europea, nel caso dell'italiano). Si possono estrarre sia espressioni monorematiche che polirematiche e l'interfaccia dei risultati di estrazione mostra i termini divisi per frequenza, quoziente di specificità, varianti e pattern. La piattaforma crea automaticamente un cloud e statistiche.
Note e Valutazioni	TermoStat è un estrattore terminologico piuttosto essenziale ma il suo uso risulta piuttosto utile, soprattutto perché identifica, spesso correttamente, varianti ortografiche e mostra per ogni occorrenza il pattern di estrazione impiegato. Non esistono liste stopword ma viene utilizzato uno strumento che stabilisce una soglia di accettabilità che permette di escludere parole ed espressioni non considerate come terminologiche nel testo.

<sup>274</sup> <http://termostat.ling.umontreal.ca/>.

Nome	<b>VocabGrabber</b>
Casa produttrice	Visual Thesaurus
Versione	-
Installazione	No (online) <sup>275</sup>
File supportati	Solo plain text (fino a 200mila caratteri)
Lingue	Inglese
Algoritmi di estrazione	statistici
Export	Possibile solo dopo iscrizione a pagamento
Entrate Terminologiche e funzionalità	I termini estratti vengono divisi rilevanza, occorrenza e familiarità. Ad ognuno viene assegnato un “subject”, una sorta di dominio automaticamente attribuito dalla piattaforma (i subject sono Arts & Literature, Geography, Math, People, Science, Social Studies, Vocabulary). Ogni termine viene presentato con una definizione e il contesto del testo. Particolarmente interessante la possibilità di creare grafi che illustrano le correlazioni semantiche tra termini, chiamata Visual Thesaurus.
Note e Valutazioni	VocabGrabber non può essere definito un estrattore terminologico ma piuttosto come un generatore di liste di parole in relazione al contesto. Visual Thesaurus è senza dubbio la funzione più interessante e la piattaforma offre

<sup>275</sup> <https://www.visualthesaurus.com/vocabgrabber/>.

	<p>anche la possibilità di creare word cloud e l'integrazione via bookmarklet ad altre applicazioni. Il numero di domini (subject) e la loro granularità sembra piuttosto arbitraria e purtroppo non soddisfa principi di metataggatura avanzati. Espressioni complesse o polirematiche non vengono riconosciute.</p>
--	---

Nome	<b>TerMine</b>
Casa produttrice	NaCTeM (National Centre for Text Mining)
Versione	-
Installazione	No (online) <sup>276</sup>
File supportati	Plain text, .txt in ASCII, pdf fino a 2 MB, URL (HTML o pdf)
Lingue	Inglese
Algoritmi di estrazione	Statistici, linguistici (C-Value/NC Value)
Export	no
Entrate Terminologiche e funzionalità	Il metodo impiega informazioni statistiche e linguistiche. Il modulo C-Value misura la frequenza e l'occorrenza, con particolare attenzione alle espressioni polirematiche; il modulo NC-Value fornisce un metodo per l'estrazione dei contesti e l'incorporazione delle informazioni dai contesti all'estrazione dei termini. Il risultato dell'estrazione viene visualizzato in un'interfaccia

<sup>276</sup> <http://www.nactem.ac.uk/software/termine/#form>.

	essenziale, modificabile anche in tabelle. I termini vengono mostrati in una lista e ordinati per occorrenza.
Note e Valutazioni	TerMine è probabilmente tra i migliori tool dedicati alle espressioni complesse. L'estrazione è sempre accurata e le polirematiche a bassa frequenza vengono sempre identificate, nonostante la presenza eventuale di altre espressioni molto più frequenti. È possibile integrarlo ad altre applicazioni tramite servizi SOAP ed offre inoltre un plugin per Protégé, per aiutare a popolare ontologie in OWL. Purtroppo non è compatibile con altre lingue.

Nome	<b>FiveFilters TermExtraction</b>
Casa produttrice	FiveFilters
Versione	1.1
Installazione	No (online <sup>277</sup> )
File supportati	Plain text
Lingue	Inglese
Algoritmi di estrazione	Statistici
Export	no

<sup>277</sup> <http://fivefilters.org/term-extraction/>.

<p>Entrate Terminologiche e funzionalità</p>	<p>Il risultato dell'estrazione terminologica è una tabella in cui compaiono i termini divisi per occorrenza. Sia espressioni semplici che polirematiche vengono mostrate, ma non è possibile effettuare alcuna esportazione o modifica nella versione on line.</p>
<p>Note e Valutazioni</p>	<p>Piuttosto basico, il software fornisce una estrazione completa ma piuttosto inaffidabile. Molti risultati sono non-termini e non è possibile filtrarli o eliminarli dall'interfaccia di visualizzazione dei risultati. Solo l'Inglese è supportato.</p>

## Bibliografia

**Allard M. G. P.**, (2012) *“Managing Terminology for Translation Using Translation Environment Tools: Towards a Definition of Best Practice”*, tesi dottorale, Università di Ottawa.

**Auger, P.**, (1999), *“L’implantation des officialismes halieutiques au Québec: essai de terminométrie”*, ed. OLF, coll. « Langues et sociétés », n. 37, Montréal.

**Auger P.**, (1989) *“La terminotique et les industries de la langue”*, Meta 34.

**Auger, P.**, (1988), *“La terminologie au Quebec et dans le monde, de la naissance à la maturité”*. Actes du sixième colloque OLF-STQ de terminologie. L’ère nouvelle de la terminologie. Quebec.

**Allemang, D., Hendler, J.** (2011). *“RDF –The basis of the Semantic Web”*. In *Semantic Web for the Working Ontologist* (2a Ed.).

**Austermühl, F.** (2001) *“Electronic Tools for Translators”*, Manchester: St. Jerome, p. 107.

**Berners-Lee T. Hendler J., Lassila O.**, (2001). *“The Semantic Web”*, Scientific American.

**Bernth, A. McCord, M. and Warburton, K.** (2003) *“Terminology extraction for global content management”*, Johns Benjamins Publishing Company, 51-69.

**Betz A.; Schmitz K.-D.** (1999). *“The Terminology Documentation Interchange Format TeDIF”*. In: *Terminology and Knowledge Engineering TKE'99*, Innsbruck, Vienna, pp. 782-792.

**Bhreathnach U., Cloke F., Caollfhlonn N. P.**, (2013) *“Terminology for the European Union, the Irish Experience: the GA IATE project”*, Fiontar, Dublin City University, ISBN 978-1-909367-66-1.

**Bilgen B.**, (2009) *“Investigating Terminology Management for Conference Interpreters”*. Elaborato di Laurea, Università di Ottawa, Canada.

**Bizer C., Heath T., Berners-Lee T.**, (2009). *“Linked Data - The Story So Far”*. *International Journal on Semantic Web and Information Systems*, Vol. 5(3).

**Bizer C., Heath T., Uyi Idehen K., Berners-Lee T.**, (2008) *“Linked Data on the Web”*, Atti Convegno WWW2008, Pechino, Cina.

**Bizer C., Heath T., Ayers D., Raimond Y.**, (2008) *“Interlinking Open Data on the Web”* Atti Convegno ESWC2008, Innsbruck, Austria.

**Bizer C., Cyganiak R., Heat T.**, (2007). *“How to Publish Linked Data on the Web”*. *Linked Data Tutorial*, Freie Universität Berlin, Germania, 27 Luglio.

**Boulanger J-C.**, (1995) *“GAUDIN, François (1993) : Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles”* coll. «Publications de l’Université de Rouen », no 182, Rouen, Université de Rouen. Meta 401.

**Bowker L.**, (2015) *“Terminology and translation”*, in *Handbook of Terminology*, a cura di Kockaert H., J., e Steurs F., John Benjamins Publishing Company.

**Bowker L.**, (2006) *"Translation memory and text"*. In Bowker L., *Lexicography, terminology and translation*, University of Ottawa Press.

**Bowker, L., Jennifer P.**, (2002). *"Working with Specialized Language: A Practical Guide to Using Corpora"*. London/New York: Routledge. DOI: 10.4324/9780203469255

**Bowker L.**, (2002) *"Computer aided-translation technology: a practical introduction"*, Ottawa, University Press.

**Cabré Castellví, M. T.**, (2010) *"Terminology and Translation"* in Handbook of Translation Studies, a cura di Y. Gambier e L. Van Doorslaer, John Benjamins.

**Cabré Castellví, M. T.**, (2005) *"La terminologie, une discipline en évolution: le passé, le présent, et quelques prospectives"*. URL : <https://www.upf.edu/pdi/df/teresa.cabre/docums/ca07passe.pdf>.

**Cabré Castellví, M. T.**, (2003), *"Theories of Terminology. Their description, prescription and explanation, in Terminology"*. John Benjamins Publishing Company.

**Cabré Castellví, M. T.**, (2000), *"Elements for a Theory of Terminology: Towards an Alternative Paradigm, in Terminology"*, John Benjamins Publishing Company.

**Cabré Castellví, M. T.**, (1998). *"Terminology Theory, Methods and Applications"*. John Benjamins: Amsterdam.

**Cimiano P.** et al. (2015). *"Linked Terminology: Applying Linked Data Principles to Terminological Resources"*, atti di Electronic lexicography in the 21st century: Linking Lexical Data in the digital age, Herstmonceux Castle (UK) 11-13 Agosto.

**Chiarcos C., et al.**, (2012), Linking linguistic resources: Examples from the Open Linguistics Working Group, in Chiarcos C., Nordhoff S., e Hellmann S., *"Linked Data in Linguistics. Representing Language Data and Metadata"* Springer, Heidelberg, p.201-216.

**Chiarcos C., et al.**, (2001) *"Towards a Linguistic Linked Open Data cloud: The Open Linguistics Working Group"*. TAL (*Traintement automatique des langues*), vol. 52, 3, pp. 245-275.

**Costa H., Corpas Pastor G., Munoz D. I.**, *"A comparative User evaluation of Terminology Management for Interpreters"*.

**D'Agostino, E., Elia, A., Vietri, S.** (2004). *"Lexicon-grammar, Electronic Dictionaries and Local Grammars in Italian"*, in Laporte E., Leclère C., Piot M., Silberstein M., (eds.). Syntaxe, Lexique et Lexique-Grammaire, Volume dédié à Maurice Gross. *Lingvisticae Investigationes Supplementa*, vol. 24, John Benjamins Publishing B.V.: Amsterdam.

**De Nicola A., Missikoff M., Navigli R.** (2009). *"A Software Engineering Approach to Ontology Building"*. Information Systems, 34(2), Elsevier.

**Dubuc R.**, (1985) *"Manuel pratique de terminologie"*, 3a edizione. Brossard, Linguatech.

**Evans V.**, (2007), *"A Glossary of Cognitive Linguistics"*, Salt Lake City: University of Utah Press.

**Faber P.**, (2015) *"Frames as framework for terminology"*, in *Handbook of Terminology*, a cura di Kockaert e Steurs, John Benjamins Publishing Company.

**Faber, P.**, (2012). *"A Cognitive Linguistics View of Terminology and Specialised Language"*. Berlin: Walter de Gruyter. DOI: 10.1515/9783110277203.

**Faber, P.**, (2011). *"The dynamics of specialised knowledge representation. Simulational reconstruction or the perception-action interface"* Terminology 17(1):9–29. DOI: 10.1075/term.17.1.02fab.

**Faber P.**, (2009) *"The cognitive shift in terminology and specialized translation."* MonTI. Monografías de Traducción e Interpretación, n. 1.

**Faber, P, Araúz P. L., Velasco J. e Reimerink A.**, (2007). *"Linking Images and Words: the description of specialized concepts"*. International Journal of Lexicography 20, no. 1.: doi:10.1093/ijl/ecl038.

**Felber H.**, (1984) *"Terminology Manual"*, United Nations Educational, Scientific and Cultural Organization: International Information Centre for Terminology, Paris.

**Frantzi, K., Ananiadou, S. and Mima, H.** (2000) *"Automatic recognition of multi-word terms"*, International Journal of Digital Libraries 3(2), pp.117-132.

**Galinski C.** (2007). *"New ideas on how to support terminology standardization projects"*, eDITion 1/2007.

**Galinski C.**, (2005), *"Semantic Interoperability and Language Resources, the role of ISO/TC 37 standards for Global Semantic Interoperability"*, Proceedings of TKE 2005 7<sup>th</sup> International Conference on Terminology and Knowledge Engineering, Terminology and Content Development, p. 11, ISBN: 87-91242-46-0.

**Galinski, C.**, (2005) *"Semantic Interoperability and Language Resources: Content Development under the Aspect of Global Semantic Interoperability"*, in Proceedings of 3<sup>rd</sup> Eastern Europe e-Gov Days 2005. 17-18 Marzo, Budapest, Ungheria.

**Gambier Y.**, (1991) *"Travail et vocabulaires spécialisés: prolégomènes à une socioterminologie"*, Meta 36, 1.

**Gaudin F.**, (2005), *"La socioterminologie"*, «Langages. La terminologie: nature et enjeux», 39 (2005), n.157, pp 80-92.

**Gaudin F.**, (1995) *"Usage sociaux des termes: théories et terrains »*. Meta 40, 2.

**Goffin R.**, (1985) *"La Science Terminologique"* Terminologie et Traduction. Meta 32, 2.

**Gómez-Pérez A., Fernández-López M., Corcho O.** (2004). *"Ontological Engineering: With Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web"*. Springer.

**Grant Clark K., Feigenbaum L., Torres E.**, (2008), *"SPARQL Protocol for RDF"*, W3C Recommendation 15 gennaio 2008, URL: <https://www.w3.org/TR/rdf-sparql-protocol/>.



**Gruber T.** (1993), "A Translation Approach to Portable Ontology Specifications". In Knowledge Acquisition, 5 (2): pp 199-220.

**Henriksen L., Povlsen C., Vasiljevs A.** (2005). "EuroTermBank – a Terminology Resource based on Best Practice". In Proceedings of LREC 2006, the 5th International Conference on Language Resources and Evaluation, Genova, Maggio 2006.

**Heat T., Bizer C.,** (2011) "Linked Data: Evolving the Web into a Global Space", Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool.

**Hellmann S., Lehmann J., Auer S., e Brummer M.,** (2013), "Integrating NLP using linked data". In Proceedings of the 12th International Semantic Web Conference, Springer, pp. 98–113.

**Hurst S.,** (2009) "Terminology, an end-to-end perspective", SDL International, <http://www.sdl.com/>.

**Ide N., Véronis J.** (1998), "Word Sense Disambiguation: The State of the Art", Computational Linguistics, 24 (1), pp 1-40.

**ISO/TC 37** (2005) [ed.]. ISO/TC 37 "Terminology and other language and content resources" (ISO/TC 37/AG N 125).

**ISO/TC 37** [ed.]. 50 Years ISO/TC 37 "Terminology and other language resources" – A history of 65 years of standardization of terminological principles and methods (ISO/TC 37 N 499).

**ISO/TC 37** [ed.]. Proposal for an MoU/MG statement concerning Semantic Interoperability and the need for a coherent policy for a framework of distributed, coordinated repositories for all kinds of content items on a world-wide scale (ISO/TC 37 N 496).

**ISO/TC 37** [ed.]. Statement on eBusiness Standards and Cultural Diversity (ISO/TC 37 N 497).

**ISO 704:2000** Terminology work – Principles and methods.

**ISO 860:1996** Terminology work – Harmonization of concepts and terms.

**ISO 1087-1:2000** Terminology work – Vocabulary – Part 1: Theory and application.

**ISO 30042** (2008). Systems to manage terminology, knowledge and content – TermBase eXchange (TBX).

**Jain P., Hitzler P., Sheth A., Vema K., Yeh Z. P.,** (2010) "Ontology Alignment for Linked Open Data", Atti Convegno di 9° International Semantic Web Conference, Shanghai, Cina.

**Jeong C. H.,** (2007) "Fundamental of Development Administration" Selangor: Scholar Press, ISBN 978- 967-5-4508-0.

**Kaguera K.,** (2002) "The dynamics of Terminology: a descriptive theory of term formation and terminological growth", John Benjamins.

**Kerremans K., Temmerman R., De Baer P.,** (2008), "Construing domain knowledge via terminological understanding", Linguistica Antverpiensia, New Series–Themes in Translation Studies.

**L'Homme M.C.** (2004), *“La Terminologie: principes et techniques”*. Montréal : Les Presses de l'Université de Montréal.

**Lassila O., Swick R. R.**, (1997) *“Resource Description Framework (RDF), Model and Syntax”*, <https://www.w3.org/TR/WD-rdf-syntax-971002/>

**Maedche A., Staab S.** (2001) *“Ontology Learning for the Semantic Web”*, in Intelligent Systems, IEEE.

**McCrae J.** et al., (2011). *“Linking lexical resources and ontologies on the semantic web with lemon. In The Semantic Web: Research and Applications”* – Atti Convegno di 8th Extended Semantic Web Conference, ESWC 2011, Part I, Heraklion, Creta, Grecia, Maggio 29-Giugno 2, pp. 245-259.

**Melby A. K.**, (2012). *“Terminology in the age of multilingual corpora”*. JosTrans, The Journal of Specialised Translation, Issue 18, luglio.

**Moro A., Raganato A., Navigli R.**, (2014) *“Entity Linking meets Word Sense Dis-ambiguation: a Unified Approach”*, Transactions of the Association for Computational Linguistics, pp 231-244.

**Moser-Mercer B.**, (1992). *“Banking on Terminology: Conference Interpreters in the Electronic Age”*. Translators' Journal, 37(3):507–522.

**Myking J.**, (2001) *“Against Prescriptivism? The Sociocritical Challenge to Terminology”*, IITF Journal 12 (1-2).

**Navigli R.**, (2016) *“Linking IATE, Eurovoc and Euramis”*, The Luxembourg BabelNet Workshop, 2-3 Marzo.

**Navigli R., Ponzetto S.P.**, (2010), *“BabelNet: Building a Very Large Multilingual Semantic Network”*, Atti Convegno di 48th Annual Meeting of the Association for Computational Linguistics Uppsala, Svezia, Luglio 11-16, pp. 216-225.

**Navigli R.**, (2009) *“Word Sense Disambiguation: A Survey”* ACM Comput. Surv. 41, 2, Articolo 10, pp 1-69.

**Noy N. F., McGuinness D. L.**, (2001) *“Ontology Development 101: A guide to Creating Your First Ontology”*, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 e Stanford Medical Informatics Technical Report SMI-2001-0880.

**Oberle D., Guarino N., Staab S.**, (2009) *“What is an ontology”*, Handbook on Ontologies, Springer.

**Oxtoby P. G.**, *“The benefits of terminology management”*, su [www.traslatorscafe.com](http://www.traslatorscafe.com) [https://www.translatorscafe.com/Cafe/images/Articles/TermX\\_Terminology\\_Tool.pdf](https://www.translatorscafe.com/Cafe/images/Articles/TermX_Terminology_Tool.pdf)

**Pavel S., Nolet D.**, (2001) *“Handbook of Terminology”*, Translation Bureau, Minister of Public Works and Government Services, Canada.

**Peristeras V.** (2013). *“Semantic Standards: Preventing Waste in the Information Industry”*, IEEE Intelligent Systems, vol.28, no. 4., p. 70-76.

**Polo L., Alvarez J. M., Azcona E. R.** (2008), *“Promoting Government Controlled Vocabularies to the Semantic Web: EUROVOC Thesaurus and CPV Product Classification Scheme”*. In Atti di Convegno di Semantic Interoperability in the European Digital Library workshop (SIEDL2008), in cooperazione con 5th European Semantic Web Conference (ESWC2008), Tenerife, Spagna, 2 Giugno.

**Rey A.**, (1998), *«Préface»*, Le MicroRobert, Langue française plus noms propres, chronologie, cartes, Paris, Le Robert.

**Rey A.**, (1995), *“Essays on Terminology”*, tradotto da Juan Sager, Amsterdam & Philadelphia, John Benjamins.

**Roche C.**, (2012) *“Ontoterminology, how to unify terminology and ontology into a single paradigm”* Atti convegno LREC 2012, International Conference on Language Resources and Evaluation.

**Roche, C., Calberg-Challot, M., Damas, L., Rouard. P.** (2009). *“Ontoterminology: A new paradigm for terminology”*. KEOD 2009. International Conference on Knowledge Engineering and Ontology Development, 5-8 Ottobre, Madeira (Portogallo).

**Sager, J.** (1990). *“A Practical Course in Terminology Processing”*. John Benjamins Publishing Company.

**Sikos, Leslie F.** (2015). *“Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data”*, ISBN 1484210492.

**Siemoneit B., McCrae J. P., Cimiano P.** (2015) *“Linking Four Heterogeneous Language Resource as Linked Data”*, in atti Convegno 4<sup>th</sup> Workshop on Linked Data in Linguistics: Resources and Applications, pp. 59-63.

**Signore O.** (2002), *“RDF per la rappresentazione della conoscenza”*, Ufficio Italiano W3C <http://www.w3c.it/papers/RDF.pdf>

**Shukair G., Loutas N., Peristeras V., Skarlss S.** (2013). *“Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema”*. In Computers in Industry, vol. 64.

**Temmerman R.** (2000). *“Towards New Ways of Terminological Description. The Sociocognitive approach”*. Amsterdam/Philadelphia: John Benjamins.

**Temmerman R.**, (1997), *“Questioning the univocity ideal. The difference between sociocognitive terminology and traditional terminology”* Hermes. Journal of Linguistics 18.

**Vasiljevs A., Schmitz K.-D.** (2006). *“Collection, harmonization and dissemination of dispersed multilingual terminology resources in online terminology databank”*, in Atti convegno of TSTT 2006, Third International Conference on Terminology, Standardization and Technology Transfer, Pechino, Agosto.

**Vasiljevs A., Rirdance S., Liedskalnins A.**, (2008) *“EuroTermBank: Towards Greater Interoperability of Dispersed Multilingual Terminology Data”*, Atti convegno The First International Conference on Global Interoperability for Language Resources, pp 214-219.

**Vellutino D.** (2015), *“Risorse linguistiche e Open Data per la comunicazione pubblica della gestione dei rifiuti urbani”*, in D. Vellutino, M.T. Zanola (a cura di), *Comunicare in Europa. Lessici istituzionali e terminologie specialistiche*, Educatt Milano, pp. 217-245.

**Vellutino D., Formisano V.**, (2016), *“MWM- Lexicon”. Studio per una risorsa terminologica bilingue sulla “Gestione dei Rifiuti Urbani”* AIDA INFORMAZIONI. Vol. 1-2. Aracne, Roma Pag.167-178, ISSN:1121-0095.

**Vellutino D., Maslias R., Rossi F.** (2016). *“Terminologie specialistiche e diffusione dei saperi, Verso l’interoperabilità semantica di IATE. Studio preliminare per il dominio «Gestione dei rifiuti urbani»”*, in *Terminologie Specialistiche e diffusione dei saperi*, a cura di Zanola M.T., Diglio C., Grimaldi C.

**Walker J., Pan. E., Johnston D., Adler Milstein J., Bates D., W., Middlenton B.**, (2005) *“The Value of Helathcare Information Exchange and Interoperability”*, Health Affairs <http://content.healthaffairs.org/content/suppl/2005/02/07/hlthaff.w5.10.DC1>

**Wiley D.**, (2008) *“Open Content Definition”*, OpenContent.org, <http://opencontent.org/definition/>

**Wright S. E., Budin G.**, (2001) *“Handbook of terminology management (Volume 2): Application-oriented Terminology Management”*, John Benjamins

**Wright S. E., Budin G.**, (1997) *“Handbook of terminology management (Volume 1): Basic Aspects of Terminology Management”*, John Benjamins.

## Documenti Unionali

Commissione Europea (2017), *“Comunicazione della Commissione al Parlamento europeo, al Consiglio, al comitato economico e sociale europeo e al Comitato delle Regioni”* (COM (2017)134).

URL: <http://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52017DC0134&from=EN>

Commissione Europea, ISA<sup>2</sup> Programme (2016) *“ISA<sup>2</sup> work programme 2016 summary”* URL: [http://ec.europa.eu/isa/library/documents/isa2-work-programme-2016-summary\\_en.pdf](http://ec.europa.eu/isa/library/documents/isa2-work-programme-2016-summary_en.pdf)

Commissione Europea, ISA<sup>2</sup> Programme (2015). *“DECISIONE (UE) 2015/2240 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 25 novembre 2015 che istituisce un programma sulle soluzioni di interoperabilità e quadri comuni per le pubbliche amministrazioni, le imprese e*

*i cittadini europei (programma ISA<sup>2</sup>) come mezzo per modernizzare il settore pubblico”.*

URL:[http://ec.europa.eu/isa/documents/isa\\_annex\\_ii\\_eif\\_en.pdf](http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf)

Commissione Europea, ISA<sup>2</sup> Programme (2015). *“Metadata management requirements and existing solutions in EU Institutions and Member States”*,

URL:[https://joinup.ec.europa.eu/site/core\\_vocabularies/governance\\_and\\_management\\_of\\_structural\\_metadata/Methodology-and-tools-for-Structural-Metadata-Management-and-Governance-for-EU-Institutions-and-Member-States\\_v1.00.pdf](https://joinup.ec.europa.eu/site/core_vocabularies/governance_and_management_of_structural_metadata/Methodology-and-tools-for-Structural-Metadata-Management-and-Governance-for-EU-Institutions-and-Member-States_v1.00.pdf)

Commissione Europea (2015), ISA Programme. *“eGovernment Core Vocabularies handbook”*

URL:<https://ec.europa.eu/isa2/sites/isa/files/publications/e-government-core-vocabularies-handbook.pdf>

Commissione Europea (2014) *“Asylum and Migration Glossary 3.0, a tool for better comparability produced by the European Migration Network”*,

URL:[https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/european\\_migration\\_network/docs/emn-glossary-en-version.pdf](https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/european_migration_network/docs/emn-glossary-en-version.pdf)

Commissione Europea (2013), ISA Programme. *“Core Public Service Vocabulary*

*Documentation”* URL:<https://joinup.ec.europa.eu/catalogue/distribution/cpsv-documentation-pdf>

Commissione Europea (2013), ISA Programme. *Process and methodology for developing semantic agreements.*

URL:[https://joinup.ec.europa.eu/community/core\\_vocabularies/document/process-andmethodology-developing-semantic-agreements](https://joinup.ec.europa.eu/community/core_vocabularies/document/process-andmethodology-developing-semantic-agreements)

Commissione Europea (2013), Proposta di direttiva del Parlamento europeo e del Consiglio relativa alla fatturazione elettronica negli appalti pubblici, COM (2013) 449 final.

URL: <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52013PC0449&from=EN>

Commissione Europea (2010), ISA Programme. *“European Interoperability Framework (EIF) for European public services”*, Annex 2 to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services',

URL:[http://ec.europa.eu/isa/documents/isa\\_annex\\_ii\\_eif\\_en.pdf](http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf)

Commissione Europea (2009), *“Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del settembre 2009, sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee (ISA)”*

URL: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:260:0020:0027:IT:PDF>

Commissione Europea, Metadata Registry. “*ELI Ontology*”  
URL:[http://publications.europa.eu/mdr/eli/documentation/ELI\\_Ontology-v1.1.pdf](http://publications.europa.eu/mdr/eli/documentation/ELI_Ontology-v1.1.pdf)

Consiglio Europeo (2012), “*Conclusioni del Consiglio che invitano all'introduzione dell'identificatore della legislazione europea (ELI) (2012/C 325/02)*” URL:<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:325:0003:0011:EN:PDF>

Commissione Europea (2007), “*FP7 in Brief, how to get involved in the EU 7<sup>th</sup> Framework Programme for Research*” URL:[https://ec.europa.eu/research/fp7/pdf/fp7-inbrief\\_en.pdf](https://ec.europa.eu/research/fp7/pdf/fp7-inbrief_en.pdf)

IATE HandBook Task Force (2017) “*IATE HandBook*”, ver. 03/05/2017

IATE HandBook Task Force (2016). “*IATE Handbook*”, ver. 13/12/2016

IATE Management Group (2016), “*IATE Best Practices*”

IATE Management Group, (2015) “*How to export data from IATE in Studio*”, EurTerm

Lider (2015), “*Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe, FP7 – 610782*” URL:<http://www.lider-project.eu/sites/default/files/D4.11-v3.0.pdf>

Parlamento europeo (2016), “*Regolamento (8<sup>a</sup> legislatura – Luglio 2016) Titolo VII, Capitolo 3: Disposizioni Generali per lo svolgimento delle sedute, articolo 158, Lingue*”  
URL:<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+RULES-EP+20160715+0+DOC+PDF+V0//IT&language=IT>

Parlamento europeo (2014), Terminology Coordination Unit. “*Terminology management*”  
URL:<http://www.termcoord.eu/wp-content/uploads/2014/11/Capturing-Terminology-Scenario-I1.pdf>

Parlamento europeo (2008), “*Multilinguismo al Parlamento: Il valore aggiunto dell'Europa*”  
URL:<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+IM-PRESS+20071017FCS11816+0+DOC+PDF+V0//IT&language=IT>

Ufficio delle Pubblicazioni (2007), Sintesi della Legislazione dell'UE. *"Allargamento 2004: la sfida di un'UE a 25"* URL:<http://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=URISERV:e50017&from=IT>

Unione Europea (2012), *"Trattato sull'Unione Europea"* (versione consolidata), Gazzetta Ufficiale C 326 del 26.10.2012, pagg. 13–390 (Maastricht, 1992) URL:<http://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:12012M/TXT&from=it>

## Sitografia

BabelNet

URL:<http://babelnet.org/>

Babelfy

URL:<http://babelfy.org/>

Commissione Europea. *ISA Programme, DCAT application profile for data portals in Europe*

URL:[https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description)

Commissione Europea. *ISA Programme, Core Criterion and Core Evidence Vocabulary*

URL:[https://joinup.ec.europa.eu/asset/criterion\\_evidence\\_cv/description](https://joinup.ec.europa.eu/asset/criterion_evidence_cv/description)

Commissione Europea. *ISA Programme, Core Location*

URL:[https://joinup.ec.europa.eu/asset/core\\_location/description](https://joinup.ec.europa.eu/asset/core_location/description)

Commissione Europea. *ISA Programme, Core Person*

URL:[https://joinup.ec.europa.eu/asset/core\\_person/description](https://joinup.ec.europa.eu/asset/core_person/description)

Commissione Europea. *ISA Programme, Core Public Organisation Vocabulary*

URL:<https://joinup.ec.europa.eu/asset/cpov/description>

Commissione Europea. *ISA Programme, Core Public Service*

URL:[https://joinup.ec.europa.eu/asset/core\\_public\\_service/description](https://joinup.ec.europa.eu/asset/core_public_service/description)

Commissione Europea. *ISA Programme, Core Public Service Vocabulary*

URL:[https://joinup.ec.europa.eu/asset/core\\_public\\_service/asset\\_release/core-public-service-vocabulary-0](https://joinup.ec.europa.eu/asset/core_public_service/asset_release/core-public-service-vocabulary-0)

Commissione Europea. *ISA Programme, Core Registered Organisation*

URL:[https://joinup.ec.europa.eu/asset/core\\_business/description](https://joinup.ec.europa.eu/asset/core_business/description)

Commissione Europea. *ISA Programme, "Facilitating the exchange of legislation data in Europe"*

URL:[http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-21action\\_en.htm](http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-21action_en.htm)

Commissione Europea. *ISA Programme, Metadata Registry*

URL:<https://joinup.ec.europa.eu/catalogue/repository/metadata-registry>

Commissione Europea. *ISA Programme. Metadata management requirements and existing solutions in EU Institutions and Member States.* URL:<https://joinup.ec.europa.eu/node/78172>

Commissione Europea, *Multilinguismo – Quale ruolo svolge l'UE nella promozione delle lingue?*

URL:[http://ec.europa.eu/education/policy/multilingualism\\_it](http://ec.europa.eu/education/policy/multilingualism_it)

Commissione Europea, *DocFinder*

URL:<https://epnet.extranet.ep.europa.eu/https://webgate.ec.testa.eu/docfinder/> (*intranet*)



Commissione Europea, *Euramis*

URL:<https://epnet.extranet.ep.europa.eu/https://webgate.ec.testa.eu/Euramis/> (*intranet*)

Commissione Europea), *Quest*

URL:<https://epnet.extranet.ep.europa.eu/https://webgate.ec.testa.eu/questmetasearch/index.php> (*intranet*)<sup>278</sup>

Eur-lex

URL:<http://eur-lex.europa.eu/homepage.html>

Eur-Lex, ELI, pagina del registro

URL:<http://eur-lex.europa.eu/eli-register/about.html>

Eur-Lex, F.A.Q.

URL:<http://eur-lex.europa.eu/content/help/faq/intro.html#help8>

Eur-Lex, numero Celex

URL:<http://eur-lex.europa.eu/content/help/faq/celex-number.html>

IATE Management Group, *EurTerm homepage*

URL:<https://webgate.ec.europa.eu/fpfis/wikis/display/iatetp/Home> (*intranet*)

IATE Management Group, *EurTerm, IATE 2*

URL:<https://webgate.ec.europa.eu/fpfis/wikis/display/iatetp/IATE+2+Task+Force> (*intranet*)

IATE Management Group, *EurTerm, IATE Public*

URL:<https://webgate.ec.europa.eu/fpfis/wikis/display/iatetp/IATE+Public> (*intranet*)

IATE Management Group, *EurTerm, IATE-Studio integration Task Force*

---

<sup>278</sup> Le URL in sitografia marcate con “*intranet*” sono raggiungibili solo tramite l’accesso alle *Intranet* delle Istituzioni Europee e quindi non accessibili pubblicamente.

URL:<https://webgate.ec.europa.eu/fpfis/wikis/display/iatetp/IATE-Studio+Integration+Task+Force> (*intranet*)

Linguistic Linked Open Data

URL:<http://linguistic-lod.org/llod-cloud>

Metadata Registry, European Legislation Identifier (ELI)

URL:<http://publications.europa.eu/mdr/eli/index.html>

Metadata Registry, European Legislation Identifier (ELI), Documentation

URL:<http://publications.europa.eu/mdr/eli/documentation/index.html>

Metadata Registry, European Legislation Identifier (ELI), Documentation, URI Templates for European Union legislation

URL:[http://publications.europa.eu/mdr/eli/documentation/uri\\_templates.html](http://publications.europa.eu/mdr/eli/documentation/uri_templates.html)

Metadata Registry, European Legislation Identifier (ELI), Documentation, URI Templates for European Union legislation (specifiche)

URL:<https://tools.ietf.org/html/rfc6570>

NIF

URL:<http://persistence.uni-leipzig.org/nlp2rdf/>

Open Content, Wiley D., OpenContent.org.

URL:<http://opencontent.org/definition/>

Parlamento europeo, *Cat4Trad*

URL:<https://epnet.extranet.ep.europa.eu/http://www.cat4trad.ep.parl.union.eu/cat4trad/CAT4TradUI.html#> (*intranet*)

Parlamento europeo, Terminology Coordination Unit

URL:<http://termcoord.eu/>

Parlamento europeo, Terminology Coordination Unit, *Cenni Storici*

URL:<http://termcoord.eu/about-us/history/>

Parlamento europeo, Terminology Coordination Unit, *DocHound*

URL:<http://termcoord.eu/DocHound/>

Parlamento europeo, Terminology Coordination Unit, *GlossaryLinks*

URL:<http://termcoord.eu/GlossaryLinks/>

Parlamento europeo, Terminology Coordination Unit, *Managing Terminology at EP*

URL:<http://termcoord.eu/about-us/managing-terminology-ep/>

Parlamento europeo, Terminology Coordination Unit, *TermFolders*

URL:<http://termcoord.eu/iate/TermFolders/>

Parlamento europeo, Terminology Coordination Unit, *TermCoord Glossaries*

URL:<http://termcoord.eu/discover/glossaries-2>

Parlamento europeo, Directorates-General

URL:[http://www.euoparl.europa.eu/the-secretary-general/en/organisation/directorates-general\\_1](http://www.euoparl.europa.eu/the-secretary-general/en/organisation/directorates-general_1)

Parlamento europeo, Directorates-General, Directorates-General for Translation

URL:<http://www.euoparl.europa.eu/the-secretary-general/en/organisation/directorate-general-for-translation>

Parlamento europeo, Multilinguismo

URL:<http://www.euoparl.europa.eu/aboutparliament/it/20150201PVL00013/Multilinguismo>

Parlamento europeo, Segretariato Generale

URL:<http://www.europarl.europa.eu/aboutparliament/it/20150201PVL00014/Segretariato-generale>

Parlamento europeo, *Translation Portal*

URL:<http://tradportal.ep.parl.union.eu/> (*intranet*)

Portale Open Data dell'Unione europea. Applicazioni

URL:<http://data.europa.eu/euodp/it/apps>

Portale Open Data dell'Unione europea. EuroVoc, il thesaurus multilingue dell'Unione Europea

URL:<http://data.europa.eu/euodp/it/data/dataset/eurovoc>

Portale Open Data dell'Unione europea. IATE

URL:<http://data.europa.eu/euodp/en/data/dataset/iate>

Portale Open Data dell'Unione europea. Linked Data

URL:<http://data.europa.eu/euodp/it/linked-data>

Translationjournal.net, Effective Terminology Management, Sanaa Benmessaoud

URL:<http://translationjournal.net/journal/47terminology.htm>

Ufficio delle pubblicazioni. EuroVoc, il thesaurus multilingue dell'Unione Europea

URL:<http://eurovoc.europa.eu/drupal/?q=it>

Unione Europea, IATE

URL:<https://iate.cdt.europa.eu/iatenew>

W3C, World Wide Web Consortium, JSON

URL:<https://www.w3.org/2001/sw/wiki/JSON-LD>

W3C, World Wide Web Consortium, Linked Data

URL:<https://www.w3.org/standards/semanticweb/data>

W3C, World Wide Web Consortium, Semantic Web

URL:[https://www.w3.org/2001/sw/wiki/Main\\_Page](https://www.w3.org/2001/sw/wiki/Main_Page)

W3C, World Wide Web Consortium, Vocabularies

URL:<https://www.w3.org/standards/semanticweb/ontology>

W3C, World Wide Web Consortium, OWL

URL: <https://www.w3.org/2001/sw/wiki/OWL>

W3C, World Wide Web Consortium, OWL Web Ontology Language Overview

URL:<https://www.w3.org/TR/owl-features/>

W3C, World Wide Web Consortium, RDF

URL:<https://www.w3.org/2001/sw/wiki/RDF>

W3C, World Wide Web Consortium, RDFS

URL:<https://www.w3.org/2001/sw/wiki/RDFS>

W3C, World Wide Web Consortium, Resource Oriented Architecture (ROA)

URL:[https://en.wikipedia.org/wiki/Resource-oriented\\_architecture](https://en.wikipedia.org/wiki/Resource-oriented_architecture)

W3C, World Wide Web Consortium, REST

<https://www.w3.org/2001/sw/wiki/REST>

W3C, World Wide Web Consortium, SKOS

URL:<https://www.w3.org/2001/sw/wiki/SKOS>

W3C, World Wide Web Consortium, SOA Architecture

URL:<https://www.w3.org/TR/ws-arch/>

W3C, World Wide Web Consortium, SOAP

URL:<https://www.w3.org/TR/soap/>

W3C, World Wide Web Consortium, SPARQL

URL:<https://www.w3.org/2001/sw/wiki/SPARQL>

WordNet, a lexical database for English

URL:<https://wordnet.princeton.edu/>

## Estrattori Terminologici Online

AlchemyAPI

URL:<http://www.alchemyapi.com/products/alchemy/language/keyword-extraction>

FiveFilters

URL:<http://fivefilters.org/term-extraction/>

TerMine

URL:<http://www.nactem.ac.uk/software/termine/#form>

TermoStat

URL:<http://termostat.ling.umontreal.ca/>

VocabGrabber

URL:<https://www.visualthesaurus.com/vocabgrabber/>



## Ringraziamenti

Al professor Elia e al professor Laudanna, grazie alla cui eccezionale disponibilità ho potuto proseguire il mio percorso di dottorato all'estero, lavorando nella cornice del Parlamento europeo.

Alla professoressa Daniela Vellutino, una delle persone più importanti della mia vita e carriera, un esempio, un faro, uno scudo. Per giorni, o forse settimane, ho pensato alle parole più adatte per esprimere la mia profonda riconoscenza. Invano. Mi limito quindi ad usare la semplicità di una sola parola: grazie. Grazie per avermi sempre guidato e per il costante sostegno. Devo a Lei una buona parte di quello di cui sono orgoglioso di essere diventato.

A Rodolfo Maslias, capo Unità di Terminology Coordination Unit, Parlamento europeo, per il costante incoraggiamento e la fiducia. Grazie per avermi fornito accesso a informazioni di prima mano e per avermi aperto le porte di IATE. Grazie per la tua amicizia e incrollabile positività. Grazie a Viola Laci-Pongracz e Ciprian Dumea, compagni di viaggio di "Terminology in Studio", per la vostra resilienza e passione.

Al team di IATE del Centro di Traduzione, in particolar modo a Paula Zorrilla Agut per la stima e l'interesse mostrato nelle nostre attività e proposte. Grazie a Maria Pia Montoro per l'insaziabile curiosità e la costante tendenza alla condivisione della conoscenza.

Grazie alla mia famiglia per l'incessante supporto. Grazie a mia madre per il costante incoraggiamento e per aver deciso di venire a Lussemburgo in missione "ripopolamento frigorifero", ma naturalmente per tanto, tanto di più. Grazie a mio padre per l'affetto, le conversazioni franche e i non-detti che significano più di ogni esplicitazione. Grazie a mia sorella per la caparbità, la forza e l'allegria, ma soprattutto per esserci. Sempre. Un grazie speciale ai miei nonni, che mi hanno incoraggiato a seguire i miei sogni e che hanno sempre riversato in me incondizionata fiducia e stima. A loro rinnovo l'invito di venire in Lussemburgo, con la flebile speranza che si lascino convincere dall'"ufficialità" di queste parole rilegate e dall'affetto di un nipote che ha già in voi il dono più grande.

Grazie ai miei amici, nuovi e vecchi, vicini e lontani per l'incoraggiamento, per la pazienza ai tanti "no", per le revisioni estemporanee e per le continue iniezioni di fiducia.