



UNIVERSITÀ DI PISA

Dipartimento di Filologia, Letteratura e  
Linguistica

Corso di Laurea Magistrale in  
Informatica Umanistica

**Annotazione semantica del *Liber Abaci* di  
Fibonacci per lo sviluppo di un modello di  
riconoscimento per le entità di dominio  
mercantile-commerciale**

RELATORI:

Prof.ssa Enrica Salvatori

Prof.ssa Maria Simi

CANDIDATO:

Letizia Ricci

Anno Accademico 2020-2021



# Indice

|  |           |
|--|-----------|
| <b>Introduzione .....</b>  | <b>5</b>  |
| <b><i>Parte 1 – Fibonacci e la sua opera .....</i></b>                   | <b>9</b>  |
| <b>1. Leonardo Bigollo Pisano, il Fibonacci.....</b>                     | <b>10</b> |
| 1.1 La questione del nome .....  | 10        |
| 1.2 Vita e opere.....  | 12        |
| 1.3 Mediatore della matematica araba .....                               | 15        |
| <b>2. Il Liber Abbaci .....</b>  | <b>19</b> |
| 2.1 L’eredità del <i>Liber Abaci</i> nelle pratiche mercantili .....     | 23        |
| 2.1.1 La lezione di Fibonacci nelle scuole d’abaco .....                 | 24        |
| 2.2 La fortuna di Leonardo tra autorità e oblio.....                     | 25        |
| 2.3 Riscoperta settecentesca.....  | 27        |
| 2.4 Edizioni cartacee.....   | 29        |
| <b><i>Parte 2 – Il progetto .....</i></b>                                | <b>32</b> |
| <b>1. Edizione digitale: il progetto <i>Fibonacci1202-2021</i> .....</b> | <b>33</b> |
| <b>2. Analisi semantica.....</b>   | <b>34</b> |
| 2.1 Annotazione dei dati .....   | 34        |
| 2.1.1 Perché annotare i dati.....  | 36        |
| 2.1.2 Annotare: strategie per farlo.....                                 | 37        |
| 2.2 Ciclo di annotazione .....   | 39        |
| <b>3. Prima fase: il cap. VIII del <i>Liber Abaci</i>.....</b>           | <b>41</b> |
| 3.1 Punto di partenza.....   | 41        |
| 3.2 Tagset e annotazione pilota.....                                     | 42        |
| 2.2.1 CATMA: tool di annotazione.....                                    | 44        |
| 3.2 Linee Guida e annotazione test.....                                  | 48        |
| 3.2.1 Il secondo annotatore .....  | 50        |
| 3.3 Valutazione delle annotazioni.....                                   | 51        |
| 3.3.1 Multi-Label classification .....                                   | 52        |
| 3.3.2 Metriche di valutazione.....                                       | 58        |
| 3.3.3 Feedback .....   | 64        |
| <b>4. Seconda fase: il taglio storico sul cap. VIII.....</b>             | <b>70</b> |
| 4.1 Considerazioni storiche.....   | 70        |
| 4.2 Modifica alle linee guida .....                                      | 71        |
| 4.3 Il <i>file gold</i> del cap. VIII.....                               | 73        |

|  |            |
|--|------------|
| <b>5. Estensione sugli altri capitoli commerciali.....</b>                     | <b>75</b>  |
| 5.1 Annotazione dei cap. IX e X.....   | 76         |
| 5.2 Creazione dei <i>file gold</i> IX e X .....                                | 78         |
| 5.2.1 UDPipe.....  | 78         |
| 5.3 I <i>file gold</i> IX e X completi .....                                   | 80         |
| <b>6. Sistemi di apprendimento automatico.....</b>                             | <b>82</b>  |
| 6.1 Train e Test dell' algoritmo .....   | 83         |
| <b>7. Riconoscimento delle entità commerciali del <i>Liber Abaci</i> .....</b> | <b>85</b>  |
| 7.1 Baseline con approccio <i>dictionary-based</i> .....                       | 85         |
| 7.1.1 Training e test set .....  | 86         |
| 7.1.2 Il Dizionario .....  | 87         |
| 7.1.3 Il Classificatore .....  | 88         |
| 7.1.4 Valutazione del classificatore .....                                     | 90         |
| 7.2 Regole euristiche di disambiguazione .....                                 | 94         |
| <b>Conclusioni e sviluppi futuri .....</b>                                     | <b>96</b>  |
| <b>Bibliografia .....</b>  | <b>99</b>  |
| <b>Sitografia .....</b>  | <b>103</b> |
| <b>Appendice .....</b>   | <b>104</b> |

# Introduzione

Leonardo Pisano, detto Fibonacci, è stato il più acuto interprete e mediatore della cultura matematica araba medievale, trasmessa in Italia attraverso le sue opere: prima fra tutte il *Liber Abaci*. Si tratta di un'opera complessa che presenta una grande varietà di argomenti: i primi 7 capitoli trattano dei numeri arabi e delle operazioni matematiche, i capitoli 8-11 hanno un approccio empirico, descrivendo l'utilizzo della matematica nelle pratiche mercantili, mentre i capitoli finali (12-15) discutono di problemi di aritmetica e algebra più astratti.

Date queste premesse può risultare sorprendente il fatto che tale opera, considerata legittimamente una delle opere più importanti per la storia della matematica occidentale, sia rimasta dimenticata per secoli. Bisogna, infatti, attendere il XIX secolo per avere la prima edizione integrale, e, addirittura i giorni nostri, per la prima edizione critica completa.

In realtà il ritardo è spiegato da due ordini di ragioni. La prima relativa alla necessità di un curatore editoriale di competenze a un tempo specializzate e ibride, esperto o esperta in filologia, paleografia, storia, matematica. In seconda istanza si deve guardare alla corposità dell'opera. Oltre alle competenze matematiche, storiche e filologiche necessarie per approcciarsi a questo trattato, il testo è conservato in 19 manoscritti diversi scaglionate tra il XIII e il XIX secolo con una corposità di più di duecento carte per manoscritto.

Il passo determinante per la diffusione degli scritti di Leonardo Fibonacci avvenne nel 1857, quando Baldassarre Boncompagni pubblicò, in due monumentali volumi in quarto, due opere cardine del matematico pisano, il primo di questi volumi contiene l'intero corpus del *Liber Abaci*<sup>1</sup>, il secondo la *Practica Geometriae*.

---

<sup>1</sup> B. Boncompagni Ludovisi, *Il Liber Abaci di Leonardo Pisano pubblicato secondo la lezione del codice Magliabechiano C. I. 2616, Badia Fiorentina, n° 73*, in Id., *Scritti di Leonardo Pisano, matematico del secolo decimoterzo pubblicati da Baldassarre Boncompagni*, Vol. I, Roma, Tipografia delle scienze matematiche e fisiche, 1857.

L'edizione di Boncompagni non era ancora, tuttavia, un'edizione critica: l'autore si limitò infatti a trascrivere diplomaticamente il codice<sup>2</sup> che aveva a disposizione e vi aderì totalmente conservando gli errori di scrittura, le incongruenze nei risultati dei calcoli matematici e, solo in alcuni casi, indicando sul testo le inesattezze più evidenti con un (*sic*).

La prima edizione critica completa ha visto la luce solo di recente, nel 2020, con cura di Enrico Giusti, coadiuvato da Paolo D'Alessandro<sup>3</sup>. Questa nuova edizione è uno strumento prezioso e un punto di arrivo fondamentale per la vasta comunità internazionale di studiosi dei diversi ambiti matematici e umanistici. Tuttavia, anche considerando il nuovo formato editoriale, per le caratteristiche intrinseche dell'opera, il volume è risultato ovviamente corposo, voluminoso e costoso: non certo alla portata di tutti, né maneggevole o di facile analisi. Proprio per questo motivo, date le possibilità che oggi offrono gli strumenti digitali di edizione e analisi del testo, è stato avviato un progetto di ricerca congiunto tra Regione Toscana e Università di Pisa per trasformare l'edizione critica a stampa in un'edizione critica digitale<sup>4</sup>. L'obiettivo è quello di facilitare l'accesso al testo e rendere ricercabili i contenuti, di modo da poter recuperare e rendere disponibile il tesoro di informazioni linguistiche, storiche, matematiche che l'opera contiene.

Il presente lavoro si inserisce all'interno di questo ampio progetto, dando in particolare un contributo dal punto di vista dell'analisi semantica, in riferimento al dominio specifico mercantile-commerciale. Le informazioni contenute nei capitoli commerciali, infatti, sono preziose non solo per il loro contributo alla storia della matematica, ma anche come testimonianza dell'economia e delle pratiche mercantili medievali. Per questo i termini del dominio sono stati individuati nel testo con lo scopo di codificare i risultati per una lettura critica del testo attraverso un modello di trattamento automatico del linguaggio.

La relazione frutto di questo contributo si divide in due parti.

---

<sup>2</sup> Codice Magliabechiano, Conv. Soppr. C. 1. 2616, ff. 1-214. Firenze, Biblioteca Nazionale Centrale.

<sup>3</sup> E. Giusti, P. D'Alessandro, *Leonardi Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, 2020.

<sup>4</sup> Responsabile del Progetto Pier Daniele Napolitani.

La prima parte, alla luce delle considerazioni appena effettuate, si propone di presentare nel dettaglio, dal punto di vista storico-umanistico, le questioni legate alla tradizione diretta e indiretta del trattato di algebra e aritmetica. Il *Liber Abaci* viene presentato nella sua completezza: prima dando informazioni sulla biografia dell'autore e sulla struttura e forma dell'opera, successivamente un rapido cenno alle fonti per comprendere la grande eredità di origine indiana e mediazione araba lasciata da Fibonacci al mondo occidentale. Vengono poi fatte considerazioni riguardo la fortuna dell'opera tra i contemporanei di Leonardo e la sua conseguente influenza a livello pratico nell'ambito mercantile e commerciale; l'oblio intermedio che ha dato luogo anche ad una serie di equivoci sulla figura di Leonardo e il ritrovato interesse nel Settecento. Infine sono analizzati i riferimenti alle opere e le edizioni che hanno riportato in auge lo studio della matematica di Fibonacci.

La seconda parte della relazione si presenta come naturale prosecuzione della precedente: infatti, dopo il grande traguardo raggiunto con l'edizione critica cartacea del Giusti, vengono presentate le motivazioni per la realizzazione dell'edizione digitale, arricchita di un lavoro prettamente informatico che va a sfruttare strumenti dell'editoria digitale avanzati e tecniche di linguistica computazionale, al fine di scoprire e rendere disponibili la ricchezza di informazioni linguistiche, storiche e scientifiche contenute nel testo.

Oltre al suo interesse scientifico, il *Liber Abaci* presenta un lessico molto particolare e specialistico non solo dell'ambito matematico, ma anche commerciale del periodo medioevale. Proprio per le sue caratteristiche e le sue argomentazioni specialistiche, il testo rappresenta una risorsa unica per gli studiosi che lavorano su questo tipo di corpora latini. Le qualità impegnative del latino, in quanto lingua morfologicamente complessa con un ordine delle parole libero, e la scarsità dei dati in latino presentano una serie di sfide per le tecniche tradizionali di riconoscimento delle entità. Risolvere tali sfide consentirebbe di riconoscere in modo affidabile campi semantici rintracciabili in vari testi latini del medesimo argomento, consentendo a storici, linguisti, classicisti, archeologi, linguisti computazionali e altri studiosi di tenere traccia delle relazioni semantiche tra testi su larga scala.

Per questo scopo il lavoro fin dalla sua progettazione si è diviso in due processi di analisi distinti, morfo-sintattico e semantico, ma che si intersecano l'un l'altro per ottenere come risultato finale dei dati sufficientemente rappresentativi e consistenti,

con cui poter addestrare un algoritmo di Machine Learning sul corpus annotato e dunque sviluppare un modello di *Natural Language Processing* (NLP) per l'annotazione (semi)automatica.

Il lavoro presentato nella relazione sviluppa il processo di analisi semantica partendo dall'individuazione del dominio di interesse storico-linguistico, che riguarda la terminologia commerciale e mercantile con tutte le attività e questioni connesse.

Per il riconoscimento di tali dati è necessario un'annotazione manuale del testo. Vengono dunque descritte le fasi di annotazione di un capitolo del *Liber Abaci*, individuato per il suo interesse linguistico e storico: il capitolo VIII. In questo capitolo del trattato sono esposte proporzioni matematiche correlate a semplici esempi di trattative d'affari presi dal mondo commerciale. Gli esempi riguardano i sistemi monetari e ponderali, nonché le principali merci di vendita e scambio del Mediterraneo medievale. Questo presuppone che il testo sia ricco di terminologia specifica del dominio e dunque si presti bene ad essere individuato come capitolo campione per l'annotazione.

Il processo riguarda: la definizione del set di tag semantici per la realizzazione dell'annotazione manuale campione, la stesura delle linee guida da seguire per annotare manualmente e un'ulteriore annotazione sullo stesso testo, svolta da un secondo annotatore. La seconda annotazione viene definita annotazione di test ed è stata compiuta per valutare l'accordo tra gli annotatori e il livello di affidabilità dell'annotazione e delle linee guida.

Lo scopo della fase di annotazione è quello di ottenere un set di tag semantici il più possibile preciso e accurato nel descrivere la terminologia mercantile e commerciale di tutto il testo del *Liber Abaci*. Questo è stato verificato con l'estensione dei tag sui termini di altri capitoli di ambito mercantile e commerciale, ossia i capitoli IX e X.

Il risultato finale è un corpus di dati annotati del *Liber Abaci* per la valutazione e realizzazione di un modello supervisionato, utile per il riconoscimento delle entità mercantili e commerciali in latino.

*Parte 1 – Fibonacci e la sua opera*

# Capitolo 1

## 1. Leonardo Bigollo Pisano, il Fibonacci

### 1.1 La questione del nome

Pur essendo il più importante matematico nell'Occidente latino del XIII secolo, dell'autore del *Liber Abbaci* si hanno ben poche notizie biografiche e, in alcuni casi, anche contraddittorie e poco chiare. Il nome stesso non è di semplice ricostruzione: in alcuni casi i suoi concittadini lo denominavano Bigollo, tra i contemporanei e nei secoli successivi era noto come Leonardo Pisano, mentre oggi è universalmente conosciuto con il nome di Leonardo Fibonacci.

Il significato del *cognomen* Bigollo, che accompagna il nome di Leonardo in alcuni documenti ufficiali dell'epoca, nonché nell'*incipit* di alcune sue opere, come il *Flos* e qualche manoscritto della *Practica Geometriae*<sup>5</sup>, è stato ed è ancora oggi motivo di dibattito. A lungo si è ritenuto che il termine avesse un'accezione spregiativa da intendersi come “bighellone” secondo Guglielmini e Boncompagni, mentre in realtà significherebbe “bilingue” secondo un'ipotesi di Bonaini, oppure “trottola”, da cui il senso metaforico di “viaggiatore”, come invece lo interpreta Milanese<sup>6</sup>.

I pochi documenti coevi, largamente utilizzati come riferimento per estrapolare informazioni sulla vita di Leonardo, riportano lo stesso *cognomen* Bigollo, che probabilmente era l'appellativo con cui Leonardo era meglio conosciuto a Pisa, città in cui visse e probabilmente nacque.

Il primo documento è un rogito del notaio Bonafidanza stilato in data 28 luglio 1226 a Pisa, in cui Leonardo compare come *procurator* del fratello Bonaccingo per

---

<sup>5</sup> R. Franci, *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, (2002), p. 302.

<sup>6</sup> N. Rozza, *La tradizione manoscritta della Pratica Geometrie di Leonardo Pisano, detto Fibonacci, e la sua lettera di dedica la magister Dominicus*, in *Spolia: Journal of medieval studies*, Filologia e letteratura latina medievale e umanistica, (2005), p. 11.

l'acquisto di beni immobili situati nelle vicinanze di Pisa e viene citato come *Leonardo Bigollo quodam Guilielmi*. Da qui emerge un'altra importante informazione, ossia il nome del padre: Guglielmo e non Bonaccio, come in passato alcuni studiosi hanno erroneamente creduto.<sup>7</sup>

L'altro documento è una delibera del Comune di Pisa, inclusa nelle aggiunte al *Constitutum Pisanum legis et usus* del 1233, in cui gli Ufficiali del Comune assegnano a *Leonardo Bigollo* un salario annuo di 20 lire per i suoi servizi resi nelle materie d'abaco. Da questo riferimento, oltre a confermare l'appellativo Bigollo, possiamo dedurre, in accordo con Bonaini, che il riferimento non può ritenersi con sfumatura spregiativa, in quanto riportato in un documento che inteso a valorizzare l'operato di questo insigne concittadino<sup>8</sup>. Si trattava quasi certamente - come era uso all'epoca - di un soprannome, teso a distinguere il nostro Leonardo, dagli altri della sua epoca, poi fissatosi nel linguaggio comune di una città che certamente conosceva e apprezzava lo studioso<sup>9</sup>.

Nei secoli successivi, caduto in disuso il termine Bigollo, il matematico verrà costantemente indicato con il nome di Leonardo Pisano. Nel 1441 Cristofano di Gherardo di Dino nella sua volgarizzazione della *Practica Geometriae* scrive: "Qui incomincia la pratica della Geometria di M<sup>o</sup> Lunardo Pisano". Con il nome di Leonardo Pisano, il matematico viene anche citato nelle opere di altri studiosi: è il caso di Luca Pacioli nella sua *Summa de Arithmetica, Geometria, Proportioni e Proportionalità* (1494), Rafael Bombelli nell'opera *L'Algebra parte maggiore dell'Aritmetica* (1572), Muzio Oddi nella sua *Dello squadro* (1625)<sup>10</sup>.

Dal Settecento in poi, con la riscoperta della figura di Leonardo e delle sue opere, si diffonde anche l'appellativo Fibonacci. Nel catalogo manoscritto della

---

<sup>7</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), pp. xiii-xiv. Dall'introduzione di Giusti sono state tratte molte delle informazioni su Fibonacci e la sua opera riportate nella presente relazione.

<sup>8</sup> *Ibidem*.

<sup>9</sup> Salvatori E., *Il sistema antroponimico a Pisa tra XI e XIII secolo*, in "Melanges de l'École Française de Rome", 106/2 (1994), pp. 487-507; Ead., *Il sistema antroponimico a Pisa nel XIII secolo: la città e il territorio*, in "Melanges de l'École Française de Rome", 107/2 (1995), pp. 427-466.

<sup>10</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), pp. xiv-xv.

Biblioteca Magliabechiana, redatto tra il 1737 e il 1747, dal bibliotecario Giovanni Targioni Tozzetti si trova la notazione: *Leonardi Pisani de filiis Bonacci (Fibonacci)*, e così a seguire nelle opere di Pietro Cossali, Giovanni Grimaldi, Guglielmo Libri e Michael Charles viene attribuito il cognome Fibonacci<sup>11</sup>. Il termine deriverebbe dalla contrazione di *filius Bonacci*, deducendo da questo il fatto che il padre di Leonardo si chiamasse Bonaccio. Tuttavia, come visto nel già citato documento notarile, il vero nome del padre era Guglielmo, mentre Bonaccio indicava il nome di un avo, poiché, secondo un uso medievale, il cognome era una derivazione formata a partire dal nome di un antenato illustre, rispetto al quale tutti i discendenti erano genericamente definiti *filii*. Tale designazione è poi valsa al matematico pisano l'appellativo di Fibonacci che tutt'oggi lo rappresenta.

## 1.2 Vita e opere

Di Leonardo non sappiamo né la data né il luogo di nascita, tuttavia dal suo appellativo 'Pisano', possiamo presumere, con un buon grado di certezza, che la città natale fosse Pisa. Nei secoli XII e XIII, epoca in cui visse Leonardo, Pisa era una grande potenza militare nel Mediterraneo e rivestiva un ruolo essenziale nel sistema delle comunicazioni nazionali e internazionali: la via Francigena la collegava alla pianura padana, il mare le apriva le rotte per tutto il Mediterraneo.

Pisa, insieme a Genova, Venezia e Amalfi, era la protagonista della grande espansione marittima e commerciale: nelle città marinare i mercanti rappresentavano la figura chiave che plasmò il nuovo mondo più cosmopolita. Il commercio con i paesi a dominazione arabo-islamica era fiorente, tanto che a metà del XII secolo Pisa aveva colonie, privilegi marittimi e favorevoli trattamenti doganali e le navi pisane si

---

<sup>11</sup> B. Boncompagni, *Della vita e delle opere di Leonardo Pisano, matematico del secolo decimoterzo*. Atti dell'Accademia Pontificia dei Nuovi Lincei V (1852), pp. 7-10, visionato come da suggerimento in nota nell'introduzione di Giusti, 2020.

muovevano per tutto il Mediterraneo. La presenza pisana nel Maghreb, e soprattutto a Tunisi e a Bugìa, era importante e consolidata in tutto il XIII e XIV secoli<sup>12</sup>.

I trattati diplomatici pisano-musulmani confermano non solo l'attività relativa a conflitti politici per il dominio sui territori del Levante e il fenomeno endemico della pirateria, ma anche i rapporti commerciali. Una testimonianza è il carteggio diplomatico tra Abu-l-Abbas Ahmed, emiro delle città algerine di Bona e Bugìa e Giovanni dell'Agnello, doge di Pisa, datato al 1366, che mostra le intense e persistenti relazioni politiche e commerciali tra Pisa e alcuni centri del Maghreb: si tratta di una promessa di sicurezza e favori verso i mercanti pisani<sup>13</sup>.

Proprio in questo quadro si inserisce la figura del padre di Leonardo: Guglielmo era, infatti, un mercante diventato poi *publicus scriba*, una sorta di funzionario doganale per il Comune di Pisa. Come riporta lo stesso Leonardo nel prologo del *Liber Abaci*, il padre esercitava questa carica diplomatica per conto dei mercanti pisani nella dogana di Bugìa<sup>14</sup> e qui si fece raggiungere dal figlio, ancora bambino, per farlo istruire sulla matematica indo-araba, approfondendo lo studio dell'arte delle nove figure indiane, quelle che noi conosciamo come numeri arabi, e le conseguenze rivoluzionarie dell'applicazione di queste ad aritmetica e geometria.

Una volta acquisite queste conoscenze, Leonardo ne rimase talmente affascinato da approfondire gli studi durante i suoi viaggi nei luoghi dove lo portava la sua attività: Egitto, Siria, Grecia, Sicilia, Provenza, in poche parole in tutto il mediterraneo, dove il Comune di Pisa disponeva di proprie sedi commerciali e in cui fioriva una matematica in parte ancora sconosciuta nell'Occidente latino.

Le conoscenze progressivamente accumulate da Leonardo negli studi e nei viaggi nel Mediterraneo orientale confluirono nei suoi scritti: primo fra tutti il *Liber*

---

<sup>12</sup> M. Tangheroni, *Pisa e il Mediterraneo all'epoca di Fibonacci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016); N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), pp. 215-218.

<sup>13</sup> E. Banfi, *Lingue d'Italia fuori d'Italia. Europa, Mediterraneo e Levante dal Medioevo all'età moderna*, Bologna, Il Mulino, (2014), p. 105.

<sup>14</sup> Attuale Béjaïa, città in Algeria.

*Abaci*, redatto nella sua prima edizione nel 1202. I manoscritti che tramandano l'opera successiva di Leonardo, la *Practica geometrie*, collocano la sua stesura al 1220, altre opere di minor corposità, come il *Flos* e il *Liber Quadratum*, sono del 1225, l'*Epistola ad magistrum Theodorum* è di data incerta, ma probabilmente collocabile negli stessi anni, mentre, secondo alcuni manoscritti, è del 1228 la seconda edizione del *Liber Abaci*. Pare inoltre che Leonardo abbia composto altre due opere andate perdute: una dal titolo *De minore guisa*, forse una versione semplificata e più compatta del *Liber Abaci*, e l'altra è il *Commento al X libro degli Elementi di Euclide*.<sup>15</sup>

Importanti furono anche i contatti che Leonardo intrattenne con la schiera di dotti della corte di Federico II. Tra questi si trova il Maestro Domenico, di cui non si hanno notizie certe - forse identificabile nella figura di un certo Domenico Ispano - che probabilmente faceva parte della corte imperiale e introdusse Leonardo all'imperatore durante il suo soggiorno a Pisa nel 1226. L'incontro con Federico II e la permanenza a Pisa sono menzionate nell'*Liber Quadratorum*<sup>16</sup>, dove nell'epistola di dedica rivolta allo stesso Federico II, Leonardo ricorda come in quell'occasione il filosofo di corte, Giovanni da Palermo, gli propose un problema matematico per la cui trattazione ebbe origine l'opera.

Federico II di Svevia, imperatore del Sacro Romano Impero e re di Sicilia, oltre che per le sue capacità politiche e militari, era conosciuto come *stupor mundi* per la sua inestinguibile sete di sapere: era, infatti, un grande promotore dell'arte, della filosofia, della letteratura e delle scienze. Per questa sua propensione alla conoscenza, commissionava opere e traduzioni ai più noti sapienti dell'epoca e discuteva delle più disparate questioni con i numerosi dotti presenti nella sua corte. Intratteneva, inoltre, un'intensa corrispondenza epistolare con gli studiosi latini, arabi, greci ed ebrei con cui non poteva prendere diretto contatto.

L'incontro con l'imperatore segnò profondamente la vita di Leonardo, che conservò sempre strette relazioni con la cerchia di Federico II. Questo è ulteriormente

---

<sup>15</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 60.

<sup>16</sup> R. Franci, *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, (2002), p. 298.

testimoniato dalle dediche delle opere in cui sono menzionati personaggi importanti della corte sveva. Oltre al Maestro Domenico, a cui è dedicata la *Practica Geometriae*, sono ricordati il filosofo imperiale Michele Scotto, destinatario della dedica della seconda edizione del *Liber Abaci* e il Maestro Teodoro a cui è indirizzata l'*Epistola*.<sup>17</sup>

Sebbene ci siano pervenute diverse informazioni relative all'attività scientifica di Leonardo, nulla sappiamo del lavoro che svolgeva per vivere. La sua presenza è attestata intorno al 1200 a Costantinopoli, dove faceva l'interprete per l'avamposto commerciale pisano.<sup>18</sup> Si può ipotizzare che avesse continuato nel corso della sua vita ad esercitare l'attività del commercio, oppure è altrettanto verosimile che esercitasse a tutti gli effetti l'insegnamento privato. Inoltre, non si può escludere che svolgesse entrambi i lavori, di fatto esempi di mercanti dediti all'insegnamento si trovano non di rado nei secoli successivi<sup>19</sup>.

Dopo il 1228, possibile anno della revisione del *Liber Abbaci*, non si hanno ulteriori notizie biografiche del matematico pisano e non è neppure possibile individuare la data e il luogo di morte. Tuttavia, grazie alla già citata delibera del Comune di Pisa del 1241 che ci indica l'attività di Leonardo a Pisa in quel periodo, possiamo fissare con questa data il *terminus post quem* per la sua morte.<sup>20</sup>

### 1.3 Mediatore della matematica araba

Nel Medioevo i paesi che si affacciano sul bacino Mediterraneo orientale assumono una posizione di grande rilievo non solo dal punto di vista economico e commerciale, ma anche culturale, rappresentando il principale ponte di trasmissione

---

<sup>17</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), p. xix.

<sup>18</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), p. 219.

<sup>19</sup> R. Franci, *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, (2002), p. 301.

<sup>20</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), p. xiv.

del sapere teorico e pratico delle civiltà antiche, in particolar modo quella greca, indiana, babilonese ed anche cinese.

In particolare gli arabi nei secoli VII e VIII furono protagonisti di un'irrefrenabile avanzata che li portò a conquistare un impero vastissimo, esteso dalle coste del Vicino Oriente, all'Africa settentrionale fino alla Spagna, gettando le basi non solo di un lungo dominio, ma anche di una variegata e ricca civiltà.

All'epoca la matematica era considerata unicamente nella sua concezione prettamente pratica legata ad esigenze di vita quotidiana. Solo con l'ascesa al potere della dinastia degli Abbasidi prese avvio il grande sviluppo delle scienze raggiunto dagli arabi. I secoli tra i X e il XIII vedono un periodo di straordinaria fioritura soprattutto nel campo matematico. Quest'epoca d'oro, densa di risultati originali, si avvia in seguito allo studio e all'assimilazione di testi greci e indiani e delle culture di altri popoli conquistati. Nell'ambito degli studi scientifici, si assiste ad un fecondo connubio fra la matematica orientale, indiana e babilonese, e quella occidentale greco-ellenistica. L'unione di questi saperi, geograficamente divisi, prende ed assimila dalla matematica indiana conoscenze aritmetiche, tra cui l'utilizzo del sistema di numerazione posizionale e l'introduzione dello zero, mentre da quella greco-ellenistica le trattazioni sulla geometria piana e solida e sui fondamenti logici e filosofici delle scienze. <sup>21</sup>

Grazie alle traduzioni effettuate nei paesi islamici dei testi greci e indiani e grazie ai contributi originali degli studiosi arabi, un ricchissimo patrimonio di manoscritti greci e arabi di ambito scientifico, raggiungono nel Medioevo l'Occidente latino, incentivando la ripresa delle attività intellettuali e la rinascita del sapere scientifico. <sup>22</sup>

In questo quadro si inserisce la figura di Leonardo, il più grande mediatore della fiorente cultura araba, appresa in prima persona in tutto il mondo mediterraneo, prima con il suo apprendistato a Bugìa, poi nei suoi viaggi. La conoscenza di Leonardo deriva non da un autore o da una scuola, bensì dalla matematica araba nel suo

---

<sup>21</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), pp. 35-38.

<sup>22</sup> *Ibidem*.

complesso studiata, assimilata e rielaborata in una *summa* del sapere matematico di origine indiana e mediazione araba, il *Liber Abaci*.

Il grande trattato matematico di Leonardo rappresenta un'opera che non ha antecedenti in Europa, in cui non mancano le acquisizioni da opere precedenti, in particolare, in alcuni tratti risultano evidenti le tracce di autori arabi. L'importanza dell'opera non risiede tanto nella sua originalità scientifica, che nel complesso risulta alquanto limitata a piccoli passaggi, quanto nella capacità di saper raccogliere, integrare e rendere disponibile nell'Occidente latino tali fondamentali conoscenze derivanti dalle scienze arabe.<sup>23</sup>

L'opera riguarda le questioni più elementari elaborate dagli autori islamici: non ci sono riferimenti alle ricerche sulla geometria archimedeica, che hanno avuto grande importanza nel progresso della matematica araba del X e XI secolo, ed anche per quanto riguarda la trattazione più strettamente algebrica e aritmetica, sembra non prendere spunto dagli autori più recenti e avanzati come al-Khayyam o al-Karaji<sup>24</sup>. Leonardo infatti si rifà più ampiamente ad una tradizione antica dei secoli IX e X, che vede come fonti al-Khwarizmi e Abu Kamil.

Al-Khwarizmi è il matematico a cui si deve la prima enunciazione del sistema di numerazione indiano e delle operazioni effettuate in questo sistema. Delle sue opere, le due principali sull'aritmetica e sull'algebra hanno esercitato una grande influenza, oltre che sugli studiosi arabi successivi, anche sullo sviluppo della matematica medievale in Occidente.

Il termine *algorithmus*, che compare nel trattato di aritmetica di al-Kawarizmi, deriva dal nome dell'autore latinizzato in *algorismus* e poi trasformato nella forma

---

<sup>23</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 59 e p. 111.

<sup>24</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), pp. 225-227. Ambrosetti riferisce che alcuni studiosi (Hughes, 1996; Giusti 2002) sono dell'opinione che Leonardo nell'opera ignori deliberatamente tutta l'evoluzione dell'algebra e dell'aritmetica arabe dei secoli più recenti; Maracchia (Maracchia 2005) propone la mancata conoscenza da parte di Leonardo dell'opera di al-Kayyam. La Hunger Parshall (Hunger Parshall, 1988) ritiene che Leonardo conoscesse l'opera di al-Karaji e la Franci (Franci, 2002) argomenta l'ipotesi che, dei 96 problemi risolti nel *Liber Abaci*, alcuni sono certamente derivati dall'opera di al-Karaji.

attuale a causa di una falsa etimologia, andando dunque a designare fino al XVII secolo il sistema di numerazione posizionale e successivamente un generico procedimento sistematico di calcolo.

Il trattato di algebra invece, composto nella prima metà del IX secolo, è a tutti gli effetti considerato l'atto di nascita della disciplina. L'algebra del matematico persiano è retorica: non usa alcun simbolo e le sue spiegazioni sono estremamente prolisse. Lo scopo principale di al-Kawarizmi era di comporre un manuale utile alla risoluzione dei problemi della vita quotidiana, tuttavia l'opera ha ottenuto una riconoscenza e una diffusione ben più ampia.<sup>25</sup>

La teoria algebrica elaborata da al-Kawarizmi viene completata ed ampliata dall'egiziano Abu Kamil nel suo trattato. L'opera, che riguarda essenzialmente le questioni teoriche sulle equazioni di primo e secondo grado, fu letta e ripresa da Leonardo in diversi problemi riportati in modo integrale nel *Liber Abaci*.<sup>26</sup> In particolare la conoscenza di Abu Kamil permise al matematico pisano di estendere quanto appreso da al-Kawarizmi.<sup>27</sup>

---

<sup>25</sup> C. S. Roero, *Algebra e Aritmetica nel Medioevo islamico*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), pp. 20-22.

<sup>26</sup> C. S. Roero, *Algebra e Aritmetica nel Medioevo islamico*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 30.

<sup>27</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 59 e pp. 109-111.

# Capitolo 2

## 2. Il Liber Abaci

L'opera preminente della produzione del Fibonacci è senza dubbio il corposo trattato di aritmetica e algebra, che la tradizione manoscritta ci ha consegnato con il titolo di *Liber Abaci*. Il termine *abacus*, che indicava originariamente lo “strumento per contare”, cioè l'abaco a tavolette, nel Medioevo muta il suo significato andando ad indicare l'“arte del contare”, in riferimento al calcolo aritmetico basato sull'uso delle figure indiane.<sup>28</sup>

Il *Liber Abaci* ha avuto due stesure: la prima risale con certezza al 1202, dal momento che tutti i manoscritti conosciuti dell'opera sono concordi nel fissare la sua composizione in tale data. Meno certa, ma ritenuta tale, è la data della seconda stesura, che è riportata solo in alcuni manoscritti ed è il 1228. A riferirsi chiaramente ad una seconda edizione è ad esempio l'incipit del manoscritto Riccardiano<sup>29</sup>: *Incipit liber abbaci compositus a Leonardo filiorum Bonaccii Pysano in anno MCCII et correctus ab eodem XXVIII*, mentre altri manoscritti, come il Magliabechiano<sup>30</sup>, riportano solo la data della prima stesura. Tuttavia anche questi manoscritti sono comunque frutto della revisione: fatto rilevato sia dalla lettera di dedica a Scotto, sia dal riferimento nel testo ad un'altra opera, la *Practica Geometrie*, composta nel 1220, dunque il testo tramandato deve essere necessariamente posteriore a questa data.

L'accurata revisione del *Liber* nasce dalla richiesta da parte di Michele Scotto di una copia del libro, circostanza che diede occasione a Leonardo, dopo diversi anni dalla composizione, di tornare sull'argomento per aggiungere alcune nozioni necessarie ed eliminarne altre superflue<sup>31</sup>.

---

<sup>28</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), pp.97-98.

<sup>29</sup> Codice Riccardiano è conservato nella Biblioteca Riccardiana di Firenze.

<sup>30</sup> Codice Magliabechiano è conservato nella Biblioteca Nazionale Centrale di Firenze.

<sup>31</sup> “Scripsistis mihi domine mi et magister Michael Scotte, summe philosophe, ut librum de numero quem dudum composui vobis transcriberem; unde vestre obsecundans postulationi, ipsum subtiliori

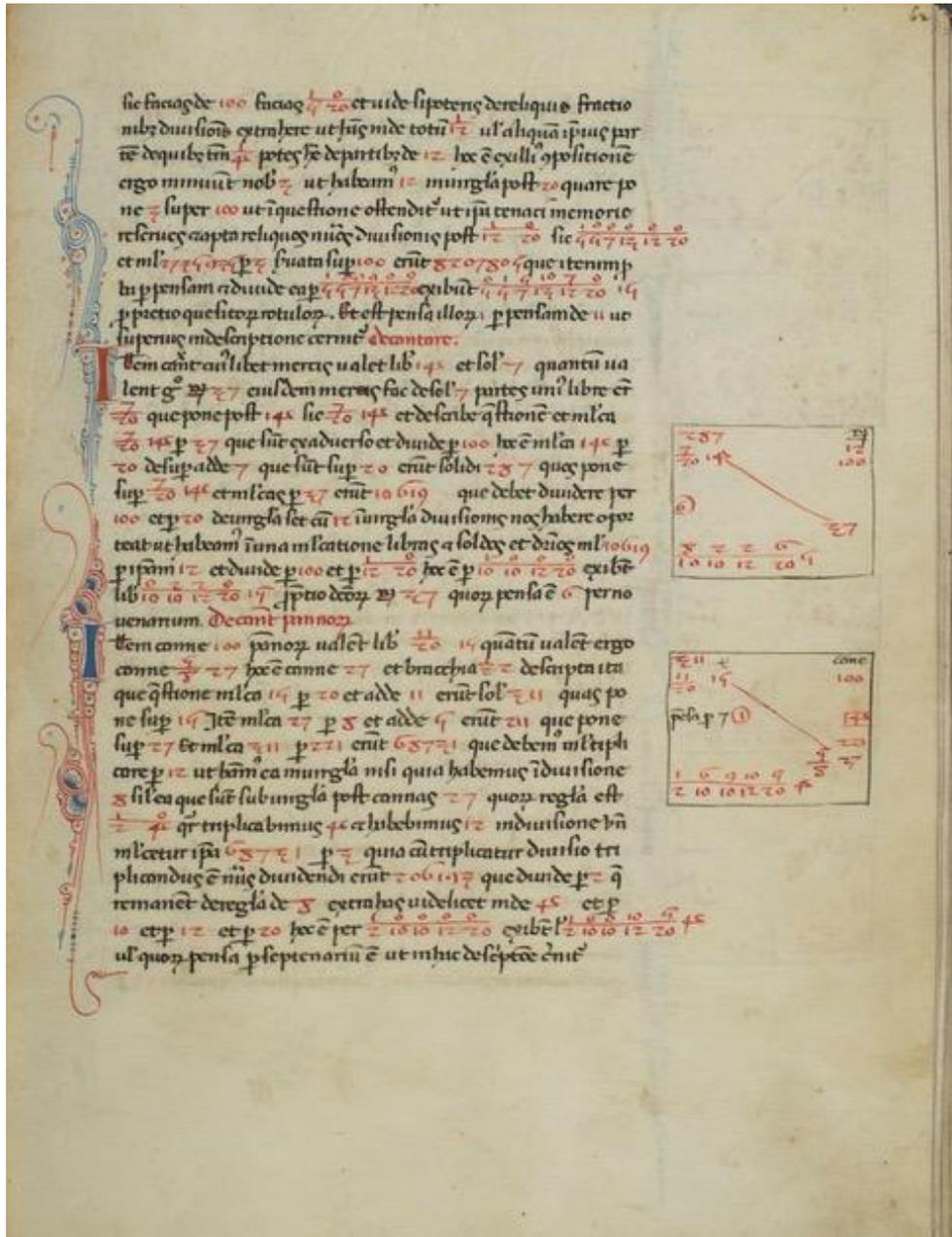


Figura 1 - Immagine del manoscritto Magl. cap. VIII.

L'obiettivo dell'opera, come indicato esplicitamente dall'autore nel prologo, è quello di dare una conoscenza più completa possibile della materia, portando anche le *gens latine* a conoscere le cifre indo-arabe, per mostrarne la superiorità rispetto ai

perscrutans indagine, ad vestrum honorem et aliorum multorum utilitatem correxi. In cuius correctione quedam necessaria addidi et quedam superflua resecavi.” (*Liber Abaci*, p.3)

metodi di calcolo macchinosi e arretrati in uso fino a quel momento nel mondo Occidentale.

Sulla base di queste considerazioni iniziali e del contenuto risulta evidente il carattere enciclopedico dell'opera, che si compone di un prologo e di 15 capitoli così suddivisi in maniera sistematica:

1. *Le nove cifre indiane, e come con esse si scrivano tutti i numeri. Quali numeri si possano tenere nelle mani, e come. Introduzione all'abaco.*
2. *La moltiplicazione degli interi.*
3. *L'addizione degli stessi.*
4. *La sottrazione dei numeri minori dai maggiori.*
5. *La divisione di numeri interi per numeri interi.*
6. *La moltiplicazione degli interi con le frazioni, e delle frazioni senza interi.*
7. *La somma, la sottrazione e la divisione degli interi con le frazioni e la riduzione delle parti di numeri in parti singole.*
8. *L'acquisto e la vendita delle merci e simili.*
9. *I baratti delle merci, l'acquisto delle monete, e alcune regole simili.*
10. *Le società fatte tra consoci.*
11. *La fusione delle monete e le regole relative.*
12. *La soluzione di questioni diverse, dette miscellanee.*
13. *La regola della doppia falsa posizione, e come con essa si risolvano pressoché tutte le questioni miscellanee.*
14. *Il calcolo delle radici quadrate e cubiche per moltiplicazione e divisione o da estrazione, e il trattato dei binomi e recisi e delle loro radici.*
15. *Le regole delle proporzioni geometriche. Questioni di algebra e almuchabala.*

In base all'argomento è possibile dividere concettualmente l'opera in quattro parti<sup>32</sup>. La prima, che copre i primi sette capitoli, è dedicata alla presentazione dei fondamenti dell'aritmetica (la notazione posizionale con le cifre indo-arabe e i relativi algoritmi per eseguire le quattro operazioni fondamentali con i numeri interi e

---

<sup>32</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), pp. 63-65.

razionali) ed è sviluppata senza alcun riferimento ad applicazioni reali pratiche, ma attraverso esempi formali con numeri via via più grandi.

La seconda parte, compresa nei capitoli dall'VIII all'XI, mostra l'utilità pratica del calcolo aritmetico, applicandola alla risoluzione di questioni e problemi relativi all'esercizio della mercatura: acquisto, la vendita e il baratto di merci, il cambi tra monete, le società, la fusione delle monete. In questa parte, che costituisce un vero e proprio manuale di pratiche commerciali, ogni problema viene svolto in tutti i suoi calcoli fin nei minimi dettagli.

I capitoli XII e XIII costituiscono la terza parte dell'opera, quella più estesa, e presentano problemi di vario genere (tra cui il "problema dei conigli" che darà luogo ai famosi numeri di Fibonacci) e la risoluzione della maggior parte di questi problemi riconducibile al metodo della doppia falsa posizione.

Infine, con la quarta parte, individuabile negli ultimi due capitoli, si passa a questioni sempre più astratte, comprendendo la trattazione dell'estrazione di radici quadrate e cubiche, i binomi e recisi, la teoria delle proporzioni geometriche e l'algebra.

L'ultima parte insieme alla prima costituiscono, nel loro complesso, quello che può essere considerato il primo trattato d'algebra scritto in lingua latina<sup>33</sup>.

Una caratteristica stilistica del *Liber Abaci*, e in generale comune a tutta la trattatistica matematica successiva, è la risoluzione completa ed elaborata di ogni problema subito dopo la sua enunciazione. Inoltre i procedimenti aritmetici e algebrici sono descritti interamente e minuziosamente a parole, gli unici simboli matematici che si trovano nel testo sono quelli relativi ai numeri; d'altro canto i segni matematici furono introdotti e divennero in uso comune solo alcuni secoli più tardi.

Ai margini delle pagine, e in particolare dei passi dedicati a calcoli di aritmetica pratica, compaiono schemi di calcolo che danno un'esemplificazione dei procedimenti: nei capitoli sull'attività della mercatura, come ad esempio nell' VIII capitolo, spesso si trova applicata, sugli esempi presentati nel testo, la "regola del tre", strumento per la risoluzione dei problemi del settore, dal momento che «in tutte le

---

<sup>33</sup> R. Franci, *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, (2002), p. 314.

operazioni commerciali si trovano sempre quattro numeri proporzionali, dei quali tre sono noti, il restante è ignoto».<sup>34</sup>

## 2.1 L'eredità del *Liber Abaci* nelle pratiche mercantili

L'impatto che il *Liber Abaci* ha avuto sulle *gens latine* va analizzato alla luce della situazione intellettuale occidentale che, dopo la caduta dell'Impero romano d'Occidente, è stata caratterizzata da un lungo periodo di decadenza culturale, nella quale furono dimenticate la maggior parte delle opere letterarie e scientifiche della cultura classica greca e romana.

Dunque un'opera tanto importante, quanto complessa per la materia trattata, come il *Liber Abaci*, si inserisce in una cultura matematica estremamente arretrata e incapace di comprendere appieno e applicare procedimenti e regole innovative che venivano presentati.

Allo stesso modo, gli esempi applicativi contenuti nel *Liber* riguardo le attività commerciali, inizialmente non produssero effetti: basti pensare che ancora nel XIII secolo si registra una certa diffidenza e resistenza proprio da parte dei mercanti all'uso delle cifre arabe rispetto ai numeri romani, dal momento che, non avendo familiarità con i nuovi numeri, li ritenevano uno strumento manipolabile con cui poter essere ingannati.

Dunque è prevedibile il fatto che il boom del calcolo con le cifre indo-arabe non seguisse immediatamente la pubblicazione del *Liber Abaci*, ma al contrario dovesse trascorrere quasi un secolo prima che la praticità e i benefici del nuovo metodo si affermassero davvero tra i mercanti.

Bisogna attendere gli inizi del XV secolo per avere riscontri concreti dell'influenza dell'opera di Leonardo sullo sviluppo della matematica in Italia e nell'Europa occidentale. Questa nuova svolta è dovuta all'espansione e

---

<sup>34</sup> “*In omnibus itaque negotiationibus quattuor numeri proportionales semper reperiuntur, ex quibus tres sunt noti, reliquus vero est ignotus.*” (*Liber Abaci*, 2020, p. 141).

all'intensificazione dei traffici commerciali che hanno portato allo sviluppo di vere e proprie imprese che gestivano, attraverso società e compagnie, estese reti nel Mediterraneo. Il commercio diventa dunque molto articolato, portando allo scambio e alla vendita di merci sempre più diversificate, come diversificato è il loro valore monetario. Necessità impellente era dunque avere un collaudato sistema di contabilità che gestisse la conduzione degli affari, l'andamento dei mercati, le informazioni sugli arrivi e le partenze delle navi e sui cambi e le quotazioni delle merci.

Dall'espandersi delle dimensioni delle imprese commerciali e dalle conseguenti necessità contabili derivano quindi le motivazioni per la diffusione delle cifre indo-arabe e dei corrispondenti metodi di calcolo, di cui il *Liber Abaci* fu grande promotore.

I mercanti iniziarono a redigere metodicamente le scritture contabili e la mercatura diventò l'unica professione non dotta che imponeva a chi la praticava il ricorso assiduo alla scrittura<sup>35</sup>.

### 2.1.1 La lezione di Fibonacci nelle scuole d'abaco

La necessità di apprendere i metodi del sistema di calcolo indo-arabo si lega alla necessità di un insegnamento sistematico delle principali nozioni in scuole apposite; per questo a partire dalla fine del Duecento vengono istituite delle vere e proprie scuole d'abaco nei principali centri attivi economicamente, in cui venivano impartiti insegnamenti sulla materia dai maestri d'abaco.

A seguito della diffusione delle scuole d'abaco, i maestri d'abaco, con la formazione di una classe sociale di insegnanti-trattatisti, possono legittimamente essere considerati i primi matematici professionisti dell'Occidente, dal momento che ottenevano il sostentamento interamente dalle loro stesse capacità intellettuali matematiche. Il primo maestro d'abaco fu probabilmente lo stesso Leonardo, incarico

---

<sup>35</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), pp. 113-114.

per cui ottenne, seppur come insegnamento a livello privato, anche il salario annuo dal Comune di Pisa.<sup>36</sup>

Attorno alle scuole fiorisce conseguentemente un'attività sistematica di produzioni di manuali, che vengono definiti trattati d'abaco, nei quali è evidente l'influenza del *Liber Abaci*. Tutti questi testi tuttavia si differenziano dal loro modello in quanto si tratta di opere divulgative e dunque meno voluminose, più elementari e semplici, con nozioni meno teoriche e astratte e scritte in lingua volgare. Il più antico trattato d'abaco che ci è pervenuto risale alla fine del XIII secolo ed è il *Livro de l'abbecho secondo la oppensione de maestro Leonardo de la chasa degli figliuogle Bonacie da Pisa*<sup>37</sup> di un anonimo maestro umbro, che già dal titolo indica come riferimento l'opera di Leonardo. I successivi trattati d'abaco trecenteschi sono in assoluta maggioranza toscani con chiari rimandi al modello pisano<sup>38</sup>.

L'affinarsi delle conoscenze matematiche, naturale conseguenza dell'assimilazione della lezione fibonacciana, via via porta alla stesura di trattati di più ampio respiro, capaci di confrontarsi con l'opera di Leonardo per mole e per quantità di nozioni descritte.

I nuovi trattati, inoltre, ampliano il raggio di conoscenze e nozioni matematiche, apportando nuovi problemi e nuovi metodi, ad esempio vediamo come l'attenzione degli abachisti si sposta precocemente dalla risoluzione delle equazioni di secondo grado, contenute nel *Liber Abaci*, a quelle di terzo e quarto grado.

## 2.2 La fortuna di Leonardo tra autorità e oblio

Il successo del *Liber Abaci* si rivelò alla lunga notevole: grazie ai mercanti pisani, che diffusero il nuovo metodo di calcolo sulle rive del Mediterraneo e nei

---

<sup>36</sup> Riferimento alla più volte citata Delibera del 1241.

<sup>37</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), p. 251.

<sup>38</sup> P. Manni, *La matematica in volgare del Medioevo (con particolare riguardo al linguaggio algebrico)*, in Gualdo R. (a cura di), *Le parole della scienza. Scritture tecniche e scientifiche in volgare (secoli XIII-XIV)*. Atti di Convegno Galatina, (2001), pp. 127-152.

contatti con commercianti del resto dell'Occidente. Le nuove tecniche matematiche infatti giunsero in Germania, in Francia, in Catalogna, nelle Fiandre, in Scandinavia, in Inghilterra e in Grecia. Il *Liber Abaci* diventò lo spunto per gli studi di matematica pratica commerciale e per l'insegnamento anche delle lingue nazionali, incoraggiando la stesura di testi di aritmetica e di algebra a livello specialistico<sup>39</sup>.

Contemporanei e posteri utilizzarono le opere fibonacciane per studiare, consultare, riprendere passi ed estrapolare esempi, problemi e metodi risolutivi, tanto da potersi dire "liberamente saccheggiate" fino alla fine del Quattrocento<sup>40</sup>. Tra gli ultimi matematici a farvi riferimento troviamo Luca Pacioli che, nella sua *Summa de arithmetica, geometria, proportioni e proportionalità* del 1494, si avvale delle scoperte dei suoi predecessori, attingendo a numerose fonti, tra cui ampiamente al *Liber Abaci* di Leonardo, per costruire una vera e propria enciclopedia del sapere abacistico.

Verso la metà del XV secolo l'invenzione della stampa cambia radicalmente la situazione: il libro a stampa ottiene subito un grande successo e si iniziano a pubblicare un gran numero di copie di testi che andavano così a sostituire il codice manoscritto. Così progressivamente gli autori che avevano preceduto l'invenzione della stampa o che non avevano utilizzato edizioni a stampa per la diffusione delle proprie opere, vengono dimenticati e sostituiti dai loro successori. Questo è il destino a cui sono andate incontro anche le opere di Leonardo che, nonostante siano state un punto di riferimento del sapere matematico di tutto il Medioevo, rimasero manoscritte e dunque progressivamente sostituite da altri testi realizzati in edizioni a stampa<sup>41</sup>.

Grazie anche all'esplicito riferimento al *Liber Abaci* nella *Summa* di Pacioli, la figura di Leonardo rimane ancora viva nella memoria del XVI secolo, tuttavia

---

<sup>39</sup> N. Ambrosetti, *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo nell'Europa medievale*, Milano, Edizioni Universitarie di Lettere, Economia, Diritto, (2008), p. 249.

<sup>40</sup> L. Pepe, *La riscoperta di Leonardo Pisano*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 161.

<sup>41</sup> E. Giusti, *Matematica e commercio nel Liber Abaci*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), pp. 116-117.

sempre più attenuata. In quel periodo ed ancora nel XVII secolo gli eruditi mescolavano informazioni di prima e seconda mano fino a dare luogo ad una serie di errori riguardo lo stesso secolo in cui visse Leonardo<sup>42</sup>. Bombelli lo nomina nella sua *Algebra* insieme a Pacioli, Tartaglia ne ha conoscenza solo attraverso quanto scritto da Pacioli, Cardano nel *De Consolatione* lo colloca qualche anno prima di Pacioli, Commadino intendeva pubblicare la *Practica Geometriae* e il suo allievo Baldi nella sua *Cronica de' matematici* parlando di Leonardo, lo colloca circa due secoli dopo, ed ancora Muzio Oddi cita direttamente Leonardo nel suo *Dello squadro*.

Non differente erano le informazioni che circolavano tra gli eruditi stranieri, basate sulle inesattezze cronologiche di Cardano e di Baldi.<sup>43</sup> Anche nel primo volume dell'*Encyclopédie* di Diderot e d'Alembert alla voce *Algèbre* Pacioli è indicato, se pur con qualche riguardo, discepolo di Leonardo nel XV secolo<sup>44</sup>.

## 2.3 Riscoperta settecentesca

A restituire Leonardo al suo secolo, ma soprattutto a riportare in auge il suo nome fu la grande stagione dell'erudizione italiana del XVIII secolo; in particolare, un catalogo dei manoscritti di argomento matematico presenti nella biblioteca Magliabechiana, redatto da Targioni Tozzetti e riportato da Francesco Antonio Zaccaria nel suo *Excursus Litterarius per Italiam ab anno MDCCXLII ad annum MDCCLII*. Nel catalogo sono segnalati due esemplari del *Liber Abaci*, uno del XIV e l'altro del XV secolo. Al codice più antico è dedicata la maggiore attenzione, riportandone il prologo con l'inequivocabile elemento di datazione dell'opera e l'indice dei capitoli.

---

<sup>42</sup> G. Loria, *Leonardo Fibonacci*, in Mieli A. (a cura di), *Gli scienziati italiani*, vol. I. Roma Nardecchia, (1921), pp. 4-12.

<sup>43</sup> A tal proposito L. Pepe, *La riscoperta di Leonardo Pisano*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), p. 163, indica che gli studiosi Vossius e Neilbronner nei loro scritti collocano Leonardo nel XV secolo.

<sup>44</sup> *Ibidem*.

Il volume di Zaccaria ebbe una vasta diffusione, ristabilendo così la corretta datazione dell'opera di Leonardo che viene poi riportata anche negli scritti di altri eruditi dell'epoca, tra cui Grimaldi, il primo ad osservare che del *Liber Abaci* ne esistevano due redazioni, datando la seconda al 1228 sulla base del codice Riccardiano, Andres e Tiraboschi<sup>45</sup>.

Successivamente l'attenzione si sposta sul problema della valutazione critica dell'opera di Leonardo e sul suo ruolo nella trasmissione del sapere matematico in Occidente, questione che fu messa in luce per la prima volta dai due storici della matematica Pietro Cossali e Giambattista Guglielmini.<sup>46</sup>

Cossali, nella sua opera di riferimento per la storia dell'algebra dal titolo *Origine, trasporto in Italia, primi progressi in essa dell'algebra* (1797), analizza con grande cura il *Liber Abaci* e ne dà una datazione certa agli inizi del XIII, dopo aver avuto modo di consultare almeno tre manoscritti dell'opera del matematico pisano (l'Ambrosiano di Milano, il Magliabechiano e Riccardiano di Firenze).

Guglielmini redige il saggio *Elogio di Lionardo Pisano* che, nella sua versione a stampa del 1813, assume il ruolo di opera di grande rilevanza e accuratezza, con particolare attenzione alla storia della matematica intesa come parte della storia delle civiltà, strettamente legata all'evoluzione politica: così anche lo studio sulla figura di Leonardo viene indagato a partire dal contesto politico e culturale del suo tempo.

Nella prima metà dell'Ottocento nell'ambito di un diffuso interesse per il Medioevo, si sviluppò la contesa tra due grandi storici della matematica del periodo, Guglielmo Libri e Michel Chasles, riguardo la storia dell'aritmetica medievale e la peculiarità dell'opera di Leonardo nella storia delle scienze matematiche.<sup>47</sup>

Chasles considerava trascurabile il debito dell'Occidente rispetto all'aritmetica araba, pensiero fortemente contestato da Libri che, invece, sosteneva l'idea che Fibonacci, Sacrobosco, Giordano Nemorario avessero diffuso in Europa l'aritmetica appresa dagli arabi, che avevano sempre chiamato indiana. Inoltre, Libri contestava

---

<sup>45</sup> *Ivi*, pp. 162-166.

<sup>46</sup> L. Pepe, *La riscoperta di Leonardo Pisano*, in E. Giusti, *Un ponte sul Mediterraneo: Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*, Firenze, Polistampa, (2016), pp. 166-168.

<sup>47</sup> *Ivi*, pp. 169-171.

anche il giudizio riduttivo di Chasles riguardo l'opera di Leonardo, ribadendo che proprio a Leonardo dobbiamo la conoscenza dell'algebra: è lui che ha introdotto o almeno diffuso in Occidente il sistema aritmetico degli indiani<sup>48</sup>.

## 2.4 Edizioni cartacee

Il rinnovato interesse per la figura di Leonardo porta, in seguito, verso la metà del XIX secolo, ad un passo definitivo per la diffusione dei suoi scritti con la pubblicazione della prima edizione a stampa ad opera di Baldassare Boncompagni. Lo studioso e storico della matematica Boncompagni iniziò a raccogliere le informazioni riguardanti Leonardo e i codici delle sue opere e a pubblicarne i risultati prima negli Atti di Accademia Pontificia dei Nuovi Lincei, *Della vita e delle opere di Leonardo Pisano*, e sul Giornale Arcadico, poi in un volume completo e articolato dal titolo *Intorno ad alcune opere di Leonardo Pisano, matematico del secolo decimoterzo* pubblicato a Roma nel 1854. Nello stesso anno iniziò inoltre ad occuparsi del primo lavoro di edizione sulle opere del matematico pisano.

Nel 1857 e nel 1862 Boncompagni pubblicò, in due monumentali volumi in quarto, due opere cardine del nostro autore: il primo di questi volumi contiene l'intero corpus del *Liber Abaci*, stampato con il titolo *Il Liber Abbaci di Leonardo Pisano pubblicato secondo la lezione del codice Magliabechiano C. I. 2616*, in *Scritti di Leonardo Pisano, matematico del secolo decimoterzo pubblicati da Baldassarre Boncompagni*.

L'*editio princeps* del *Liber Abaci* pubblicata da Boncompagni, tuttavia non è un'edizione critica: lo studioso non fece una ricerca scientifica, ampia e approfondita, sui manoscritti che tramandano l'opera, ma si limitò a trascrivere diplomaticamente il codice Magliabechiano che aveva tra le mani e vi aderisce totalmente conservando gli

---

<sup>48</sup> Come riporta R. Franci, *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, (2002), p. 304: "Nous devons la connaissance de l'algèbre, c'est lui qui a introduit ou au moins répandu chez le Chrétiens, le système arithmétique des Hindous".

errori di scrittura e le incongruenze nei risultati dei calcoli matematici, che non si curò - o in alcuni casi forse non fu in grado - di controllare e correggere; solamente gli errori più evidenti sono segnalati con un (*sic*) sul testo. Dunque, il frutto del lavoro di Boncompagni risulta poco affidabile non solo dal punto di vista filologico, ma anche storico e matematico.<sup>49</sup>

A distanza di più di un secolo, con pubblicazione postuma nel 2002, l'americano Laurence Sigler<sup>50</sup> propone la prima traduzione dell'opera latina in una lingua moderna, come l'autore stesso afferma nell'introduzione alla sua edizione. La versione inglese è tratta dall'edizione di Boncompagni, di cui Sigler corregge gran parte degli errori, anche se spesso la traduzione si discosta anche di molto dal testo originale.

Per avere la prima edizione critica bisogna aspettare i giorni nostri: infatti è nel 2020 che Enrico Giusti, coadiuvato da Paolo D'Alessandro,<sup>51</sup> pubblica l'edizione critica completa del *Liber Abaci*.

Si tratta del frutto di un importante lavoro filologico che ha previsto la valutazione di tutti i manoscritti disponibili, completi e quasi completi, per un totale di diciannove, scaglionati fra il XIII e il XIX secolo, e la collazione dei più significativi. Anche dal punto di vista matematico, è stato fatto un accurato controllo di tutti i calcoli, correggendo gli errori quando possibile o riportando le incongruenze nelle note. Il risultato è un testo attendibile sia dal punto di vista matematico che da quello storico e linguistico<sup>52</sup>.

Questa nuova edizione è uno strumento prezioso e un punto di arrivo fondamentale per la vasta comunità internazionale di studiosi dei diversi ambiti matematici e umanistici. Tuttavia, guardando al nuovo formato editoriale, per le

---

<sup>49</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), p. xxvii-xxviii.

<sup>50</sup> Sigler L. E., *Fibonacci's Liber Abaci. A translation into Modern English of Leonardo Pisanos's book of calculation*. New York, Springer, 2002.

<sup>51</sup> *Leonardi Bigolli Pisani vulgo Fibonacci, Liber Abbaci*, edito da Enrico Giusti adiuante Paolo D'Alessandro, Firenze, Olschki, 2020.

<sup>52</sup> E. Giusti, D'Alessandro P., *Leonardo Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki, (2020), p. xxvii-xxviii.

caratteristiche intrinseche dell'opera, non poteva che risultare un volume compatto e voluminoso: si tratta di un volume di cm 17x24, cxviii-824 pp. 22 tavole f.t. a colori pp., inoltre il testo è stampato su carta indiana, rilegato in seta con impressioni dorate che vanno ad impreziosire un'opera della tal portata. Dunque si tratta di un'edizione non certo maneggevole e di facile analisi e, anche per il costo, è chiaramente indirizzata ad una nicchia di mercato estremamente ristretta.

L'opera in sé è corposa, complessa e presenta alcuni problemi riguardo non solo la comprensione dei metodi di calcolo aritmetico, ma anche la corretta interpretazione e contestualizzazione dei termini legati al mondo della mercatura medievale, utilizzati da Leonardo nell'illustrare problemi matematici.

Per questo motivo l'edizione critica a stampa ha tardato ad essere pubblicata nella ricerca di un editore che avesse le competenze matematiche, storiche e filologiche per affrontare il compito di realizzare un'edizione completa basata sull'intera tradizione manoscritta del *Liber Abaci*.

## *Parte 2 – Il progetto*

# Capitolo 1

## 1. Edizione digitale: il progetto

### *Fibonacci1202-2021*

Il grande risultato editoriale di Giusti ha poi aperto le porte ad un nuovo progetto che prende il nome di *Fibonacci1202-2021*. Si tratta di un progetto di ricerca congiunto, tra il Museo Galilei di Firenze e l'Università di Pisa e diretto da Pier Daniele Napolitani del Dipartimento di Matematica di Pisa, per trasformare l'edizione critica cartacea in un'edizione critica digitale.

L'obiettivo della realizzazione di un'edizione digitale è quello di facilitare l'accesso ad un testo tanto corposo e complesso, quanto importante per studi di vari ambiti umanistici, così da poter rendere ricercabili i contenuti per recuperare il tesoro di informazioni linguistiche, storiche, matematiche che l'opera contiene.

All'interno di questo progetto complessivo, si inserisce uno studio approfondito sull'applicazione di metodologie di linguistica computazionale al *Liber Abaci*, che permettano di sviluppare sistemi di *Natural Language Processing* (NLP) per l'estrazione automatica di informazioni. L'elaborazione automatica del testo è un processo delicato a causa delle complesse caratteristiche del linguaggio stesso, dunque i contenuti testuali vengono sottoposti ad analisi morfologica, sintattica e semantica.

Nel caso specifico del *Liber Abaci*, a causa della peculiarità linguistica del testo, gli strumenti standard di analisi non possono essere utilizzati senza un'ampia soglia di errore e una notevole perdita di prestazione; pertanto, si è resa necessaria l'annotazione semantica manuale del testo attraverso un *tool* di annotazione digitale per codificare i risultati, di modo da ottenere dati annotati di alta qualità.

# Capitolo 2

## 2. Analisi semantica

Il campo dell'analisi semantica è innovativo e ha molto da dare allo studio dei testi: consente di fare ricerche nei contenuti testuali, servendosi dei concetti o domini e non delle parole.

Attraverso l'annotazione semantica, il lavoro sul testo risulta dunque più fine e complesso rispetto a quello semplicemente basato sulla ricerca delle occorrenze delle parole, genericamente rintracciate in quanto segnali della presenza di uno specifico dominio semantico.

Inoltre, formalizzare in modo appropriato un'interpretazione ed esprimerla attraverso l'annotazione di uno specifico dominio d'interesse, consente di creare strumenti di riconoscimento automatico delle entità, utili per realizzare tecniche avanzate di interrogazione, che aprono prospettive ad un approccio e studio delle opere nel campo digitale.

La realizzazione dell'analisi semantica su un'opera di rilevante interesse è un lavoro di grande importanza a livello di disponibilità di risorse per gli studiosi e consente un approccio analitico ai testi.

### 2.1 Annotazione dei dati

I fattori da considerare prima di iniziare ad annotare sono: il contesto di applicabilità del modello, ossia la valutazione dell'ambito semantico del testo in questione, la quantità e la qualità dei dati disponibili e il task da portare a termine (clustering, analisi delle associazioni, classificazione, regressione). In particolare, in base alla quantità e qualità dei dati e al task che si desidera svolgere, esistono algoritmi che consentono al computer di imparare a svolgere un task a partire da un campione di dati rappresentativo. Ci sono tre tipi di algoritmi di apprendimento automatico:

- supervisionato, in cui viene addestrato un modello prendendo come variabile di input un corpus di dati annotati (training set), e una variabile di output, ovvero i risultati reali associati ai dati. L'algoritmo dunque cerca di approssimare al meglio la funzione di relazione tra le due variabili, al fine di predire l'output corretto in presenza di nuovi dati;
- non supervisionato, che utilizza dati non annotati da cui estrarre informazioni, con l'implicazione che il sistema debba individuare caratteristiche comuni da cui inferire ipotesi per evidenziare pattern;
- semi-supervisionato, a metà tra le due modalità sopra presentate: algoritmi che combinano informazioni derivanti sia da corpora annotati che da dati non annotati.

Gli algoritmi supervisionati, sono quelli più utilizzati nella maggior parte dei task di Machine Learning e, a fronte di performance più accurate, richiedono dati ben annotati.

Il processo di annotazione rende esplicita l'informazione implicitamente contenuta in un testo scritto in linguaggio naturale. Annotare significa associare delle etichette ai dati, che permettono di fornire qualche tipo di informazione sui dati stessi. I dati con codifica esplicita di informazioni rappresentano il punto di partenza da cui un algoritmo può apprendere un task e riprodurlo.

Tra i più comuni task di annotazione nell'ambito di NLP si trova l'annotazione linguistica, che si interessa di taggare il linguaggio a diversi livelli di analisi, come l'identificazione di elementi grammaticali (*Part Of Speech Tagging*) e sintattici (*Parsing* sintattico). Inoltre è possibile l'annotazione delle entità, ossia la *Named Entity Recognition* (NER), che riguarda il processo di individuazione e classificazione di luoghi, persone, oggetti e simili presenti in un testo; o la *Keyword Tagging*, ovvero l'annotazione di parole chiave all'interno di un testo. Ed ancora, possono esserci casi in cui è utile la *Sentiment Analysis*, in cui ad essere annotate sono emozioni, sentimenti, opinioni (positive, negative, neutre) presenti in un testo.

### 2.1.1 Perché annotare i dati

Annotare i dati permette di rendere esplicita e *machine readable* l'informazione relativa alla struttura linguistica del testo. Avere dei corpora annotati permette di usare il computer per ricerche ed elaborazioni più avanzate dei dati linguistici. Infatti, i corpora annotati sono impiegati per addestrare modelli computazionali basati su algoritmi di apprendimento automatico supervisionato (corpus annotato di training). L'annotazione indica in modo chiaro alla macchina i pattern possibili e questo risulta molto utile nei task di classificazione avanzati e nelle costruzioni di complessi modelli di predizione.

In alcuni casi le informazioni da estrarre dai testi sono proprie di un particolare contesto, per questo l'accuratezza degli strumenti di annotazione diminuisce al cambiare del dominio del testo da annotare<sup>53</sup>. Le caratteristiche specifiche del linguaggio specialistico o di un particolare dominio influiscono negativamente sul risultato dell'analisi. In questi casi è dunque necessario provvedere all'annotazione manuale. Nonostante lo sforzo in termini di costo e tempo che potrebbe richiedere la loro produzione, acquisire i dati annotati permette di avere una descrizione accurata del contesto analizzato.

Se da una parte l'annotatore umano garantisce dati annotati in modo accurato, dall'altra i dati raccolti potrebbero non essere precisi o sempre coerenti. Per questo è importante intervallare il processo di annotazione con fasi di revisione.

Sebbene l'annotazione manuale possa risultare costosa, in caso di progetti singoli *ad hoc* è indispensabile e può risultare meno dispendiosa rispetto allo sviluppo di uno strumento di annotazione automatico soprattutto in rapporto alla qualità dei dati annotati.

---

<sup>53</sup> D. Gildea, *Corpus variation and parser performance*, in *Proceeding of Empirical Methods in Natural Language Processing*, Pittsburgh, (2001), PA.

## 2.1.2 Annotare: strategie per farlo

I testi comprendono diversi tipi di dati e, prima di attuare il processo di annotazione, è necessario strutturare i passi che comporranno il task.

1. Definire uno schema di annotazione: si tratta di fissare un repertorio di categorie per l'annotazione, corrispondenti ai tratti da rappresentare sul testo. Lo schema di annotazione definisce il contenuto dell'annotazione, non il modo in cui l'annotazione è proiettata sul testo.
2. Stabilire delle linee guida: è il primo fondamentale step di lavoro effettivo sul testo per garantire dati annotati coerenti e consistenti. Si tratta di indicare delle convenzioni da seguire durante il processo di annotazione, di modo da stabilire le entità che devono essere taggate, il modo con cui devono essere taggate, ossia delle etichette chiare e comprensibili, e i limiti delle entità da taggare, ossia la quantità di testo che le etichette possono comprendere e la quantità di etichette che ciascuna parte di testo può presentare. Sebbene le convenzioni vadano stabilite prima di intraprendere il lavoro di annotazione, queste possono essere modificate in corso d'opera a patto che venga mantenuta la coerenza nel processo.
3. Scelta del formato di annotazione: visto che un'entità di dominio può essere composta da una singola parola, ma anche da un gruppo di parole consecutive, è necessario una notazione capace di identificare correttamente anche le entità formate da più parole.  
Il formato IOB2 (*Inside Outside Beginning*) è uno standard per l'annotazione di entità suddivise in token, in cui il prefisso B rappresenta l'inizio di un'entità, I etichetta tutti i successivi token di un'entità e O che indica un token non appartenente ad un'entità.
4. Metodi di annotazione: l'annotazione può essere svolta in modo manuale da uno o più annotatori con l'ausilio di tool, in alternativa può contemplare l'utilizzo della programmazione con *data programming*, metodo che presuppone la realizzazione di script che prendono le informazioni da

risorse linguistiche precise per il task e, attraverso le istruzioni, annotano i dati. Tuttavia i dati annotati in questo modo non risultano del tutto precisi ed accurati, ma presentano errori consistenti ed è necessaria comunque una revisione da parte di un annotatore umano.

5. Valutazione dell'annotazione: stabilire il grado di accordo tra le annotazioni svolte da annotatori differenti, definito *Inter-Annotator Agreement* (IAA).

In generale, l'annotazione di dati implica la formulazione di giudizi soggettivi: i giudizi di un annotatore, dunque, possono divergere anche di molto da quelli di un altro. Per questo è fondamentale stabilire fino a che punto tali giudizi siano riproducibili e attendibili e questo si concretizza nel calcolo delle metriche di valutazione dell'IAA.

Oltre alle considerazioni sul processo di annotazione, una buona strategia è focalizzare i requisiti del buon annotatore. L'annotatore si occupa di preparare e annotare i dati che servono per allenare un modello. Le competenze che richiede tale compito sono di tipo linguistico-culturale, in quanto, per padroneggiare i dati da taggare, serve una conoscenza degli aspetti del linguaggio e del dominio specifico dell'ambito culturale analizzato.

L'annotatore inoltre è la figura che conosce ed è in grado di interpretare i dati. Il suo ruolo non si esaurisce al termine del processo di annotazione, ma rimane una figura di riferimento anche in fase di training e testing del modello di NLP. Per questo motivo l'annotatore dovrebbe avere delle competenze informatiche e in ambito di trattamento automatico del linguaggio, che gli permettano non solo di dotarsi di strumenti necessari per l'annotazione, ma anche di indirizzare il lavoro dell'èquipe verso una struttura di dati rappresentativi e consistenti, utile per la realizzazione di un modello per l'annotazione (semi)automatica.

## 2.2 Ciclo di annotazione

Una volta fissato lo schema di annotazione, ossia una descrizione teorica del fenomeno da codificare, e i metodi con cui codificare tale fenomeno, viene attuato lo sviluppo del ciclo di annotazione, fase che si inserisce nel quadro più ampio del ciclo MATTER.<sup>54</sup>

Il ciclo di annotazione, chiamato anche ciclo MAMA (Model-Annotate-Model-Annotate), richiede più iterazioni di modellazione e annotazione per l'identificazione delle entità e delle categorie corrette da associarvi.

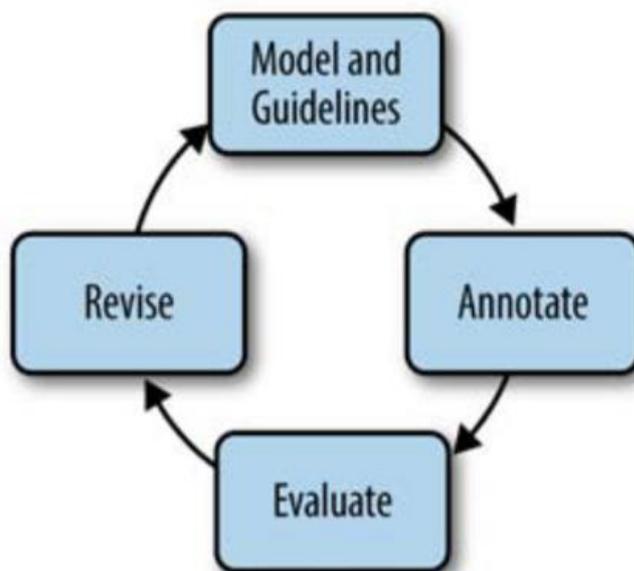


Figura 2 - Ciclo MAMA (Pustejovsky & Stubbs, 2012).

Il ciclo MAMA prevede quattro fasi: la definizione del modello e delle linee guida di annotazione, utili a indirizzare gli annotatori nel compito di identificazione degli elementi; l'annotazione sul testo, svolta da due annotatori in modo separato; la valutazione dell'annotazione, ottenuta dal calcolo delle metriche di valutazione sul confronto delle due annotazioni; ed infine, in caso di disaccordo, una fase di revisione,

---

<sup>54</sup> J. Pustejovsky, A. Stubbs, *Natural Language Annotation for Machine Learning*, (2012), O'Reilly Media, pp.22-30.

frutto del confronto tra le versioni dei due annotatori per ottenere un modello comune di annotazione.

La revisione del modello di annotazione innesca un nuovo ciclo MAMA, finché non si ottiene una versione definitiva.

Una volta ottenuta un'annotazione coerente ed uniforme dai due annotatori, secondo le specifiche aggiornate e le linee guida revisionate, si può creare un corpus standard, definito *gold standard*. Questo è il corpus utilizzato per l'apprendimento automatico.

# Capitolo 3

## 3. Prima fase: il cap. VIII del *Liber Abaci*

### 3.1 Punto di partenza

Il *Liber Abaci* è un corposo trattato medievale di aritmetica e geometria, pietra miliare per la storia della matematica. L'opera, per la sua fase tarda del latino, per le sue caratteristiche e per il suo contenuto molto specialistico, rappresenta una risorsa unica per gli studiosi che lavorano su corpora latini. La ricchezza delle informazioni e dei contenuti si identificano nella peculiarità di un lessico ben preciso che costella tutto il testo.

Oltre alla sua importanza scientifica, il *Liber Abaci* presenta un lessico molto particolare, non solo relativo alla matematica ma anche per il commercio e la finanza, che difficilmente è rappresentato nei corpora per il latino attualmente disponibili.

Dunque per colmare questa lacuna, il punto di partenza del lavoro di analisi semantica sul *Liber Abaci* è stato l'individuazione, come ambito specifico di interesse linguistico e storico, l'argomento mercantile e commerciale dei capitoli VIII, IX e X. Argomento che si affianca alla dissertazione matematica come esemplificazione dei concetti nell'uso pratico del commercio.

Le informazioni contenute in questi capitoli sono preziose non solo per chi è interessato al contributo dell'opera per la storia della scienza, ma anche come testimonianza storica dell'economia e del commercio del mondo mediterraneo nel XIII secolo. Gli esempi di matematica pratica del *Liber Abaci* richiamano le suggestioni di un mercato medievale: si parla del commercio di pepe, noce moscata, zafferano, stoffe, pelli, olio, conigli, oro, argento, rame e falso argento (argento misto a stagno); del cambio di valute che vanno dalla lira, ai massamutini, dai bisanti alle monete dei vari enti emissari; ed ancora delle unità di misura di cui si trovano le relative equivalenze a Pisa, Genova, Firenze, Bologna, Venezia, Taranto, Palermo, Messina, Siria, Alessandria, Cipro, Barcellona, Provenza e la Barberia; facendo anche un accenno alle

dogane con i dazi sulle transizioni e alle società e soci che si occupano di capitale e investimenti<sup>55</sup>.

Valutando la ricchezza di queste informazioni, l'attività di annotazione semantica è stata contestualizzata sul dominio mercantile-commerciale. Il lavoro ha dunque preso avvio con l'individuazione del capitolo VIII come testo pilota su cui sviluppare i passi del processo di annotazione manuale.

### 3.2 Tagset e annotazione pilota

Prima di effettuare l'annotazione semantica è stato necessario stabilire le entità da taggare e il modo in cui devono essere taggate. Per questo è stato preso come riferimento un tagset, ossia un insieme predefinito di categorie, progettato per fornire supporto nell'identificazione semantica delle entità. Nello specifico è stato usato il tagset usato dal sistema USAS (*UCREL Semantic Analysis System*)<sup>56</sup>. Il sistema di analisi semantica dell'UCREL (*University Center for Computer Corpus Research on Language*) è un framework per l'analisi semantica automatica di testi, inizialmente sviluppato per l'inglese, poi esteso ad altre lingue. Il tagset utilizzato ha una struttura multilivello e consente di classificare termini in base a 21 campi semantici con ulteriori sotto-categorie di specificazione.

Questo tagset è diventato quasi uno standard per la definizione di campi semantici, tanto da essere stato utilizzato in diversi progetti di ricerca e pubblicazioni fin dal 1990, non solo per l'inglese contemporaneo, ma anche per altre lingue e su testi storici<sup>57</sup>. Il framework USAS è stato esteso per coprire molte lingue tra cui cinese,

---

<sup>55</sup> S. Cuomo, *Fibonacci's Liber Abaci. A translation into Modern English of Leonardo Pisano's book of calculation by Laurence E. Sigler*, in *Aestimatio: Critical Reviews in the History of Science*, 2004, vol. 1, pp. 24-26.

<sup>56</sup> <http://ucrel.lancs.ac.uk/usas/>

<sup>57</sup> D. Archer, *Exploring verbal aggression in English historical texts using USAS*, in *Diachronic Corpus Pragmatics*, (2014), pp. 277-302; R. Sprugnoli, S. Tonelli, G. Moretti, S. Menini, *A little bit of bella pianura: Detecting Code-Mixing in Historical English Travel Writing*, in *Proceedings of the Fourth Italian Conference on Computational Linguistics CliC-it 2017*, (2017), pp. 304-309.

olandese, italiano, portoghese, spagnolo e maltese. Per queste lingue sono stati compilati lessici semantici traducendoli dal tagset standard inglese originario. A causa dell'inevitabile ambiguità delle traduzioni e della corrispondenza tra le lingue, i lessici hanno avuto necessità di miglioramenti manuali ove possibile. Anche nel caso del nostro progetto, l'applicazione del tagset prevede riadattamenti e modifiche puntuali su alcuni campi semantici.

Il fatto di utilizzare, per la realizzazione del nostro tagset, uno schema già esistente e consolidato, come quello dell'UCREL, si basa sul principio di riutilizzo e ha l'obiettivo di ovviare all'esistenza di una varietà di schemi tra loro incompatibili, di modo da incentivare l'interscambio e la riproducibilità del modello su domini simili per la creazione di risorse compatibili e più facilmente utilizzabili dalla comunità internazionale di studiosi.

Il tagset, che è stato utilizzato nella fase di annotazione del progetto è, dunque, una versione ridotta del tagset dell'UCREL, realizzata selezionando solo i tag considerati rilevanti per l'obiettivo di ricerca del progetto. Dunque valutando un primo approccio agli elementi da annotare, dei 21 principali campi di discorso identificati nel set originale, ne sono stati presi in considerazione 7, nello specifico *I: money and commerce in industry, M: movement, location, travel and transport, N: numbers and measurement, O: substances, materiales, objects and equipment, S: social actions, states and processes, T: Time, Z: names and grammar.*

Sotto sono riportati i tag e sottotag di ogni campo, presi in considerazione come possibili categorie con cui identificare entità specifiche del testo, come, ad esempio, monete, materiali, prodotti, merci, unità di misura, condizioni commerciali, mestieri, persone, luoghi e specificazioni temporali.

**I1 Money generally**

I1.1 Money: Affluence

I1.2 Money: Debts

I1.3 Money: Price

**I2 Business**

I2.1 Business: Generally

I2.2 Business: Selling

**I3 Work and employment**

I3.1 Work and employment:  
Generally

I3.2 Work and employment:  
Professionalism

**M7 Places**

**N1 Numbers**

**N3 Measurement**

N3.1 Measurement: General

N3.2 Measurement: Size

N3.3 Measurement: Distance

N3.4 Measurement: Volume

N3.5 Measurement: Weight

|   |                                    |
|---|------------------------------------|
| N3.6 Measurement: Area                              | <b>O2</b> <b>Objects generally</b> |
| N3.7 Measurement: Length & height                   | <b>S2</b> <b>People</b>            |
| N3.8 Measurement: Speed                             | S2.1 People: Female                |
| <b>N5</b> <b>Quantities</b>                         | S2.2 People: Male                  |
| <b>O1</b> <b>Substances and materials generally</b> | <b>T1</b> <b>Time</b>              |
| O1.1 Substances and materials generally: Solid      | T1.1 Time: General                 |
| O1.2 Substances and materials general: Liquid       | T1.1.1 Time: General: Past         |
| O1.3 Substances and materials generally: Gas        | T1.1.1 Time: General: Present      |
|   | T1.1.1 Time: General: Future       |
|   | <b>Z1</b> <b>Personal names</b>    |
|   | <b>Z2</b> <b>Geographical name</b> |
|   | <b>Z3</b> <b>Other proper name</b> |

Il tagset selezionato è stato dunque utilizzato per considerare l'effettiva presenza dei campi semantici all'interno del *Liber Abaci*. Questo è stato valutato attraverso l'annotazione pilota del capitolo VIII.

## 2.2.1 CATMA: tool di annotazione

Per l'annotazione pilota è stato impiegato il tool Catma<sup>58</sup> (Computer Assisted Textual Markup and Analysis), uno strumento online gratuito e open source per l'annotazione di testi, sviluppato dall'Università di Amburgo.

La piattaforma incorpora quattro funzionalità visibili nell'interfaccia: 'Project' consente di creare un progetto che può essere condiviso con altri utenti per un lavoro collaborativo di annotazione; 'Tags' in cui creare il proprio set di tag con nome, attributi (opzionali) e colore identificativo del tag da inserire nel testo per una resa grafica e intuitiva dell'annotazione; 'Annotate' che consente di caricare il testo in formato *plain text* e realizzare l'annotazione e 'Analyze' che comprende un generatore di *query* con cui costruire opzioni di ricerca dei tag su una combinazione di domande predefinite.

---

<sup>58</sup> <https://app.catma.de/catma/>

L'utente dunque può importare il proprio file testuale, creare manualmente il tagset specifico per la sua annotazione e lavorare sul file inserendo i tag nel testo.

Su Catma è possibile l'annotazione solo delle entità e non quella delle relazioni tra entità, tuttavia un vantaggio della piattaforma è la grande libertà concessa sulle entità, dal momento che sono consentiti tag multipli e tag annidati.

Alla fine del processo, le annotazioni possono essere esportate in formato XML-TEI. Catma permette un approccio descritto provocatoriamente dai suoi creatori come “non dogmatico”, in quanto contrasta con i sistemi più consolidati per la rappresentazione del testo come il markup XML o la TEI.

L'output è un file con markup standoff, ossia i dati di annotazione sono posti al di fuori del testo da annotare. Questo implica l'applicazione di etichette ad un insieme di caratteri del testo, indicanti la parola o l'insieme di parole da taggare, e l'utilizzo di tali etichette come identificatori per memorizzare l'interpretazione esternamente.

Questo approccio offre tutta la libertà della tradizionale annotazione manuale e consente annotazioni multiple, sovrapposte e persino tassonomicamente contraddittorie da parte di uno o più annotatori<sup>59</sup>.

---

<sup>59</sup> J. Horstmann, *Undogmatic Literary Annotation with CATMA*, in *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*, (2020), Berlin, Boston: De Gruyter, pp.163-164.

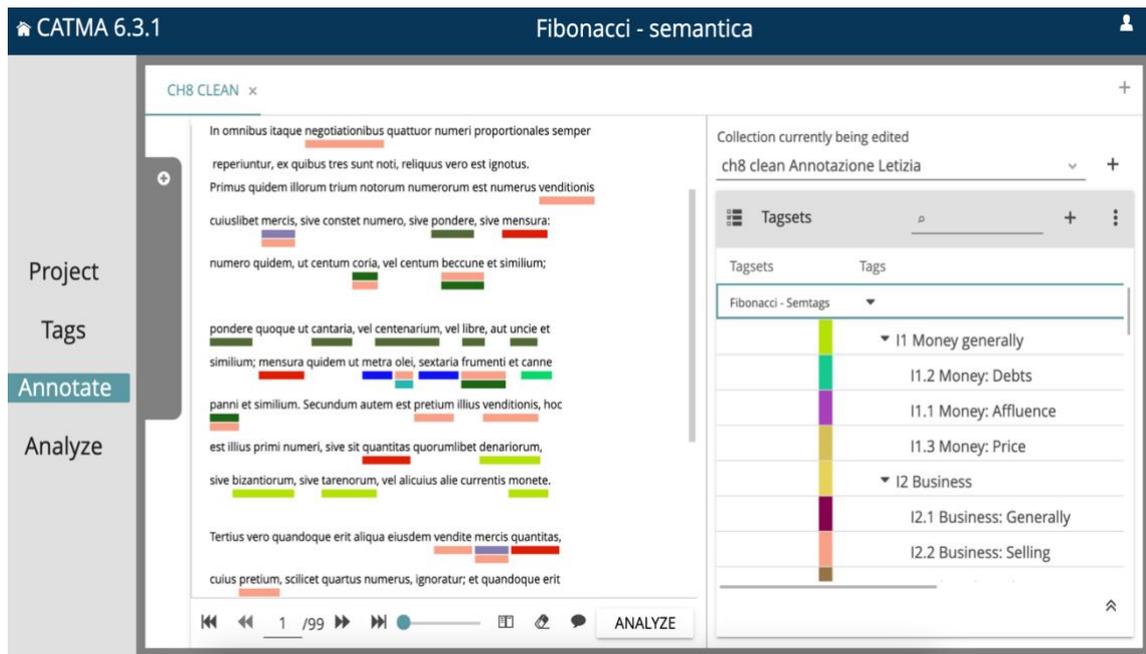


Figura 3 - Interfaccia di Catma: annotazione pilota sul cap. VIII.

Proprio per questa caratteristica di Catma, la realizzazione dell'annotazione sul cap. VIII ha permesso una resa "multiforme" delle entità commerciali e finanziarie taggate, ossia i concetti sono stati espressi con più categorie con l'idea di indicare il significato semantico attraverso tag multipli dal più generico al più dettagliato.

Questo consente di presentare in maniera più puntuale il significato semantico dei termini: ad esempio l'entità *merx* è stata individuata con tag multipli, espressa come un generico prodotto attraverso il tag <O2 Objects generally>, con funzione di oggetto di vendita indicato dal tag <I2.2 Business: Selling>.

Un caso particolare è l'annotazione di monete e unità di misura che hanno una specificazione geografica, in quanto indicano il valore che assume quella moneta o quell'unità di misura in una determinata area. Dal momento che tra questi due tipi di entità c'è un legame intrinseco, si è scelto di interpretare la relazione che le lega attraverso l'uso di entità annidate. Questo è il caso, ad esempio, di *rotuli gerovi* che indicano insieme l'unità di peso <N3.5 Measurement: Weight>, mentre *gerovi* è anche una specificazione geografica del valore che assumeva il rotulo a Genova e dunque indicata anche come <Z2 Geographical names>; allo stesso modo *denari pisani* indica

la valuta che era utilizzata a Pisa e dunque <I1 Money generally>, con la specificazione dell'applicazione geografica attraverso il tag <Z2 Geographical names>.

Questa specificazione consente, inoltre, di distinguere tali indicazioni geografiche riferite a monete e unità di misura da indicazioni geografiche di luoghi in senso stretto, indicate come stato in luogo o moto a luogo: ad esempio *in Pisae* che ha come tag <M7 Places>.

La piattaforma risulta intuitiva nell'attività pratica di annotazione: ad ogni tag semantico è associato un colore e sul testo è visibile la parola sottolineata in base al colore del tag scelto per tale parola. Tuttavia, in caso di annotazioni multiple o annidate, il sistema non applica l'ordine di marcatura con cui l'annotatore inserisce i tag sulle parole, ma risulta un ordine casuale, che non rispetta la disposizione dei tag dal più generico al più dettagliato. Questo implica che nel file risultante la stessa parola ottenga marcature non omogenee.

Un altro limite di Catma è quello di non presentare una funzionalità che permetta di annotare in maniera automatica varie occorrenze di una stessa parola, cosa che avrebbe velocizzato di molto il lavoro. Invece, il tool consente solo di taggare manualmente le parole che si trovano scorrendo pagina per pagina, dal momento che non è possibile neppure una semplice ricerca che individui tutte le occorrenze sul testo.

Nell'interfaccia la funzione 'Analyze' consente la ricerca per ottenere informazioni generali, attraverso delle *query* preimpostate, sulla frequenza delle parole taggate o sui tag utilizzati. Tuttavia Catma non permette alcun collegamento al testo, dunque, individuando con la ricerca possibili errori, l'unico modo per attuare la correzione è intervenire scorrendo il testo fino all'occorrenza della parola annotata in modo sbagliato.

Dai risultati dell'annotazione possiamo considerare che i tag più utilizzati sono: <I1 Money generally> (frequenza 872), con il quale si annotano termini di vari tipi di monete; i sottotag di <N3 Measurement> (972), che annotano varie unità di misura, di cui le più comuni sono le unità di peso; <I2.2 Business: Selling> (291), che indica termini riferiti alla vendita commerciale; <Z2 Geographical names> (281) termini che ricorrono solitamente con unità di misura o monete. Il testo non contiene alcuni tag e sottotag come <I3 Work and employment>, <S2 People>, <T1 Time>.

<Z1 Personal Names> e <Z3 Other proper names>, che sono stati dunque esclusi dai campi semantici.

L' esportazione dell'annotazione svolta sul cap. VIII è possibile solamente nel formato XML-TEI. Ai fini del progetto questo formato necessita una conversione, deve dunque essere rielaborato per ottenere i dati in un formato standard per NLP come ad esempio IOB.

## 3.2 Linee Guida e annotazione test

Le Linee Guida sono una spiegazione del tagset per l'annotazione dell'ambito mercantile e commerciale del *Liber Abaci*, redatte sulla base dell'annotazione pilota del capitolo VIII. La loro stesura ha lo scopo di indicare come svolgere l'annotazione ad altri possibili annotatori esterni. Indicano le convenzioni da seguire durante l'annotazione, in particolare presentano le categorie da prendere in considerazione e le entità da individuare con tali categorie, specificando i criteri di annotazione in casi dubbi. I tag possono riferirsi a sostantivi e, solo in casi specifici, ad aggettivi, poiché le categorie grammaticali contemplate nell'annotazione sono esclusivamente queste due.

Nel caso di espressioni costituite da più parole viene indicato un tag complessivo dell'espressione generale e un tag specifico per ogni parola che rientri nelle categorie semantiche d'interesse.

Le Linee Guida sono, inoltre, corredate da esempi prototipici che mostrano specifiche parole da annotare e il modo in cui annotarle con i relativi tag.

I tag semantici sono composti da:

- una lettera maiuscola che indica il campo del discorso generale,
- una o più cifre che indicano la suddivisione del campo,
- una definizione letterale del tag.

In base alle questioni emerse dall'annotazione campione, le Linee Guida spiegano il modo in cui applicare i tag. Per ogni campo semantico sono indicati i tag e sotto tag e di ciascuno viene fornita la spiegazione e gli esempi prototipici.

In caso di situazioni di annotazione dove è necessario l'utilizzo di tag multipli o annidati, le Linee Guida spiegano che è necessario applicare prima il tag più generico e poi il più specifico. Nel campo semantico del tag da applicare per primo, sotto alla definizione del tag, viene riportata una nota con il riferimento al campo semantico del secondo tag da applicare.

Ad esempio, per rendere i tag annidati delle monete, le Linee Guida riportano quanto segue:

## **I: MONEY & COMMERCE IN INDUSTRY**

### **I1 Money generally → Termini relativi al denaro in generale**

ESEMPI PROTOTIPICI: tipi monete, soldi (*libra, soldus, denarius, pisaninus, bizantius, libara denariorum, tarenus, miliarensis, granum, moneta, massamutinus, yperperum, marca argentum, uncia aurum, cambium (monetarum), toloneum, kartus*)

#### **I1.1 Money: Affluence → Termini relative a ricchezza e prosperità**

ESEMPI PROTOTIPICI: guadagno, flusso di cassa, affluenza (*lucrum, proficuum*)

#### **I1.2 Money: Debts → Termini relative al livello di debito**

ESEMPI PROTOTIPICI: debito, spesa, fallimento, pagamento

#### **I1.3 Money: Price → Termini relative al costo/valore**

ESEMPI PROTOTIPICI: affare, prezzo, valore, sconto

**NOTA:** vedi sezione Z NAMES & GRAMMATICAL WORDS (di cui Z2).

Il rimando è al tag:

## Z: NAMES & GRAMMATICAL WORDS

**Z1 Personal names → Sostantivi che contraddistinguono/identificano una persona (es. nome e/o cognome, titolo)**

ESEMPI PROTOTIPICI: nomi e cognomi, duca, cardinale, signore

**Z2 Geographical names → Sostantivi che contraddistinguono/identificano un luogo specifico (es. il nome di una strada, di una città, di un paese, di un continente, ecc.)**

ESEMPI PROTOTIPICI: nome di una città, paese, e riferimenti ad essi (*forfori, gerovi, pisanus, venetianus, bononinus, mergulienis, podiensis, barcellonaiensis, yperperus, Messana, Garbum, Sicilia, Panormus, Pisae, Provincia, Alexandria, Syria, Bugea, Septis*),

**NOTA:** Z2 generalmente collegato a unità di misura o monete va indicato in un unico tag.

ESEMPIO: *rotuli gerovi* riferito all'unità di peso *rotuli* con il valore usato a Genova, dunque:

*rotuli gerovi* - N3.5 Measurement: Weight & *gerovi* - Z2 Geographical names

*soldus ianuinarum* – I1 Monei generally & *gerovi* - Z2 Geographical names

Le Linee Guida, così redatte, sono state dunque utilizzate per effettuare l'annotazione test.

### 3.2.1 Il secondo annotatore

L'annotazione test è un'annotazione di controprova, eseguita da un annotatore estraneo alle fasi di sviluppo dell'annotazione. Il secondo annotatore ha lavorato sullo stesso testo del capitolo VIII, caricato sul tool Catma e l'annotazione è stata svolta avendo a disposizione lo stesso tagset e le Linee Guida come ausilio. La necessità di allestire una seconda annotazione esterna deriva dall'esigenza di stabilire fino a che punto i giudizi del primo annotatore siano affidabili e riproducibili.

Lo scopo dell'annotazione test è quello di mostrare la comprensibilità delle linee guida e l'effettiva corrispondenza tra termini e tag, con cui esprimerne il concetto.

Dal momento che il dominio delle entità da rintracciare nel testo è espresso da termini commerciali e mercantili medievali, il secondo annotatore è stato individuato nell'équipe di lavoro nella figura dello storico medievalista con competenze mirate sull'argomento, così da valutare la scelta di termini da annotare e cogliere il significato storico delle entità.

Oltre alle questioni sulle entità già sollevate, il secondo annotatore presenta ulteriori questioni da indagare e rendere attraverso l'annotazione semantica. In particolare, nota la resa troppo uniforme tra le entità geo-politiche, riferite a monete, e le entità geografiche, riferite a pesi; questo fattore non permette di evidenziare la sottile distinzione che le contraddistingue. Inoltre, il secondo annotatore mostra la scarsa rilevanza dal punto di vista storico della resa semantica delle tipologie di merci, indicate semplicemente con il riferimento allo stato della materia in cui si trova il prodotto (liquido, solido, gassoso).

Le questioni saranno poi approfondite e spiegate nella seconda fase del lavoro sul capitolo VIII.

### 3.3 Valutazione delle annotazioni

Le due annotazioni sono state analizzate per valutarne l'accordo tra gli annotatori e il livello di affidabilità dell'annotazione. La procedura di validazione è prassi ormai consolidata in Linguistica Computazionale e prende il nome di *Inter-Annotator Agreement* (IAA).

Nell'ambito computazionale l'IAA è usata per passare dal materiale annotato ad un *gold standard*, ossia un insieme di dati sufficientemente consistenti che serva come risorsa valida per il processo di *training* e *testing* di sistemi automatici che svolgano il medesimo compito del linguista.

L'assunzione di base della valutazione è che lo schema di annotazione risulti attendibile se due o più annotatori sono in accordo nell'assegnare una determinata categoria allo stesso termine in analisi. L'attendibilità si configura dunque come prerequisito per dimostrare la validità di uno schema di codifica e un ampio consenso tra gli annotatori viene considerato a garanzia della precisione intrinseca del processo di annotazione<sup>60</sup>.

---

<sup>60</sup> M. J. Warrens, *Inequalities between multi-rater kappas*, in *Advances in Data Analysis and Classification*, 4, (2010), pp. 271–286.

I coefficienti di *agreement* vengono usati per assicurare la bontà della procedura di annotazione e del materiale annotato: un alto livello di IAA fa sì che il fenomeno sia considerato consistente e sistematico. Va tenuto in considerazione il fatto che l'*agreement* tra due o più annotatori può variare in relazione al livello di esperienza, per questo risulta ragionevolmente più basso il valore dei coefficienti in un gruppo misto di annotatori formato da esperti e non esperti del dominio.

Il calcolo dei coefficienti di *agreement* prevede un insieme di item, ossia le parole che sono state annotate, un insieme di categorie, che indicano i tag utilizzati per annotare e un insieme di annotatori che assegnano una categoria a ciascun item. Gli indici utilizzati per il calcolo sono detti “misure di Kappa” (il più noto è K di Cohen) e si fondano su tre assunzioni: gli item soggetti a valutazione sono indipendenti l'uno dall'altro; gli annotatori operano in autonomia; le categorie usate sono mutualmente esclusive ed esaustive<sup>61</sup>.

Nel caso specifico del progetto, per la valutazione delle annotazioni si ha una situazione più complessa a livello semantico che non si riduce ad una classificazione ad etichetta singola. Nell'annotazione del cap. VIII *del Liber Abaci*, infatti, sono presenti item con categorie multiple e item con categorie annidate, derivanti dalle considerazioni di ambito storico che si vuol far emergere dall'analisi. Ad esempio la referenziazione geopolitica di monete e unità di peso deve essere dunque considerata un problema di multi-label classification.

### 3.3.1 Multi-Label classification

La classificazione nell'apprendimento automatico è la tecnica che consiste nell'identificare a quali categorie appartengono gli item presi in analisi.

I problemi di classificazione si distinguono in tre tipologie: la classificazione binaria, è il problema per cui si stabilisce se un item appartiene o meno ad una

---

<sup>61</sup> G. Gagliardi, *Inter-Annotator Agreement in linguistica: una rassegna critica*, in *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*, (2018), pp. 206-212.

categoria; la classificazione multi-class, per cui un item viene classificato con una sola categoria, scelta tra un insieme di categorie possibili; la classificazione multi-label, invece, è il problema di categorizzazione che ammette per ogni item l'assegnazione di più di una categoria, presa nell'insieme di categorie possibili.

I metodi di classificazione multi-label sono sempre più utili in diversi ambiti, ne sono un esempio la categorizzazione della musica, in cui una canzone può appartenere a più generi musicali, e la classificazione semantica degli scenari, che permette di individuare più classi concettuali come tramonto e spiaggia nello stesso scenario<sup>62</sup>. Nella classificazione multi-label dunque ogni item è associato ad un insieme di etichette.

I metodi esistenti per la classificazione multi-label si dividono in due categorie principali: metodi di trasformazione del problema, quei metodi in cui il problema di classificazione multi-label viene trasformato in uno o più problemi di classificazione a etichetta singola; metodi di adattamento dell'algoritmo che prevedono algoritmi di apprendimento specifici per gestire direttamente i dati multi-label<sup>63</sup>.

Nel caso specifico del *Liber Abaci*, utilizzare la classificazione multi-label permette di tenere in considerazione gli item con categorie multiple presenti nelle annotazioni (caso di studio: moneta o peso con referenziazione geopolitica, tipo di merce). Alle due annotazioni di cui vogliamo valutare l'accordo è stato dunque applicato un metodo di adattamento dell'algoritmo che prevede per ogni item taggato, di indicare tutte le categorie possibili, segnalando quelle effettivamente utilizzate. La *tabella 1* è un esempio semplificato del metodo utilizzato.

---

<sup>62</sup> G. Tsoumakas, I. Katakis, *Multi-Label Classification: An Overview*, in *International Journal of Data Warehousing and Mining (IJDWM)*, (2007), IGI Global, vol.3, pp.1-13.

<sup>63</sup> *Ibidem*.

|           | I1 | I2.2 | N3.5 | O1.1 | O2 | Z2 |
|-----------|----|------|------|------|----|----|
| rotuli    | 0  | 0    | 1    | 0    | 0  | 0  |
| gerovi    | 0  | 0    | 1    | 0    | 0  | 1  |
| merx      | 0  | 1    | 0    | 0    | 1  | 0  |
| piper     | 0  | 1    | 0    | 1    | 0  | 0  |
| soldus    | 1  | 0    | 0    | 0    | 0  | 0  |
| pisaninus | 1  | 0    | 0    | 0    | 0  | 1  |

*Tabella 1 - Item e categorie rappresentati secondo il metodo.*

Questo tipo di struttura può essere realizzata con la trasformazione delle annotazioni in due matrici che hanno dimensione  $i$  per  $c$ , dove  $i$  sono il numero degli item annotati e  $c$  il numero di categorie utilizzabili.

Tuttavia va considerato che ciascun annotatore può aver annotato parole in più o in meno, dunque le due matrici avrebbero dimensioni differenti. Questo fattore potrebbe influenzare le prestazioni del confronto tra matrici multi-label, alterandone i risultati. Per questo motivo prima di calcolare i coefficienti di valutazione è necessario ristrutturare le matrici considerando i casi di non annotazione: quando nel confronto tra le due annotazioni, risulta un item mancante nell'una o nell'altra, questo item viene aggiunto con la categoria "NOT".

La costruzione di queste due matrici prevede una procedura di trasformazione dell'output di Catma attraverso specifici script in linguaggio Python, creati e applicati secondo i seguenti passi:

1. `dizionari.py` prende le due annotazioni di output date dal tool Catma in formato XML-TEI e le riporta in due dizionari che hanno come chiave il range della posizione del carattere di inizio e fine della parola annotata nel testo e come valore la categoria corrispondente.

Applicato ai file del progetto: `Annotazione_Letizia.xml` >  
`diz_Annotazione_Letizia` e `Annotazione_Salvatori.xml` >  
`diz_Annotazione_Salvatori`;

2. `ricostruzione_dizionari.py` va a ristrutturare i dizionari con l'unione delle chiavi dei due dizionari di annotazione.

Applicato ai file del progetto: `diz_Annotazione_Letizia` >  
`diz_Letizia_complex` e `diz_Annotazione_Salvatori` >  
`diz_Salvatori_complex`;

3. `normalizzazione_diz.py` va a normalizzare i dizionari, controllando le chiavi ed eliminando quelle spurie.

Applicato ai file del progetto: `diz_Letizia_complex` >  
`diz_Letizia_normalizzato` e `diz_Salvatori_complex` >  
`diz_Salvatori_normalizzato`.

Lo script `dizionari.py` va a prendere le informazioni delle annotazioni dai file XML-TEI: i dati del tagset sono rappresentati nell'elemento `<encodingDesc>` del `<teiHeader>` attraverso il tag `<fsDesc>` che contiene una serie di attributi informativi, in particolare `@xml:id` che è l'identificativo univoco del tag. I riferimenti alle annotazioni sul testo, invece, sono contenuti negli elementi `<fs>` di `<text>`: ogni annotazione in `<fs>` ha un identificativo univoco `@xml:id` e un attributo `@type` che punta al tag specificato nella descrizione della codifica nell'intestazione. Nel `<body>` il frammento di testo di riferimento è indicato da due elementi `<ptr>`, questi hanno un attributo `@target` che contiene il range della posizione di inizio e fine. Un segmento di testo annotato è racchiuso da un elemento `<seg>`, che fa riferimento all'annotazione nel suo attributo `@ana`, corrispondente all'id delle annotazioni (elemento `<fs>`).

Tutte queste informazioni vengono raccolte per ciascuna annotazione e riorganizzate in un dizionario che, nella sua forma finale, ha come chiavi il range della parola annotata e per valori la lista di tag associati. Il risultato è un un dizionario come mostra l'esempio:

```
{'[49, 56]': ['I2.2 Business: Selling', 'O2 Objects generally'], '[96, 111]': ['I2.2 Business: Selling'], ...}
```

Lo script `ricorstruzione_dizionari.py` va a creare una lista che contiene l'unione delle categorie, aggiungendo anche la categoria "NOT", che indica il caso di una mancata annotazione da parte di uno dei due annotatori.

```
#lista con l'unione delle categorie possibili
def class_union(dict):
    union_class = []
    for i in dict.keys():
        union_class.append(dict[i][0])
    union_class.append('NOT')
    union_class = set(union_class)
    return union_class
```

*Figura 4 - Snippet per la costruzione della lista con l'unione delle categorie.*

Si ottiene un totale di 15 categorie:

```
{'N3 Measurement', 'I2.2 Business: Selling', 'N3.7 Measurement: Length & height', 'O1.2 Substances and materials generally: Liquid', 'O1.1 Substances and materials generally: Solid', 'I1 Money generally', 'N5 Quantities', 'M7 Places', 'O2 Objects generally', 'Z2 Geographical names', 'N3.4 Measurement: Volume', 'N3.1 Measurement: General', 'N3.5 Measurement: Weight', 'N3.2 Measurement: Size', 'NOT'}.
```

Inoltre lo script va a ristrutturare la dimensione dei dizionari con l'unione delle chiavi, ossia in ciascun dizionario risultano presenti tutte le chiavi possibili. Le chiavi non presenti in uno dei due dizionari (dunque parole non annotate) vengono aggiunte nella ristrutturazione con valore "NOT".

```
def dict_complex(list, dict):
    for key in list:
        if key in dict.keys():
            continue
```

```

else:
    d = {key : ['NOT']}
dict.update(d)
return dict

```

*Figura 5 - Snippet per ristrutturazione dizionari.*

Il risultato è un dizionario come nel frammento di seguito:

```

{'[41, 48]': ['NOT'], '[49, 56]': ['I2.2 Business: Selling',
'O2 Objects generally'], '[96, 111]': ['I2.2 Business:
Selling'], ...}

```

Lo script `normalizzazione_diz.py` va a “pulire” i range delle chiavi: in ciascun dizionario, se ci sono nello stesso dizionario due chiavi *close*, cioè chiavi che hanno il range della posizione del carattere di inizio e fine che varia di 1 o 2 posizioni, si tratta presumibilmente della stessa parola che è stata annotata in maniera poco precisa da uno dei due annotatori sull’interfaccia di Catma.

```

def trova_close(dict):
    close = []
    for e1 in dict.items():
        key1 = e1[0].replace('[', '').replace(']', '')
        for e2 in dict.items():
            key2 = e2[0].replace('[', '').replace(']', '')
            start1 = int(key1.split(',')[0])
            end1 = int(key1.split(',')[1])
            start2 = int(key2.split(',')[0])
            end2 = int(key2.split(',')[1])
            res = abs(start1 - start2) <= 2 and abs(end1 - end2) <= 2
            if res and (start1 != start2 or end1 != end2):
                v = e1, e2
                if v[0][1] != ['NOT']:
                    close += [v]
    return close

```

*Figura 6 - Snippet che trova le chiavi close nel dizionario.*

Dunque, se si trova nel dizionario una chiave che ha una corrispondente *close*, delle due viene eliminata la chiave con valore “NOT” da entrambi i dizionari e mantenuta quella corretta con la categoria corrispondente.

```
def del_spuriouskey(dict):
    elimina = []
    for key in dict.keys():
        num = key.replace('[', ' ').replace(']', ' ')
        start = int(num.split(',')[0])
        end = int(num.split(',')[1])
        if end - start == 1:
            elimina += [key]
    for e in elimina:
        if e in dict.keys():
            del dict[e]
    return dict
```

Figura 7 - Snippet che va a eliminare le chiavi close nel dizionario.

A questo punto si ottengono come risultato finale due dizionari con le stesse chiavi, ossia gli item annotati, i cui valori corrispondono alle categorie date rispettivamente dai due annotatori e, in caso di mancata annotazione, alla categoria “NOT”.

Il dizionario dell’annotazione campione è la previsione delle annotazioni, ossia come dovrebbe risultare l’annotazione fatta da un annotatore esterno, mentre il dizionario dell’annotazione test è l’effettiva applicazione da parte di un annotatore esterno.

### 3.3.2 Metriche di valutazione

Per il calcolo delle metriche è stata utilizzata la libreria *scikit-learn*<sup>64</sup>. Si tratta di una libreria open source di apprendimento automatico del linguaggio in Python,

---

<sup>64</sup> <https://scikit-learn.org/stable/index.html>

progettata per operare con le librerie NumPy e SciPy, che implementa diversi algoritmi che trattano le funzionalità relative a problemi di apprendimento multiplo, inclusa la classificazione e la regressione multi-label.

La classificazione multi-label di *scikit-learn*, in linea con il metodo di adattamento dell'algoritmo, propone un'attività di classificazione che associa ogni item a  $n$  categorie possibili. Formalmente ad ogni categoria viene associata un'uscita binaria: le categorie con cui è stato etichettato un item vengono indicate con 1 e le categorie non utilizzate con 0. Dunque una rappresentazione dei dati da cui partire per il calcolo delle metriche con *scikit-learn* è una matrice binaria  $m$  per  $n$ , come l'esempio mostra per una matrice  $y$  con 3 item a cui è possibile assegnare 4 categorie:

```
y = np.array([[1, 0, 0, 1], [0, 0, 1, 1], [0, 0, 0, 0]])
```

Da questa struttura dei dati è dunque possibile calcolare con la libreria considerata le metriche di riferimento per i casi di multi-label classification.

Nel caso specifico del progetto, in primo luogo vengono create due matrici di dimensione *numero\_item* per *numero\_categorie*, riempite con tutti 0. Successivamente per ogni item viene sostituito 0 con 1 per ogni categoria associata a quel item in ciascun dizionario delle annotazioni, ottenendo così le due matrici per il calcolo delle *evaluation metrics*.

La classificazione multi-label richiede metriche differenti rispetto a quelle utilizzate dalla tradizionale classificazione a etichetta singola<sup>65</sup>. Infatti la libreria *scikit-learn* presenta il calcolo della *Kappa di Cohen* come indice di concordanza per casi di problemi binari o multi-class, esplicitando che non può essere applicato a casi di classificazione multi-label. Per questi casi la letteratura di riferimento propone i seguenti parametri: *Hamming Loss* (Schapire & Singer, 2000), *accuracy*, *precision* e *recall* (Godbole & Sarawagi, 2004) e Boutelli et al. (2004) forniscono una versione più generalizzata dell'*accuracy* calcolata in relazione ad un parametro chiamato *forgiveness rate*.

---

<sup>65</sup> G. Tsoumakas, I. Katakis, *Multi-Label Classification: An Overview*, in *International Journal of Data Warehousing and Mining (IJDM)*, (2007), IGI Global, vol.3, pp.1-13.

Il modulo *sklearn.metrics* implementa funzioni specifiche per la multi-label classification, in particolare permette di calcolare la *multilabel confusion matrix*, una matrice di confusione per ogni categoria utilizzata nelle due annotazioni, dove l'annotazione pilota presenta le previsioni, mentre l'annotazione test del secondo annotatore indica l'effettivo utilizzo.

| CATEGORIA |          | TRUE     |          |
|-----------|----------|----------|----------|
|           |          | POSITIVE | NEGATIVE |
| PREDICTED | POSITIVE | TP       | FN       |
|           | NEGATIVE | FP       | TN       |

Nella matrice di confusione vengono riportati i valori numerici di:

- *True Positive (TP)*: numero di item correttamente previsti come da annotare con quella categoria;
- *True Negative (TN)*: numero di item correttamente previsti come non da annotare con quella categoria;
- *False Positive (FP)*: numero di item erroneamente previsti come da annotare con quella categoria;
- *False negative (FN)*: numero di item erroneamente previsti come non da annotare con quella categoria.

Inoltre, con la libreria è possibile calcolare i valori di metriche come:

- *accuracy* o *exact matches*, la misura del rapporto tra le osservazioni correttamente previste e le osservazioni totali;

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- *Precision* che indica il rapporto tra le osservazioni positive previste correttamente sul totale delle osservazioni positive previste;

$$Precision = \frac{TP}{TP + FP}$$

- *Recall* che determina il rapporto tra le osservazioni positive previste correttamente e tutte le osservazioni della classe effettiva;

$$Recall = \frac{TP}{TP + FN}$$

- *F1 score* che indica la media armonica tra *Precision* e *Recall*;

$$F1\ score = 2 \frac{Precision * Recall}{Precision + Recall}$$

- *Jaccard score* che calcola la similarità tra due insiemi: definito come il rapporto tra la dimensione dell'intersezione e la dimensione dell'unione di due insiemi; sia T l'insieme delle annotazioni effettive (annotazione test) e P l'insieme di annotazioni previste (annotazione pilota):

$$Jaccard = \frac{|T \cap P|}{|T \cup P|}$$

- *Hamming Loss* calcolata tra due insiemi di campioni, quantifica la distanza tra i due. Confrontando i due insiemi, indica la media delle *Hamming Loss* sui singoli campioni, ossia il rapporto tra il numero delle categorie differenti e il numero totale delle categorie.

$$HL_{campione} = \frac{categorie\ differenti}{totale\ categorie}$$

Il calcolo delle metriche è stato attuato con `iaa.py`. In prima istanza lo script va a creare le due matrici inizialmente composte da tanti 0 quante sono le categorie di annotazione, poi riempite con un 1 in base al tag corrispondente per quella parola.

```
# crea matrice: per ogni posizione lista di tanti 0 quanti le categorie
def create_matrix(list1, list2):
    matrix = {}
    for elem1 in list1:
        matrix[elem1] = {}
        for elem2 in list2:
```

```

        matrix[elem1][elem2] = 0
    return matrix

def matrix_dict(dict, matrix):
    for y in dict.keys():
        for x in matrix.keys():
            if y == x:
                if len(dict[y]) == 1:
                    matrix[x][dict[y][0]] = 1
                if len(dict[y]) == 2:
                    matrix[x][dict[y][0]] = 1
                    matrix[x][dict[y][1]] = 1
                if len(dict[y]) == 3:
                    matrix[x][dict[y][0]] = 1
                    matrix[x][dict[y][1]] = 1
                    matrix[x][dict[y][2]] = 1
    return matrix
# RESTITUISCE: matrix_Letizia e matrix_Salvatori

```

Figura 8 - Snippet per la creazione e riempimento delle matrici.

Per ottenere un quadro più completo è stata calcolata inoltre una misura di accuratezza meno restrittiva, per cui si valuta l'intersezione non vuota tra gli insiemi dei due annotatori, ossia *at-least-one accuracy score*, che contempla almeno una categoria in comune sul totale delle categorie possibili.

```

correct = 0
for i in range(0, len(y_PRED)):
    l = y_PRED[i,:]
    e = y_TRUE[i,:]
    for k in range(0, len(l)):
        if l[k] == 1 and e[k] == 1:
            correct += 1
            break
atleastone = correct/len(y_PRED)
print ('Atleast one accuracy: ', atleastone)

```

Figura 9 - Snippet per il calcolo di *at-least-one accuracy*.

In prima istanza, è stato calcolato il numero totale di occorrenze delle categorie per ciascun dizionario, di modo da poter confrontare le differenze tra i valori:

|                    |   |    |    |     |     |    |    |    |    |   |     |     |   |   |    |
|--------------------|---|----|----|-----|-----|----|----|----|----|---|-----|-----|---|---|----|
| Annotazione pilota | 8 | 23 | 34 | 134 | 320 | 5  | 9  | 58 | 10 | 3 | 341 | 23  | 0 | 0 | 33 |
| Annotazione test   | 4 | 1  | 33 | 115 | 257 | 14 | 64 | 0  | 7  | 2 | 295 | 158 | 0 | 1 | 31 |

I valori delle metriche di valutazione calcolate con *scikit-learn* sulle due matrici delle annotazioni mostrano i seguenti risultati:

At-least-one accuracy score: 0.6529100529100529

Categorical accuracy: 0.691005291005291

Accuracy score: 0.6698412698412698

Precision score: 0.6743256743256744

Recall score: 0.6873727087576375

F1 score: 0.680786686838124

Jaccard score: 0.5160550458715596

Hamming loss: 0.04465608465608466

In generale i risultati ottenuti non mostrano un decisivo grado di accordo tra i due annotatori: tutte le metriche mostrano un valore intorno allo 0.6 su una scala che va da 0 totale disaccordo a 1 totale accordo, con il valore più basso 0.51 per l'indice di similarità tra i due insiemi di annotazioni. *Hamming loss* invece indica la distanza tra i due insiemi e dunque più il valore è vicino allo 0 migliore è il risultato: in questo caso il valore è 0.04.

La discrepanza tra le annotazioni può essere dovuta a due motivazioni:

1. le linee guida non risultano abbastanza chiare,
2. le categorie non rispecchiano a pieno i concetti semantici da designare.

Si è reso dunque necessario una fase di confronto tra i due annotatori per risolvere i *disagreement*, Questa condizione presuppone un procedimento per mettere a confronto le due annotazioni, in modo che gli annotatori arrivino a concordare uno

stile di annotazione condiviso, sia in caso di una variazione redazionale alle linee guida, sia per una modifica concettuale alle categorie.

### 3.3.3 Feedback

Il confronto ha previsto il trasferimento delle annotazioni al formato conllu e IOB2. Questo processo è funzionale e necessario nell'ultima fase del lavoro, ossia l'allenamento del classificatore per il riconoscimento automatica delle entità. Il formato però ha anche il vantaggio di avere una rappresentazione strutturata che consente una semplice visualizzazione e comparazione dei termini e tag associati.

Per questo è stato utilizzato il *file gold dataset*<sup>66</sup> ottenuto da un lavoro di analisi sintattica del cap. VIII svolto in parallelo dall'équipe, che ha fornito come risultato il file tokenizzato del testo. Su questo file chiamato `ch8_gold.conllu` sono state riportate le due annotazioni in colonne affiancate, in modo da confrontare e valutare visivamente su quali parole si differenziano le categorie scelte dai due annotatori.

Il file è in formato conllu, standard per NLP, specifico per l'output di annotazione linguistica. Nel *file gold* ogni riga rappresenta un token del testo con una serie di campi separati da tabulazione che ne precisano il lemma e il POS tagging. L'inserimento delle annotazioni è stato ottenuto con `IOBgold.py`. Lo script prende i due dizionari e sostituisce al range della posizione di inizio e fine la parola stessa che si trova in quella posizione sul testo originale. In questo modo il risultato è un dizionario che ha per chiave la parola e per valore la categoria o le categorie associate a quella parola nell'annotazione.

---

<sup>66</sup> <http://dialogo.di.unipi.it/LiberAbaci/>

```

with open('ch8_clean.txt', 'r') as f:
    ch8 = f.read()
#dizionario indicizzato per parola
def diz_word(dict):
    diz = {}
    for i in dict.items():
        num = i[0].replace('[', '').replace(']', '')
        range = num.split(',')
        init = int(range[0])
        end = int(range[1])
        word = ch8[init:end]
        if word in diz.keys():
            diz[word] += [i]
        else:
            diz[word] = [i]
    return diz

```

*Figura 10 - Snippet che crea il dizionario indicizzato per parola.*

A questo punto si scorre `ch8_gold.conllu` e, quando si incontra una parola che corrisponde alla parola del dizionario, si inserisce sul file la categoria in una nuova cella separata da tabulazione sulla riga della parola corrispondente.

```

lista = []
with open('ch8_gold.conllu', 'r') as f:
    for line in f:
        line = line.strip()
        if line == "":
            print(line)
        elif line.startswith("#"):
            print(line)
        else:
            fields = line.split("\t")
            word = fields[1]
            nominativo = fields[2]
            lista += [nominativo]
            word = word.replace("uenditionis", "venditionis")
            word = word.replace("uenditione", "venditione")
            word = word.replace("geroui", "gerovi")
            word = word.replace("gerouinos", "gerovinos")
            word = word.replace("gerouinus", "gerovinus")

```

```

word = word.replace("geroui", "gerovi")
iob = inside(word, d.keys())
if word in d.keys():
    t = controllo(word, lista, iob)
    print("\t".join(fields[0:4]) + "\t" + "\t".join(t))
else:
    print("\t".join(fields[0:4]) + "\tO")

```

Figura 11 – Snippet per l’inserimento dei tag sul gold.

Lo stesso procedimento avviene per entrambi i dizionari le cui categorie, per le parole corrispondenti, vengono sistemate su due celle contigue del file gold.

Nell’inserimento ad ogni categoria è stato inoltre applicato il formato IOB: O per le parole prive di riferimento semantico, B e I per contrassegnare le parole che rappresentano un unico blocco semantico (es. *nux muscata*, *denarius pisanus*, *rotuli gerovi*). Questo è stato ottenuto con un doppio controllo: se le parole nel dizionario erano già state annotate con un unico tag, sul file la prima riporta B e la seconda I. Per le parole che presentano tag multipli, B e I sono stati inseriti una sola volta, presentandoli come un unico blocco di indicazione di tale parola.

Ci sono anche casi per cui parole con un unico blocco semantico sono state annotate separatamente, questo dovuto principalmente alla sintassi latina (es. *rotuli 7 gerovi*). In questa casistica rientrano tutti i concetti di monete e unità di misure che hanno una specificazione geopolitica, dunque per inserire il formato IOB è stato applicato un controllo sul file: se nel file gold si trovano le parole che indicano monete e unità di misura (*soldus*, *denarius*, *rotulus*, *cantare*, *libra*) e sono seguite ad una distanza che va da 0 a 3 posizioni da parole che indicano una specificazione geopolitica (*pisanus*, *gerovus*, *forforus*, *barcellonensius*) la prima prende il tag B, la seconda I.

```

#Ritorna B, I o O in base alle parole del dizionario (se chiave con due
parole: prima parola B, seconda parole I)
def inside(word, dictkeys):
    for w in dictkeys:
        wl = w.split()
        if len(wl) > 0 and wl[0] == word:
            return "B"
        elif len(wl) > 1:

```

```

        for w2 in wl[1:]:
            if w2 == word:
                return "I"

    return "O"
#Crea il tag (I-[tag_parola], B-[tag_parola]) in base alle parole del file
gold: fa un controllo su parole che devono essere I
def controllo(word, list, iob):
    for i,item in enumerate(list):
        if item in [",", ".", ":", ";", "?", ",", "]:# spurious characters
            tags = "O"
        elif ((item == 'gerovi' or
              item == 'forfori' or
              item == 'gerovinus' or
              item == 'forforinus') and
              (list[i-1]=='rotulus' or
              list[i-2]=='rotulus')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'pisanus' and (list[i-1]=='cantare' or
              list[i-2]=='cantare')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'pisanus' and (list[i-1]=='libra' or
              list[i-2]=='libra')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'pisanus' and (list[i-1]=='rotulus' or
              list[i-2]=='rotulus')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'pisaninus' and (list[i-1]=='denarius' or
              list[i-2]=='denarius')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'pisaninus' and (list[i-1]=='soldus' or
              list[i-2]=='soldus')):
            tags = ['I-' + str(d[word][0][1])]
        elif (item == 'cantare' and list[i-1]=='de' and
              (list[i-2]=='denarius' or list[i-3]=='denarius')):
            tags = ['I-' + str(d[word][0][1])]
        else:
            tags = [iob + '-' + str(d[word][0][1])]
    return tags

```

*Figura 12 – Snippet per l'inserimento di IOB sui tag.*

Il risultato è dunque un file gold conllu che presenta l'analisi sintattica e semantica come mostra il seguente frammento del file:

```

22- karatos karatus NOUN- B-['I1 Money generally']- B-['N3.5 Measurement: Weight']
23- 6 6 NUM 0 0
24- , , PUNCT 0 0
25- qui qui PRON 0 0
26- karatus karatus NOUN- B-['N3.5 Measurement: Weight']- 0
27- ponderat pondero VERB 0 0
28- abbas abba NOUN- B-['N3.5 Measurement: Weight']- 0
29- 3 3 NUM 0 0
30- , , PUNCT 0 0
31- scilicet scilicet- CCONJ- 0 0
32- grana granum NOUN- B-['I1 Money generally']- B-['N3.5 Measurement: Weight']
33- 3 3 NUM 0 0
34- . . PUNCT 0 0

# sent_id = 1480 - - - - -
1- De de ADP 0 0
2- forfori forfori NOUN- B-['Z2 Geographical names', 'N3.5 Measurement: Weight'] B-['N3.5 Measurement: Weight', 'M7 Places']
3- ad ad ADP 0 0
4- geroui geroui NOUN- B-['N3.5 Measurement: Weight', 'Z2 Geographical names'] B-['N3.5 Measurement: Weight', 'M7 Places']
5- . . PUNCT 0 0

# sent_id = 1481 - - - - -
1- Item item ADV 0 0
2- si si SCONJ 0 0
3- de de ADP 0 0
4- rotulis rotulus NOUN- B-['N3.5 Measurement: Weight']- B-['N3.5 Measurement: Weight']
5- forfori forfori NOUN- I-['Z2 Geographical names', 'N3.5 Measurement: Weight']- I-['N3.5 Measurement: Weight', 'M7 Places']
6- 453 453 NUM 0 0
7- rotulos rotulus NOUN- B-['N3.5 Measurement: Weight']- B-['N3.5 Measurement: Weight']
8- geroui geroui NOUN- B-['N3.5 Measurement: Weight', 'Z2 Geographical names'] B-['N3.5 Measurement: Weight', 'M7 Places']
9- facere facio VERB 0 0
10- uolueris uolo VERB 0 0
11- , , PUNCT 0 0
12- describes describo VERB 0 0
13- forforinos forforinus ADJ B-['N3.5 Measurement: Weight', 'Z2 Geographical names'] 0

```

Figura 13 - Snippet file gold con entrambe le annotazioni.

Il file risultante, tuttavia, non ha una corrispondenza precisa tra i termini dei dizionari derivati dalle annotazioni e i token del *file gold*: per i token rintracciabili nel dizionario viene inserito il tag corrispondente indipendentemente dalla posizione. Riportare le annotazioni sul gold in maniera puntuale per posizione richiederebbe di riprodurre esattamente la tokenizzazione tra i due file, mentre in questo caso il risultato è un file gold in formato IOB con l'allineamento tra i token e i tag corrispondenti, utile per il confronto e la valutazione delle annotazioni.

A questo punto il file risultante è stato trasferito su Excel per rendere più semplice e immediato il controllo e la modifica durante l'incontro tra i due annotatori. Le differenze tra le categorie scelte riguardano principalmente:

- mancata annotazione di parole rilevanti

- variazione di annotazione tra pesi e monete
- disaccordo tra le annotazioni geopolitiche
- differenza di annotazioni per tipologie di merci.

Già dall'esempio di *Figura 13* risultano evidenti alcuni disaccordi che sono stati valutati puntualmente in merito alle questioni appena elencate.

In particolare, alcuni disaccordi sono derivati dalla difficoltà, in diverse parti del testo, di distinguere le unità di peso dalle monete, come, per esempio, *granum*, *libbra*, *uncia*, in quanto nel Medioevo i conti e le somme dei valori si facevano col peso dei metalli<sup>67</sup>. Queste sono state contestualizzate nel testo e risolte puntualmente.

Le altre questioni sono analizzate in modo significativo nel prossimo capitolo.

---

<sup>67</sup> E. Martinori, *La moneta: vocabolario generale*, (1915), Roma, Istituto Italiano di numismatica.

# Capitolo 4

## 4. Seconda fase: il taglio storico sul cap. VIII

### 4.1 Considerazioni storiche

Le questioni sollevate dal confronto tra gli annotatori hanno rivelato la necessità di considerare più nel dettaglio le questioni storiche che emergono dal testo sul dominio specifico mercantile-commerciale. Proprio per questo motivo si è ritenuto indispensabile una modifica profonda su due principali questioni.

- *Indicazioni geopolitiche.* Nel testo pesi e monete spesso vengono indicati con i valori specifici nelle varie zone geografiche di utilizzo. Per quanto riguarda le monete, la specificazione geografica designa l'entità politica emittente, ad esempio *libra pisana* indica la lira coniata a Pisa e *denarius regalis* la moneta coniata dalla corona di Francia; mentre per i pesi indica l'equivalenza del peso relativa ad un determinato territorio, come *rotuli forfori* è l'unità di peso dei rotoli utilizzata a Firenze e così via<sup>68</sup>. Questa distinzione concettuale di rilevanza storica tra entità governativa geopolitica per le monete e semplice indicazione geografica per i pesi viene dunque sottolineata anche a livello di annotazione semantica.
- *Tipologie di merci.* Le distinzioni pensate nella prima fase del lavoro andavano a considerare lo stato materiale in cui si presenta la merce (liquido, solido, aeriforme) che risulta poco utile a fini storici. Per questo si è valutato di considerare la funzione che hanno le merci nel commercio (da mangiare, da

---

<sup>68</sup> E. Martinori, *La moneta: vocabolario generale*, (1915), Roma, Istituto Italiano di numismatica.

bere, da fare vestiti). Tuttavia questa strada risulta piuttosto complessa in quanto un tipo di merce come il cuoio può essere usata per realizzare calzature, pergamene, scudi, tende, vestiti o ancora il pepe per insaporire i cibi ma anche, e soprattutto, per conservarli. Dunque dal punto di vista storico risulta più utile individuare la natura della merce (agricola, alimentare, animale, manifatturiera).

Queste considerazioni, poi, sono state presentate alla luce dei tag semantici per la ristrutturazione delle Linee Guida.

## 4.2 Modifica alle linee guida

Le Linee Guida sono risultate buone per l'indicazione di tag singoli, come, ad esempio, *venditio*, *negotiatio*, *pretium*, termini associati al tag <I2.2 Business: Selling> e *miliarum*, *decina*, *centum* indicati con <N5 Quantities>.

I tag annidati di monete e pesi hanno avuto bisogno di una modifica che permettesse di distinguerli secondo la valenza geografico-politica. Le monete si identificano come valute emesse da una specifica entità governativa geo-politica:

<denarius <regalium> G1.1 Gouvernement etc.> I1 Money generally,  
<bizantius <saracenus> G1.1 Gouvernement etc.> I1 Money generally,  
<denarius <pisaninus> G1.1 Gouvernement etc.> I1 Money generally,

mentre i pesi con identificazione geografica:

<rotuli <forfori> Z2 Geographical names> N3.5 Measurement: Weight,  
<libra <pisana> Z2 Geographical names> N3.5 Measurement: Weight,  
<uncia <Panormi> Z2 Geographical names> N3.5 Measurement: Weight.

Per quanto riguarda le tipologie di merci, sono stati mantenuti i tag multipli che comprendono la specificazione di oggetto di vendita con il tag <I2.2 Business: Selling> e la precisazione della natura della merce.

In particolare sono state divise le merci per gruppi di prodotti, assimilabili in base alla loro funzione:

- *frumentum, saccarum, nux muscatum, zafferanus, caseus, piper* da <O1.1 Substances and materials generally: Solid> sono stati convertiti in <F4 Farm and Horticulture>, in quanto prodotti alimentari derivanti dall' attività agricola;

- *oleum* si differenzia dai precedenti perché alimento di produzione liquido, tuttavia mantenere il tag <O1.2 Substances and materials generally: Liquid> renderebbe solo l'idea dello stato materiale dell'olio, dunque per questo viene assimilato ai prodotti dell'attività agricola con <F4 Farm and Horticulture>;

- *conilius* è considerato negli esempi di carico delle navi valutato sul volume che occupa la merce, il coniglio potrebbe aver avuto utilità per la produzione di pelli, ma anche come alimento, per questo si è reso utile valutarlo non come <F4 Farm and Horticulture>, che ne rivela solo la natura alimentare, ma con il tag <L2 Living creatures generally> che indica la tipologia animale della merce;

- *pannus, coto, petia, pannellus, linum, fustanium* dal generic oggetto solido sono stati identificati nel tag <B5 Clothes and personal belonging>, che comprende prodotti della tessitura, stoffe e vestiti;

- *rame, aurum, argentum, stagnum* sono utilizzati come prodotti commerciali. Il tagset di riferimento non propone un campo semantico specifico per i metalli, per cui sono stati definiti con il tag <O1.1 Substances and materials generally: Solid>;

- *corium, beccuna* sono la tipologia di merce di più difficile identificazione, poiché si tratta di prodotti semi-lavorati che all'epoca avevano vari utilizzi. Mantenendo il tag <O1.1 Substances and materials generally: Solid> che potrebbe generalizzare la natura di questi prodotti, si avrebbe tuttavia una sovrapposizione con la categoria dei metalli. Per questo è stata presa in considerazione la categoria <C1 Arts and Crafts> che rende l'idea di prodotti di artigianato e lavorati.

Altre questioni analizzate sono quelle relative alle società e al flusso monetario:

- *entheca, proficuum, lucrum, capitale* sono termini relativi a entrate di denaro e dunque resi con il tag <I1.1 Money: Affluence>, *capitale* che prima era indicato con <I2.1 Business Generally> è stato accorpato a questo gruppo in quanto identifica l'insieme delle somme in denaro conferito dai soci e destinato a impieghi produttivi per ottenere profitto;
- *societas, socius* che erano identificate da <I2.1 Business: Generally> hanno mantenuto questa valenza, aggiungendo un'ulteriore specificazione <S5 Groups and affiliation>, in modo che il doppio tag indichi sia la natura di attività produttiva, sia l'idea dell'associazione di persone in affari.

### 4.3 Il *file gold* del cap. VIII

Il file Excel prodotto dal confronto tra i due annotatori, `gold_ristrutturato.tsv`, è stato esportato in tsv, formato che mantiene la struttura tabellare con il carattere di tabulazione come separatore. Il file contiene su una colonna i tag originari, derivanti dall'annotazione con la prima redazione delle Linee Guida, e su un'altra, tutte le modifiche analizzate e ragionate nella seconda fase.

A questo punto è stato aggiornato il file `conllu`: si scorre `gold_ristrutturato.tsv` riga per riga, quando si incontra una riga che presenta i tag, si suddivide la stringa in base alle tabulazioni e si prende la parte di stringa posta dopo la quarta tabulazione, che corrisponde alla colonna dei tag modificati.

Deve inoltre essere riorganizzato il formato IOB sui tag, così da ottenere un'indicazione Beginning, Inside, Outside per ciascun tag. Inizialmente era stato pensato un unico blocco di tag per ciascuna entità indicato con il formato IOB. La modifica è stata resa necessaria soprattutto per la resa dei tag annidati, inizialmente redatti nel seguente modo:

```
2      uncia uncia NOUN B-['I1 Money generally']
3      pisana pisanus ADJ I-['I1 Money generally', 'G1.1
Government etc.]
```

La notazione di *uncia pisana* in questo modo indicherebbe: *uncia* con B come inizio dell'entità moneta e *pisana* con I intesa come parte interna dell'entità moneta, ma anche dell'entità governativa, che invece non ha inizio in nessun altro termine precedente.

Dunque la resa corretta è:

```
2   uncia uncia NOUN B-['I1 Money generally']
3   pisana pisanus ADJ I-['I1 Money generally'], B-['G1.1
Government etc.]
```

distinguendo i tag e rendendoli singolarmente.

Il risultato finale è `goldVIII.conllu` che presenta tutte le rielaborazioni della seconda fase sul capitolo VIII.

# Capitolo 5

## 5. Estensione sugli altri capitoli commerciali

Il lavoro di analisi semantica è stato poi esteso agli altri capitoli commerciali con l'obiettivo di valutare la copertura dei campi semantici individuati nella fase precedente.

La fase di annotazione di un nuovo testo di dominio commerciale del *Liber Abaci*, è stata dunque fatta unicamente sulla base delle Linee Guida aggiornate.

Il *tool* di annotazione Catma produce annotazioni che sappiamo confrontare, ma che non sappiamo trasferire in maniera affidabile, in quanto manca l'aggancio tra la posizione delle parole nel testo e quelle sul file conllu. Una possibile soluzione è quella di avere un testo da annotare già tokenizzato, tuttavia Catma, per quanto riguarda l'esportazione delle annotazioni, non rispetta la tokenizzazione del file in input.

La sostituzione di Catma con un nuovo strumento di annotazione, in grado di rispettare la tokenizzazione e di esportare il file direttamente in formato conllu, avrebbe richiesto una nuova valutazione del tool da parte dell'èquipe e un costo in termini di tempo per adeguare le funzionalità dello strumento al progetto.

Per questo motivo, la soluzione è stata quella di mantenere Catma che restituisce il file di output annotato in XML-TEI - da cui ottenere i dizionari di annotazioni attraverso lo script già realizzato e utilizzato per il capitolo VIII - e, invece, modificare il file gold su cui trasferire le annotazioni. La modifica ha riguardato i dati contenuti nel file gold, poiché oltre alla presenza di tokenizzazione, lemmatizzazione e POS, è necessaria l'aggiunta del range della posizione di inizio e fine della parola sul testo.

## 5.1 Annotazione dei cap. IX e X

L'annotazione semantica è stata applicata alla parte iniziale dei capitoli IX e X: i due frammenti testuali sono composti da un numero di token che va da 3100 a 3900 (compresi i segni di punteggiatura) per evitare di spezzare le frasi. Questi file sono stati importati su Catma e annotati con il tagset ristrutturato e sulla base delle linee guida aggiornate nella fase precedente del lavoro.

Il frammento iniziale del capitolo IX riguarda il baratto tra merci varie e scambio tra diverse monete. La terminologia si sovrappone perfettamente alle linee guida redatte e, solo in alcuni casi, arricchisce il lessico dei campi semantici esistenti: ad esempio *emptio* come sinonimo di *venditio* e dunque riferito al tag <I2.2 Business: Selling>; *bolsonalie* che, come viene spiegato poi nella seconda parte del capitolo, sono monete vendute a peso o a valore dell'argento che contengono, viene associato al tag <I1 Money generally>; *bombicus*, come sinonimo di *coto*, merce tessile e dunque reso con <I2.2 Business: Selling><B5 Clothes and personal belongings> e infine un altro tipo di merce, ossia *mastica*, una resina vegetale prodotta dal lentisco e dunque associata a <I2.2 Business: Selling><F4 Farming & Horticulture>.

Nel testo si trovano inoltre due riferimenti a persone Tolomeo e Ameto che risultano rilevanti dal punto di vista storico. Il primo è sicuramente Claudio Tolomeo di cui viene citato anche il nome dell'opera da lui scritta: l'Almagesto, un grande trattato astronomico-matematico; il secondo, indicato sul testo come *Ametus filius Joseph*, è probabilmente Ahmad ibn Yusuf, matematico e astronomo arabo, che scrisse un trattato sulle proporzioni e sul teorema di Menelao.<sup>69</sup> I due personaggi vengono indicati con il tag <Z1 Personal names> e *Almagestus* con <Z3 Other proper name>, inteso come nome proprio di un'opera.

---

<sup>69</sup> L. E. Sigler, *Fibonacci's Liber Abaci. A translation into Modern English of Leonardo Pisano's book of calculation*, (2002), New York, Springer, p. 623.

| CATEGORIA                             | FREQ. |
|---------------------------------------|-------|
| B5 Clothes and personal belongings    | 11    |
| F4 Farming & Horticulture             | 43    |
| G1 Government, Politics               | 41    |
| I1 Money generally                    | 88    |
| I2 Business                           | 119   |
| N3 Measurement                        | 80    |
| O1 Substances and materials generally | 6     |
| O2 Objects generally                  | 31    |
| Z1 Personal names                     | 3     |
| Z2 Geographical names                 | 4     |
| Z2 Geographical names                 | 1     |

*Tabella 2 – Categorie con relativa frequenza del cap. 9.*

Il frammento del capitolo X tratta delle società create tra due o più soci, delle quote di capitale messe dai consoci e del profitto che ne ricavano in base alla quota. I termini si riferiscono tutti all'ambito finanziario (*socius, societas, lucrum, capitale*) e monetario (*soldus, tarenum, libra, denarius*) già individuati nel capitolo VIII e riportati sulle Linee Guida.

| CATEGORIA                 | FREQ. |
|---------------------------|-------|
| G1 Government, Politics   | 3     |
| I1 Money generally        | 204   |
| I2 Business               | 68    |
| N3 Measurement            | 3     |
| S5 Groups and affiliation | 65    |

*Tabella 3 – Categorie con relativa frequenza del cap. 10.*

Una volta annotate le due parti di testo, è stato applicato lo script (`dizionari.py`) precedentemente descritto, con cui si ottengono i relativi dizionari

che hanno come chiave il range di inizio e fine del termine e come valore il tag semantico corrispondente.

## 5.2 Creazione dei *file gold* IX e X

I file gold dei capitoli IX e X sono stati realizzati in modo autonomo rispetto a quello già predisposto del capitolo VIII. L'obiettivo è quello di avere, come file di partenza per riportare i tag semantici, un file gold con l'analisi sintattica in cui viene indicato anche il range di inizio e fine della parola sul testo originale.

In questo modo avendo il range indicato sia sui dizionari delle annotazioni, sia sul file gold di partenza, confrontando i due valori, può essere attuato un trasferimento delle annotazioni sul gold in maniera sistematica e affidabile.

Per ottenere il file gold di partenza è stato utilizzato il tool UDPipe<sup>70</sup>.

### 5.2.1 UDPipe

UDPipe è un tool per l'analisi sintattica di testi che permette di realizzare tokenizzazione, lemmatizzazione, POS tagging e analisi delle dipendenze. Attraverso il tool è possibile indicare:

- il formato del file in input, nel nostro caso è un formato testuale semplice,
- il formato del file output, utile ai fini del progetto è ottenere un file nello standard conllu;
- il tipo di analisi da svolgere, nello specifico ci interessa la tokenizzazione.

Attuando la tokenizzazione, tra le opzioni è possibile definire anche il campo TokenRange, che specifica l'indice di inizio e fine del token a livello di documento, quindi del file testuale dato in input.

---

<sup>70</sup> <https://lindat.mff.cuni.cz/services/udpipe/>

Con questa funzionalità sono stati dati in input i frammenti *plain text* dei due capitoli del *Liber Abaci* da analizzare, ottenendo come output, in formato conllu, il file tokenizzato con lemma, analisi morfologica e range del token nel seguente modo:

```
# sent_id = 1
1- Incipit incipio VERB  _ _ _ _ _ TokenRange=0:7
2- capitulum capitulum NOUN  _ _ _ _ _ TokenRange=8:17
3- nonum nonus ADJ  _ _ _ _ _ TokenRange=18:23
4- de de ADP  _ _ _ _ _ TokenRange=24:26
5- baractis baractum NOUN  _ _ _ _ _ TokenRange=27:35
6- mercium merx NOUN  _ _ _ _ _ TokenRange=36:43
7- atque atque CCONJ  _ _ _ _ _ TokenRange=44:49
8- earum is PRON  _ _ _ _ _ TokenRange=50:55
9- similium similius ADJ  _ _ _ _ _ TokenRange=56:65
10- . . PUNCT  _ _ _ _ _ TokenRange=00:00

# sent_id = 2
1- Hoc hic DET  _ _ _ _ _ TokenRange=67:70
2- itaque itaque CCONJ  _ _ _ _ _ TokenRange=71:77
3- capitulum capitulum NOUN  _ _ _ _ _ TokenRange=78:87
4- in in ADP  _ _ _ _ _ TokenRange=88:90
5- tres tres NUM  _ _ _ _ _ TokenRange=91:95
6- partes pars NOUN  _ _ _ _ _ TokenRange=96:102
7- dividere divido VERB  _ _ _ _ _ TokenRange=103:111
8- decrevi decerno VERB  _ _ _ _ _ TokenRange=112:120
9- , , PUNCT  _ _ _ _ _ TokenRange=00:00
10- ut ut SCONJ  _ _ _ _ _ TokenRange=121:123
11- quicquid quisquis DET  _ _ _ _ _ TokenRange=124:132
12- lector lector NOUN  _ _ _ _ _ TokenRange=133:139
13- in in ADP  _ _ _ _ _ TokenRange=140:142
14- hoc hic DET  _ _ _ _ _ TokenRange=143:146
15- audire audio VERB  _ _ _ _ _ TokenRange=147:153
16- desiderat desidero VERB  _ _ _ _ _ TokenRange=154:163
17- citius cito ADV  _ _ _ _ _ TokenRange=164:170
18- reperire reperio VERB  _ _ _ _ _ TokenRange=171:179
19- valeat valeo VERB  _ _ _ _ _ TokenRange=180:187
20- . . PUNCT  _ _ _ _ _ TokenRange=00:00
```

Figura 14 - File output di UDPipe (*gold\_cap9.conllu*).

Il file è stato corretto manualmente in quanto il tool non distingueva la punteggiatura come un unico token, ma la riportava associata alla parola precedente. La modifica ha riguardato solamente la distinzione tra parole e segni di interpunzione, mentre non è stato cambiato manualmente il valore dei TokenRange: posto a valore 0 per la punteggiatura e invariato quello relativo alla parola.

Questa incongruenza è stata poi tenuta in considerazione in fase di confronto tra i range dei dizionari e i range del file gold.

## 5.3 I file gold IX e X completi

Per ottenere i file gold dei capitoli IX e X completi delle annotazioni è stato realizzato con lo script `tag_insert.py` che confronta i valori dei range nei dizionari delle annotazioni e nei file gold.

Per ciascun capitolo scorrendo il dizionario e scorrendo il file gold, se si trovano range uguali o che differenziano di 1 nel valore di fine range, il tag, che indica il valore nel dizionario relativo a quel range, viene inserito nel file gold. Con lo scarto di 1 tra i range si va a considerare il valore che identifica un segno di interpunzione calcolato come un unico token con la parola precedente del file risultante dalla fase di tokenizzazione di UDPipe.

```
with open('gold_cap10.conllu', 'r') as f:
    d = {}
    for line in f:
        line = line.strip()
        if line == "" or line.startswith("#"):
            continue
        else:
            fields = line.split("\t")
            range = fields[9].replace('TokenRange=', '').replace(':', ',')
            start1 = int(range.split(',')[0])
            end1 = int(range.split(',')[1])
            for e in diz.items():
                key2 = e[0].replace('[', '').replace(']', '')
                start2 = int(key2.split(',')[0])
                end2 = int(key2.split(',')[1])
                if (start1 == start2 and end1 == end2) or
                    (start1 == start2 and abs(end1 - end2) == 1):
                    d[range] = e[1]
```

*Figura 15 – Snippet che confronta i range.*

Il tag corrispondente per il range da inserire nel file gold viene posizionato nel campo libero dopo la quarta tabulazione con il formato IOB. Il risultato finale è un file gold come mostra il seguente frammento:

```

# sent_id = 10
1- Verbi- verbum- NOUN- 0
2- gratia- gratia- NOUN- 0
3- :- :- PUNCT- 0
4- brachia- brachium- NOUN- B-['N3.7 Measurement: Length & height']
5- 20- 20- NUM 0
6- panni- pannus- NOUN- B-[I2.2 Business: Selling], B-[B5 Clothes and personal belongings]
7- valeant- valeo- VERB- 0
8- libras- libra- NOUN- B-['I1 Money generally']
9- 3- 3- NUM 0
10- pisanorum- pisanorum- NOUN- I-[I1 Money generally], B-[G1.1 Government etc.]
11- ,- ,- PUNCT- 0
12- et- et- CCONJ- 0
13- rotuli- rofero- NOUN- B-['N3.5 Measurement: Weight']
14- 42- 42- NUM 0
15- cotonis- coto- NOUN- B-[B5 Clothes and personal belongings], B-[I2.2 Business: Selling]
16- valeant- valeo- VERB- 0
17- libras- libra- NOUN- B-['I1 Money generally']
18- 5- 5- NUM 0
19- similiter- similiter- ADV- 0
20- pisanorum- pisanorum- NOUN- I-[I1 Money generally], B-[G1.1 Government etc.]
21- .- .- PUNCT- 0

# sent_id = 11
1- Queritur- queror- VERB- 0
2- pro- pro- ADP- 0
3- brachiis- brachium- NOUN- B-['N3.7 Measurement: Length & height']
4- 50- 50- NUM 0
5- panni- pannus- NOUN- B-[I2.2 Business: Selling], B-[B5 Clothes and personal belongings]
6- quot- quot- DET- 0
7- rotuli- rotulus- NOUN- B-['N3.5 Measurement: Weight']
8- cotonis- coto- NOUN- B-[I2.2 Business: Selling], B-[B5 Clothes and personal belongings]
9- habebuntur- habeo- VERB- 0
10- .- .- PUNCT- 0

```

*Figura 16 – Snippet file gold (goldIX.conllu).*

I tag semantici inseriti nel file gold vengono completati con l'indicazione di Inside, Outside e Beginning, inserita attraverso il controllo sulle parole specifiche individuate come blocco semantico unico.

# Capitolo 6

## 6. Sistemi di apprendimento automatico

Dopo avere esaminato ed attuato la metodologia dell'annotazione, si analizza ora l'ambito computazionale all'interno del quale vengono utilizzati tali dati annotati, ovvero quello dell'apprendimento automatico del linguaggio.

Le informazioni che vengono utilizzate per codificare uno specifico fenomeno devono essere sufficientemente ricche da catturare il comportamento desiderato dall'algoritmo che vogliamo addestrare. I dati annotati sono dunque utilizzati per sviluppare algoritmi di Machine Learning supervisionati con lo scopo di insegnare al computer a prendere decisioni per lo svolgimento di un determinato task a partire da un campione rappresentativo dei dati. Tali algoritmi richiedono che le categorie assegnabili agli input siano già definite a priori e che i dati di input utilizzati per l'addestramento siano già provvisti dell'etichetta corretta.

I task per cui vengono utilizzati determinati algoritmi supervisionati sono ad esempio: regressione, alberi di decisione e classificazione. In particolare, la classificazione è il task che prevede l'identificazione delle etichette per le singole entità da un insieme di dati predisposti.

Per quanto riguarda il caso specifico del progetto, il nostro sistema rientra nella categoria della classificazione per algoritmi di ML supervisionati, in quanto richiede per l'addestramento del classificatore l'utilizzo di dati, nello specifico lemmi delle entità commerciali, su cui determinare la categoria o le categorie corrispondenti.

Il processo di sviluppo per addestrare e testare un algoritmo di identificazione e classificazione può essere individuato nella metodologia del ciclo MATTER<sup>71</sup>, già indicata nella sua prima fase di sviluppo per l'annotazione. Le sue fasi vengono mostrate nel dettaglio in *Figura 17*.

---

<sup>71</sup> J. Pustejovsky, A. Stubbs, *Natural Language Annotation for Machine Learning*, (2012), O'Reilly Media, pp. 29-30.

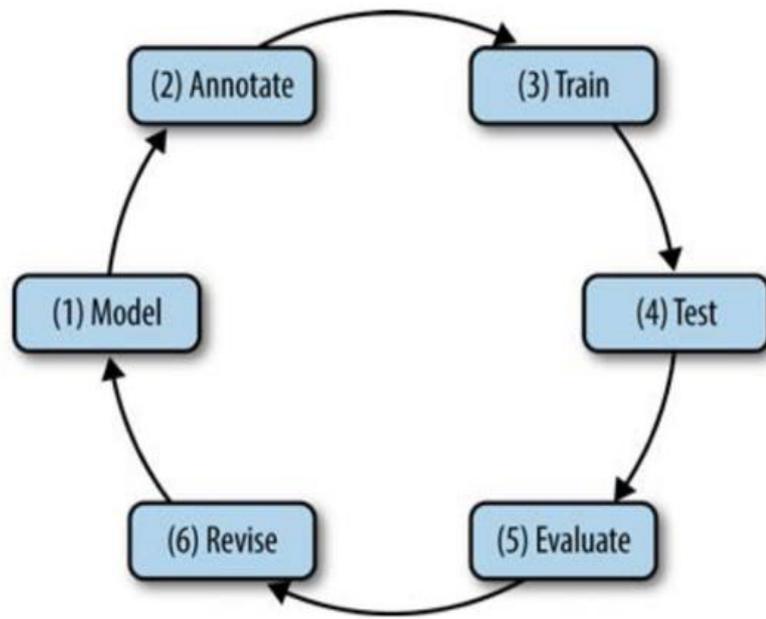


Figura 17 - Ciclo MATTER (Pustejovsky & Stubbs, 2012).

Il ciclo prevede le prime due fasi, Modello e Annotazione, già prese in analisi per il processo di annotazione, seguite da altre quattro che identificano il cuore del modello di apprendimento supervisionato:

- *Train*, fase in cui l'algoritmo viene addestrato sul corpus annotato;
- *Test*, fase di prova del sistema addestrato su un nuovo campione di dati;
- *Evaluate*, riguarda la valutazione delle performance del sistema;
- *Revise*, in caso scarsa corrispondenza dei dati nella valutazione, fase di revisione del modello e delle specifiche di annotazione al fine di rendere l'annotazione più robusta per addestrare il modello.

## 6.1 Train e Test dell'algoritmo

Le fasi di apprendimento automatico si sviluppano, a livello pratico, nella suddivisione del corpus in due parti: una parte sarà il train, l'altra il test.

La fase di train prevede la selezione di dati annotati del corpus, definiti *training set*, che sono utilizzati per identificare l'approccio migliore di addestramento del modello. Questo permette di creare un classificatore che riconosca ed etichetti le entità del corpus.

Nella fase di test vengono selezionati una serie di dati non annotati del corpus, il *test set*, diversi da quelli di training, su cui applicare il classificatore per annotare ciascun token del set di dati. Alla fine di questa fase si hanno in output i dati annotati.

In ultima istanza, viene utilizzato un *gold standard*, formato dagli stessi dati del *test set* annotati manualmente, con cui valutare le performance del modello addestrato.

Il metodo per valutare le prestazioni dell'algoritmo è il confronto dell'output del modello applicato sui dati del *test set* con il *gold standard*. Un modo utile per esaminare i risultati è creare la matrice di confusione<sup>72</sup>, in cui ogni colonna rappresenta le etichette predette, ossia quelle sul gold standard, e ogni riga le etichette reali, quelle inserite nel test dall'algoritmo. Le annotazioni corrette sono indicate dai valori sulla diagonale, mostrando così graficamente se vi è accordo nella classificazione delle entità nelle diverse categorie. La matrice di confusione restituisce infatti la rappresentazione dell'accuratezza di classificazione statistica. Dalla matrice è possibile definire dunque l'*accuracy*, il rapporto tra le risposte corrette sulle risposte fornite nell'etichettatura svolta dal classificatore. Per valutare le prestazioni è, inoltre, utile il calcolo di altre metriche come *Precision*, *Recall* e *F1*.

---

<sup>72</sup> J. Pustejovsky, A. Stubbs, *Natural Language Annotation for Machine Learning*, (2012), O'Reilly Media, pp.171-174.

# Capitolo 7

## 7. Riconoscimento delle entità commerciali del *Liber Abaci*

L'approccio iniziale per il riconoscimento automatico delle entità mercantili-commerciali nel *Liber Abaci* ha riguardato lo sviluppo di una baseline, ossia un modello semplice di base su cui, poi, andare a creare soluzioni più complesse per ottenere un risultato migliore.

Nel caso del nostro progetto si ha a che fare con un dominio specifico su cui sviluppare il modello di apprendimento automatico supervisionato. In generale, scegliere una baseline permette uno sviluppo in prima istanza semplice, di cui valutare i limiti e stabilire delle regole euristiche che vadano a identificare strategie e procedimenti inventivi per sviluppare una procedura adeguata a risolvere il task in maniera il più possibile precisa.

### 7.1 Baseline con approccio *dictionary-based*

La baseline che è stata sviluppata è *dictionary-based*, ossia l'approccio propone la creazione di un dizionario dei lemmi con rispettive categorie, ottenuto su una parte del corpus di dati annotati.

Sulla base del dizionario è stato realizzato un classificatore che andasse a identificare i lemmi e a classificarli con la specifica categoria in una porzione differente del corpus formato da dati non annotati.

Vediamo nel dettaglio le fasi di questo processo.

## 7.1.1 Training e test set

Il corpus di dati annotati composto da `goldVIII.conllu`, `goldIX.conllu` e `goldX.conllu`, è stato diviso in due parti:

- `train.conllu` che comprende quasi interamente i dati annotati del cap. VIII e le prime quattro frasi annotate del cap. IX e X;
- `test_gold.conllu` formato, invece, dalle restanti parti annotate dei cap. IX e X e le ultime 25 frasi del cap. VIII.

```
train.conllu x ...
1 # sent_id = 752
2 1- Incipit incipio VERB 0
3 2- capitulum capitulum NOUN 0
4 3- octauum octauus ADJ 0
5 4- de de ADP 0
6 5- reperiendis reperio VERB 0
7 6- pretiis pretium NOUN B-I2.2
8 7- mercium merx NOUN B-I2.2, B-02
9 8- per per ADP 0
10 9- maiorem magnus ADJ 0
11 10- guisam guisa NOUN 0
12 11- . . PUNCT 0
13
14 # sent_id = 753
15 1- In in ADP 0
16 2- omnibus omnis DET 0
17 3- itaque itaque PART 0
18 4- negotiationibus negotiatio NOUN B-I2.2
19 5- quattuor quattuor NUM 0
20 6- numeri numerus NOUN 0
21 7- proportionales proportionalis ADJ 0
22 8- semper semper ADV 0
23 9- reperiuntur reperio VERB 0
24 10- , , PUNCT 0
25 11- ex ex ADP 0
26 12- quibus qui PRON 0
27 13- tres tres NUM 0
28 14- sunt sum AUX 0
29 15- noti notus ADJ 0
30 16- , , PUNCT 0
31 17- reliquus reliquus DET 0
32 18- uero uero ADV 0
33 19- est sum AUX 0
34 20- ignotus ignotus ADJ 0
35 21- . . PUNCT 0
36
37 # sent_id = 754

test_gold.conllu x ...
1 # sent_id = 1478
2 1- Et et CCONJ 0
3 2- sic sic ADV 0
4 3- eos is PRON 0
5 4- describes describo VERB 0
6 5- in in ADP 0
7 6- questione quaestio NOUN 0
8 7- , , PUNCT 0
9 8- et et CCONJ 0
10 9- pones pono VERB 0
11 10- rotulos rotulus NOUN B-N3.5
12 11- 347 347 NUM 0
13 12- geroui geroui NOUN I-N3.5, B-Z2
14 13- sub sub ADP 0
15 14- rotulis rotulus NOUN B-N3.5
16 15- 6 6 NUM 0
17 16- geroui geroui NOUN I-N3.5, B-Z2
18 17- ut ut SCONJ 0
19 18- hic hic ADV 0
20 19- ostenditur ostendo VERB 0
21 20- , , PUNCT 0
22 21- et et CCONJ 0
23 22- multiplica multiplico VERB 0
24 23- 13 13 NUM 0
25 24- per per ADP 0
26 25- 347 347 NUM 0
27 26- et et CCONJ 0
28 27- diuide diuido VERB 0
29 28- per per ADP 0
30 29- 6 6 NUM 0
31 30- : : PUNCT 0
32 31- exhibunt exeo VERB 0
33 32- rotuli rotulus NOUN B-N3.5
34 33- forfori forfori NOUN I-N3.5, B-Z2
35 34- 5 5 NUM 0
36 35- / / SYM 0
37 36- 6 6 NUM 0
```

Figura 18 - Snippet dei file `train.conllu` e `test_gold.conllu`.

## 7.1.2 Il Dizionario

A questo punto è stato creato un dizionario sulla base del file `train.conllu`: scorrendo il file per ogni lemma, se ad esso è associato un tag diverso da “O”, nel dizionario viene inserito il lemma con i relativi tag (uno o più tag in base al modo in cui è stata annotata l’entità).

```
lista = []
dictionary = {}
with open('train.conllu', 'r') as f:
    for line in f:
        line = line.strip()
        line = line.strip()
        if line == "" or line.startswith("#"):
            continue
        else:
            fields = line.split("\t")
            lemma = fields[2]
            tag = fields[4]
            if tag != "O":
                lista += [[lemma, tag]]
                dictionary[lemma] = [tag]
for e in lista:
    if e[0] in dictionary.keys() and e[1] not in dictionary.values():
        dictionary[e[0]] += [e[1]]
```

*Figura 19 – Snippet del codice per la creazione del dizionario.*

Il risultato finale è un dizionario che ha per chiave tutti i lemmi annotati e per valore tutte le possibili categorie di quei lemmi, come mostra l’esempio:

```
{'pretium': ['B-I2.2'], 'merx': ['B-I2.2, B-O2'], ... ,
'pisanus': ['I-N3.7, B-Z2', 'B-I1, B-G1.1', 'B-N3.7, B-Z2', 'I-
N3.5, B-Z2', 'I-I1, B-G1.1', 'B-N3.5, B-Z2'], ... , 'soldus':
['B-I1', 'B-N3.5'],...}
```

### 7.1.3 Il Classificatore

Una volta ottenuto il dizionario, viene creato un file su cui testare l'applicazione delle categorie: `test_sys.conllu`, ossia un file costruito su `test_gold.conllu` con la struttura tokenizzata del testo, i relativi lemmi, tuttavia privo dei tag. Il file viene riempito con i tag corrispondenti ai lemmi attraverso un classificatore.

```
def prendoTag(lemma, dict):
    for x in dict.items():
        if x[0] == lemma:
            tag = [x[1][0]]
            return tag
#crea e riempie test_sys.conllu
with open('test_gold.conllu', 'r') as f:
    for line in f:
        line = line.strip()
        if line == "":
            print(line)
        elif line.startswith("#"):
            print(line)
        else:
            fields = line.split("\t")
            lemma = fields[2]
            if lemma in dictionary.keys():
                t = prendoTag(lemma, dictionary)
                print("\t".join(fields[0:4]) + "\t" + "\t".join(t))
            else:
                print("\t".join(fields[0:4]) + "\tO")
```

*Figura 20 – Snippet del classificatore.*

Il classificatore scandisce `test_gold.conllu` e riporta riga per riga token, lemma e analisi POS sul nuovo file `test_sys.conllu`. Per ogni lemma del file controlla se tale lemma è presente nelle chiavi del dizionario, in caso positivo nel file associa al lemma la prima categoria riportata nel dizionario, altrimenti il tag “O”. Il risultato è il seguente:

```

test_sys.conllu x
1 # sent_id = 1478
2 1 Et et CCONJ 0
3 2 sic sic ADV 0
4 3 eos is PRON 0
5 4 describes describo VERB 0
6 5 in in ADP 0
7 6 questione quaestio NOUN 0
8 7 , , PUNCT 0
9 8 et et CCONJ 0
10 9 pones pono VERB 0
11 10 rotulos rotulus NOUN B-N3.5
12 11 347 347 NUM 0
13 12 geroui geroui NOUN I-N3.5, B-Z2
14 13 sub sub ADP 0
15 14 rotulis rotulus NOUN B-N3.5
16 15 6 6 NUM 0
17 16 geroui geroui NOUN I-N3.5, B-Z2
18 17 ut ut SCONJ 0
19 18 hic hic ADV 0
20 19 ostenditur ostendo VERB 0
21 20 , , PUNCT 0
22 21 et et CCONJ 0
23 22 multiplica multiplico VERB 0
24 23 13 13 NUM 0
25 24 per per ADP 0
26 25 347 347 NUM 0
27 26 et et CCONJ 0
28 27 diuide diuido VERB 0
29 28 per per ADP 0
30 29 6 6 NUM 0
31 30 :- :- PUNCT 0
32 31 exhibunt exeo VERB 0
33 32 rotuli rotulus NOUN B-N3.5
34 33 forfori forfori NOUN I-N3.5, B-Z2
35 34 5 5 NUM 0

```

Figura 21 – Snippet del file test\_sys.conllu.

Un approccio meno naïf potrebbe essere quello per cui il classificato se trova il lemma del file test nel dizionario, inserisca, al posto della prima categoria presentata, la categoria con cui tale lemma risulta con frequenza maggiore nel file train. Questa è un'implicazione del fatto che il dizionario venga costruito con una struttura più elaborata, ossia presenti le categorie di ogni lemma non in ordine casuale, ma in ordine di frequenza.

## 7.1.4 Valutazione del classificatore

Per valutare le prestazioni dell'algoritmo viene fatto il confronto dell'output del modello applicato sui dati di `test_sys.conllu` con `test_gold.conllu`. Questo viene realizzato con `valutazione.py`: lo script scorre i due file annotati e crea rispettivamente due liste, `y_true` e `y_pred`, che contengono la lista dei tag utilizzati. In questo modo si ottengono due liste simmetriche, ordinate per token, di cui confrontare la corrispondenza dei tag.

```
y_pred = []
with open('test_gold.conllu', 'r') as f:
    for line in f:
        line = line.strip()
        if line == "" or line.startswith("#"):
            continue
        else:
            fields = line.split("\t")
            tag = fields[4]
            lemma = fields[2]
            y_pred += [tag]

y_true = []
with open('test_sys.conllu', 'r') as f:
    for line in f:
        line = line.strip()
        if line == "" or line.startswith("#"):
            continue
        else:
            fields = line.split("\t")
            tag = fields[4]
            lemma = fields[2]
            y_true += [tag]
```

*Figura 22 – Snippet della creazione delle liste per la valutazione.*



| NER tags       | # ERRORS |
|----------------|----------|
| B-I1           | 178      |
| 0              | 61       |
| B-I2.2, B-F4   | 44       |
| B-I1, B-G1.1   | 27       |
| B-N3.5         | 16       |
| B-N3.7         | 15       |
| B-I2.2, B-B5   | 15       |
| B-N3.5, B-Z2   | 13       |
| I-I1, B-G1.1   | 13       |
| B-I2.2         | 5        |
| B-I1.1         | 5        |
| B-I2.1, B-S5   | 3        |
| B-N3.4         | 3        |
| B-M7           | 3        |
| B-Z1           | 3        |
| B-G1.1         | 1        |
| B-Z3           | 1        |
| B-I2.2, B-01.1 | 1        |
| I-I1           | 1        |
| B-Z2, B-N3.5   | 1        |
| I-N3.5, B-Z2   | 1        |

Figura 24 – Analisi degli errori per categoria.

L'accuracy è molto alta ma poco significativa, mentre le metriche di Precision, Recall e F1 sono molto variabili per i diversi tag. Il calcolo di queste tre metriche consente di fare un'analisi degli errori più fine:

| NER          | PREC | RECALL | F1   |
|--------------|------|--------|------|
| B-I1         | 0.88 | 0.08   | 0.14 |
| B-I1.1       | 1.00 | 0.92   | 0.96 |
| B-I2.1, B-S5 | 0.98 | 0.95   | 0.96 |
| B-I2.2       | 0.96 | 0.81   | 0.88 |
| B-I2.2, B-B5 | 1.00 | 0.21   | 0.35 |
| B-I2.2, B-C1 | 1.00 | 1.00   | 1.00 |
| B-I2.2, B-F4 | 1.00 | 0.14   | 0.24 |
| B-I2.2, B-02 | 0.94 | 1.00   | 0.97 |
| B-M7         | 1.00 | 0.40   | 0.57 |
| B-N3.1       | 0.02 | 1.00   | 0.04 |
| B-N3.5       | 0.46 | 0.89   | 0.60 |
| B-N5         | 0.33 | 1.00   | 0.50 |
| I-I1, B-G1.1 | 0.14 | 0.19   | 0.16 |
| I-N3.5       | 1.00 | 1.00   | 1.00 |
| I-N3.5, B-Z2 | 0.48 | 0.92   | 0.63 |
| 0            | 0.98 | 0.99   | 0.98 |
| TOT          | 0.76 | 0.72   | 0.63 |

Figura 25 – Precision, Recall, F1 per categoria.

Si nota, ad esempio, che il tag B-I1 ha una *Recall* molto bassa, consistente col fatto che su questo tag ci sono 178 errori, mentre solo 15 volte è stato riconosciuto correttamente. Questa circostanza è dovuta al fatto che la classificazione attraverso il dizionario riporta sul `test_sys.conllu` il primo dei tag relativi al lemma specifico. Spesso token con tag B-I1, in base al contesto della frase, possono essere, invece, resi con il tag 'B-N3.5' o 'B-N3.5, B-Z2' o 'B-I1, B-G1.1' e rispettivi formati I.

Considerando alcuni lemmi come *soldus* e *granum* che possono riferirsi, in base al contesto, sia all'unità di peso sia al valore monetario, il classificatore li assegnerà sempre e solo ad un tag, nello specifico B-N3.5, mentre il riferimento a I1 non si verificherà mai. Nel confronto con `test_gold.conllu`, dunque, i casi in cui ha valore monetario saranno tutti espressi con un errore di classificazione.

Le differenze principali che si notano nel confronto tra i due file riguardano sostanzialmente tre tipologie di casi:

- tag non annotati nel `train_sys.conllu`, si tratta di tag che non sono stati trovati nel file di train e che, dunque, il classificatore non riconosce;
- distinzione tra B e I, sul file `test_sys.conllu` sono presenti entità individuate come B, mentre in realtà si tratta di token interni ad un'entità e dunque I. Questo è dovuto al fatto che il classificatore prende il primo tag del dizionario e non ha una regola secondo cui capire quando si tratta di token iniziali o interni ad un'entità;
- tag ambigui, ossia tag differenti associati allo stesso termine nei due file di test. Questi riguardano principalmente entità che identificano unità di misura e monete, in quanto nel Medioevo non c'era una distinzione netta tra questi due concetti: spesso per indicare il costo di una merce ci si basava sul peso in monete e viceversa. Un esempio è il seguente:

```
18 uncia uncia B-N3.5
19 ponderat pondero O
20 denarios denarius B-N3.5
```

In questo caso *denarius* viene esplicitamente indicato come peso, in alcuni casi è a tutti gli effetti una moneta, in altri risulta di difficile interpretazione. Allo stesso modo possono risultare ambigue le specificazioni geografiche, in quanto termini come *pisanus* se riferito a una moneta indica l'ente geo-politico emissario, mentre se specifica un'unità di misura è indicazione geografica.

Per gli ultimi due casi è necessario dunque individuare regole euristiche di disambiguazione, con cui determinare procedimenti che consentano la corretta classificazione dei token in base all'analisi del contesto.

## 7.2 Regole euristiche di disambiguazione

Per individuare dunque la corretta classificazione di tag ambigui bisogna valutare il contesto in cui si trova il termine. Questo può essere attuato analizzando lemma e POS dei token precedenti a quello considerato per la classificazione.

Nel caso della distinzione tra B e I sui tag, è possibile individuare come regola il fatto che i token interni all'entità, e identificati dunque con I, sono generalmente aggettivi che presentano lo stesso tag del termine precedente collegato, come viene mostrato nell'esempio:

```
14   libras      libra      NOUN  B-N3.5
15   22   num.arab.  NUM    O
16   pisanorum  pisanus   ADJ    I-N3.5, B-Z2
```

L'entità *libras pisanorum* è un'unica entità, caratteristica che viene resa con l'indicazione del tag N3.5 su entrambi i token con la specificazione B per il nome *libras*, in quanto termine di inizio dell'entità, e I per l'aggettivo *pisanorum*, in quanto termine interno alla stessa entità.

Dunque, scorrendo il file `test_sys.conllu`, l'algoritmo quando trova un aggettivo ne guarda il tag e controlla se il token direttamente precedente o, a ritroso,

quelli ancora precedenti hanno lo stesso tag. L'aggettivo dunque risulta l'elemento interno dell'entità e dunque va cambiato da B a I.

Nel caso di tag ambigui, si potrebbe individuare un pattern ricorrente, guardando ai tag dei termini precedenti. Se il token precedente ha un tag diverso da "O" ne riporti lo stesso tag sul token corrente, altrimenti guarda al token ancora precedente. Tuttavia, non sempre i tag dei token precedenti hanno correlazione semantica con quello corrente.

Sarebbe dunque più adeguato creare una lista di lemmi per cui se questi si trovano nelle posizioni precedenti al token considerato, questo assumerà uno specifico tag. Ad esempio nel file `test_sys.conllu` i lemmi *soldus*, *denarius*, *libra*, sono associati tutti al tag N3.5 relativo al peso, se questi hanno, tra i token precedenti o successivi, lemmi come *socius*, *consocius*, *societas*, *lucrum*, *valeo*, allora il tag N3.5 deve essere sostituito con I1, in quanto relativi all'ambito monetario.

Questo viene mostrato nell'esempio:

```
14 brachia brachium NOUN B-N3.7
15 20 20 NUM O
16 panni pannus NOUN B-I2.2, B-B5
17 valeant valeo VERB O
18 libras libra NOUN B-N3.5
```

Il termine *libra* si trova vicino al lemma *valeo*, che significa "valere" ad indicazione del valore monetario, dunque il tag va sostituito con B-I1.

## Conclusioni e sviluppi futuri

Il processo che ha portato al riconoscimento delle entità nel *Liber Abaci* ha previsto varie fasi di tipo sperimentale, dovute all'ambito particolare di analisi. Si tratta, infatti, del dominio commerciale e mercantile in latino medievale presente nel testo e alla classificazione di tali entità.

Per quanto riguarda le categorie, il lavoro si è concentrato sull'identificazione di entità espresse da un tagset generico, che sono state riadattate per la rappresentazione dello specifico dominio considerato, utilizzando spesso tag multipli e annidati. La scelta di una resa più dettagliata e precisa delle entità attraverso tag complessi ha portato a complicazioni che possono essere individuate nella fase di valutazione del modello automatico. In generale, nel caso standard, si ha un classificatore multi-classe che attribuisce ciascuna entità a una sola tra più categorie, dunque per ogni parola si ha un solo tag. Nel nostro caso, invece, abbiamo tag multipli (multi-label) e situazioni che presentano entità annidate, dunque la scelta risulta più complessa.

Nel progetto il criterio da usare potrebbe essere che se i tag non coincidono, viene valutato un riconoscimento parziale per cui almeno uno dei tag è uguale, oppure, per una valutazione più stringente, considerare i doppi tag non come due tag separati, ma come un formato unico composto da un massimo di due tag che, per quell'entità, si devono ripetere sempre in quel modo.

Nelle fasi di lavoro, dunque, sperimentale è stato anche l'approccio per il riconoscimento automatico delle entità, questo perché non è stato possibile usare modelli già esistenti, in quanto rivolti ad altri generi e formati di *Named Entity Recognition*, in particolare al riconoscimento di luoghi, persone e date espressi con tag unico.

Per questo il nostro approccio ha previsto l'utilizzo di una baseline, che permette uno sviluppo semplice per il riconoscimento delle entità, di cui valutarne i limiti e su cui poter evidenziare e stabilire delle regole euristiche. Da queste regole,

poi, si possono valutare e identificare strategie e procedimenti inventivi per sviluppare una procedura adeguata al tipo di entità considerate.

La baseline con approccio *dictionary-based* è sicuramente il passo iniziale del lavoro di riconoscimento delle entità mercantili-commerciali nel *Liber Abaci*, in quanto circoscritto ad un metodo semplice e poco preciso, che va a basarsi su un numero esiguo di dati raccolti.

Come eventuale sviluppo del lavoro svolto è possibile andare ad accrescere il lessico commerciale e mercantile con cui fare una classificazione più accurata. Questo si può ottenere andando a valutare risorse linguistiche per il latino da cui cercare di ricavare altri termini di dominio, quali ulteriori nomi di monete, unità di misura, prodotti commerciali, per arricchire il dizionario utile alla classificazione, con l'obiettivo di aumentare il richiamo nella valutazione.

Inoltre, una seconda fase di sviluppo di un modello supervisionato può essere quello di passare da una baseline *dictionary-based* a un classificatore standard già esistente. Potrebbe essere valutato un sistema di NER basato su *Machine Learning*, che fornisca funzionalità progettate per il riconoscimento di entità e molte opzioni per la definizione di classificatori. L'idea è quella di prendere un classificatore basato sul sistema di NER che deve essere opportunamente riallenato per adattarlo alla nostra tipologia di dati.

Può essere avviato uno studio per capire in che formato il sistema vuole i file di allenamento e quali *features* vengono ricavate dai file. Queste informazioni sono contenute in un file di configurazione in cui, in un certo formato, si specificano tutte le *features*.

Nel caso del nostro progetto, le entità sono di un dominio particolare e in un formato a doppio tag, quindi anche le *features* morfologiche che vengono utilizzate in Stanza non sono quelle ideali per questo task.

Bisogna dunque lavorare sulla configurazione ridefinendo le *features* e rifare l'allenamento del modello per i nostri scopi. La configurazione deve prendere in considerazione le regole euristiche, presentate nell'ultimo paragrafo della relazione, di modo da sviluppare *features* specifiche per il nostro task, che mirino allo sviluppo di

un modello *sequence labeling* che guarda la sequenza dei token e valuta il contesto semantico di applicazione.

Il processo prevede poi una fase di allenamento e la costruzione di un modello che deve essere provato sul test.

Le attività svolte e i possibili sviluppi del progetto aprono le porte ad un lavoro innovativo nel campo dell'analisi semantica per il latino medievale, che facilitano un approccio globale a un capolavoro di così grande rilevanza storico-culturale come il *Liber Abaci*, che viene per la prima volta interrogato con metodi automatici.

# Bibliografia

- Ambrosetti N., 2008. *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo dell'Europa medievale*. Milano, Edizioni Universitarie di Lettere, Economia, Diritto.
- Archer D., 2014. *Exploring verbal aggression in English historical texts using USAS*, in Taavitsainen I., Jucker A. H., Tuominen J., *Diachronic Corpus Pragmatics*, pp. 277-302.
- Boncompagni B., 1852. *Della vita e delle opere di Leonardo Pisano matematico del secolo decimoterzo*. Roma, Tipografia delle Belle Arti.
- Boutell M. R., Luo J., Shen X., Brown C. M., 2004. *Learning multi-label scene classification*, in *Pattern Recognition*, vol. 37, no. 9, pp. 1757-1771.
- Cuomo S., 2004. *Fibonacci's Liber Abaci. A translation into Modern English of Leonardo Pisano's book of calculation by Laurence E. Sigler*, in *Aestimatio: Critical Reviews in the History of Science*, vol. 1.
- Erdmann A., Brown C., Hoseph B., Janse M., Ajaka P., Elsner M., de Marneffe M. C., 2016. *Challenges and Solutions for Latin Named Entity Recognition*. The Ohio State University Columbus, OH, USA.
- Franci R., 2002. *Il Liber Abaci di Leonardo Pisano 1202-2002*, in *Bollettino dell'Unione Matematica Italiana*, Serie 8, Vol. 5-A, pp. 293-328. Unione Matematica Italiana.
- Gagliardi G., 2018. *Inter-Annotator Agreement in linguistica: una rassegna critica*, in *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*, pp. 206-212.
- Gildea D., 2001. *Corpus variation and parser performance*, in *Proceeding of Empirical Methods in Natural Language Processing*. Pittsburgh, PA.
- Giusti E., 2016. *Matematica e commercio nel Liber Abaci*, in Giusti E. (a cura di), *Un ponte sul Mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*. Firenze, Edizioni Polistampa.

- Giusti E., 2017. *The twelfth Chapter of Fibonacci's Liber Abaci in its 1202 Version*. Bollettino di Storia delle Scienze Matematiche XXXVII.
- Giusti E., D'Alessandro P., 2020. *Leonardi Bigolli Pisani, vulgo Fibonacci, Liber Abbaci*, Biblioteca di Nuncius, Studi e testi LXXIX, Firenze, Olschki.
- Godbole S., Sarawagi S., 2004. *Discriminative Methods for Multi-labeled Classification*, in (eds) *Advances in Knowledge Discovery and Data Mining*. PAKDD 2004. Lecture Notes in Computer Science, vol 3056. Springer, Berlin, Heidelberg.
- Horstmann J., 2020. *Undogmatic Literary Annotation with CATMA*, in *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*. Berlin, Boston: De Gruyter, pp. 157-176.
- Hughes B. B., 1996. *Arabic Algebra. Victim of Religious and Intellectual Animus*, in M. Folkerts (a cura di), *Mathematische Probleme im Mittelalter: der lateinische und arabische Sprachbereich*. Wiesbaden, 1996.
- Hunger Parshall K., 1988. *The art of Algebra from Al-Khwarizmi to Viète: a study in the Natural Selection of Ideas*, in *History of Science*, 26, pp. 129-164.
- Jurafsky D., Martin J. H., 2018. *Speech and Language Processing. An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Pearson International Edition.
- Lenci A., Montemagni S., Pirrelli V., 2016. *Testo e computer. Elementi di linguistica computazionale*. Carrocci Editore.
- Maccagni C., 1988. *Leonardo Fibonacci e il rinnovamento delle matematiche*, in *L'Italia e i paesi mediterranei: vie di comunicazione e scambi commerciali e culturali al tempo delle repubbliche marinare*, Atti di Convegno Internazionale di studi (Pisa 6-7 giugno 1987). Pisa, Nistri-Lischi e Pacini Editore.
- Maracchia S., 2005. *Storia dell'algebra*. Napoli, Liguori editore.
- Martinori E., 1915. *La moneta: vocabolario generale*. Roma, Istituto Italiano di numismatica.

- Pepe L., 2016. *La Riscoperta di Leonardo Pisano*, in Giusti E. (a cura di), *Un ponte sul Mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*. Firenze, Edizioni Polistampa.
- Pustejovsky J., Stubbs A., 2012. *Natural Language Annotation for Machine Learning*. O'Reilly Media.
- Ricci L., Grotto F., Fantoli M., Sprugnoli R., Passarotti M., Salvatori E., Simi M., 2022. *From Close to Distant Reading. Towards the Computational Analysis of "Liber Abaci"*, in *AIUCD 2022 - Book of Abstracts*. Lecce, AIUCD.
- Roero C. S., 2016. *Algebra e Aritmetica nel Medioevo islamico*, in Giusti E. (a cura di), *Un ponte sul Mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*. Firenze, Edizioni Polistampa.
- Rozza N., 2005. *La tradizione manoscritta della Pratica Geometrie di Leonardo Pisano, detto Fibonacci, e la sua lettera di dedica la magister Dominicus*, in *Spolia: Journal of medieval studies*, Filologia e letteratura latina medievale e umanistica.
- Salvatori E., 1994. *Il sistema antroponimico a Pisa tra XI e XIII secolo*, in "Melanges de l'École Française de Rome", 106/2, pp. 487-507.
- Salvatori E., 1995. *Il sistema antroponimico a Pisa nel XIII secolo: la città e il territorio*, in "Melanges de l'École Française de Rome", 107/2, pp. 427-466.
- Schapire R. R., Singer Y., 2000. *Boostexter: a boosting-based system for text categorization*, in *Machine Learning*, vol. 39, no. 2/3, pp. 135-168.
- Sigler L. E., 2002. *Fibonacci's Liber Abaci. A translation into Modern English of Leonardo Pisano's book of calculation*. New York, Springer.
- Sprugnoli R., Tonelli S., Moretti G., Menini S., 2017. *A little bit of bella pianura: Detecting Code-Mixing in Historical English Travel Writing*, in *Proceedings of the Fourth Italian Conference on Computational Linguistics CliC-it 2017*, pp. 304-309.
- Straka M., Strakova J., 2017. *Tokenizer, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe*, in *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pp. 88-99. Vancouver, Canada. Association of Computer Linguistics.

- Tangheroni M., 2016. *Pisa e il Mediterraneo all'epoca di Fibonacci*, in Giusti E. (a cura di), *Un ponte sul Mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*. Firenze, Edizioni Polistampa.
- Tsoumakas G., Katakis I., 2007. *Multi-Label Classification: An Overview*, in *International Journal of Data Warehousing and Mining (IJDWM)*. IGI Global, vol.3, pp.1-13.
- Ulivì E., 2016. *Scuole e maestri d'abaco in Italia tra Medioevo e Rinascimento*, in Giusti E. (a cura di), *Un ponte sul Mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della matematica in Occidente*. Firenze, Edizioni Polistampa.
- Warrens M. J., 2010. *Inequalities between multi-rater kappas*, in *Advances in Data Analysis and Classification*, 4 (4), pp. 271–286.

# Sitografia

UCREL, <<http://ucrel.lancs.ac.uk/>> (sito visitato il 10 settembre 2021).

CATMA, <<https://app.catma.de/catma/>> (sito visitato il 30 agosto 2021).

Libreria Scikit-learn, <<https://scikit-learn.org/stable/index.html>> (sito visitato il 11 ottobre 2021).

LabCD. Fibonacci 1202-2021, Dataset annotato cap. 8,

<<http://dialogo.di.unipi.it/LiberAbaci/>> (sito visitato il 19 ottobre 2021).

UDPipe, <<https://ufal.mff.cuni.cz/udpipe>> (sito visitato il 30 novembre 2021).

Stanford NER, <<https://nlp.stanford.edu/software/CRF-NER.html>> (sito visitato il 20 dicembre 2021).

# Appendice

## UCREL Semantic Tagset.

|  |   |   |
|--|---|---|
| <p><b>A GENERAL &amp; ABSTRACT TERMS</b></p> <p>A1 General</p> <p>A1.1.1 General actions, making etc.</p> <p>A1.1.2 Damaging and destroying</p> <p>A1.2 Suitability</p> <p>A1.3 Caution</p> <p>A1.4 Chance, luck</p> <p>A1.5 Use</p> <p>A1.5.1 Using</p> <p>A1.5.2 Usefulness</p> <p>A1.6 Physical/mental</p> <p>A1.7 Constraint</p> <p>A1.8 Inclusion/Exclusion</p> <p>A1.9 Avoiding</p> <p>A2 Affect</p> <p>A2.1 Affect: Modify, change</p> <p>A2.2 Affect: Cause/Connected</p> <p>A3 Being</p> <p>A4 Classification</p> <p>A4.1 Generally kinds, groups, examples</p> <p>A4.2 Particular/general; detail</p> <p>A5 Evaluation</p> <p>A5.1 Evaluation: Good/bad</p> <p>A5.2 Evaluation: True/false</p> <p>A5.3 Evaluation: Accuracy</p> <p>A5.4 Evaluation: Authenticity</p> <p>A6 Comparing</p> <p>A6.1 Comparing: Similar/different</p> <p>A6.2 Comparing: Usual/unusual</p> <p>A6.3 Comparing: Variety</p> <p>A7 Definite (+ modals)</p> <p>A8 Seem</p> <p>A9 Getting and giving; possession</p> <p>A10 Open/closed; Hiding/Hidden; Finding; Showing</p> <p>A11 Importance</p> <p>A11.1 Importance: Important</p> <p>A11.2 Importance: Noticeability</p> <p>A12 Degree</p> <p>A12.1 Degree: Non-specific</p> <p>A12.2 Degree: Maximizers</p> <p>A12.3 Degree: Boosters</p> <p>A12.4 Degree: Approximators</p> <p>A12.5 Degree: Compromisers</p> <p>A12.6 Degree: Diminishers</p> <p>A12.7 Degree: Minimizers</p> <p>A14 Excluzivizers/particularizers</p> <p>A15 Safety/Danger</p> <p><b>B THE BODY &amp; THE INDIVIDUAL</b></p> <p>B1 Anatomy and physiology</p> <p>B2 Health and disease</p> <p>B3 Medicines and medical treatment</p> <p>B4 Cleaning and personal care</p> <p>B5 Clothes and personal belongings</p> <p><b>C ARTS &amp; CRAFTS</b></p> <p>C1 Arts and crafts</p> <p><b>E EMOTIONAL ACTIONS, STATES &amp; PROCESSES</b></p> <p>E1 General</p> <p>E2 Liking</p> <p>E3 Calm/Violent/Angry</p> <p>E4 Happy/sad</p> <p>E4.1 Happy/sad: Happy</p> <p>E4.2 Happy/sad: Contentment</p> <p>E5 Fear/bravery/shock</p> <p>E6 Worry, concern, confident</p> <p><b>F FOOD &amp; FARMING</b></p> <p>F1 Food</p> <p>F2 Drinks</p> <p>F3 Cigarettes and drugs</p> <p>F4 Farming &amp; Horticulture</p> <p><b>G GOVT. &amp; THE PUBLIC DOMAIN</b></p> <p>G1 Government, Politics &amp; elections</p> <p>G1.1 Government etc.</p> <p>G1.2 Politics</p> <p>G2 Crime, law and order</p> <p>G2.1 Crime, law and order: Law &amp; order</p> <p>G2.2 General ethics</p> <p>G3 Warfare, defence and the army; Weapons</p> <p><b>H ARCHITECTURE, BUILDINGS, HOUSES &amp; THE HOME</b></p> <p>H1 Architecture, kinds of houses &amp; buildings</p> <p>H2 Parts of buildings</p> <p>H3 Areas around or near houses</p> <p>H4 Residence</p> <p>H5 Furniture and household fittings</p> | <p><b>I MONEY &amp; COMMERCE</b></p> <p>I1 Money generally</p> <p>I1.1 Money: Affluence</p> <p>I1.2 Money: Debts</p> <p>I1.3 Money: Price</p> <p>I2 Business</p> <p>I2.1 Business: Generally</p> <p>I2.2 Business: Selling</p> <p>I3 Work and employment</p> <p>I3.1 Work and employment: Generally</p> <p>I3.2 Work and employment: Professionalism</p> <p>I4 Industry</p> <p><b>K ENTERTAINMENT, SPORTS &amp; GAMES</b></p> <p>K1 Entertainment generally</p> <p>K2 Music and related activities</p> <p>K3 Recorded sound etc.</p> <p>K4 Drama, the theatre &amp; show business</p> <p>K5 Sports and games generally</p> <p>K5.1 Sports</p> <p>K5.2 Games</p> <p>K6 Children's games and toys</p> <p><b>L LIFE &amp; LIVING THINGS</b></p> <p>L1 Life and living things</p> <p>L2 Living creatures generally</p> <p>L3 Plants</p> <p><b>M MOVEMENT, LOCATION, TRAVEL &amp; TRANSPORT</b></p> <p>M1 Moving, coming and going</p> <p>M2 Putting, taking, pulling, pushing, transporting &amp;c.</p> <p>M3 Movement/transportation: land</p> <p>M4 Movement/transportation: water</p> <p>M5 Movement/transportation: air</p> <p>M6 Location and direction</p> <p>M7 Places</p> <p>M8 Remaining/stationary</p> <p><b>N NUMBERS &amp; MEASUREMENT</b></p> <p>N1 Numbers</p> <p>N2 Mathematics</p> <p>N3 Measurement</p> <p>N3.1 Measurement: General</p> <p>N3.2 Measurement: Size</p> <p>N3.3 Measurement: Distance</p> <p>N3.4 Measurement: Volume</p> <p>N3.5 Measurement: Weight</p> <p>N3.6 Measurement: Area</p> <p>N3.7 Measurement: Length &amp; height</p> <p>N3.8 Measurement: Speed</p> <p>N4 Linear order</p> <p>N5 Quantities</p> <p>N5.1 Entirely; maximum</p> <p>N5.2 Exceeding; waste</p> <p>N6 Frequency etc.</p> <p><b>O SUBSTANCES, MATERIALS, OBJECTS &amp; EQUIPMENT</b></p> <p>O1 Substances and materials generally</p> <p>O1.1 Substances and materials generally: Solid</p> <p>O1.2 Substances and materials generally: Liquid</p> <p>O1.3 Substances and materials generally: Gas</p> <p>O2 Objects generally</p> <p>O3 Electricity and electrical equipment</p> <p>O4 Physical attributes</p> <p>O4.1 General appearance and physical properties</p> <p>O4.2 Judgement of appearance (pretty etc.)</p> <p>O4.3 Colour and colour patterns</p> <p>O4.4 Shape</p> <p>O4.5 Texture</p> <p>O4.6 Temperature</p> <p><b>P EDUCATION</b></p> <p>P1 Education in general</p> <p><b>Q LINGUISTIC ACTIONS, STATES &amp; PROCESSES</b></p> <p>Q1 Communication</p> <p>Q1.1 Communication in general</p> <p>Q1.2 Paper documents and writing</p> <p>Q1.3 Telecommunications</p> <p>Q2 Speech acts</p> <p>Q2.1 Speech etc: Communicative</p> <p>Q2.2 Speech acts</p> <p>Q3 Language, speech and grammar</p> <p>Q4 The Media</p> <p>Q4.1 The Media: Books</p> <p>Q4.2 The Media: Newspapers etc.</p> <p>Q4.3 The Media: TV, Radio &amp; Cinema</p> <p><b>S SOCIAL ACTIONS, STATES &amp; PROCESSES</b></p> <p>S1 Social actions, states &amp; processes</p> <p>S1.1 Social actions, states &amp; processes</p> | <p>S1.1.1 General</p> <p>S1.1.2 Reciprocity</p> <p>S1.1.3 Participation</p> <p>S1.1.4 Deserve etc.</p> <p>S1.2 Personality traits</p> <p>S1.2.1 Approachability and Friendliness</p> <p>S1.2.2 Avarice</p> <p>S1.2.3 Egoism</p> <p>S1.2.4 Politeness</p> <p>S1.2.5 Toughness; strong/weak</p> <p>S1.2.6 Sensible</p> <p>S2 People</p> <p>S2.1 People: Female</p> <p>S2.2 People: Male</p> <p>S3 Relationship</p> <p>S3.1 Relationship: General</p> <p>S3.2 Relationship: Intimate/sexual</p> <p>S4 Kin</p> <p>S5 Groups and affiliation</p> <p>S6 Obligation and necessity</p> <p>S7 Power relationship</p> <p>S7.1 Power, organizing</p> <p>S7.2 Respect</p> <p>S7.3 Competition</p> <p>S7.4 Permission</p> <p>S8 Helping/hindering</p> <p>S9 Religion and the supernatural</p> <p><b>T TIME</b></p> <p>T1 Time</p> <p>T1.1 Time: General</p> <p>T1.1.1 Time: General: Past</p> <p>T1.1.2 Time: General: Present; simultaneous</p> <p>T1.1.3 Time: General: Future</p> <p>T1.2 Time: Momentary</p> <p>T1.3 Time: Period</p> <p>T2 Time: Beginning and ending</p> <p>T3 Time: Old, new and young; age</p> <p>T4 Time: Early/late</p> <p><b>W THE WORLD &amp; OUR ENVIRONMENT</b></p> <p>W1 The universe</p> <p>W2 Light</p> <p>W3 Geographical terms</p> <p>W4 Weather</p> <p>W5 Green issues</p> <p><b>X PSYCHOLOGICAL ACTIONS, STATES &amp; PROCESSES</b></p> <p>X1 General</p> <p>X2 Mental actions and processes</p> <p>X2.1 Thought, belief</p> <p>X2.2 Knowledge</p> <p>X2.3 Learn</p> <p>X2.4 Investigate, examine, test, search</p> <p>X2.5 Understand</p> <p>X2.6 Expect</p> <p>X3 Sensory</p> <p>X3.1 Sensory: Taste</p> <p>X3.2 Sensory: Sound</p> <p>X3.3 Sensory: Touch</p> <p>X3.4 Sensory: Sight</p> <p>X3.5 Sensory: Smell</p> <p>X4 Mental object</p> <p>X4.1 Mental object: Conceptual object</p> <p>X4.2 Mental object: Means, method</p> <p>X5 Attention</p> <p>X5.1 Attention</p> <p>X5.2 Interest/boredom/excited/energetic</p> <p>X6 Deciding</p> <p>X7 Wanting; planning; choosing</p> <p>X8 Trying</p> <p>X9 Ability</p> <p>X9.1 Ability: Ability, intelligence</p> <p>X9.2 Ability: Success and failure</p> <p><b>Y SCIENCE &amp; TECHNOLOGY</b></p> <p>Y1 Science and technology in general</p> <p>Y2 Information technology and computing</p> <p><b>Z NAMES &amp; GRAMMATICAL WORDS</b></p> <p>Z0 Unmatched proper noun</p> <p>Z1 Personal names</p> <p>Z2 Geographical names</p> <p>Z3 Other proper names</p> <p>Z4 Discourse Bin</p> <p>Z5 Grammatical bin</p> <p>Z6 Negative</p> <p>Z7 If</p> <p>Z8 Pronouns etc.</p> <p>Z9 Trash can</p> <p>Z99 Unmatched</p> |
|--|---|---|

## Dizionario lemma-tags

```
{'pretium': ['B-I2.2'], 'merx': ['B-I2.2, B-O2'],  
'negotiatio': ['B-I2.2'], 'uenditio': ['B-I2.2'], 'numerus':  
['B-N3.1'], 'pondus': ['B-N3.5'], 'mensura': ['B-N3.1'],  
'centum': ['B-N5'], 'corium': ['B-I2.2, B-C1'], 'becuna': ['B-  
I2.2, B-C1'], 'cantare': ['I-N3.5', 'B-N3.5'], 'centenarium':  
['B-N3.5'], 'libra': ['B-I1', 'B-N3.5'], 'uncia': ['B-I1', 'B-  
N3.5'], 'metrum': ['B-N3.4'], 'oleum': ['B-I2.2, B-O1.2'],  
'sextarium': ['B-N3.4'], 'frumentum': ['B-I2.2, B-F4'],  
'canna': ['B-N3.7'], 'pannus': ['B-I2.2, B-B5'], 'quantitas':  
['B-N3.1'], 'denarius': ['B-I1', 'B-N3.5'], 'byzantius': ['B-  
I1'], 'tarenus': ['B-I1'], 'moneta': ['B-I1'], 'merces': ['B-  
I2.2, B-O2'], 'rotulus': ['B-N3.5'], 'soldus': ['B-I1', 'B-  
N3.5'], 'toloneum': ['B-I1.1'], 'cambium': ['B-I1.1'],  
'marca': ['B-N3.5'], 'argentum': ['B-I2.2, B-O1.1'], 'aurum':  
['B-I2.2, B-O1.1'], 'balla': ['B-N3.5'], 'torscellus': ['B-  
N3.7'], 'pisanus': ['I-N3.7, B-Z2', 'B-I1, B-G1.1', 'B-N3.7,  
B-Z2', 'I-N3.5, B-Z2', 'I-I1, B-G1.1', 'B-N3.5, B-Z2'],  
'carruba': ['B-N3.5'], 'granum': ['B-I1', 'B-N3.5'],  
'quattuor': ['B-N5'], 'mercatio': ['B-I2.2'], 'decena': ['B-  
N5'], 'piper': ['B-I2.2, B-F4'], 'pisae': ['B-M7', 'B-N3.5, B-  
Z2', 'I-N3.5, B-Z2'], 'massamatinus': ['B-I1'], 'bracchium':  
['B-N3.7'], 'miliarium': ['B-N3.4'], 'caseus': ['B-I2.2, B-  
F4'], 'carica': ['B-N3.5'], 'prouincia': ['B-M7', 'B-N3.7, B-  
Z2', 'I-N3.5, B-Z2'], 'ruptus': ['B-N5'], 'sicilia': ['I-N3.7,  
B-Z2', 'B-M7', 'B-N3.7, B-Z2'], 'messana': ['B-N3.5, B-Z2',  
'I-N3.5, B-Z2'], 'panormus': ['I-N3.5, B-Z2'], 'garbum': ['I-  
N3.7, B-Z2', 'B-M7', 'I-I1, B-G1.1'], 'miliarensis': ['B-I1'],  
'pannellus': ['B-I2.2, B-B5'], 'zaffaranum': ['B-I2.2, B-F4'],  
'nux': ['B-I2.2, B-F4'], 'muscatus': ['I-I2.2, B-F4'],  
'saracenus': ['I-I1, B-G1.1'], 'hyperperus': ['B-I1'],  
'karatus': ['B-N3.5'], 'linum': ['B-I2.2, B-B5'], 'syria':  
['I-N3.7, B-Z2', 'B-M7', 'B-N3.7, B-Z2'], 'alexandria': ['B-  
M7'], 'constantinopolitanus': ['B-N3.4, B-Z2', 'B-N3.7, B-  
Z2'], 'constantinopolis': ['B-M7', 'B-N3.7, B-Z2'],  
'cantarium': ['B-N3.5'], 'geroui': ['B-N3.5, B-Z2', 'I-N3.5,  
B-Z2'], 'forfori': ['B-N3.5, B-Z2', 'B-Z2, B-N3.5', 'I-N3.5,  
B-Z2'], 'gerouinus': ['B-N3.5, B-Z2', 'I-N3.5, B-Z2'],  
'forforinus': ['B-N3.5, B-Z2', 'I-N3.5, B-Z2'], 'imperialis':  
['B-I1', 'I-I1, G1.1', 'I-I1', 'B-I1, B-G1.1', 'I-I1, B-G1.1'],  
'OB-I1'], 'pisaninus': ['B-I1, B-G1.1', 'I-I1, B-G1.1'],  
'ianuinus': ['B-I1, B-G1.1', 'I-I1, B-G1.1'], 'merguliensis':  
['B-I1, B-G1.1', 'I-N3.5, B-Z2', 'I-I1, B-G1.1'], 'regalis':  
['B-I1', 'I-I1, B-G1.1'], 'barcellonaensis': ['B-I1, B-G1.1',  
'I-I1, B-G1.1'], 'podiensis': ['B-I1, B-G1.1', 'I-I1, B-  
G1.1'], 'bononinus': ['B-I1, B-G1.1', 'I-I1, B-G1.1'],  
'uenetianus': ['B-N3.5, B-Z2', 'B-I1, B-G1.1', 'I-I1, B-  
G1.1'], 'cantera': ['I-N3.5'], 'stagnum': ['B-I2.2, B-O1.1'],  
'bolsonalia': ['B-I1'], 'palmus': ['B-N3.7'], 'ianua': ['I-  
N3.7, B-Z2', 'B-N3.7, B-Z2'], 'obolus': ['B-I1'], 'longitudo':  
['B-N3.7'], 'fustaneum': ['B-I2.2, B-B5'], 'petia': ['B-I2.2,  
B-B5'], 'societas': ['B-I2.1, B-S5'], 'lucrum': ['B-I1.1'],  
'entheca': ['B-I2.1'], 'socius': ['B-I2.1, B-S5'], 'lucror':  
['B-I1.1'], 'capitale': ['B-I2.1'], 'proficuum': ['B-I1.1'],  
'baractum': ['B-I2.2'], 'baractus': ['B-I2.2'], 'emptio': ['B-  
I2.2'], 'baracte': ['B-I2.2'], 'equus': ['B-L2'], 'ordeum':  
['B-F4'], 'consocius': ['B-I2.1, B-S5']}
```