

# Il corpus del Digesto: approcci e metodi computazionali per la creazione di risorse linguistiche

Alessandra Cinini<sup>1</sup>, Paola Marongiu<sup>2</sup>, Eva Sassolini<sup>3</sup>

<sup>1</sup>Istituto di Linguistica Computazionale "A. Zampolli", (CNR - Pisa), Italia, [alessandra.cinini@cnr.it](mailto:alessandra.cinini@cnr.it)

<sup>2</sup>Istituto di Linguistica Computazionale "A. Zampolli", (CNR - Pisa), Italia, [paola.marongiu@cnr.it](mailto:paola.marongiu@cnr.it)

<sup>3</sup>Istituto di Linguistica Computazionale "A. Zampolli", (CNR - Pisa), Italia, [eva.sassolini@cnr.it](mailto:eva.sassolini@cnr.it)

## ABSTRACT<sup>1</sup> (ITALIANO)

Il contributo vuole discutere della creazione, elaborazione e uso di corpora testuali bilingui (latino-italiano) paralleli a supporto della traduzione specialistica in ambito giuridico.

L'occasione che ci ha permesso di affrontare i diversi aspetti sottesi a questo tipo di iniziative è stato il progetto di traduzione in italiano del Digesto di Giustiniano, del quale il nostro gruppo di ricerca è partner tecnologico nel progetto PRIN 2022 PNRR "TESTO". Parallelamente alla creazione del corpus bilingue sincronizzato sono stati implementati una serie di strumenti di supporto alla traduzione da offrire agli esperti di dominio: memorie di traduzione (*Translation Memory*) disponibili online in modalità contrastiva dei libri già tradotti; annotazione linguistica dei testi con l'obiettivo di predisporre un glossario terminologico estratto che rappresenti un punto di partenza per la creazione di un lessico giuridico bilingue.

**Chiavi:** Linguistica Computazionale; sincronizzazione di testi; risorse linguistiche di dominio; estrazione di termini

## ABSTRACT (ENGLISH)

*The Corpus of the Digest: Computational Approaches and Methods for the Creation of Linguistic Resources.*

The paper aims to describe the creation, development and use of parallel bilingual (Latin-Italian) text corpora to support specialised translation in the legal field. The PRIN 2022 PNRR project 'TESTO' on the Italian translation of Justinian's Digest, in which our research group is a technology partner, allowed us to address the various aspects underlying this type of initiative. In addition to the creation of the parallel bilingual corpus, we implemented a series of translation support tools to be offered to domain experts: translation memories of the texts already translated, available online in contrastive mode; linguistic annotation of the texts to prepare an extracted terminological glossary, identified as a starting point for the creation of a bilingual legal lexicon.

**Keywords:** Computational linguistics; texts alignment; domain linguistic resources; term extraction

## 1. INTRODUZIONE

Il progetto TESTO ha come obiettivo primario il completamento della traduzione in lingua italiana del Digesto di Giustiniano. Il contesto è quindi quello della traduzione specialistica in ambito giuridico, affrontata da un gruppo/team di traduttori con competenze di dominio. La traduzione ha evidenziato in passato e presenta anche oggi diversi aspetti peculiari che ne hanno dilatato i tempi. Infatti, da un lato le grandi dimensioni del corpus testuale (50 libri per migliaia di pagine), dall'altro le lingue coinvolte e la natura specialistica dei contenuti rappresentano ancora oggi una sfida dal punto di vista metodologico. L'obiettivo dei traduttori e del prof. Sandro Schipani, che ha guidato per anni il progetto, è sempre stato quello di creare un ponte comunicativo tra culture diacronicamente molto distanti, permettendo allo stesso tempo la valorizzazione di profili comuni. La traduzione del Digesto richiede la comprensione del testo originario per garantire la salvaguardia della struttura logico-argomentativa e delle sfumature linguistiche anche nel testo in traduzione (Addis 2023). Nello scenario descritto la presenza di risorse di supporto alla traduzione (memorie di traduzione e/o lessici terminologici bilingui) è fortemente auspicabile, ma di difficile reperimento poiché sistemi di supporto consolidati adatti al compito specifico non sono facilmente disponibili. Con questo obiettivo l'Istituto di Linguistica Computazionale "A. Zampolli" del CNR (CNR-ILC) ha offerto il proprio contributo tecnologico nella realizzazione di un corpus parallelo bilingue a partire dai libri già tradotti, che ha rappresentato poi la base dalla quale poter derivare altre risorse linguistiche a supporto della traduzione. Si è trattato di un processo incrementale che vede oggi il corpus parallelo coprire circa l'80% dell'intera opera. Le problematiche legate alla traduzione di un testo come il Digesto hanno richiesto la ricerca di soluzioni specifiche i cui risultati verranno analizzati per determinare quali strumenti software possano supportare adeguatamente il processo di traduzione, senza però sostituire il traduttore umano nei

<sup>1</sup> Nel quadro di un'elaborazione comune, Eva Sassolini, come responsabile delle attività scientifiche del progetto, ha curato l'impianto, la revisione complessiva dell'articolo e la redazione dei §§ 1 e 4, Alessandra Cinini, come responsabile della banca dati, ha invece curato i §§ 2 e 3, Paola Marongiu, vincitrice di un assegno di ricerca sul progetto, ha curato i §§ 2.1, 3.1 e 4.1.

domini specialistici. Inoltre, sarà discusso l'approccio utilizzato che combina l'intervento manuale agli strumenti computazionali.

## 2. LA RISORSA BILINGUE: CONTESTO E PROBLEMATICHE DI TRADUZIONE

Il Digesto di Giustiniano (*Digesta* nel suo nome latino), insieme con il *Codex Iustiniani*, le *Iustiniani Institutiones* e le *Novellae Constitutiones*, fanno parte del *Corpus Iuris Civilis*, un progetto realizzato sotto l'ordine dell'imperatore Giustiniano (527-566 d.C.) tra il 529 e il 534 d.C., e che raccoglie un ricco numero di testi della giurisprudenza romana (Banchich et al. 2015; Dingley, 2016). Il Digesto rappresenta la parte più importante del *Corpus Iuris Civilis*, poiché offre la più ricca collezione di testi giuridici romani. Il Digesto comprende infatti estratti selezionati da 1528 scritti prodotti da 39 giuristi romani, coprendo 800 anni di storia giuridica romana (Dingley, 2016: 5; Ribary & McGillivray, 2020). Gli scritti dei giuristi sono organizzati in 50 libri, suddivisi in cinque gruppi a seconda dell'argomento: Diritto Pubblico (Libro I); Diritto Privato (Libri II-XLVII); Diritto Penale (Libro XLVIII); Procedura d'Appello e Tesoro (Libro XLIX); Diritto Municipale, Diritto Specializzato e Definizioni (Libro L). I libri sono suddivisi in 432 titoli, ogni titolo è diviso in leggi e queste ultime in paragrafi.

Lo studio del Digesto per i giuristi di oggi serve per comprendere e sviluppare un sistema di principi sui quali si basa il diritto di ogni Stato, indipendentemente dai rispettivi sistemi giuridici (Schipani 2005: xxvii).

### 2.1 Il testo latino e la traduzione

Il tipo di latino utilizzato nel Digesto è altamente tecnico-specialistico e appartiene ai cosiddetti linguaggi di dominio, o lingue speciali (Cortelazzo 1994; Mazzini 2009). Di conseguenza, presenta caratteristiche specifiche a livello lessicale, sintattico e stilistico. Nella traduzione di testi specialistici, il traduttore si trova sempre a mediare tra una traduzione *source-oriented* (che punta al rispetto del testo e della struttura della lingua di partenza) e una *target-oriented* (che è invece orientata al fruitore del testo in traduzione, alla lingua e alla cultura target). Nel primo caso si adottano traduzioni estremamente fedeli al testo fonte, e quindi si opta per l'utilizzo di calchi o prestiti per esprimere concetti tecnici che non sono presenti nella cultura e/o nel sistema giuridico della lingua target. Nel secondo caso invece concetti specifici del testo fonte vengono interpretati nella lingua target, utilizzando ad esempio neologismi, parafrasi e rielaborazioni più o meno invasive della sintassi della lingua di partenza (Paolucci 2013; Paolucci 2017:325;328). Al contesto descritto si aggiungono altri due fattori di complessità. Il primo è legato alla lingua del testo fonte, ovvero il latino, una lingua storica e quindi non più parlata, utilizzata per descrivere un sistema legale non più esistente; il secondo riguarda lo status peculiare del Digesto, che si colloca a metà strada tra un'opera letteraria e un testo di dominio.

Particolare attenzione è posta inoltre sulla traduzione dei termini tecnici e sulla coerenza terminologica lungo tutta l'opera tradotta. Per quanto riguarda il primo aspetto, nonostante si cerchi di rispettare la sintassi di base del testo latino, la lingua esprime i termini/concetti in modo più conciso rispetto all'italiano, utilizzando talvolta delle ellissi o sottintendendo il referente; inoltre esistono concetti non presenti di fatto nella lingua italiana, che devono essere quindi resi attraverso una parafrasi. Ne sono mostrati esempi in Figura 1.

| Tipo aggiunta    | Definizione  | Esempio   |
|------------------|--|---|
| Morfosintattica  | Veicola informazioni grammaticali che nel latino vengono espresse a livello morfologico, o che sono richieste dalla costruzione sintattica in italiano | (LAT) Si heres <b>praecepto fundo</b> rogatus sit hereditatem restituere [...]<br>(ITA) Se l'erede, <a cui> <b>sia stato prelegato un fondo</b> , sia stato richiesto <mediante fedecommesso> di dare l'eredità [...]   |
| Stilistica       | Aggiunge elementi non necessari alla grammaticalità della frase in italiano, ma utili al miglioramento della traduzione dal punto di vista stilistico  | (LAT) Ius civile est, quod in totum a <b>naturali vel genitum</b> recedit [...]<br>(ITA) Il diritto civile... non si discosta in tutto dal <b>diritto naturale</b> o <dal diritto> delle genti [...]  |
| Di completamento | Aggiunge pochi elementi/dettagli che aiutano il lettore nella comprensione del (con)testo, a volte includendo elementi dedotti dal contesto precedente | (LAT) [...] ab intestato <b>rogatusque sit</b> restituere hereditatem [...]<br>(ITA) [...] e <b>sia stato richiesto &lt;per fedecommesso&gt;</b> di dare l'eredità [...]  |
| Informativa      | Aggiunge contenuto per favorire la comprensione del lettore, esplicitando informazioni non presenti nel testo latino                                   | (LAT) [...] nisi liberalitatem tantum ad priorem <b>fideicommissarium heres voluit pertinere</b> .<br>(ITA) [...] a meno che l'erede < <b>che pur ha ugualmente diritto di trattenerla</b> > non abbia voluto < <b>e dichiarato</b> > che <questa sua> liberalità spetti solo al primo fedecommissario. |

Figura 1. Tipologie di aggiunte individuate nel Digesto

In questi casi, i traduttori adottano un approccio interpretativo che si esplica anche con aggiunte, cioè delle inserzioni testuali segnalate graficamente attraverso delle parentesi uncinate (< >) che esplicitano le informazioni non direttamente riconducibili al testo latino (o greco). Questa pratica è finalizzata a facilitare la comprensione del testo, ma produce nei testi un disallineamento non trascurabile. Per questa ragione è stata condotta un'approfondita analisi quantitativa e qualitativa di tali interventi allo scopo di individuare strategie per il loro trattamento nella fase di allineamento automatico dei testi. Dall'analisi condotta è emerso che le aggiunte più rilevanti dal punto di vista dell'approccio alla traduzione sono certamente quelle informative. È infatti attraverso queste ultime che il traduttore specialista veicola concetti tecnici, spesso connessi al contesto legale romano, che quindi non trovano un diretto corrispondente in italiano.

### 3. GLI STRUMENTI IMPLEMENTATI

Nella linguistica computazionale, i corpora rivestono un ruolo centrale. In particolare, quelli paralleli sono fondamentali per diverse applicazioni, tra le quali la traduzione automatica, l'estrazione di terminologie e la disambiguazione semantica (Fathi, 2015). Per la costruzione del corpus bilingue lo studio della letteratura in ambito di sincronizzazione dei testi ha mostrato come le più diffuse procedure software di allineamento automatico di testi implementino algoritmi basati su metodi statistici, che tipicamente lavorano a livello di frase o sequenza di caratteri. Quando però il parallelismo stringente tra testo e traduzione si perde a favore di un'estensione/semplificazione dei contenuti, gli stessi algoritmi di allineamento producono un numero maggiore di porzioni non allineate, sulle quali è necessario intervenire con strategie diverse. Tipicamente, un tale approccio produce buoni risultati solo per lingue tipologicamente affini e per testi *well-behaved* (bene educati), ma non è adatto al caso del corpus di dominio descritto. Anche quello che utilizza il metodo linguistico (o *rule-based*) non era utilizzabile in assenza di un lessico bilingue in grado di stabilire corrispondenze traduttive univoche tra parole/lemmi. La situazione descritta ha orientato le scelte verso un metodo 'misto', in cui l'allineamento automatico viene corretto stabilendo delle 'ancore' linguistiche in corrispondenza di una lista di parole/elementi per i quali sono stati identificati degli equivalenti bilingui (Marinai et al. 1992). Per la parallelizzazione del corpus del Digesto sono state prima generate liste allineate di parole target e corrispondenti elementi source, tra i quali è stata individuata e annotata una rete di ancore, in modo da permettere la parallelizzazione dei contesti relativi, visualizzabili a richiesta in modalità contrastiva. Successivamente, la componente aleatoria del legame tra ancore (non si tratta di veri traduttori) ha richiesto un puntuale lavoro di revisione manuale delle corrispondenze individuate automaticamente, e ha rappresentato un test di valutazione delle procedure di estrazione di candidati termini. Il corpus testuale allineato, che in forma di concordanze contrastive è possibile consultare online<sup>2</sup>, si compone dei primi 36 libri ed è strutturato in 72 archivi indicizzati (uno per ciascun libro e lingua) contenenti 1.044.055 occorrenze per la lingua italiana e 635.260 occorrenze per la lingua latina, per un totale di quasi 1.700.000 parole. Questa funzionalità, assimilabile alle memorie di traduzione, è stata utilizzata da subito come supporto al lavoro dei traduttori per perseguire l'uniformità dei criteri di traduzione e la coerenza terminologica. Andando oltre lo stretto ambito dei traduttori, sfruttando le corrispondenze tra unità testuali che sono traduzione una dell'altra è possibile infine disporre di risorse bilingue da poter riutilizzare per l'addestramento di sistemi di traduzione automatica di testi di dominio. Nella maggior parte dei casi, la corrispondenza è già possibile a livello di frase: nei primi 36 libri del Digesto le frasi allineate sono 24.213, mentre le unità non allineate a livello di frase risultano 1.358. Sono inoltre in fase di studio strategie per sfruttare le unità già allineate per una sperimentazione sulla traduzione automatica basata su un approccio di Machine Learning (Deep Learning) da applicare ai libri ancora da tradurre.

#### 3.1 Studi preparatori per il glossario

Le problematiche nell'allineamento dei testi hanno influito sulla costruzione di un sistema automatico in grado di estrarre una corrispondenza bilingue tra termini. La costruzione di un lessico terminologico bilingue porta con sé diverse sfide. Mentre per le lingue moderne più diffuse la procedura di estrazione di termini nella lingua source può avvalersi sia di risorse terminologiche già esistenti da modificare/integrare da cui partire, sia di tecniche di NLP (lemmatizzazione e *term extraction*), la presenza di una lingua storica riduce la disponibilità di tali risorse e strumenti. Inoltre, per quanto riguarda l'individuazione dei candidati traduttori, l'operazione può risultare complessa nei casi in cui un termine non trovi un corrispettivo nella lingua di arrivo e sia stata utilizzata per esso un'espressione polirematica e/o anche più estesa (si vedano gli esempi in figura 1). Sulla base dell'analisi delle traduzioni, abbiamo rilevato 4 livelli di complessità nella riconduzione dei termini dalla lingua source alla lingua target.

<sup>2</sup> <http://dbtvm1.ilc.cnr.it/digesto>

Il primo livello riguarda i casi più semplici di polisemia. Si veda l'esempio 1: recuperando i contesti del lemma *insula* emergono dai testi due traduzioni diverse: "isola" (1a) e "casamento" (1b), con diversi ambiti di applicazione.

### 1) **Insula**

- a) (LAT) *ULPIANUS libro nono ad edictum. Insulae Italiae pars Italiae sunt et cuiusque provinciae.* (Dig. 5, I, 9)  
(ITA) ULPIANO, nel libro nono All'editto. Le isole dell'Italia sono parte dell'Italia, come quelle di ogni provincia <lo sono della stessa>.
- b) (LAT) *Idem scribit, si aream emerò et **insulam** in ea aedificaverò, recte me Publiciana usurum.* (Dig. 6, II, 11, 9)  
(ITA) Lo stesso <Pomponio> scrive che, se avrò comprato l'area, ed avrò edificato su di essa un **casamento**, correttamente mi servirò della Publiciana.

I casi di polisemia includono anche quei lemmi che utilizzati in un contesto non specialistico assumono generalmente uno o più sensi comuni, ma che quando compaiono nel discorso legale specializzano il proprio significato per indicare un concetto distintivo di questo ambito. Si veda a questo proposito l'esempio 2, che illustra il caso del verbo latino *caveo*. Nel discorso non specialistico, questo lemma si può tradurre con "stare in guardia, guardarsi da (rispetto/contro qualcosa o qualcuno)", "stare all'erta", oppure "evitare" (2a); tuttavia, nel contesto giuridico e in particolare nel Digesto, questo termine viene utilizzato con il senso tecnico di "decretare, disporre, stabilire", nel caso di leggi ed editti (2b), oppure "prestare (stipulazione di) garanzia" (2c) (OLD, s.v. "caveo").

### 2) **Caveo**

- a) (LAT) *Si fullo vestimenta polienda acceperit eaque mures roserint, ex locato tenetur, quia debuit ab hac re **cavere**.* (Dig. 19, II, 13, 6)  
(ITA) Se un tintore abbia ricevuto dei vestiti per pulirli ed i topi li abbiano rosicchiati, è tenuto <con l'azione> da locazione, poiché avrebbe dovuto **cautelarsi** riguardo a ciò.
- b) (LAT) *Oratione divi Marci **cavetur**, ut, si senatoris filia libertino nupsisset, nec nuptiae essent:* (Dig. 23, II, 16)  
(ITA) Nell'orazione del divo Marco <Aurelio letta in senato> **si dispone** che, se la figlia di un senatore avesse sposato un liberto, <queste> non sarebbero state nozze:
- c) (LAT) *Tutor et curator, ut rem salvam fore pupillo **caveant**, mittendi sunt in municipium, quia necessaria est satisdatio.* (Dig. 2, VIII, 8, 4)  
(ITA) Il tutore e il curatore, affinché **prestino garanzia** al pupillo a salvaguardia dell'integrità del suo patrimonio, devono essere fatti andare al municipio, perché è una dazione di garanti necessaria.

Il caso di *caveo* "prestare (stipulazione di) garanzia" (2c) è utile anche per illustrare un'ulteriore complicazione nella compilazione del glossario, che è rappresentata dalla presenza di espressioni polirematiche, sia in latino che in italiano. Alcune di queste infatti, specialmente in un contesto specialistico, assumono grande rilevanza e diffusione. Nell'esempio 2c un unico termine nella lingua source viene tradotto con una polirematica nella lingua target. Esistono anche casi nei quali la polirematica si trova nella lingua source, e viene tradotta con un unico termine nella lingua target. Questo è il caso di *aes alienum* (lingua source) che viene tradotto come "debito" (lingua target), illustrato nell'esempio 3.

### 3) **Aes alienum**

- (LAT) *Si **aere alieno** dominico exhauriatur peculium servi, res tamen in causa peculiaris manent:* (Dig. 15, I, 4, 5)  
(ITA) Se il peculio del servo si esaurisca per **debiti** nei confronti del padrone, le cose, tuttavia, restano ugualmente nel peculio:

Infine, sono stati individuati casi di polirematiche a incastro o concatenate, illustrate nell'esempio 4.

### 4) **Ab intestato**

- a) (LAT) *Haec oratio pertinet ad alimenta, quae testamento vel codicillis fuerint relicta sive ad testamentum factis sive **ab intestato**.* (Dig. 2, XV, 8, 2)  
(ITA) Questa orazione riguarda gli alimenti che siano stati lasciati nel testamento o nei codicilli, sia testamentari sia **in assenza di testamento**.

- b) (LAT) *Si quis ex his personis, quae ad **successionem ab intestato** non admittuntur, de inofficioso egerit (nemo enim eum repellit) et casu optinuerit, non ei prosit victoria, sed his qui habent ab intestato successionem:* (Dig. 5, II, 6, 1)  
(ITA) Se una delle persone, che non sono ammesse alla **successione intestata**, abbia agito per testamento inofficioso (infatti, nessuno la respinge), e per un caso abbia vinto, questa non giova a lei, ma a coloro che hanno titolo alla successione intestata:

La collocazione *ab intestato* significa letteralmente “da (parte di) qualcuno che non ha fatto testamento”. Questa viene tradotta solitamente nel testo italiano attraverso la polirematica “in assenza di testamento” (4a), o addirittura si conserva la locuzione latina *ab intestato*, tuttora utilizzata nel linguaggio giuridico moderno. Questa espressione compare in casi di successione nei quali il defunto non ha potuto, per qualche motivo, lasciare testamento. Quando si accompagna a sostantivi che si riferiscono alla cessione dei beni, come *successio* “successione” o *hereditas* “eredità”, *ab intestato* forma, con tali sostantivi, un’ulteriore espressione specialistica del dominio giuridico, che viene tradotta con “successione/eredità legittima” o “intestata”, in quanto in assenza di testamento si applica la legge che regola le successioni. In questo scenario l’intervento della correzione/revisione manuale del lessico/glossario è imprescindibile. Ma è proprio qui che la linguistica computazionale può offrire un supporto fattivo e importante, non solo per cercare strategie in grado di ridurre l’intervento umano allo stretto necessario, ma anche per verificare le ipotesi con campagne di estrazione mirate e puntuali sui casi più ambigui. Sulla base delle analisi condotte, abbiamo optato per un processo incrementale, in cui si parte dalla lemmatizzazione del testo latino e dalla selezione di termini validati e/o da validare da parte degli esperti, per poi definire un primo insieme dei termini source. Per quelli target, date le problematiche descritte e l’esperienza maturata in fase di allineamento, per il collegamento delle ‘ancore’ linguistiche si sta studiando un modo di individuare il ‘senso’ di una traduzione identificandolo tra un nucleo di parole ricorrenti. Questo procedimento si basa sull’ipotesi distributiva di Harris (1968), che afferma che parole semanticamente simili tendono a comparire in contesti linguistici simili. Si tratterebbe quindi di estendere la ricerca del termine traducendo a un insieme di più parole che hanno legami semantici stabili.

#### **4. PROSPETTIVE FUTURE**

Il Progetto è tuttora in corso, così come il processo di traduzione. Nell’ambito delle attività svolte si ritengono rilevanti sia il protocollo di trattamento, allineamento, standardizzazione del formato di rappresentazione del corpus bilingue, sia l’estrazione di risorse linguistiche di dominio. Ogni aspetto di tali attività ha richiesto uno studio specifico e un continuo lavoro di riallineamento/adattamento degli strumenti, in particolare per l’allineamento dei testi e per l’estrazione di terminologia bilingue. Quest’ultimo aspetto resta ancora una sfida aperta perché la traduzione non è conclusa e i libri che sono ancora in traduzione potrebbero introdurre cambiamenti nella terminologia, vista la strutturazione ad argomento dell’opera.

##### **4.1 Un confronto tra modelli neurali generativi e traduzione umana**

La traduzione automatica di norma è poco applicata nei contesti di traduzione di dominio. È tuttavia intento del progetto esplorare ambiti non usuali. Con questa intenzione si è deciso di provare comunque un approccio automatico, ora che i modelli di Machine Learning hanno preso il posto di quelli statistici. In tempi recentissimi, lo sviluppo del settore dell’Intelligenza Artificiale (IA) ha portato all’introduzione di modelli estremamente avanzati che possono essere applicati alla traduzione automatica. La svolta in questo senso è rappresentata dalla Traduzione Automatica Neurale (NMT, Neural Machine Translation), che consiste nell’utilizzo di reti neurali artificiali: sottoposte a grandi quantità di dati linguistici in fase di training, queste rielaborano l’informazione appresa in maniera complessa attraverso i vari strati della rete. Nel campo della NMT sono attualmente estremamente popolari i Large Language Models (LLM), tra i quali uno degli esponenti più in vista è il modello GPT, sviluppato dall’azienda OpenAI (2023), i quali generano automaticamente una risposta a partire da un prompt impartito dall’utente (Angelucci 2024:187-190). Il motivo del successo di questi modelli risiede nell’abbandono della statica e limitativa rappresentazione vettoriale delle parole a favore di quella contestuale. Le innovazioni prodotte dalle ultime ricerche hanno cambiato le tecniche disponibili, dotandole della capacità di valutare il contesto della parola, ovvero utilizzando informazioni legate alle parole adiacenti ad essa. Al fine di confutare le affermazioni degli esperti della traduzione, che restano dubbiosi sull’utilizzo di questi strumenti, sono stati condotti alcuni esperimenti pilota su GPT-4, confrontando l’output del modello con la traduzione realizzata dagli specialisti su diversi passaggi del Digesto. Per effettuare gli esperimenti è stato utilizzato il modello GPT-4, in accesso gratuito, al quale è stato sottoposto il seguente prompt: “Traduci il seguente passaggio dal latino all’italiano: [PASSAGGIO IN LATINO]”. Al modello sono state sottoposte frasi dalla lunghezza e complessità

variabili. GPT raggiunge già degli ottimi risultati nei compiti di traduzione, anche quando la lingua di partenza è il latino (Volk et al., 2024). Tuttavia, il nostro obiettivo era testare le performance del modello sulla traduzione di un testo specialistico, di materia giuridica e di età antica. I primi risultati hanno evidenziato come anche i modelli più recenti trovino difficoltà in tutti quei casi di utilizzo delle aggiunte che abbiamo ampiamente descritto in precedenza (sezione 2.1). A conferma di questo riportiamo qui di seguito alcuni passaggi del Digesto (esempi in a); le traduzioni degli specialisti (esempi in b); e le traduzioni di GPT-4 (esempi in c).

#### 5) **Manus**

- a) (LAT) *Manumissiones quoque iuris gentium sunt. est autem manumissio de **manu** missio, id est datio libertatis*: (Dig. 1, 1, 4)
- b) (ITA: Dig.) Anche le manomissioni appartengono al diritto delle genti. La mano-missione è la dismissione dalla **<sofferenza a quella potestà chiamata> "mano"**: si tratta, cioè, della concessione della libertà.
- c) (ITA: chatGPT) Anche le manomissioni appartengono al diritto delle genti. Ora, la manomissione è la liberazione dalla **mano**, cioè la concessione della libertà.

#### 6) **Status/libertinus**

- a) (LAT) *Homo liber, qui se vendidit, manumissus non ad suum statum revertitur, quo se abdicavit, sed efficitur libertinae condicionis*. (Dig. 1, V, 21)
- b) (ITA: Dig.) L'uomo libero, che si vendette, se viene manomesso non ritorna al proprio **stato <originario>**, dal quale ha abdicato, ma assume la **condizione di liberto**.
- c) (ITA: chatGPT) Un uomo libero, che si è venduto come schiavo, una volta manomesso, non ritorna al suo **stato originario**, che ha rinunciato, ma diventa di **condizione libertina**.

Gli esempi 1 e 2 offrono una panoramica di pregi e difetti del modello GPT, se confrontato con la traduzione realizzata dagli specialisti. In particolare, l'esempio 2 permette di confrontare la traduzione di *status* e *libertinus*. Il primo termine si riferisce allo stato giuridico del soggetto, un concetto che esiste anche nella lingua italiana, e non ha quindi bisogno di essere ulteriormente specificato attraverso una perifrasi. È interessante notare anche che il traduttore opta per un'aggiunta che specifica lo stato giuridico al quale ritorna il soggetto (<originario>, esempio 2b); d'altra parte GPT-4 inserisce lo stesso tipo di informazione anche se questa non è direttamente reperibile nel testo, probabilmente su indicazione del verbo *revertitur* da *revertor* "ritornare", che suggerisce un'idea di ripristino di una situazione precedente. Diverso è il caso di *libertinus*, aggettivo che si riferisce alla condizione di liberto. "Condizione di liberto" è infatti la formula scelta dai traduttori per questo aggettivo nell'esempio 2b. Diversamente, il modello GPT-4 opta per "libertino", che però risulta essere una scelta errata, in quanto in italiano questo aggettivo ha subito un mutamento semantico, e fa oggi riferimento ad atteggiamenti licenziosi e a una condotta disordinata.

## RINGRAZIAMENTI

Progetto di ricerca PRIN 2022 PNRR P20224NJLK - SH2 - TESTO "Translating, Encoding, Sharing The Origins. From the Littera Florentina to an open-access Italian translation of Justinian's Digest", nell'ambito del Piano Nazionale Ripresa e Resilienza (PNRR) M4C2 - Investimento 1.1 "Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN)- finanziato dall'Unione europea - NextGeneration EU" - CUP B53D23032480001.

## BIBLIOGRAFIA

- Addis, F. (2023). Sistema, codice e formazione nel canone di Sandro Schipani. Roma e America: diritto romano comune: rivista di diritto dell'integrazione e unificazione del diritto in Europa e in America Latina: 44, 2023, 239-254.
- Angelucci, M. (2024). Computabilità e traduzione interlinguistica: considerazioni sull'impiego dell'Intelligenza Artificiale in ambito traduttivo. *Digitalita*, 19(1), 183-197.
- Banchich, T. M., Marenbon, J. & Reid, C. J. (2015). The Revival of Roman Law and Canon Law. In Miller Jr, F. D., & Biondi, C. A. (eds.) *A Treatise of Legal Philosophy and General Jurisprudence: Volume 6: A History of the Philosophy of Law from the Ancient Greeks to the Scholastics*. Dordrecht: Springer. 251-265.
- Dingledy, F. W. (2016). The Corpus Juris Civilis: A Guide to its History and Use. *Legal Reference Services Quarterly*. DOI: 10.1080/0270319X.2016.1239484.
- Fawi, Fathi. (2015). Costituzione di un corpus giuridico parallelo italiano-arabo. In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015*, 125-129.
- Harris Z. S. (1968) *Mathematical Structures of Language*, New York, Wiley.
- Marinai E., Peters C., Picchi E. (1992). A Project for Bilingual Reference Corpora. *Acta Linguistica Hungarica*. vol. 41. n.1/4. 191-204. <http://www.jstor.org/stable/44308294>.

- OLD = Glare, P. G. W. (ed.). 1997[1982]. Oxford Latin dictionary. Oxford: Oxford University Press.
- OpenAI. 2023. GPT-4 technical report.
- Paolucci, S. (2013). Strategia estraniante e strategia addomesticante nella traduzione dei testi giuridici. *Linguistica* 53(2), 73-89.
- Paolucci, S. (2017). Foreignising and domesticating strategies in translating legal texts. *International Journal of legal discourse*, 2(2), 243-263.
- Pezzato E. (2022). Brevi note sulle inscriptiones del Digesto. *CULTURA GIURIDICA E DIRITTO VIVENTE*, 10, pp. 1-14. <https://dx.doi.org/10.14276/2384-8901/3403>.
- Ribary, M., & McGillivray, B. (2020). A Corpus Approach to Roman Law Based on Justinian's Digest. *Informatics*, 7(4), 44. <https://doi.org/10.3390/informatics7040044>.
- Volk, M., Fischer, D. P., Fischer, L., Scheurer, P., Ströbel, P., Sprugnoli, R., & Passarotti, M. (2024). LLM-based machine translation and summarization for Latin. Third Workshop on Language Technologies for Historical and Ancient Languages -- LT4HALA (at LREC/COLING), Torino, 25 May 2024.