



UNIVERSITÀ DI PISA

Corso di Laurea Magistrale in Informatica Umanistica

TESI DI LAUREA MAGISTRALE

**BibliBERT: un *Neural Language Model*  
per l'esegesi biblica**

**Studente** *Luca Baù (matr. 626070)*

**Relatore** *Alessandro Lenci*

Anno Accademico 2020-21

# Indice

<b>Indice delle abbreviazioni</b>	<b>vii</b>
<b>Abbreviazioni dei libri biblici</b>	<b>viii</b>
<b>Introduzione</b>	<b>1</b>
<b>1 Modelli computazionali del significato</b>	<b>3</b>
1.1 Similarità, contesto e <i>corpora</i> . . . . .	4
<b>2 Introduzione generale alla Bibbia</b>	<b>6</b>
2.1 L'origine e le caratteristiche del testo . . . . .	8
2.2 La traduzione della Bibbia . . . . .	13
2.3 La numerazione dei versetti . . . . .	14
2.4 Gli strumenti per lo studio della Bibbia . . . . .	15
<b>3 Le fasi del progetto</b>	<b>17</b>
3.1 La scelta del modello di partenza . . . . .	17
3.2 La raccolta dei testi . . . . .	18
3.3 La tokenizzazione . . . . .	19
3.4 L'addestramento del modello . . . . .	21
3.5 L'estrazione delle rappresentazioni . . . . .	23
3.6 Le applicazioni . . . . .	24
3.6.1 I significati delle parole e dei lemmi . . . . .	24
3.6.2 La visualizzazione delle rappresentazioni . . . . .	26
3.6.3 Gli strumenti per il confronto . . . . .	29
3.6.4 Il confronto tra i versetti . . . . .	30

<b>4</b>	<b>I risultati – parole di interesse teologico</b>	<b>32</b>
4.1	Legge . . . . .	32
4.2	Amore e carità . . . . .	40
4.3	Spirito . . . . .	49
4.4	Segno . . . . .	56
4.5	Davide . . . . .	63
4.6	Signore . . . . .	69
4.7	Dodici . . . . .	76
<b>5</b>	<b>I risultati – parole suggerite dal modello</b>	<b>82</b>
5.1	Parola . . . . .	86
5.2	Rivolta . . . . .	92
5.3	Figlio . . . . .	99
<b>6</b>	<b>I risultati – ricerca dei versetti</b>	<b>104</b>
6.1	I versetti del Nuovo Testamento con maggiore disaccordo . . . . .	105
6.2	I versetti dell’Antico Testamento con maggiore disaccordo . . . . .	110
6.3	Il confronto con le antifone . . . . .	114
	<b>Conclusioni</b>	<b>120</b>
	<b>A Spirito: versetti nel <i>cluster Q</i></b>	<b>122</b>
	<b>Bibliografia</b>	<b>130</b>

# Elenco delle figure

3.1	Andamento dell'errore del modello durante l'addestramento. . . . .	23
4.1	Legge — confronto tra le principali sezioni della Bibbia . . . . .	34
4.2	Legge — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	34
4.3	Legge — visualizzazione in uno spazio bidimensionale . . . . .	36
4.4	Amore — distribuzione delle occorrenze nei <i>cluster</i> . . . . .	42
4.5	Amore — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	42
4.6	Amore — visualizzazione in uno spazio bidimensionale . . . . .	47
4.7	Spirito — confronto tra le principali sezioni della Bibbia . . . . .	50
4.8	Spirito — confronto tra le traduzioni . . . . .	50
4.9	Spirito — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	53
4.10	Spirito — visualizzazione in uno spazio bidimensionale . . . . .	53
4.11	Segno — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	57
4.12	Segno — confronto tra le principali sezioni della Bibbia . . . . .	60
4.13	Segno — confronto tra le traduzioni . . . . .	60
4.14	Segno — visualizzazione in uno spazio bidimensionale . . . . .	62
4.15	Davide — distribuzione delle occorrenze nei <i>cluster</i> . . . . .	65
4.16	Davide — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	65
4.17	Davide — visualizzazione in uno spazio bidimensionale . . . . .	68
4.18	Geova — distribuzione delle occorrenze nei <i>cluster</i> . . . . .	71
4.19	Confronto tra i <i>cluster</i> di <i>Geova</i> e quelli del termine parallelo in NR06 . . . . .	71
4.20	Geova — distanza del trasporto ottimo tra le sezioni della Bibbia	73
4.21	Geova — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	73
4.22	Geova — visualizzazione in uno spazio bidimensionale . . . . .	75

4.23	12 — distribuzione delle occorrenze nei <i>cluster</i> . . . . .	78
4.24	12 — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	78
4.25	12 — visualizzazione in uno spazio bidimensionale . . . . .	81
5.1	Parola — confronto tra le principali sezioni della Bibbia . . . . .	87
5.2	Parola — confronto tra le traduzioni . . . . .	87
5.3	Parola — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	89
5.4	Parola — visualizzazione in uno spazio bidimensionale . . . . .	89
5.5	Rivolta — confronto tra le principali sezioni della Bibbia . . . . .	93
5.6	Rivolta — confronto tra le traduzioni . . . . .	93
5.7	Rivolta — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	95
5.8	Rivolta — visualizzazione in uno spazio bidimensionale . . . . .	95
5.9	Figlio — distanza del coseno tra i centri dei <i>cluster</i> . . . . .	100
5.10	Figlio — distribuzione delle occorrenze nei <i>cluster</i> . . . . .	100
5.11	Figlio — visualizzazione in uno spazio bidimensionale . . . . .	103
6.1	Distanza media tra le traduzioni dei versetti del NT . . . . .	109
6.2	Distanza media tra le traduzioni dei versetti dell'AT . . . . .	109
6.3	Distanza media tra l'antifona ed i versetti corrispondenti (CEI74) . . . . .	115
6.4	Distanza media tra l'antifona ed il versetto più simile . . . . .	115

# Elenco delle tabelle

2.1	Panoramica dei libri biblici . . . . .	9
3.1	<i>Labeling</i> dei <i>cluster</i> tramite le parole co-occorrenti . . . . .	28
3.2	<i>Labeling</i> dei <i>cluster</i> tramite le parole del vocabolario . . . . .	28
4.1	Legge — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{J}$ . . . . .	38
4.2	Amore — confronto tra le traduzioni dei versetti in cui occorrono le parole in $\mathcal{E}$ . . . . .	43
4.3	Spirito — risultati del riempimento del <i>token</i> mascherato . . . . .	55
4.4	Segno — esempi dai <i>cluster</i> di dimensioni ridotte ( $\mathcal{L} - \mathcal{Q}$ ) . . . . .	57
4.5	Segno — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{K}$ . . . . .	58
4.6	Davide — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{P}$ (Ezechiele ed Ebrei, traduzione CEI08) . . . . .	66
4.7	Dodici — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{E}$ (traduzione CEI08) . . . . .	79
5.1	Parole con distanza del trasporto ottimo maggiore tra le traduzioni CEI . . . . .	83
5.2	Parole con distanza del trasporto ottimo maggiore tra due traduzioni . . . . .	83
5.3	Parole con distanza del trasporto ottimo maggiore tra NT e deuterocanonici . . . . .	85
5.4	Parole con distanza del trasporto ottimo maggiore tra NT e AT ebraico . . . . .	85
5.5	Parola — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{C}$ . . . . .	90
5.6	Parola — parole della Bibbia più simili per <i>cluster</i> . . . . .	92

5.7	Rivolta — versetti in cui occorrono le parole del <i>cluster</i> $\mathcal{F}$ (significati diversi) . . . . .	96
5.8	Rivolta — parole della Bibbia più simili per <i>cluster</i> . . . . .	98
5.9	Figlio — parole della Bibbia più simili per <i>cluster</i> . . . . .	101
6.1	Versetti del NT con maggiore disaccordo . . . . .	106
6.2	Versetti dell'AT con maggiore disaccordo . . . . .	111
6.3	Antifone più distanti dal versetto di partenza . . . . .	116
6.4	Antifone più distanti dal versetto riconosciuto dal sistema . . . . .	117
A.1	Spirito — confronto tra le traduzioni dei versetti in cui occorrono le parole in $\mathcal{Q}$ . . . . .	122

# Indice delle abbreviazioni

AT	Antico Testamento
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CEI	Conferenza episcopale italiana
CEI08	<i>La Sacra Bibbia</i> , traduzione della CEI, edizione 2008
CEI74	<i>La Sacra Bibbia</i> , traduzione della CEI, edizione 1974
GPU	<i>Graphics Processing Unit</i>
LXX	Traduzione greca dell'Antico Testamento, detta dei Settanta
NR06	<i>La Bibbia. Nuova Riveduta</i> , edizione 2006
NT	Nuovo Testamento
OCR	<i>Optical Character Recognition</i>
par.	parallelo, per i passi biblici
s.	secolo
Tf-idf	<i>Term frequency – inverse document frequency</i>
TILC	<i>La Sacra Bibbia, Parola del Signore</i> , traduzione interconfessionale in lingua corrente
TNW17	<i>Traduzione del nuovo mondo delle Sacre Scritture</i> , edizione 2017
UMAP	<i>Uniform Manifold Approximation and Projection for Dimension Reduction</i>



# Abbreviazioni dei libri biblici

1Cor	Prima lettera ai Corinzi di Paolo
1Cr	Primo libro delle Cronache
1Gv	Prima lettera di Giovanni
1Mac	Primo libro dei Maccabei
1Pt	Prima lettera di Pietro
1Re	Primo libro dei Re
1Sam	Primo libro di Samuele
1Tm	Prima lettera a Timoteo di Paolo
1Ts	Prima lettera ai Tessalonicesi di Paolo
2Cor	Seconda lettera ai Corinzi di Paolo
2Cr	Secondo libro delle Cronache
2Gv	Seconda lettera di Giovanni
2Mac	Secondo libro dei Maccabei
2Pt	Seconda lettera di Pietro
2Re	Secondo libro dei Re
2Sam	Secondo libro di Samuele
2Tm	Seconda lettera a Timoteo di Paolo
2Ts	Seconda lettera ai Tessalonicesi di Paolo
3Gv	Terza lettera di Giovanni

Ab	Libro del profeta Abacuc	Gl	Libro del profeta Gioele
Abd	Libro del profeta Abdia	Gn	Libro del profeta Giona
Ag	Libro del profeta Aggeo	Gs	Libro di Giosuè
Am	Libro del profeta Amos	Gv	Vangelo di Giovanni
Ap	Apocalisse di Giovanni	Is	Libro del profeta Isaia
At	Atti degli apostoli	Lam	Libro delle Lamentazioni
Bar	Libro del profeta Baruc	Lc	Vangelo di Luca
Col	Lettera ai Colossesi di Paolo	Lv	Libro del Levitico
Ct	Cantico dei Cantici	Mc	Vangelo di Marco
Dn	Libro del profeta Daniele	Mi	Libro del profeta Michea
Dt	Libro del Deuteronomio	Ml	Libro del profeta Malachia
Eb	Lettera agli Ebrei	Mt	Vangelo di Matteo
Ef	Lettera agli Efesini di Paolo	Na	Libro del profeta Naum
Es	Libro dell'Esodo	Ne	Libro di Neemia
Esd	Libro di Esdra	Nm	Libro dei Numeri
Est	Libro di Ester	Os	Libro del profeta Osea
Ez	Libro del profeta Ezechiele	Pr	Libro dei Proverbi
Fil	Lettera ai Filippesi di Paolo	Qo	Libro di Qoelet
Fm	Lettera a Filemone di Paolo	Rm	Lettera ai Romani
Gal	Lettera ai Galati di Paolo	Rt	Libro di Rut
Gb	Libro di Giobbe	Sal	Libro dei Salmi
Gc	Lettera di Giacomo	Sap	Libro della Sapienza
Gd	Lettera di Giuda	Sir	Libro del Siracide
Gdc	Libro dei Giudici	Sof	Libro del profeta Sofonia
Gdt	Libro di Giuditta	Tb	Libro di Tobia
Gen	Libro della Genesi	Tt	Lettera a Tito di Paolo
Ger	Libro del profeta Geremia	Zc	Libro del profeta Zaccaria

# Introduzione

«L'accostamento del mondo dell'informatica alla Bibbia può suscitare ancora oggi qualche imbarazzo»<sup>1</sup>. Esordisce così l'appendice dedicata agli strumenti informatici per lo studio della Bibbia di un manuale di studi biblici pubblicato non molti anni fa.

La Bibbia è il libro più stampato, venduto e probabilmente studiato. L'approccio allo studio della Bibbia non è limitato a quello teologico ma è stato affrontato anche con il ricorso ai metodi delle scienze umane. Tra di esse spiccano sicuramente filologia e critica testuale ma non sono le uniche.

Nonostante la presenza di un padre gesuita tra i pionieri della linguistica computazionale, ad oggi non sembrano esserci grandi applicazioni dei modelli computazionali più recenti al mondo biblico, specialmente in italiano.

Se natura e Bibbia sono due libri scritti dallo stesso autore, per citare la metafora resa celebre da un illustre pisano, le rappresentazioni quantitative della Bibbia create da un modello distribuzionale del linguaggio possono essere visti come un'opera di traduzione di un libro nella lingua dell'altro, che Galileo diceva fosse fatta di numeri e simboli matematici.

Uscendo dalla metafora, gli strumenti della linguistica computazionale possono essere certamente di aiuto per lo studio e la comprensione della Bibbia come per qualsiasi altro testo. Questo lavoro si prefigge di applicarne alcuni per provare ad affrontare la Scrittura in un modo diverso cercando di interpretare alcune delle risposte che un simile strumento può dare.

Tra i modelli computazionali del linguaggio è stato scelto un modello di BERT (*Bidirectional Encoder Representations from Transformers*) già addestrato sull'italiano su cui è stato eseguito il *fine-tuning* sul testo biblico.

---

<sup>1</sup>Barbaglia 2014, p. 727.

Con le rappresentazioni vettoriali estratte dal modello si studierà il cambiamento semantico di alcune parole tra le diverse parti del testo biblico e tra alcune traduzioni che si distinguono per l'approccio seguito, per il contesto in cui sono nate o per gli anni in cui sono state realizzate.

Delle stesse parole, usando algoritmi di *clustering* e vari criteri di *labeling*, si cercheranno le diverse sfumature di significato e quelle caratteristiche del testo biblico.

Infine, partendo dai vettori delle parole, si otterranno le rappresentazioni vettoriali dei versetti in cui è diviso il testo biblico che verranno usate per il confronto tra le traduzioni e per creare uno strumento di ricerca della provenienza di citazioni bibliche non letterali.

# Capitolo 1

## Modelli computazionali del significato

Tra i modelli di rappresentazione del significato, quelli distribuzionali (*Distributional Semantic Models*) hanno ricevuto un grande impulso dalla crescente quantità di testi disponibili in rete e dalle crescenti «capacità dei *computer* di gestire ed elaborare quantità di dati sempre più estesi»<sup>1</sup>.

Questi modelli rappresentano quantitativamente il significato attraverso sequenze ordinate di numeri (vettori). Il principio su cui si basa questo approccio è l'ipotesi distribuzionale che lega il significato di un'espressione ai contesti in cui viene usata<sup>2</sup>.

Prendendo i grandi *corpora* come campioni di uso del linguaggio si possono costruire rappresentazioni del significato che consentano di investigare la relazione tra significati e contesti, e di affrontare la dinamicità e plasticità del significato<sup>3</sup>.

I primi *Distributional Semantic Models* usavano rappresentazioni vettoriali in cui ad ogni dimensione corrispondeva un determinato contesto ed il valore era semplicemente la frequenza di occorrenza dell'oggetto lessicale da rappresentare insieme al contesto.

La frequenza di co-occorrenza non sempre è indicativa della rilevanza di un contesto ed è possibile e ragionevole sostituirla con misure di salienza che tengano conto ad esempio della frequenza del contesto.

---

<sup>1</sup>Lenci, Montemagni e Pirrelli 2005, p. 15.

<sup>2</sup>Harris 1954.

<sup>3</sup>Lenci 2018.

Queste rappresentazioni vettoriali esplicite sono caratterizzate da vettori di grandi dimensioni e da dati sparsi: non tutte le parole occorrono con tutti i contesti e viceversa.

Raccogliendo le rappresentazioni vettoriali dei diversi oggetti lessicali in matrici di co-occorrenza è anche possibile ricavare matematicamente delle rappresentazioni più dense di dimensioni minori. Si perde in questo modo la possibilità di interpretare chiaramente un contesto ma si ottengono nuove dimensioni che potrebbero rappresentare aspetti latenti del significato<sup>4</sup>.

Lo sviluppo di reti neurali con architetture più articolate, sempre insieme all'incremento della potenza di calcolo e dei dati a disposizione, ha aperto la strada ad un loro uso crescente. Addestrando una rete neurale a predire il contesto dato un oggetto lessicale o viceversa, questa diventa di fatto capace di generare vettori distribuzionali per rappresentare i significati<sup>5</sup>.

Infine, i modelli più recenti non generano più vettori di parole-tipo da usare per ogni contesto ma associano una rappresentazione unica ad ogni uso della data parola includendo in sé anche gli aspetti derivati dal determinato contesto in cui occorre<sup>6</sup>.

## 1.1 Similarità, contesto e *corpora*

L'approccio distribuzionale al significato si basa sul concetto di similarità tra i significati, assunto come chiaro ed auto-evidente. In realtà la similarità tra i significati è un concetto piuttosto vago che non permette da solo di trarre conclusioni su quale sia l'esatta relazione tra essi. Le informazioni che si possono trarre da questi modelli sono comunque un prezioso punto di partenza.

Per misurare la similarità tra due significati è possibile usare una qualsiasi misura di distanza, quella che viene generalmente usata in letteratura è la distanza (o similarità) del coseno che dipende dall'angolo che si forma tra i due vettori. Questa misura, calcolabile qualsiasi sia il numero delle dimensioni del vettore,

---

<sup>4</sup>Landauer e Dumais 1997.

<sup>5</sup>Mikolov et al. 2013.

<sup>6</sup>Devlin et al. 2019.

ha il vantaggio di non tenere conto della lunghezza del vettore che può essere influenzata fortemente dalla frequenza della parola più che dal suo significato.

Anche il termine *contesto* rischia di essere vago ma è invece determinante e si collega all'interpretazione della similarità.

La prima idea di contesto a venire in mente è probabilmente costituita dalle parole che si trovano nel testo ad una certa distanza dalla parola da rappresentare. Però, è possibile scegliere di considerare solo le parole precedenti o seguenti o di andare a distanze diverse nei due versi, con conseguenze differenti sugli aspetti di significato che vengono colti a seconda della lingua del testo. Si può anche decidere se fermarsi o meno al confine di frase ma, soprattutto, di non considerare la distanza lineare ma quella di dipendenza sintattica che richiede però l'annotazione sintattica del testo. Nei contesti sintattici è possibile decidere di considerare solo alcuni tipi di relazione o di inserire l'informazione del tipo di relazione sintattica.

La scelta del contesto condiziona inevitabilmente il risultato e gli aspetti del significato che vengono evidenziati.

Una delle caratteristiche chiave di BERT<sup>7</sup> sta proprio nel considerare il contesto (non annotato) per intero a tutti i livelli e di essere addestrato anche per andare esplicitamente oltre i confini di frase con l'addestramento a predire la frase successiva.

Infine, un ultimo aspetto veramente determinante per la qualità delle rappresentazioni ottenute sono i testi di partenza. Anche all'interno della stessa lingua una diversa selezione dei testi darà luogo a rappresentazioni molto diverse. La grande mole di dati oggi a disposizione e la quantità di dati necessari per un buon addestramento di una rete neurale ha fatto propendere sempre di più verso l'attenzione alla quantità che alla rappresentatività ed al bilanciamento dei testi.

Quando si deve lavorare con varietà specifiche della lingua ma non si possiedono grandi quantità di dati specifici, si ottengono spesso risultati migliori addestrando la rete neurale anche su grandi quantità di dati generici piuttosto che solamente sui pochi specifici.

---

<sup>7</sup>Devlin et al. 2019.

# Capitolo 2

## Introduzione generale alla Bibbia<sup>1</sup>

La Bibbia non è un libro scritto ad una sola mano in un solo tempo. Il nome stesso viene dal greco βιβλία plurale di *libro* e, passando per il latino, diventa poi un singolare ed infine il nome proprio della raccolta dei libri sacri ebraici e cristiani.

Una prima grande suddivisione del testo è quella tra Antico (AT) e Nuovo Testamento (NT), cioè tra le Scritture che vengono dalla tradizione ebraica e quelle redatte in ambiente cristiano. Questa prima grande divisione non è sufficiente né a descrivere la complessità dei libri né a classificare chiaramente tutti i libri.

Esistono infatti diversi canoni (liste dei libri inclusi), specialmente per l'Antico Testamento. Nella grande traduzione greca detta dei Settanta (LXX), nata ad Alessandria d'Egitto nel terzo secolo avanti Cristo è presente una lista più ampia di libri, la quasi totalità dei quali è stata accolta nel canone cristiano per l'Antico Testamento. Quando, dopo la distruzione del tempio di Gerusalemme (71 dopo Cristo), anche gli ebrei stilano una loro lista, scelgono una lista più ristretta che include solamente i libri scritti originalmente in ebraico, anche per distinguersi dalla scelta fatta dai primi cristiani che avevano adottato la LXX come riferimento, come faranno, per gli stessi motivi, le comunità della Riforma.

Con il canone cambiano anche l'ordine dei libri e i criteri per raccogliarli in grandi sezioni. Il canone ebraico suddivide i libri in Torah, Profeti e Scritti,

---

<sup>1</sup>Il capitolo fa spesso riferimento a Boscolo 2011; Fabris et al. 2014.



quello greco in libri storici, sapienziali e profetici. Sarebbe possibile individuare altri criteri per suddividere i libri in base al periodo o al luogo di composizione, allo stile letterario, alla tradizione in cui si inseriscono, solo per citarne alcuni. Il canone del Nuovo Testamento è abbastanza condiviso (non senza discussioni su alcuni libri) e generalmente i libri sono divisi in Vangeli, Atti degli apostoli, Epistolario ed Apocalisse.

In questo lavoro sarà adoperato l'ordine dei libri secondo l'uso più familiare in Italia che è quello del canone cristiano cattolico delle Scritture. Come grandi suddivisioni della Bibbia, verranno usate quelle più comuni negli studi biblici e teologici in Italia, dividendo sia Antico che Nuovo Testamento in quattro parti.

Per i libri dell'Antico Testamento si usa in genere la suddivisione del canone greco:

- *Pentateuco*: i primi cinque libri dell'Antico Testamento, corrispondenti alla Torah (Genesi, Esodo, Levitico, Numeri, Deuteronomio);
- *Libri storici*: i successivi sedici libri nel canone greco, quelli presenti nel canone ebraico sono divisi tra profeti anteriori e scritti (Giosuè, Giudici, Rut, 1-2 Samuele, 1-2 Re, 1-2 Cronache, Esdra, Neemia, Tobia, Giuditta, Ester, 1-2 Maccabei);
- *Libri sapienziali*: i successivi sette del canone greco, i primi cinque dei quali inseriti tra gli scritti nel canone ebraico (Giobbe, Salmi, Proverbi, Qohelet, Cantico, Sapienza, Siracide);
- *Libri profetici*: gli ultimi diciotto del canone greco, quasi tutti inseriti tra i profeti posteriori del canone ebraico (Isaia, Geremia, Lamentazioni, Baruc, Ezechiele, Daniele e i 12 profeti minori).

I libri del Nuovo Testamento si suddividono prevalentemente secondo l'attribuzione all'autore:

- *Vangeli sinottici ed Atti*: i Vangeli redatti da Marco, Matteo e Luca che hanno una struttura e contenuto simile e procedono quasi paralleli, insieme all'altro libro attribuito a Luca, cioè gli Atti degli Apostoli;<sup>2</sup>

---

<sup>2</sup>Per brevità, nell'analisi, ci si riferirà a Sinottici per identificare tutti e quattro i libri, quando si vorrà escludere il libro degli Atti degli apostoli si esplicherà il termine Vangeli.

- *Corpus paolino*: le lettere tradizionalmente attribuite a Paolo, includendo quelle di attribuzione discussa fino alla lettera agli Ebrei per la quale c'è un consenso degli studiosi nell'affermare che non sia attribuibile a Paolo;
- *Lettere cattoliche*: le lettere scritte dagli apostoli Giacomo, Pietro e Giuda; dell'elenco farebbero parte anche le lettere scritte da Giovanni, inserite però nella prossima ed ultima sezione;
- *Corpus giovanneo*: gli scritti tradizionalmente attribuiti a Giovanni, cioè il quarto Vangelo, le tre lettere e l'Apocalisse.

## 2.1 L'origine e le caratteristiche del testo

I più antichi brani della Bibbia hanno trovato forma scritta con il sorgere della monarchia nel decimo secolo avanti Cristo, gli ultimi del Nuovo Testamento verso la fine del primo dopo Cristo. Prima della scrittura molti brani hanno avuto una lunga fase orale e in molti casi il redattore ha messo insieme materiale proveniente da diverse tradizioni.

Datate un libro dell'Antico Testamento può voler dire individuare un intervallo di tempo entro il quale è avvenuta la redazione finale, ma all'interno dello stesso libro è possibile riconoscere sezioni che provengono da tempi molto diversi. Anche i Vangeli, scritti meno di un secolo dopo gli eventi che raccontano, sono preceduti da una fase orale, e la ricerca filologica ha individuato contributi di fonti diverse.

Le differenze nell'origine del testo non sono soltanto temporali ma anche geografiche: gli scritti più antichi hanno origine alla corte di Gerusalemme, ma quando il regno si divide, anche i discepoli dei profeti che operano nel regno del nord raccolgono e scrivono le loro parole; durante l'esilio babilonese, i circoli sacerdotali accrescono il bagaglio di testi, mentre il secondo libro dei Maccabei è scritto probabilmente nella comunità giudaica alessandrina. La collocazione geografica dell'origine dei libri del Nuovo Testamento è molto più varia e include la Siria, l'Asia minore, la Grecia fino forse a Roma.

Un'altra utile classificazione dei testi biblici è il genere letterario, la Bibbia, come una piccola biblioteca, ne contiene un'ampia varietà: opere storiche, racconti didattici, testi poetici o apocalittici, ammonimenti profetici, parabole, detti

sono una lista non esaustiva. Spesso questa varietà è presente anche all'interno dello stesso libro. Cercare però di classificare i testi secondo il genere letterario in modo rigido rischia di essere controproducente, va evitato il riduzionismo, cioè la tentazione di voler ridurre un testo ad una forma tipica, ma nemmeno si devono moltiplicare le suddivisioni, né si deve dimenticare che chi scrive adotta con libertà le convenzioni letterarie senza restarne totalmente schiavo. Solo con queste precauzioni la classificazione dei generi letterari può risultare utile per un approccio all'opera.<sup>3</sup>

I libri biblici, quindi, nonostante in questo lavoro siano l'unità omogenea più piccola considerata, non si possono considerare pienamente tali e la scelta di questo lavoro è un'approssimazione accettabile finché si riconosce come tale.

Può essere utile per il lavoro da compiere fare una breve panoramica dei libri biblici. Non può essere questo il luogo di una trattazione esaustiva del dibattito sulla datazione e la collocazione dell'origine dei libri biblici, ma può essere utile avere un'idea di ciò che compone le grandi sezioni di testo individuate, per notare altre possibili vicinanze o differenze di massima tra i vari libri.

Tabella 2.1: Panoramica dei libri biblici

..... <i>PENTATEUCO</i> .....						
<b>Libro</b>	<b>Altro</b>		<b>Composizione</b>		<b>Dimensioni</b>	
	<b>gruppo</b>	<b>lingua</b>	<b>luogo</b>	<b>data</b>	<b>cap.</b>	<b>ver.</b>
Genesi	Torah	ebraico	Giudea (Babilonia)	V s. <sup>4</sup> (XIII-VI) <sup>5</sup>	50	1533
Esodo	Torah	ebraico	Giudea (Babilonia)	V s. (XIII-VI)	40	1213
Levitico	Torah	ebraico	Giudea (Babilonia)	V s. (XIII-VI)	27	859
Numeri	Torah	ebraico	Giudea (Babilonia)	V s. (XIII-VI)	36	1288
Deuteronomio	Torah	ebraico	Giudea (Babilonia)	V s. (XIII-VI)	34	959

<sup>3</sup>Cfr. Alonso Schökel 1989, p. 21.

<sup>4</sup>Si intende avanti Cristo per l'Antico Testamento ed ovviamente dopo Cristo per il Nuovo.

<sup>5</sup>Le informazioni di composizione indicate tra parentesi si riferiscono alle fonti precedenti.

Libro	Altro		Composizione		Dimensioni	
	gruppo	lingua	luogo	data	cap.	ver.
..... <i>LIBRI STORICI</i> .....						
Giosuè	Profeti (anteriori) <sup>6</sup>	ebraico	Giudea (Babilonia)	V s. (VII-VI)	24	658
Giudici	Profeti (anteriori)	ebraico	Giudea (Babilonia)	V s. (VII-VI)	21	618
Rut	Scritti (meghillot <sup>7</sup> )	ebraico	Giudea	V s.	4	85
1 Samuele	Profeti (anteriori)	ebraico	Giudea (Babilonia)	V s. (VII-VI)	31	810
2 Samuele	Profeti (anteriori)	ebraico	Giudea (Babilonia)	V s. (VII-VI)	24	695
1 Re	Profeti (anteriori)	ebraico	Giudea (Babilonia)	V s. (VII-VI)	22	816
2 Re	Profeti (anteriori)	ebraico	Giudea (Babilonia)	V s. (VII-VI)	25	719
1 Cronache	Scritti	ebraico	Giudea	IV-III s.	29	942
2 Cronache	Scritti	ebraico	Giudea	IV-III s.	36	822
Esdra	Scritti	ebraico	Giudea	IV-III s.	10	280
Neemia	Scritti	ebraico	Giudea	IV-III s.	13	406
Tobia	Deutero <sup>8</sup>	greco (aramaico)	Giudea	III s.	14	249
Giuditta	Deutero	greco (ebraico)	Giudea	II s.	16	340
Ester	Deutero	greco	Egitto	II s.	10	260
1 Maccabei	Deutero	greco (ebraico)	Giudea	II s.	16	924
2 Maccabei	Deutero	greco	Egitto	II s.	15	555

<sup>6</sup>La classificazione alternativa indicata tra parentesi indica una possibile ulteriore suddivisione

<sup>7</sup>Significa letteralmente cinque rotoli: sono cinque libri che vengono letti, essenzialmente a scopo liturgico, durante certe festività.

<sup>8</sup>Per deutero canonici, esclusi dal canone ristretto o ebraico ma presenti in quello della LXX e della Bibbia cattolica.

Libro	Altro		Composizione		Dimensioni	
	gruppo	lingua	luogo	data	cap.	ver.
..... <i>LIBRI SAPIENZIALI</i> .....						
Giobbe	Scritti	ebraico	Giudea	VI s. (XI-X)	42	1070
Salmi	Scritti	ebraico	Giudea	III s. (XI-X)	150	2461
Proverbi	Scritti	ebraico	Giudea	V s. (XI-X)	31	915
Qohelet	Scritti (meghillot)	ebraico	Giudea	III s.	12	222
Cantico dei Cantici	Scritti (meghillot)	ebraico	Giudea	V-III s. (X)	8	117
Sapienza	Deutero	greco	Egitto	I s.	19	435
Siracide	Deutero	greco (ebraico)	Giudea	II s.	51	1380
..... <i>LIBRI PROFETICI</i> .....						
Isaia	Profeti (posteriori)	ebraico	Giudea (VIII-VI)	V s.	66	1292
Geremia	Profeti (posteriori)	ebraico	Giudea (VII-VI)	V s.	52	1364
Lamentazioni	Scritti (meghillot)	ebraico	Giudea	VI s.	5	154
Baruc (cap. 6)	Deutero	greco greco	Antiochia Giudea	II s. II s.	5 1	141 72
Ezechiele	Profeti (posteriori)	ebraico	Giudea (Babilonia)	V s. (VI)	48	1273
Daniele (aggiunte)	Scritti Deutero	ebraico greco	Giudea Egitto	II s. II s.	12	357
Osea	Profeti (minori)	ebraico	Israele	VIII s.	14	197
Gioele	Profeti (minori)	ebraico	Giuda	VII-VI s.	3	73

<b>Libro</b>	<b>Altro</b>	<b>Composizione</b>			<b>Dimensioni</b>	
	<b>gruppo</b>	<b>lingua</b>	<b>luogo</b>	<b>data</b>	<b>cap.</b>	<b>ver.</b>
Amos	Profeti (minori)	ebraico	Giudea	VIII s.	9	146
Abdia	Profeti (minori)	ebraico	Giudea	VI s.	1	21
Giona	Profeti (minori)	ebraico	Giudea	VI s.	4	48
Michea	Profeti (minori)	ebraico	Giudea	VIII-VII s.	7	105
Naum	Profeti (minori)	ebraico	Giudea	VII s.	3	47
Abacuc	Profeti (minori)	ebraico	Giudea	VII-VI s.	3	56
Sofonia	Profeti (minori)	ebraico	Giudea	VII s.	3	53
Aggeo	Profeti (minori)	ebraico	Giudea	VI s.	2	38
Zaccaria	Profeti (minori)	ebraico	Giudea	VI-IV s.	14	211
Malachia	Profeti (minori)	ebraico	Giudea	V s.	4	55
..... <i>VANGELI E ATTP</i> <sup>9</sup> .....						
Matteo	Sinottici	greco	Antiochia	80-90	28	1071
Marco	Sinottici	greco	Roma	65-70	16	678
Luca	Sinottici	greco	Grecia	80-90	24	1151
Giovanni	Giovanni	greco	Efeso	100	21	879
Atti	Sinottici	greco	Grecia	80-90	28	1007
..... <i>LETTERE DI PAOLO</i> .....						
Romani		greco	Corinto	57-58	16	433
1 Corinzi		greco	Efeso	55-56	16	437
2 Corinzi		greco	Macedonia	56-57	13	257

<sup>9</sup>Continuiamo a seguire l'ordine dei libri come si presenta su una Bibbia cattolica, nel lavoro poi sarà seguita la classificazione alternativa, dove indicata

Libro	Altro		Composizione		Dimensioni	
	gruppo	lingua	luogo	data	cap.	ver.
Galati		greco	Efeso	56-57	6	149
Efesini		greco	Roma	62(-100) <sup>10</sup>	6	155
Filippesi		greco	Efeso	56-57	4	104
Colossesi		greco	Roma	62(-100)	4	95
1 Tessalonicesi		greco	Corinto	51	5	89
2 Tessalonicesi		greco	Corinto	51(-100)	3	47
1 Timoteo		greco	Roma	63(-100)	6	113
2 Timoteo		greco	Roma	62(-100)	4	83
Tito		greco	Roma	63(-100)	3	46
Filemone		greco	Cesarea	62	1	25
Ebrei		greco		70	13	303
..... <i>LETTERE CATTOLICHE</i> .....						
Giacomo		greco		62(-90)	5	108
1 Pietro		greco		64(-80)	5	105
2 Pietro		greco		125	3	61
1 Giovanni	Giovanni	greco	Efeso	100	5	105
2 Giovanni	Giovanni	greco	Efeso	100	1	13
3 Giovanni	Giovanni	greco	Efeso	100	1	14
Giuda		greco		80-90	1	25
..... <i>APOCALISSE</i> .....						
Apocalisse	Giovanni	greco	Patmos	100	22	404

## 2.2 La traduzione della Bibbia

Mentre in ambito ebraico la lettura della Bibbia è stata a lungo tempo principalmente nella lingua originale, almeno in ambito liturgico, in ambiente cristiano l'accesso al testo biblico è stato, molto presto e per quasi tutti i suoi lettori, mediato da un traduttore.

<sup>10</sup>Nel Nuovo Testamento la seconda data indicata tra parentesi è quella individuata nel caso in cui la redazione attuale sia da attribuire non all'autore indicato ma alla comunità dei suoi discepoli e fatta dopo la sua morte.

Questo, insieme alla scarsità di risorse nelle lingue bibliche (ebraico antico e greco ellenistico) e alle prospettive di uso del modello realizzato ha fatto propendere per l'uso della lingua italiana, che mette a disposizione sia una scelta di modelli di base che di traduzioni diverse del testo.

Le traduzioni non sono tutte uguali, prima di tutto perché la traduzione è il risultato finale di un processo di comprensione del testo, quindi traduzioni nate in contesti confessionali diversi avranno inevitabilmente alcune differenze, quali e quante sono domande a cui potrebbe rispondere, almeno in parte, un sistema come questo.

Una seconda differenza sta anche nello stile traduttivo: le traduzioni a *equivalenza formale* tendono a rendere lo stesso termine sempre allo stesso modo cercando di mantenere il più possibile immutata anche la struttura della frase. Un approccio opposto è quello a *equivalenza dinamica* in cui si cerca di rendere il senso della frase sacrificando la corrispondenza tra le parole in lingua originale e quelle della traduzione.<sup>11</sup>

Una terza differenza è tra le traduzioni primarie, cioè quelle che partono dai testi originali e quelle secondarie che sono traduzioni di testi già tradotti. Sia per fare un buon confronto che per ottenere il più possibile dai testi è conveniente quindi fare un *mix* bilanciato dei diversi approcci.

## 2.3 La numerazione dei versetti

Per lavorare con il testo è utile la divisione dei libri in capitoli e versetti a cui è abituato chi è familiare con la Bibbia. Anche questa divisione non è univoca e ci sono traduzioni più vicine nel testo che nella numerazione. Il problema è concentrato soprattutto nell'Antico Testamento ma non è limitato alle ovvie differenze legate alla scelta del canone, ci sono libri che presentano un testo più lungo se si prende come riferimento quello greco.

Questa confusione è dovuta soprattutto al fatto che la suddivisione in capitoli e versetti è molto posteriore alla scrittura dei testi e non sempre segue criteri chiari. I manoscritti più antichi presentano una *scripto continua* del testo, la

---

<sup>11</sup>Nida 1995.



cui unica suddivisione chiara era quella tra i diversi libri. I capitoli sono stati identificati a partire del XIII secolo per facilitare il lavoro dei copiatori della Volgata. I versetti nel XIV secolo con almeno due grandi riferimenti diversi, uno che si basa sull'originale ebraico dell'Antico Testamento ed uno sul testo greco.

In particolare il libro dei Salmi, tenendo ferma la divisione in 150 capitoli, confrontando tradizione greca ed ebraica, presenta uno spostamento di uno o due capitoli nella quasi totalità del testo<sup>12</sup>.

Infine, le traduzioni più attente al significato che al testo originale non sempre rispettano queste suddivisioni, quando si presenta la necessità di cambiare l'ordine dei versetti per rendere meglio una frase in lingua corrente, procedono raggruppando il testo di più versetti senza indicarne la suddivisione.<sup>13</sup>

## 2.4 Gli strumenti per lo studio della Bibbia

Per lo studio e la comprensione del testo biblico sono disponibili numerosi strumenti: concordanze, dizionari, sinossi, commentari, ecc.; spesso sono il frutto di un lavoro di raccolta manuale delle conoscenze acquisite da parte di uno o più studiosi.

Una concordanza è un'opera che elenca in ordine alfabetico le parole della Bibbia, indicando per ognuna i passi biblici dove essa ricorre. Questo strumento permette di ritrovare un passo della Bibbia che non si conosce a memoria oppure di studiare sistematicamente un argomento. Il limite di uno strumento come questo è che non sempre è facile ricollegare un argomento preciso a ciò che si sta cercando e la parola esatta può sfuggire o può dipendere dalla traduzione di riferimento. A volte magari ciò che si cerca è dovuto all'unione di brani diversi il che rende ancor più difficile che la ricerca vada a buon fine.

Come riferimento per i risultati del modello che sarà realizzato, sarà usata la *Concordanza pastorale della Bibbia*.

È una presentazione sistematica delle parole, delle idee e dei nomi principali, (...) ogni articolo, a seconda dei casi, viene scomposto in suddivisioni

---

<sup>12</sup>Cfr. Scaiola 1997, pp. 119–121.

<sup>13</sup>Cfr. "Sussidi per la lettura" 2014, p. 12.

che corrispondono ai diversi significati del termine o dell'idea riportata, o agli specifici contesti che gli danno, nella Bibbia, un valore particolare. (...) È inoltre, un indice analogico: riferisce cioè le rassomiglianze, le equivalenze e le connessioni che richiamano fra loro i contesti analizzati, soprattutto nei casi di sinonimia.<sup>14</sup>

Probabilmente riportare questa definizione di dizionario è superfluo: «collezione di parole corredate della loro corretta grafia, etimologia, sinonimi e contrari, significato o traduzione; presentate solitamente in ordine alfabetico o altro che ne faciliti la consultazione»; ma tra gli strumenti utili per la comprensione e lo studio della Bibbia ci sono anche i *dizionari teologici*, «i cui articoli pongono l'accento sulla portata dottrinale e spirituale dei termini biblici»<sup>15</sup> riflettendo la comprensione del significato specifico che tali termini hanno nel contesto biblico alla luce di quanto è stato intuito e compreso da chi si è confrontato con la Bibbia nei secoli della sua storia. Un esempio datato ma autorevole è il *Dizionario di Teologia Biblica* coordinato da Padre Xavier Leon-Dufour, che sarà adottato come riferimento per formulare ipotesi e confrontare i risultati.

Sinossi ed indici dei passi paralleli aiutano a recuperare brani simili, in particolare le sinossi presentano il testo dei Vangeli in colonne parallele, permettendo di evidenziare somiglianze e differenze.

Esistono alcuni sussidi informatici per lo studio della Bibbia, principalmente mettono insieme testi in lingua originale, apparati critici, diverse traduzioni, note, concordanze, dizionari, grammatiche, atlanti ed altri strumenti senza dover consultare numerosi volumi e senza dover cercare i riferimenti grazie a collegamenti ipertestuali. Quello che cambia, in pratica, con l'uso di questi strumenti è soprattutto il luogo di lavoro ed il supporto materiale utilizzato.

---

<sup>14</sup>Passelecq e Poswick 2012.

<sup>15</sup>Leon-Dufour 1976, p. XIII.

# Capitolo 3

## Le fasi del progetto

### 3.1 La scelta del modello di partenza

Tra i *Distributional Semantic Models* presi in esame nel capitolo 1 l'uso di un modello BERT<sup>1</sup> dà la possibilità di partire da un modello che abbia già conoscenze linguistiche senza il bisogno di riaddestrare il modello da zero.

Non sono molti i modelli di *transformer* disponibili addestrati sull'italiano, i più utilizzati sono quelli della Biblioteca Nazionale Bavarese<sup>2</sup>.

Vista la presenza di molti nomi propri nel testo biblico, i quali potrebbero essere oggetto di interesse, sembra preferibile l'uso di un modello *case-sensitive*, scartando quelli che non tengono conto della differenza tra lettere maiuscole e minuscole.

Tra le alternative rimaste la scelta è caduta sul modello BERT addestrato su una maggiore quantità di dati: un *download* di Wikipedia, alcuni testi dal *corpus* OPUS<sup>3</sup> e dalla parte italiana del *corpus* OSCAR<sup>4</sup> per un totale di più di 13 miliardi di *token*.

Scelto un modello di partenza, è possibile riaddestrarlo con dei testi scelti per ottenere un modello specializzato nel compito che dovrà svolgere o per adattare il modello generico al dominio a cui verrà applicato. Per farlo, tra i testi disponibili sono state selezionate cinque diverse traduzioni italiane dell'intero testo biblico.

---

<sup>1</sup>Devlin et al. 2019.

<sup>2</sup>Schweter 2020.

<sup>3</sup>Tiedemann 2012.

<sup>4</sup>Ortiz Suárez, Sagot e Romary 2019.

Le versioni scelte rappresentano diversi approcci alla traduzione: la Nuova Riveduta (NR06), nata in contesto protestante, vuole essere più aderente alla forma del testo originale, la traduzione interconfessionale in lingua corrente (TILC) vuole adottare una forma più vicina al linguaggio italiano comune, le traduzioni CEI si prefiggono di trovare un equilibrio tra questi due approcci, la traduzione del nuovo mondo (TNW17), di riferimento per i Testimoni di Geova, è una traduzione secondaria dall'inglese. Dove possibile, si è cercato di usare la revisione disponibile più recente dei testi (2006 per la Riveduta, 2017 per la traduzione del nuovo mondo), le traduzioni ufficiali della CEI (CEI08 e CEI74) si sono usate entrambe, quella completata nel 1974 per l'addestramento e la più recente, completata nel 2008, come *test set*.

## 3.2 La raccolta dei testi

Le versioni del testo biblico digitalizzato disponibili *on-line* sono ormai numerose, alcuni portali accolgono più traduzioni per ciascuna delle lingue che ospitano, soprattutto l'inglese ma anche nelle altre lingue europee e non. Con uno *script* in Python<sup>5</sup> è stato possibile navigare tra le pagine contenenti il testo ordinato in libri e capitoli ed estrarlo. Con l'uso delle espressioni regolari, il testo è stato ripulito dalle note e suddiviso in versetti.

Come è stato già accennato nel capitolo 2 non c'è un riferimento unico e condiviso per la divisione del testo in versetti. Sarebbe stato possibile a questo punto prendere una numerazione di riferimento come standard per semplificare il confronto tra i paralleli, ma il vantaggio ottenuto non compenserebbe la perdita di granularità per alcune traduzioni (il riferimento comune avrebbe dovuto considerare le porzioni più ampie di testo delle traduzioni più "libere") e sarebbe stato necessario comunque tenere traccia dello spostamento per poter fornire il riferimento reale al versetto per la data traduzione e non solo il riferimento standardizzato.

Una seconda standardizzazione del testo che è stata presa in considerazione e poi scartata è stata quella dei nomi propri, ogni traduttore infatti usa un proprio

---

<sup>5</sup>Cfr. Van Rossum e Drake 2009.

criterio per la traslitterazione o la traduzione dei nomi. Per quanto uniformare permetta di ottenere più occorrenze dello stesso nome senza disperderli tra le varie forme, è sensato supporre che la stessa variabilità (se non una maggiore) sia presente tra possibili utenti di una ricerca e addestrare più forme dello stesso nome diventerebbe un vantaggio piuttosto che un *handicap*.

La standardizzazione a cui invece si è deciso di procedere è stata quella degli accenti. Prima di tutto si è notato che la scelta se facilitare la pronuncia di alcuni nomi propri con gli accenti varia tra edizioni diverse della stessa traduzione, inoltre variano anche i criteri sulla posizione dell'accento per facilitare una lettura più simile al suono originale in ebraico, in greco o in latino o più familiare in italiano. In secondo luogo, si è riscontrata una simile variabilità nella scelta della distinzione degli omografi: cambia la scelta della forma base e di quella accentata. Questo si è notato non soltanto tra le traduzioni ma anche rispetto al vocabolario del tokenizzatore. Tutto considerato è stato ritenuto preferibile definire una funzione che riportasse le lettere accentate alla forma senza accento, lasciando come si presentano nel testo solo le lettere accentate alla fine di una parola.

### 3.3 La tokenizzazione

Ad ogni modello BERT è associato un tokenizzatore che traduca il testo in un *input* accettabile, associando a ciascuna sequenza di caratteri riconosciuta un identificativo numerico da dare in ingresso alla rete neurale.

In questo modo, il numero di sequenze identificabili è finito a fronte di una quantità potenzialmente infinita di combinazioni di caratteri che danno origine a parole che si potrebbero trovare in un testo. Il tokenizzatore deve avere quindi un meccanismo che permetta di accettare anche degli *input* con parole non presenti nel vocabolario preventivato.

I modelli *transformer* hanno dei tokenizzatori di tipo *subword* che nel caso di parole non conosciute le scompongono in parti. In questo modo la rappresentazione della parola è riconducibile ad una qualche funzione delle rappresentazioni delle parti riconosciute, considerando ogni singola parte come associata ad un vettore. Questa semplificazione dovrebbe funzionare particolarmente bene quan-

do le parti riconosciute corrispondono ai morfemi, potendosi anche avvantaggiare in fase di addestramento di un maggior numero di occorrenze, il che dovrebbe contribuire a creare rappresentazioni migliori.

I modelli BERT usano un algoritmo chiamato *WordPiece*<sup>6</sup> per addestrare il tokenizzatore. Una volta diviso il testo in parole in base a delle regole predefinite (pre-tokenizzazione), questo algoritmo inizia includendo ogni carattere presente nei dati di addestramento e progressivamente impara regole di associazione che massimizzino la verosimiglianza dei dati una volta aggiunta al vocabolario.

Usando il tokenizzatore del modello di partenza scelto, si nota presto che divide i segni di punteggiatura (compresi apostrofi e trattini) dalle sequenze di lettere e numeri, ciò significa ad esempio che le parole composte separate da trattino daranno origine ad almeno tre *token* e che la lettera «l» non è distinta dall'articolo «l'».

Il numero di elementi del vocabolario corrisponderà ad una delle dimensioni della matrice che traduce la sequenza di numeri fornita dal tokenizzatore nel livello di *input* del modello. Per aggiungere nuove parole nel vocabolario bisognerà anche intervenire su questo parametro del modello, che poi dovrà essere riaddestrato, altrimenti fornirebbe per quelle parole rappresentazioni basate sull'inizializzazione casuale del modello.

Nel modello adottato, per un errore nella selezione dei parametri in fase di addestramento, il vocabolario ha di fatto 1000 spazi liberi<sup>7</sup>. Questo è sembrato un numero sufficiente di possibili aggiunte al vocabolario da fare prima del riaddestramento.

Per scegliere quali parole aggiungere al vocabolario bisogna tenere conto che il tokenizzatore tenderà a preferire queste parole aggiunte a ciò che ha già presente nel vocabolario anche nel caso di sequenze più lunghe nel testo fornito. Inoltre questa divisione non sarà considerata dal tokenizzatore una scomposizione in parti di parola ma in parole diverse, rendendo così più complesso rintracciare queste scomposizioni per ricomporre le rappresentazioni. Per evitare questo inconveniente non sono state considerate tra le possibili aggiunte al vocabolario tutte le parole

---

<sup>6</sup>Descritto per la prima volta in Schuster e Nakajima 2012.

<sup>7</sup>Il vocabolario contiene 31102 elementi ma il modello è predisposto per un vocabolario di 32102. Cfr. <https://github.com/dbmdz/berts/issues/7>

che fossero parte di termini già presenti nel vocabolario o nelle cinque traduzioni bibliche scelte.

Per identificare i candidati all'aggiunta al vocabolario, è stato tokenizzato tutto il testo e sono state identificate le parole scomposte in parti e ne è stata calcolata la frequenza per traduzione. Sarebbe stato possibile un semplice confronto con il vocabolario ma in questo modo c'è stata la possibilità di tener traccia della scomposizione. Usando più testi per il *corpus* di addestramento avrebbe senso usare una misura come *Tf-idf* per scegliere le parole, ma in questo caso è possibile e preferibile trattare la Bibbia come un unico testo usando una misura più semplice come la frequenza nel testo.

Essendo la traduzione più recente della CEI quella più comunemente usata in Italia è stato dato un peso maggiore alle occorrenze di questa versione e si è riservata una piccola quota per parole presenti principalmente o quasi esclusivamente in una sola traduzione (ad esempio *Geova* è legato all'uso confessionale ed è assente nelle altre traduzioni).

Sono stati usati altri due accorgimenti particolari: considerando che i tokenizzatori di questo tipo permettono di ottenere rappresentazioni più significative quando le parti identificate corrispondono ai morfemi, alle parole segmentate in questo modo sono state preferite altre (tra le appena meno frequenti) ritenute particolarmente rilevanti o di possibile interesse per possibili analisi successive.

Ad esempio sono state inserite nel vocabolario *aspersione*, *lodino* e *Cafarnao*<sup>8</sup>, con una totale di 50 occorrenze ciascuna, mentre non sono state incluse *contribuzione* (diviso in contribu + zione), *ricevettero* (diviso in ricevette + ro), *rispondimi* (diviso in rispondi + mi), tutte con frequenza totale 60.

### 3.4 L'addestramento del modello

Il *fine-tuning* di un modello pre-addestrato per adattarlo ad un dominio si può fare riaddestrandolo su un compito di *language modeling*. Nel caso di modelli BERT il compito è quello di *masked language modeling*: il modello deve predire alcuni *token* che sono sostituiti nell'*input* da un *token* speciale *[MASK]*. Avendo

---

<sup>8</sup>Usato solo nelle traduzioni della CEI e in quella interconfessionale, è il nome di un villaggio della Galilea in cui si è svolti alcuni episodi raccontati nei vangeli.

accesso anche alla frase intera, in addestramento il modello confrontare i due *output* per correggere i pesi.

Il *training set* realizzato per questo secondo addestramento consiste nel testo biblico fornito versetto per versetto in quattro traduzioni diverse. Per il *test set* è stata utilizzata la traduzione più recente della CEI, sempre considerando ogni versetto una frase.

Per l'addestramento sono stati usati gli stessi parametri suggeriti nel *tutorial* disponibile insieme alla documentazione del pacchetto *transformers*<sup>9</sup> e generalmente seguiti in letteratura:

- 15% di *token* mascherati,
- tasso di apprendimento  $2 \cdot 10^{-5}$ ,
- ottimizzatore Adam con  $\beta = (0.9, 0.999)$  ed  $\varepsilon = 10^{-8}$ .

Per la ripetibilità dell'esperimento e per poter eventualmente riprendere il lavoro sospeso è stato fissato il *random feed* usato per il mascheramento.

Il rischio di eccessiva specializzazione e di instabilità in seguito al riaddestramento ha portato, già a partire da Devlin et al. 2019, a scegliere un numero molto ridotto di epoche di addestramento. Nel caso in cui il *corpus* per il *fine-tuning* sia particolarmente ristretto rispetto a quello usato per il pre-addestramento Mosbach, Andriushchenko e Klakow 2021 hanno però riscontrato che la causa dell'instabilità sembra più legata al numero di iterazioni che alla dimensione del *dataset* consigliando quindi di aumentare il numero di epoche lasciando un tasso di apprendimento piuttosto basso. Con questo procedimento hanno ottenuto risultati migliori con un numero di epoche molto superiori alle 3 precedentemente stabilite come ottimali, arrivando a consigliare di continuare l'addestramento fino ad ottenere *Training loss* vicini allo zero.

Come si vede in Figura 3.1, al termine delle 50 epoche di addestramento del modello realizzato non si sono presentate grandi instabilità e si è raggiunta una *Training loss* di 0.71. Considerate anche le risorse GPU a disposizione è stato ritenuto un risultato soddisfacente, quasi dimezzato rispetto al risultato che

---

<sup>9</sup>[https://github.com/huggingface/notebooks/blob/master/examples/text\\_classification.ipynb](https://github.com/huggingface/notebooks/blob/master/examples/text_classification.ipynb)



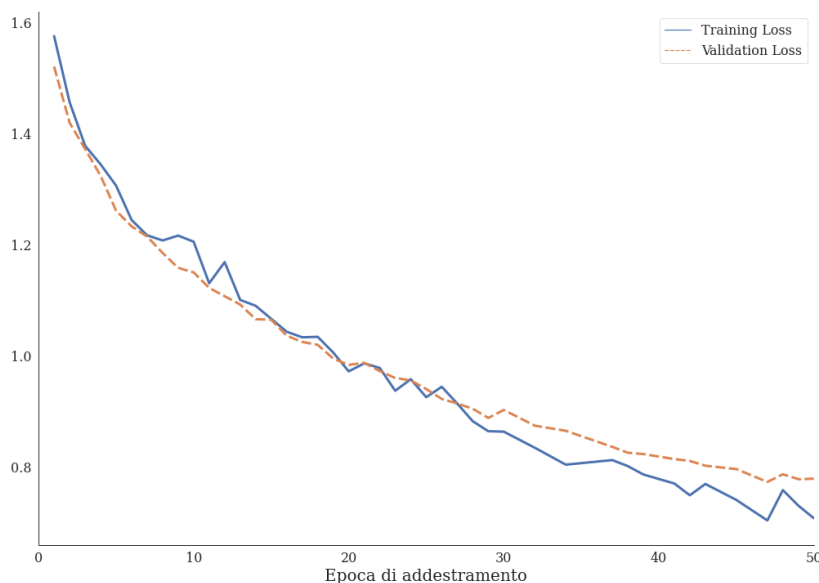


Figura 3.1: Andamento dell'errore del modello durante l'addestramento.

avremmo ottenuto fermandosi alla terza epoca (1.38). Non si è quindi proceduto con ulteriori epoche di addestramento. Dal grafico si nota inoltre che per quanto continuo entrambi a diminuire, a partire da circa metà dell'addestramento iniziano a distanziarsi gli errori sul *corpus* di addestramento rispetto a quelli sul *corpus* di verifica su cui poi si concentrerà parte del lavoro.

### 3.5 L'estrazione delle rappresentazioni

Una volta ottenuto il modello adattato al dominio della Bibbia, sono state estratti i vettori per le parole versetto per versetto delle cinque traduzioni selezionate. Inizialmente sono state salvate anche tutte le rappresentazioni degli strati nascosti per ogni *token*.

Successivamente, è stata ricavata una rappresentazione unica per le parole che erano state spezzate dal tokenizzatore. La scelta del metodo per realizzare questa nuova rappresentazione è caduta sulla media. La somma è un'operazione semplice da realizzare e altri metodi più complessi in genere non hanno dato risultati sensibilmente migliori. La decisione di dividere la somma per il numero di *token* che componevano la parola è legata principalmente all'intenzione di ottenere anche delle rappresentazioni dei versetti per evitare che queste parole abbiano un peso maggiore delle altre.

Infine, per realizzare una rappresentazione delle frasi lo spunto è arrivato da Reimers e Gurevych 2019 che usano come algoritmo di *pooling* una semplice media dei *token* che costituiscono la frase. Un'altra soluzione presa in considerazione è stata l'uso del *token*  $[CLS]$ <sup>10</sup> che rappresenta l'inizio della frase ed il cui vettore dipende dalle altre parole che la compongono.

Lavorare con tutti e 12 i livelli, oltre ad essere antieconomico (sia in termini di tempo di elaborazione che di spazio occupato), non dovrebbe portare a grandi benefici in termini di qualità della rappresentazione del significato delle parole, in quanto diversi esperimenti hanno portato a concludere che l'informazione semantica risieda nei livelli superiori<sup>11</sup>. Per rappresentare il significato, quindi, una scelta comune alla quale è sembrato opportuno uniformarsi è quella di utilizzare gli ultimi quattro livelli<sup>12</sup>, anche in questo caso mediandoli.

## 3.6 Le applicazioni

### 3.6.1 I significati delle parole e dei lemmi

BERT crea una diversa rappresentazione per ogni parola-*token*, volendo lavorare con le parole-tipo, con i lemmi o anche con gruppi di sinonimi, è necessario agire sulle rappresentazioni per ottenere qualcosa di interpretabile. In questo caso fare direttamente una media significherebbe perdere il vantaggio fornito da BERT nel trattare la polisemia.

Piuttosto, quello che serve in questo caso è un criterio per raggruppare i *token* che rappresentino usi simili e solo dopo, per ciascuno di questi significati, trovare una rappresentazione che può essere la somma (o la media) che non è altro che il centroide dei vettori.

Montariol, Martinc e Pivovarova 2021 hanno definito una procedura di *clustering* che cerca di mediare tra l'uso di risorse e l'interpretabilità del risultato.

Per ridurre l'uso di memoria ed il tempo di elaborazione dell'algoritmo di *clustering* si limita la lista dei vettori ad un numero massimo (200 nell'articolo). Questa lista si ottiene scorrendo i vettori ed aggiungendo alla lista (finché c'è

---

<sup>10</sup>Devlin et al. 2019.

<sup>11</sup>Tenney, Das e Pavlick 2019.

<sup>12</sup>Wiedemann et al. 2019; Montariol, Martinc e Pivovarova 2021.

spazio) soltanto quelli che siano più distanti di una determinata soglia (nell'articolo distanza del coseno maggiore di 0.01) da tutti i presenti nella lista, gli altri vettori vengono mediati con il più vicino già presente nella lista.

Confrontando due algoritmi di *clustering* usati in letteratura: *k-means*<sup>13</sup> e *affinity propagation*<sup>14</sup> hanno ottenuto risultati migliori con il secondo, che tende a produrre un numero minore di *cluster* quando gli usi di una parola non sono molti.

Per diminuire il numero di *cluster* quando è eccessivo e per escludere quelli che, contenendo solo poche istanze, non rappresentano un uso specifico della parola o che sono particolarmente rari, hanno proposto una successiva fase per filtrare e riunire i *cluster*. Selezionando inizialmente solo quelli con un numero di istanze sopra una certa soglia<sup>15</sup>, riunisce ricorsivamente i *cluster* la cui distanza è minore della distanza media meno due volte la deviazione standard. Si ripete la stessa operazione unendo anche i *cluster* precedentemente esclusi ed eventualmente ignorando alla fine quelli rimasti troppo piccoli.

Per applicare questo procedimento al presente lavoro sono stati necessari alcuni accorgimenti. Nella prima fase di creazione della lista da fornire all'algoritmo di *clustering* è risultato presto evidente che in molti casi il limite di 200 si esaurisse prima di arrivare all'ultima traduzione, ed in alcuni casi prima di arrivare al Nuovo Testamento, già nella prima traduzione fornita. Inizialmente le soluzioni prese in esame sono state quelle di agire sui due parametri: la distanza minima ed il numero di elementi. Intervenire sulla distanza minima avrebbe avuto effetti anche su quei casi in cui non sarebbe stato necessario intervenire ma aumentare semplicemente la dimensione della lista non avrebbe garantito di non incorrere nuovamente nel problema senza perdere i vantaggi sui tempi e sull'uso della memoria dell'utilizzo di un numero ridotto di elementi da fornire all'algoritmo di *clustering*.

Una prima soluzione intermedia provata è stata quella di aumentare la dimensione massima della lista con lo scorrere del testo, riservando di fatto alcuni *slot* alle sezioni esaminate dopo. Questa si è rivelato un buon compromesso applicata

---

<sup>13</sup>Usato da Giulianelli, Del Tredici e Fernández 2020.

<sup>14</sup>Usato da Martinc et al. 2020.

<sup>15</sup>Derivata da Schlechtweg et al. 2020.

alle diverse traduzioni, mentre per le sezioni di testo è risultato più difficile trovare un equilibrio tra intervalli abbastanza piccoli da poter garantire l'ingresso in lista ad usi specifici in quelle sezioni ed un numero sufficientemente grande di *slot* liberi perché non fossero occupati subito dalle primissime istanze della parola.

Per affrontare questo problema si è pensato di non scorrere le rappresentazioni nell'ordine in cui si presentano nel testo ma dopo averle rimescolate casualmente. Per non caricare in memoria tutte le rappresentazioni contemporaneamente, l'algoritmo ha scorso le traduzioni una per una, aumentando ogni volta la dimensione della lista, e all'interno di ogni traduzione, scorrendo le parole in ordine casuale, associando ad ogni traduzione un diverso generatore di numeri casuali per non fornire i versetti sempre nello stesso ordine.

### 3.6.2 La visualizzazione delle rappresentazioni

Per visualizzare gli *embedding* ottenuti, l'ispirazione è arrivata dal progetto *context-atlas*<sup>16</sup> che usa *UMAP*<sup>17</sup> per ridurre il numero di dimensioni dei vettori alle due rappresentabili su un piano cartesiano ottenendo buoni risultati in termini di visibilità, significatività ed uso di risorse.

Per capire quali significati identifichino i singoli *cluster* individuati, esplorare tutti i versetti in cui appare la parola che appartiene al *cluster* sarebbe anche fattibile per le parole meno frequenti ma diventa proibitivo al crescere del numero di versetti da esaminare.

Per rendere interpretabili i *cluster* ottenuti per le singole parole, sono stati provati due diversi approcci di *labeling* dei *cluster*: un confronto “interno” con le parole dei versetti che contengono il termine da esaminare, ed uno “esterno” con le rappresentazioni delle parole di un *corpus* usato come esempio di lingua *standard*.

Il primo approccio<sup>18</sup> consiste nell'estrarre le parole ed i bigrammi più discriminanti per ogni *cluster*. Considerando il *cluster* come documento e tutti i *cluster*

---

<sup>16</sup>Coenen et al. 2019.

<sup>17</sup>McInnes, Healy e Melville 2020.

<sup>18</sup>Seguito ad esempio da Montariol, Martinc e Pivovarova 2021.

come *corpus* si può usare come misura di informatività la  $Tf-idf$ <sup>19</sup>, escludendo le parole grammaticali e quelle presenti in più dell'80% dei *cluster*.

Usare un *corpus* diverso per etichettare i *cluster* ha il vantaggio di non essere limitato alle parole presenti nel testo biblico, il rischio, invece, è quello di non riuscire a distinguerli con etichette diverse. Questo, in alcuni casi, potrebbe anche essere un sintomo dell'eccessivo numero di *cluster* ottenuti. Per differenziare i *cluster*, le etichette di questo tipo inserite nelle immagini sono state scelte in modo che provenissero tutte da *token* diversi e facendo in modo di massimizzare la somma delle similarità del coseno tra centro del *cluster* ed etichetta.

Cercando un testo che possa avere un vocabolario abbastanza ampio senza avere dimensioni proibitive, la scelta è caduta su un dizionario della lingua italiana. Dei tre presi in esame e disponibili *on-line*: il *Sabatini Coletti*<sup>20</sup>, il *Vocabolario Treccani*<sup>21</sup> ed il *Dizionario libero* di *Wikipedia*<sup>22</sup>, la scelta è caduta sul terzo perché la formattazione è più facilmente riconoscibile e il progetto *Wikimedia* prevede la realizzazione periodica di archivi dei vari *database*, con il doppio vantaggio di avere un riferimento stabile per il testo scaricato<sup>23</sup> e di non dover bombardare i *server* che ospitano i dizionari scaricandoli pagina per pagina.

Dall'archivio scaricato sono state estratte le varie definizioni in italiano delle parole (escludendo quelle che indicavano soltanto lemma e categorie grammaticali delle forme flesse) e le frasi di esempio. Sono state create tre tipologie di rappresentazione delle parole del dizionario:

- partendo dalle frasi di esempio<sup>24</sup>, estraendo le rappresentazioni delle parole, il problema di questa opzione è che non a tutte le definizioni è associato un esempio e che a volte sono molto brevi, frasi incomplete o citazioni di testi lontani nel tempo;

---

<sup>19</sup>Considerando l'insieme dei versetti che contengono le parole inserite nel *cluster* come documento e l'insieme dei versetti che contengono le parole in esame come *corpus*.

<sup>20</sup>[dizionari.corriere.it/dizionario\\_italiano/](http://dizionari.corriere.it/dizionario_italiano/)

<sup>21</sup>[www.treccani.it/vocabolario/](http://www.treccani.it/vocabolario/)

<sup>22</sup>[it.wiktionary.org](http://it.wiktionary.org)

<sup>23</sup>Per il dizionario italiano sono disponibili su [dumps.wikimedia.org/itwiktionary](https://dumps.wikimedia.org/itwiktionary). La versione scaricata, al tempo la più aggiornata è stata quella del 20 ottobre 2021.

<sup>24</sup>Come fanno ad esempio Hu, Li e Liang 2019.

- partendo dalle singole definizioni usando lo stesso procedimento applicato per gli *embedding* dei versetti, questa scelta ha il difetto di dover poi confrontare frasi e parole anche se rappresentati con vettori che hanno le stesse dimensioni (le 768 di ogni livello di BERT);
- creando frasi del tipo **parola : definizione** ed estraendo solo la rappresentazione della parola, in questo modo si evita di dover confrontare frasi e parole ma si analizza una frase artificiale che non corrisponde in pieno ad un uso normale della lingua italiana.

Se si prende ad esempio la parola *Canto* con l’iniziale maiuscola come è usata nella traduzione più recente della CEI, l’algoritmo di *clustering* divide le 29 occorrenze in tre *cluster* che indicheremo con le lettere maiuscole *A*, *B* e *C*. Questi, seguendo il primo approccio, si possono descrivere con le parole o i bigrammi co-occorrenti più rilevanti. In questo modo si otterranno le etichette in Tabella 3.1.

Tabella 3.1: *Labeling* dei *cluster* tramite le parole co-occorrenti

---

<i>A</i> :	salite, Signore, casa, te, occhi, Israele, né, com, invano, monti
<i>B</i> :	Salmo, figli, coro, Core, maestro, monti, Maskil, egli, Asaf, te_siedi
<i>C</i> :	maestro, coro, Salmo, figli, Core, Asaf, Maskil, corda, strumenti, angoscia_gridato

---

Descrivendo i *cluster* in riferimento al vocabolario si ottengono le etichette in Tabella 3.2.

Tabella 3.2: *Labeling* dei *cluster* tramite le parole del vocabolario

<i>label</i> e similarità	frase di provenienza
..... <i>Dalle frasi di esempio estratte dal vocabolario</i> .....	
<i>A</i> : Canto (0.7892)	<b>Canto</b> di 'gallo' sul far della sera, o brutto tempo o gente straniera
<i>B</i> : canto <sup>25</sup> (0.5359)	il <b>canto</b> diciassettesimo del Purgatorio si svolge sulla terza cornice
<i>C</i> : Rime <sup>26</sup> (0.4865)	Eccoti o mar, solenne ed infinito. (Olindo Guerrini: <b>Rime</b> di Argia Sbolenti, In mare, v. 1)

<sup>25</sup>La parola *Canto* selezionata per *A* sarebbe stata l’etichetta più vicina (0.7795), ma quella inserita massimizza la somma delle similarità del coseno.

<sup>26</sup>Le parole indicate per *A* e *B* avrebbero una maggiore similarità, rispettivamente 0.7727 e

..... <i>Dalle definizioni del vocabolario</i> <sup>27</sup> .....		
$\mathcal{A}$ : cucù	(0.5779)	canto del cuculo
$\mathcal{B}$ : liutista	(0.5874)	compositore di melodie per liuto
$\mathcal{C}$ : stanza	(0.5664)	strofa di un componimento
..... <i>Dalle parole definite dal vocabolario</i> .....		
$\mathcal{A}$ : Cantico dei Cantici	(0.5976)	<b>Cantico dei Cantici</b> : ventiduesimo libro della Bibbia, composto di otto capitoli
$\mathcal{B}$ : sceneggiato <sup>28</sup>	(0.6293)	<b>sceneggiato</b> : adattato al teatro, al cinema, alla televisione
$\mathcal{C}$ : lamento	(0.6187)	<b>lamento</b> : nel melodramma italiano del XVII° secolo, scena che precede la conclusione tragica dell'opera

---

### 3.6.3 Gli strumenti per il confronto

Gli strumenti per misurare i cambiamenti semantici diacronici possono essere utili anche per valutare se la stessa parola è usata in modi diversi in documenti contemporanei<sup>29</sup>. Le varie traduzioni della Bibbia o le diverse sezioni abbastanza omogenee in cui è divisibile ben si prestano a questo tipo di analisi. In particolare possono essere utili i modelli che si concentrano nella comparazione tra due intervalli di tempo che determinano se avviene un cambiamento senza cercare di spiegarne il tipo<sup>30</sup>.

In letteratura sono presenti numerose misure per quantificare questo cambiamento<sup>31</sup>, Montariol, Martinc e Pivovarova 2021 applicano ai *cluster* d'uso delle parole nei diversi intervalli di tempo il problema del trasporto ottimo<sup>32</sup>.

---

0.5214, ma questa è quella selezionata dall'algoritmo di assegnamento ottimo.

<sup>27</sup>In questo caso il confronto viene fatto con il vettore dell'intera frase.

<sup>28</sup>La parola *lamento*, selezionata per  $\mathcal{C}$ , sarebbe più vicina (0.6592) ma è stata assegnata questa per i motivi già indicati.

<sup>29</sup>Pražák et al. 2020, p. 246.

<sup>30</sup>Cfr. Tang 2018, p. 662.

<sup>31</sup>Ad esempio Giulianelli, Del Tredici e Fernández 2020 ne propongono e valutano tre usando come *benchmark* Gulordava e Baroni 2011 e Frermann e Lapata 2016. Pražák et al. 2020 usano la similarità del coseno usando la media come soglia o confrontando i *nearest neighbours*.

<sup>32</sup>Cfr. Solomon 2018, usando in *Python* il pacchetto sviluppato da Flamary et al. 2021

Il problema del trasporto ottimo cerca di minimizzare il costo necessario per spostare delle masse da alcuni luoghi in altri. Prendendo come funzione di costo una misura di distanza e come masse le frequenze relative è possibile applicarlo al confronto tra distribuzioni di probabilità definite in uno spazio geometrico, come è quello degli *embedding* di 768 dimensioni creati da BERT.

Un algoritmo che risolva questo tipo di problemi in tempi ragionevoli al crescere dei punti e delle dimensioni dello spazio è stato reso possibile solamente di recente tramite tecniche di calcolo numerico.

Presi gli *embedding* di una parola in due sezioni del *corpus* e applicato l'algoritmo di *clustering* si possono ottenere, per ciascuna delle sezioni di testo, le frequenze relative nei *cluster* e la posizione dei centroidi.

Come funzione di costo viene usata quindi la matrice delle distanze del coseno tra i centroidi dei *cluster* delle occorrenze della prima sezione e quelli della seconda. Come massa di partenza e di arrivo si considerano le due distribuzioni di frequenza. L'uso della frequenza relativa è necessario perché la massa totale deve rimanere costante.

Il risultato ottenuto in questo modo tiene conto contemporaneamente delle distribuzioni nei singoli *cluster* e della loro posizione nello spazio semantico.

Nell'articolo citato, con questa misura si ottengono buoni risultati in termini di correlazione con i giudizi umani e considerano i risultati superiori allo 0.3 una differenza alta tra i significati in due intervalli di tempo.

Come è uso comune, per visualizzare il cambiamento nelle varie sezioni di testo, sarà adottato un diagramma a barre della distribuzione di frequenza relativa nei *cluster*.

#### 3.6.4 Il confronto tra i versetti

Ottenute le rappresentazioni dei vari versetti come descritto sopra (cfr. 3.5), queste si possono usare ad esempio per trovare quelli per cui le traduzioni sono più o meno concordi. Come misura del disaccordo tra i traduttori è possibile usare la distanza del coseno.

Per procedere al confronto ritorna il problema precedentemente accantonato dell'allineamento dei versetti. Limitandoci inizialmente al Nuovo Testamento, il



problema dovrebbe riguardare solo i versetti raggruppati dal traduttore, presenti soltanto nella traduzione interconfessionale (ad esempio Mt 2, 9–10).

Nel caso dei versetti raggruppati, si può fare prima una somma degli *embedding* dei versetti ed usarla per misurare la distanza.

Per l'Antico Testamento, invece, per riallineare i versetti si può cercare il più simile all'interno dello stesso libro e poi calcolare le similarità relative.

Infine, il sistema dovrebbe essere in grado, partendo da una citazione distorta di ricondurre al versetto. Questa è una applicazione molto utile, dato che cercando un versetto per come si ricorda, si parte spesso da parafrasi o si usano sinonimi, e senza parole chiave esatte la ricerca nel testo non dà risultati.

Per vedere quanto il sistema funzioni, serve una certa quantità di citazioni con rumore da confrontare con i risultati del sistema di ricerca. Versetti, antifone e responsori usati nella liturgia partono spesso da citazioni bibliche ma differiscono dalle traduzioni di riferimento per ragioni metriche, per aiutare la comprensione inserendo qualche riferimento al contesto da cui sono estratti o sono il risultato di una qualche composizione di più riferimenti biblici. Quando la fonte di queste brevi frasi può anche essere extra-biblica, i libri liturgici riportano anche il versetto o i versetti che li hanno ispirati. Estruendo queste informazioni dai libri liturgici<sup>33</sup> e controllando quante volte il riferimento trovato è esatto ed il risultato della misura di similarità, si può stimare l'efficacia del sistema.

---

<sup>33</sup>I testi digitalizzati della maggior parte dei libri liturgici sono scaricabili da <https://liturgico.chiesacattolica.it/category/documenti/libri-liturgici/>. Dopo un esame visivo è stato possibile estrarre i dati usando espressioni regolari.

# Capitolo 4

## I risultati – parole di interesse teologico

### 4.1 Legge

L'ebraico *tôrah* possiede un significato più largo, meno strettamente giuridico, del greco *nómos* con cui l'hanno tradotto i Settanta. Designa un “insegnamento” dato da Dio agli uomini per regolare la loro condotta. Si applica innanzitutto all'insieme legislativo che la tradizione dell'AT collegava a Mosè. Fondandosi su questo senso della parola, classico nel giudaismo, il NT chiama “la legge” tutta l'economia di cui questa legislazione era la parte principale, in opposizione al regime di grazia inaugurato da Gesù Cristo; tuttavia parla anche della “legge di Cristo”.<sup>1</sup>

Esaminare la parola *legge* è interessante per vedere se i significati diversi che sono attribuiti al termine dagli studiosi a seconda del libro in cui si trovano sono percepiti anche dal modello. Un'altra curiosità può essere se da qualcuno dei *cluster* esca fuori la sfumatura di significato meno giuridica e più vicina all'insegnamento, presente nel testo ebraico. La *Concordanza pastorale della Bibbia* individua quattro significati principali per la parola *legge*: legge dell'alleanza, natura della legge, legge nuova ed osservanza della legge; suddividendo ciascuno in otto o nove contesti d'uso specifici<sup>2</sup>.

---

<sup>1</sup>Grelot 1976, p. 589.

<sup>2</sup>Cfr. Passeleq e Poswick 2012, pp. 704–709.

Usando la stessa traduzione CEI del 2008 come riferimento, includendo sia la parola con l’iniziale maiuscola che con quella minuscola, il modello ottiene 15 *cluster* di cui uno con solo 3 occorrenze del termine.

Confrontando prima di tutto le due forme con e senza lettera maiuscola, si ottiene una distanza del trasporto ottimo di 0.31, piuttosto alta, appena inferiore a quella tra Antico Testamento ebraico e il Nuovo. L’84% delle occorrenze di *Legge* è incluso nel *cluster*  $\mathcal{C}$  (il più grande individuato dal modello) che include anche una occorrenza con l’iniziale minuscola in Romani 2, 14: «Quando i pagani, che non hanno la *Legge*, per natura agiscono secondo la *Legge*, essi, pur non avendo *Legge*, sono **legge** a se stessi.», probabilmente il modello è stato influenzato dalla presenza delle altre tre occorrenze con l’iniziale maiuscola, nonostante il traduttore sembra aver voluto segnalare il diverso significato del termine con la diversa grafia. Il *cluster*  $\mathcal{I}$  che presenta prevalentemente forme minuscole, ospita anche un *Legge* in Atti 22, 3<sup>3</sup>. Tutti gli altri presentano solo una delle due forme.

In Figura 4.1b si può confrontare la distribuzione delle parole nei *cluster* per le otto macro-sezioni di testo biblico. Per ognuna di esse, disposte sull’asse orizzontale, l’altezza di ogni rettangolo rappresenta la frequenza relativa delle occorrenze accolte in ciascun *cluster*. I *cluster* sono identificabili dal colore e, dove l’altezza lo consente (tralasciando quelli con meno del 2,5% delle occorrenze), dall’etichetta del nome del *cluster* che sarà sempre indicato con le lettere maiuscole dell’alfabeto latino.

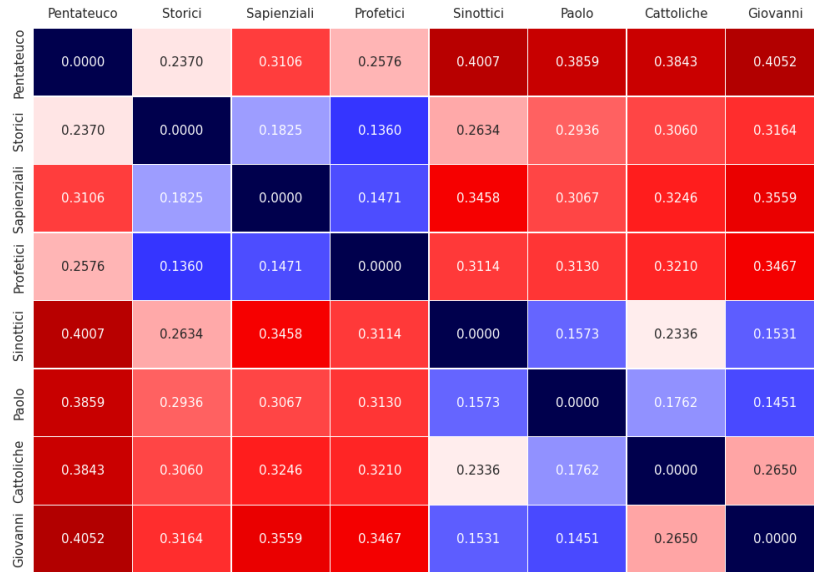
Per realizzare la legenda affiancata al grafico, è stato preso il testo dei versetti al cui interno si trovano le parole incluse in ciascun *cluster*. Da questi testi sono state estratte dieci tra le parole e i bigrammi più salienti secondo la *Tf-idf*, come spiegato sopra al paragrafo 3.6.2.

Dal grafico si nota che alcuni degli usi identificati sono specifici di determinate sezioni di testo. In particolare più di metà degli usi sono caratteristici dell’Antico Testamento ( $\mathcal{H}, \mathcal{K} - \mathcal{O}$ ) o del Nuovo ( $\mathcal{A} - \mathcal{C}$ ).

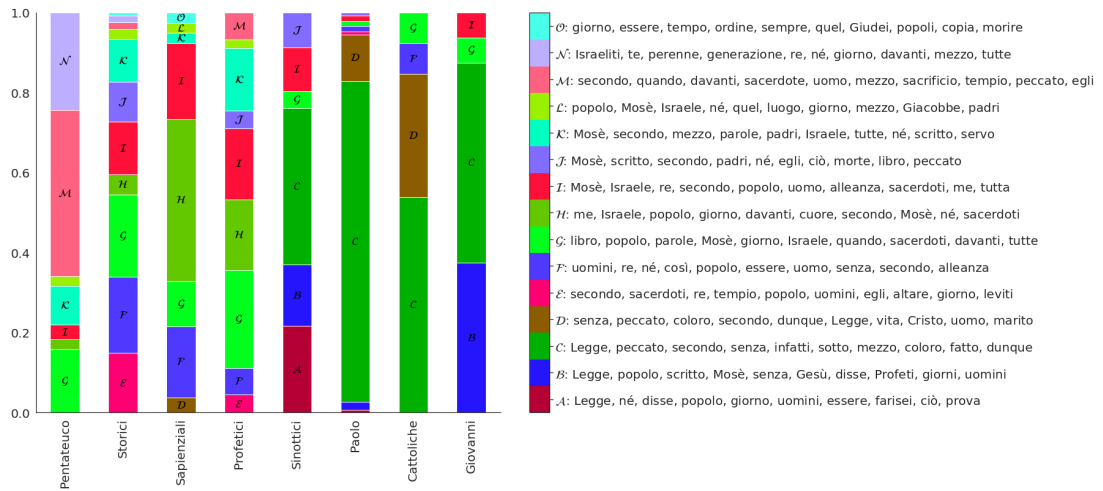
Alcuni *cluster* sono presenti solo in piccoli gruppi di libri, ad esempio  $\mathcal{M}$  include occorrenze quasi esclusivamente dalla parte centrale del Pentateuco,  $\mathcal{A}$

---

<sup>3</sup>«Io sono un Giudeo, nato a Tarso in Cilicia, ma educato in questa città, formato alla scuola di Gamaliele nell’osservanza scrupolosa della **Legge** dei padri, pieno di zelo per Dio, come oggi siete tutti voi.»



(a) distanza del trasporto ottimo



(b) distribuzione delle occorrenze nei cluster

Figura 4.1: Legge — confronto tra le principali sezioni della Bibbia



Figura 4.2: Legge — distanza del coseno tra i centri dei cluster

dei Sinottici,  $\mathcal{D}$  della letteratura sapienziale più tarda e dell'epistolario del Nuovo Testamento. Allo stesso modo alcuni libri sono caratterizzati da uno o due usi dominanti, ad esempio: Levitico e Numeri da  $\mathcal{M}$  e  $\mathcal{N}$ , Maccabei da  $\mathcal{E}$  e  $\mathcal{F}$ , i Salmi da  $\mathcal{H}$ .

L'osservazione delle differenze tra Antico e Nuovo Testamento, come di quella del Pentateuco dal resto dell'Antico Testamento (in particolare dai libri Sapienziali), è confermata dalla misura della distanza nel problema del trasporto ottimo.

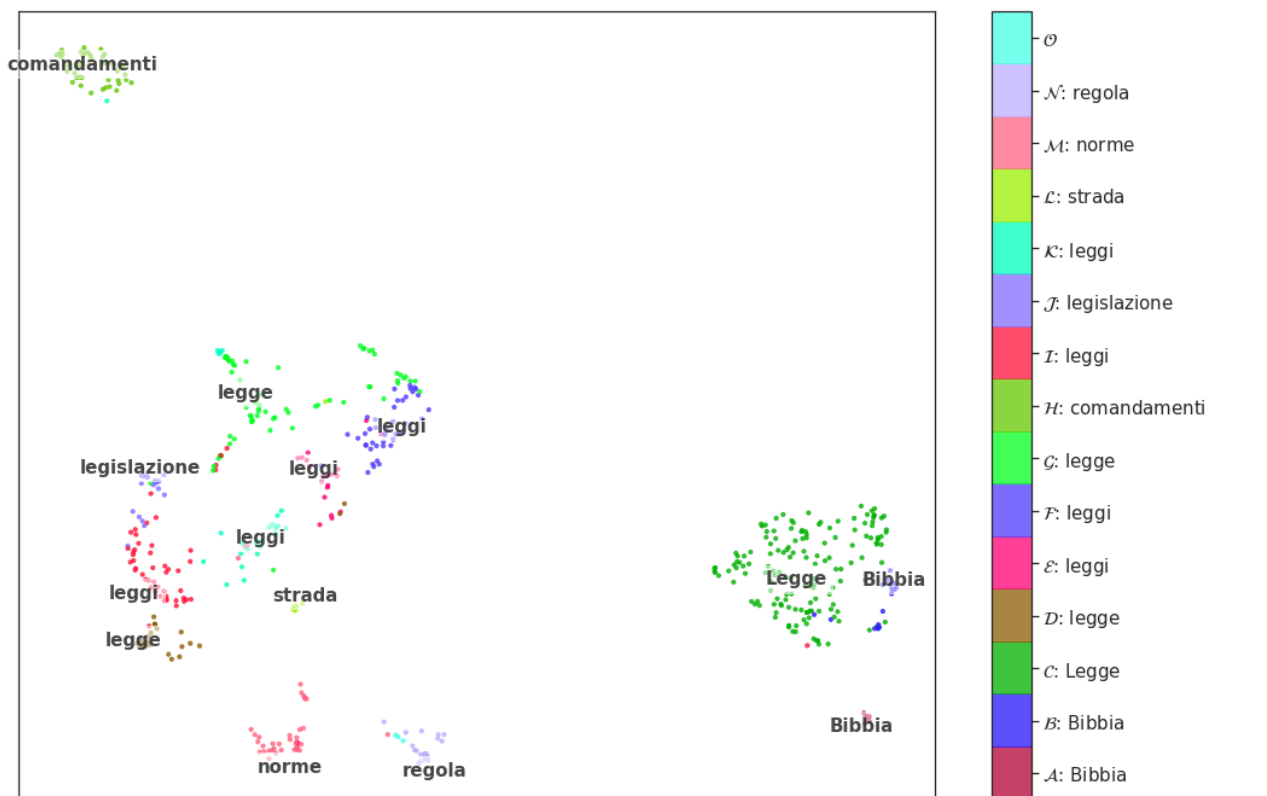
Le distanze, nel confronto tra Antico e Nuovo Testamento sono tutte piuttosto alte, all'interno del Nuovo Testamento il gruppo di libri che si distingue è quello delle lettere Cattoliche, probabilmente a causa della minore quantità di testo che le compone, anche se non sorprende che il più vicino sia l'epistolario paolino (0.1762).

In Figura 4.1a, per ogni sezione di testo, disposte sulle righe e sulle colonne, si può trovare la distanza del trasporto ottimo rispetto alle altre. Il colore varia in base alla distanza, dal blu scuro della distanza nulla ottenuta confrontando una sezione con sé stessa, al rosso più intenso delle più distanti.

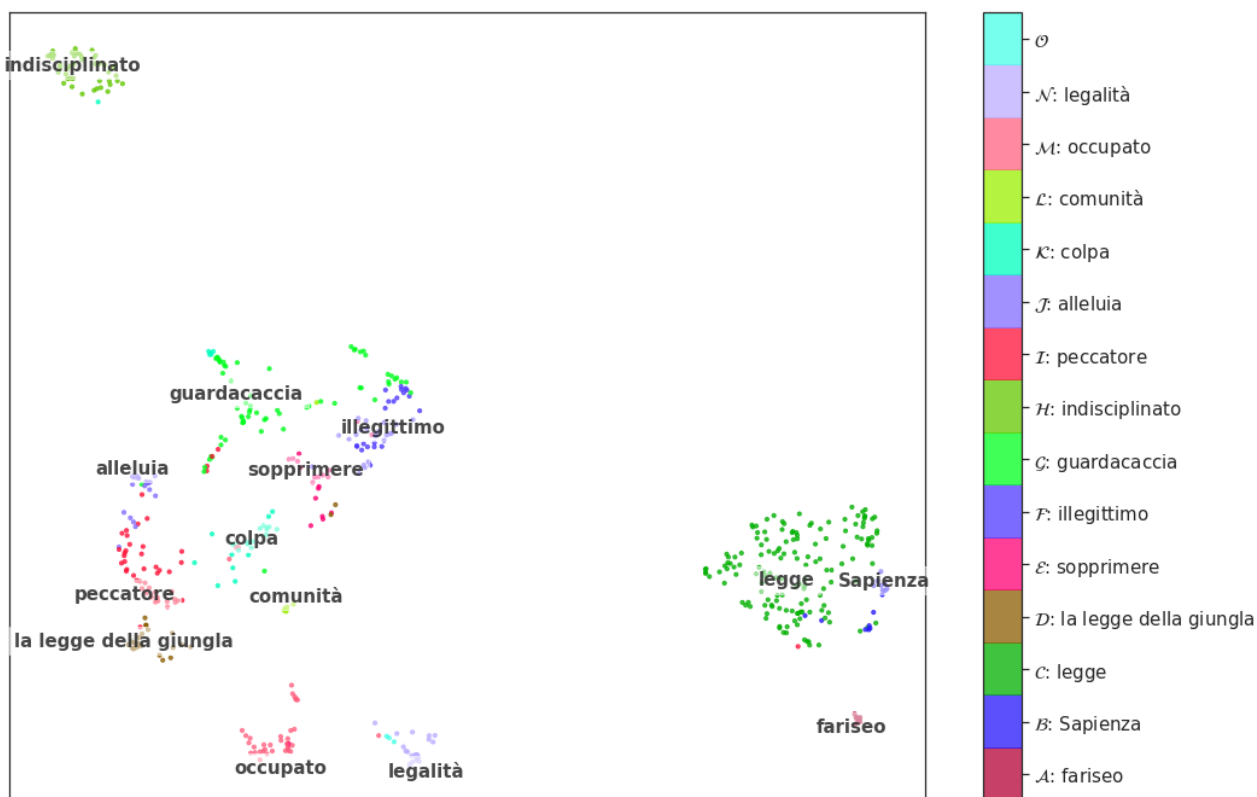
In Figura 4.2, sono riportate allo stesso modo le distanze del coseno tra i centroidi dei *cluster* di dimensioni rilevanti individuati.

In Figura 4.3 sono rappresentate le posizioni dei vettori di tutte le occorrenze della parola esaminata. Le 768 dimensioni degli *embedding* sono state riportate in due dimensioni con *UMAP*, come descritto sopra (3.6.2). Ogni punto rappresenta un vettore ed il colore indica il *cluster* di appartenenza. I centroidi dei *cluster* di dimensioni rilevanti sono etichettati con la parola estratta dagli esempi o dalle definizioni del vocabolario secondo quanto indicato nella didascalia. Dove più l'etichetta fosse contesa da più centroidi sono stati assegnati riferimenti unici in modo tale da minimizzare la somma delle distanze. Questo non toglie che la stessa parola possa ricorrere più volte come etichetta nel caso in cui provenga da definizioni o esempi diversi. Per agevolare la lettura dell'immagine, le parole sono riportate anche in legenda accanto ai colori e le lettere che identificano i *cluster*.

Confrontando i significati più simili ai centri dei *cluster* non si nota traccia del significato più ampio e vicino ad insegnamento. Come era lecito attendersi,



(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 4.3: Legge — visualizzazione in uno spazio bidimensionale

invece, la maggior parte dei riferimenti è proprio di natura giuridica: *norme, regole, diritto, legislazione, codice civile*.

Sono ricorrenti, invece, riferimenti alla Bibbia e ad alcuni suoi libri (Sapienza, Malachia, apocalisse, pericope, alleluia<sup>4</sup>) nel *cluster*  $\mathcal{B}$ , caratteristico dell'uso con la lettera maiuscola e del Nuovo Testamento, e in misura minore in  $\mathcal{J}$  (usato in particolare da libri storici e sinottici) e in  $\mathcal{G}$  avvicinato alla definizione di *libro*, il cui centro risulta vicino, tra le rappresentazioni ricavate dagli esempi, a *legge* come terza persona singolare del verbo leggere<sup>5</sup>. La presenza di *libro* e delle parole associate ai *cluster*  $\mathcal{B}$  e  $\mathcal{J}$  fa dubitare che questa associazione sia semplicemente il frutto di una confusione del modello dovuta all'omografia, ma permette di affermare che sia un riferimento significativo.

Nel *cluster*  $\mathcal{A}$  (tipico dei sinottici) spuntano anche diversi riferimenti a termini caratteristici della cultura ebraico-cristiana come *fariseo* e *sadduceo*.

Tra i *cluster* che, esplorando i significati più simili, si distinguono, ci sono i termini *parola* e *comandamenti* per descrivere (come in Figura 4.3a)  $\mathcal{H}$ , che identifica un uso caratteristico dell'Antico Testamento, comune in particolare nella letteratura sapienziale. Interessanti anche i riferimenti a *strada* e *comunità* per  $\mathcal{L}$  (nell'Antico Testamento), e a *colpa* e *peccato* in  $\mathcal{I}$  (il più trasversale tra i libri biblici) e in parte nei vicini  $\mathcal{J}$  e  $\mathcal{K}$ .

Confrontando i centri dei *cluster* con le parole del dizionario nella frase che le definisce, si identificano tre macro-gruppi di significati:

1. «*legge*: diritto di legge: attuazione della legge e/o moderazione» è il più simile per ben 10 *cluster*:  $\mathcal{A} - \mathcal{D}$ ,  $\mathcal{F} - \mathcal{I}$ ,  $\mathcal{K} - \mathcal{L}$
2. «*legge*: relazione determinata e costantemente valida fra quantità variabili, elementi o grandezze in un determinato fenomeno» in tre casi:  $\mathcal{E}$ ,  $\mathcal{M}$ ,  $\mathcal{N}$  (etichettati in Figura 4.3b rispettivamente con *sopprimere*, *occupato* e *legalità*)

---

<sup>4</sup>Dalla definizione: «nel rito cattolico, grido di esultanza; nella Bibbia è anche parte del libro dei salmi composto da re Davide».

<sup>5</sup>Come in «L'uomo che pratica una sola classe sociale, è come lo studioso che non legge altro che un libro».

3. «*disposto dalla legge*: fissato, stabilito, deliberato, comandato, ordinato, prescritto, deciso» per un solo gruppo di occorrenze:  $\mathcal{J}$  (etichettato in Figura 4.3b con *alleluia*).

Questi macro-gruppi possono servire per ordinare le occorrenze e realizzare una concordanza che tenga conto dei significati della parola. Il *cluster*  $\mathcal{J}$  rappresenta, da solo, il terzo di questi significati individuati. In Tabella 4.1 sono riportati i versetti al cui interno è presente la parola *legge* se il vettore è incluso in  $\mathcal{J}$ . Di ogni versetto è riportata la provenienza ed è evidenziata in neretto la parola *legge* quando è inclusa nel *cluster* indicato.

Tabella 4.1: Legge — versetti in cui occorrono le parole del *cluster*  $\mathcal{J}$

<b>LEGGE 3,</b>	<i>disposto dalla l.:</i>	fissato, stabilito, deliberato, comandato, ordinato, prescritto, deciso
Gs 8, 31	come aveva ordinato Mosè, servo del Signore, agli Israeliti, secondo quanto è scritto nel libro della <b>legge</b> di Mosè ...	
Gs 23, 6	Siate forti nell'osservare e mettere in pratica quanto è scritto nel libro della <b>legge</b> di Mosè, senza deviare da esso né a destra né a sinistra,	
1Re 2, 3	Osserva la legge del Signore, tuo Dio, procedendo nelle sue vie ed eseguendo le sue leggi, i suoi comandi, le sue norme e le sue istruzioni, come sta scritto nella <b>legge</b> di Mosè, perché tu riesca in tutto quello che farai e dovunque ti volgerai,	
2Re 14, 6	Ma non fece morire i figli degli uccisori, secondo quanto è scritto nel libro della <b>legge</b> di Mosè, ove il Signore prescrive ...	
2Cr 23, 18	... perché offrirono olocausti al Signore, come sta scritto nella <b>legge</b> di Mosè, fra gioia e canti, secondo le disposizioni di Davide.	
2Cr 25, 4	Ma non fece morire i loro figli, secondo quanto è scritto nel libro della <b>legge</b> di Mosè, ove il Signore prescrive: ...	
2Cr 31, 3	Una parte dei beni del re era per gli olocausti del mattino e della sera, gli olocausti dei sabati, dei noviluni e delle feste, come sta scritto nella <b>legge</b> del Signore.	
2Cr 35, 26	Le altre gesta di Giosia, le sue opere di pietà secondo ciò che è scritto nella <b>legge</b> del Signore,	
Esd 3, 2	... e costruirono l'altare del Dio d'Israele, per offrirvi olocausti, come è scritto nella <b>legge</b> di Mosè, uomo di Dio.	



- Tb 1, 8 ... La portavo loro ogni tre anni e la si consumava insieme, come vuole la **legge** di Mosè e secondo le raccomandazioni di Debora, moglie di Ananiel, la madre di nostro padre, poiché mio padre, morendo, mi aveva lasciato orfano.
- Tb 6, 13 ... egli incorrerebbe nella morte secondo la prescrizione della **legge** di Mosè, poiché egli sa che prima di ogni altro spetta a te avere sua figlia. ...
- Tb 7, 14 Chiamò poi la madre di lei e le disse di portare un foglio e stese l'atto di matrimonio, secondo il quale concedeva in moglie a Tobia la propria figlia, in base al decreto della **legge** di Mosè. Dopo di ciò cominciarono a mangiare e a bere.
- Bar 2, 2 Non era mai avvenuto sotto la volta del cielo quello che egli ha fatto a Gerusalemme, secondo ciò che è scritto nella **legge** di Mosè,
- Dn 9, 11 Tutto Israele ha trasgredito la tua legge, si è allontanato per non ascoltare la tua voce; così si è riversata su di noi la maledizione sancita con giuramento, scritto nella **legge** di Mosè, servo di Dio, perché abbiamo peccato contro di lui.
- Lc 2, 23 come è scritto nella **legge** del Signore: Ogni maschio primogenito sarà sacro al Signore -
- Lc 24, 44 Poi disse: " Sono queste le parole che io vi dissi quando ero ancora con voi: bisogna che si compiano tutte le cose scritte su di me nella **legge** di Mosè, nei Profeti e nei Salmi ".
- At 13, 38 ... Da tutte le cose da cui mediante la **legge** di Mosè non vi fu possibile essere giustificati,
- At 28, 23 ... Dal mattino alla sera egli esponeva loro il regno di Dio, dando testimonianza, e cercava di convincerli riguardo a Gesù, partendo dalla **legge** di Mosè e dai Profeti.
- 1Cor 9, 9 Nella **legge** di Mosè infatti sta scritto: Non metterai la museruola al buo che trebbia. Forse Dio si prende cura dei buoi?
- 

Gli usi raggruppati in questo *cluster* riguardano la *legge di Mosè* (19 occorrenze) o *del Signore* (2Cr 31, 3 e 35, 26; Lc 2, 33) ma non *legge del Signore* in primo libro dei Re 2, 3. Come risulta anche dalle parole chiave selezionate in Figura 4.3a, ci si riferisce alla legge come scritta, o come libro.

Fanno parziale eccezione al riferimento alla scrittura le citazioni dal libro di Tobia, anche se *secondo la prescrizione della legge* e soprattutto *in base al decreto della legge* fanno pensare alla legge come codice, quindi a qualcosa di non tanto distante. In Luca 24, 44 e Atti 28, 23 il collegamento al libro lo si evince dall'affiancamento a Profeti e Salmi, due parti della Bibbia. L'unico versetto che sembra estraneo a questi riferimenti è Atti 13, 38.

### Risultati principali

Tra i significati individuati in questa analisi, sono presenti alcuni riferimenti meno strettamente giuridici, ma riferiti all'uso di *legge* come sinonimo di Bibbia e come libro e non quelli legati al mondo dell'insegnamento individuati dagli studiosi del testo ebraico. Al contrario, l'opposizione tra Antico e Nuovo Testamento è stata confermata sia dal problema del trasporto ottimo che dal confronto delle frequenze nei *cluster*. L'analisi ha permesso di individuare un cambiamento nel significato della parola anche all'interno sia dell'Antico che del Nuovo Testamento.

## 4.2 Amore e carità

ἀγάπη; termine assolutamente preponderante nel NT (φιλία compare 1 volta e φιλαδελφία 5 volte); preponderante anche il verbo ἀγαπάω (rispetto a φιλέω: solo 25 volte). La Nova Vulgata traduce ἀγάπη mescolando senza un criterio chiaro i termini latini *caritas* e *dilectio*; analogamente la Bibbia CEI 1971<sup>6</sup> alterna (in Paolo) *amore* e *carità*. Si è scelto il seguente orientamento: in Giovanni (Vangelo e lettere) ἀγάπη è sempre tradotto con *amore*; in Paolo, se indica il rapporto Dio-uomo o uomo-Dio, si traduce con *amore*; se indica il rapporto fra uomini, *amore* (quando equivale a: affetto fraterno) ma *carità* (quando esprime la virtù in quanto tale).<sup>7</sup>

Le discussioni sulla traduzione di ἀγάπη sono numerose in ambito biblico. Può essere interessante il confronto in particolare tra le due traduzioni successive della

<sup>6</sup>La data si riferisce alla traduzione del Nuovo Testamento, poi confluito nella versione del 1974.

<sup>7</sup>Costa 1998, p. 131.

CEI per vedere se effettivamente c'è stato un cambiamento sensibile, esplorando insieme gli usi delle tre parole *affetto*, *amore* e *carità* più frequentemente usate per rendere il termine greco<sup>8</sup>.

Il nostro sistema ha due strumenti per il confronto: è possibile calcolare la distanza tra le due traduzioni nel problema del trasporto ottimo, oppure fare un confronto diretto tra le parole che sono state cambiate nell'aggiornamento.

La distanza tra le due traduzioni è meno di 0.1: non c'è un grande cambiamento nel significato dei tre termini tra i due testi. Analizzando la distribuzione delle parole nei *cluster* nelle due versioni come in Figura 4.4a, si notano il *cluster*  $\mathcal{E}$  presente soltanto nella traduzione più recente ed il  $\mathcal{G}$  molto più esteso che nella precedente, al contrario  $\mathcal{F}$ ,  $\mathcal{I}$  e i *cluster* di grandezza trascurabile ( $\mathcal{N} - \mathcal{O}$ ) sono più ampi nella traduzione CEI del 1974.

Confrontando invece la distribuzione dei tre termini nei *cluster* notiamo che sono nettamente distinti: in nessun caso ci sono istanze di più di una forma. In particolare, dalla Figura 4.4b notiamo anche che *affetto* e *carità* sono divisi in due sole accezioni, mentre per *amore* sono presenti più sfaccettature.

Dal confronto delle distanze tra i centroidi (in Figura 4.5) si nota che i due significati di *carità* ( $\mathcal{A} - \mathcal{B}$ ) sono molto vicini tra loro e distanti dal resto, similmente i significati di *amore* sono divisi in due gruppi ben distinti:  $\mathcal{C} - \mathcal{D}$  ed  $\mathcal{E} - \mathcal{K}$ . I due *cluster* di *affetto* ( $\mathcal{L} - \mathcal{M}$ ) sono invece piuttosto distanti tra loro (0.7227) con il secondo abbastanza distante da tutti gli altri. Guardando le scelte fatte nella *Concordanza pastorale della Bibbia* si nota che viene identificato un solo significato per *affetto*, due (con un totale di 18 sfumature) per *amore* (in generale e in riferimento a Dio) ed uno (ma con all'interno 6 sfumature) per *carità*.

Prendendo i versetti in cui è avvenuto un cambiamento tra le tre traduzioni di  $\acute{\alpha}\gamma\acute{\alpha}\pi\eta$  tra il 1974 ed il 2008 e confrontando le due rappresentazioni, la distanza del coseno è mediamente inferiore a 0.1; confrontando i versetti escludendo la parola in questione, non si nota un grande cambiamento nella comprensione (una variazione media della distanza dell'8%, pari a 0.003).

Riguardo alla questione del cambiamento tra le due traduzioni si può concludere che, in generale, non ha provocato un grande spostamento di significato ma

---

<sup>8</sup>Cfr. Ghidelli 1998, p. 148.

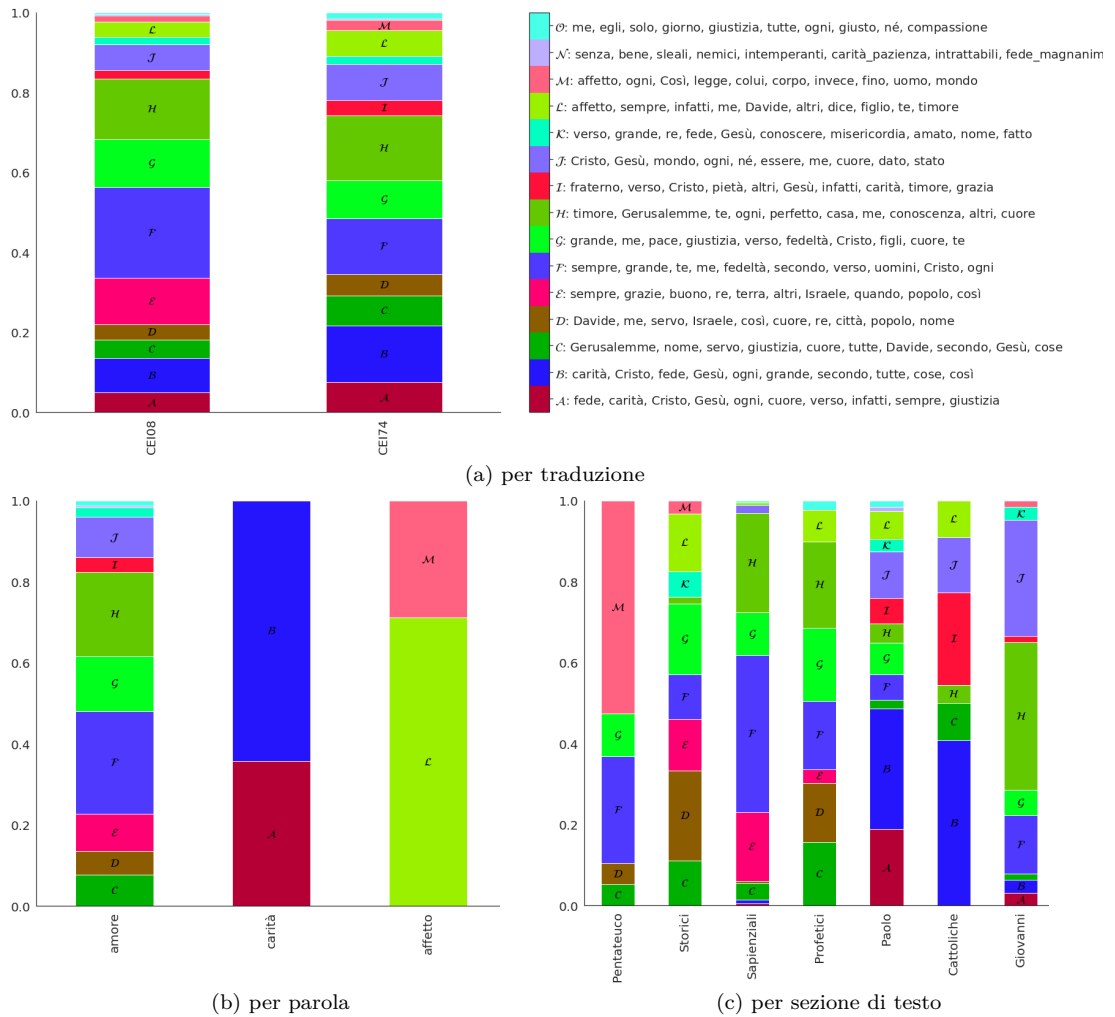


Figura 4.4: Amore — distribuzione delle occorrenze nei cluster

	A	B	C	D	E	F	G	H	I	J	K	L	M	
A-	0.0000	0.0517	0.6627	0.6727	0.5534	0.5212	0.4619	0.5420	0.4807	0.4541	0.4949	0.5279	0.8642	-0.5239
B-	0.0517	0.0000	0.6350	0.6563	0.5143	0.4784	0.4467	0.4754	0.4442	0.4205	0.4991	0.5327	0.8697	-0.5020
C-	0.6627	0.6350	0.0000	0.0625	0.4651	0.3958	0.3698	0.4491	0.4645	0.3898	0.4591	0.5845	0.8754	-0.4844
D-	0.6727	0.6563	0.0625	0.0000	0.4679	0.4025	0.3693	0.4577	0.4817	0.4234	0.4496	0.5651	0.9126	-0.4934
E-	0.5534	0.5143	0.4651	0.4679	0.0000	0.0820	0.1979	0.2507	0.2747	0.2315	0.2459	0.3791	0.9016	-0.3803
F-	0.5212	0.4784	0.3958	0.4025	0.0820	0.0000	0.1088	0.1543	0.1849	0.1671	0.1828	0.3271	0.8887	-0.3245
G-	0.4619	0.4467	0.3698	0.3693	0.1979	0.1088	0.0000	0.0963	0.1161	0.1407	0.1467	0.3184	0.8685	-0.3034
H-	0.5420	0.4754	0.4491	0.4577	0.2507	0.1543	0.0963	0.0000	0.1495	0.1656	0.2267	0.4483	0.8694	-0.3571
I-	0.4807	0.4442	0.4645	0.4817	0.2747	0.1849	0.1161	0.1495	0.0000	0.1502	0.1919	0.3944	0.8624	-0.3496
J-	0.4541	0.4205	0.3898	0.4234	0.2315	0.1671	0.1407	0.1656	0.1502	0.0000	0.1516	0.4018	0.8294	-0.3271
K-	0.4949	0.4991	0.4591	0.4496	0.2459	0.1828	0.1467	0.2267	0.1919	0.1516	0.0000	0.2824	0.8704	-0.3501
L-	0.5279	0.5327	0.5845	0.5651	0.3791	0.3271	0.3184	0.4483	0.3944	0.4018	0.2824	0.0000	0.7227	-0.4570
M-	0.8642	0.8697	0.8754	0.9126	0.9016	0.8887	0.8685	0.8694	0.8624	0.8294	0.8704	0.7227	0.0000	-0.8612

Distanza media

Figura 4.5: Amore — distanza del coseno tra i centri dei cluster

che puntualmente è abbastanza forte da cambiare qualcosa nella comprensione che il modello ha della singola occorrenza.

Prima di esplorare i cambiamenti a livello di porzione di testo, sempre interessanti per lo studio del testo biblico nell’equilibrio tra la percezione del testo come un tutt’uno e le evidenti differenze stilistiche e di epoca di composizione, alcune considerazioni di questo primo confronto meritano di essere approfondite.

Il primo dato che suscita curiosità è il *cluster*  $\mathcal{E}$ , presente soltanto nella traduzione CEI del 2008, che presenta un numero non trascurabile di occorrenze di *amore*, come si nota in Figura 4.4a. Confrontando i versetti in cui occorrono i termini inclusi nel *cluster* nelle due traduzioni, scopriamo che la traduzione del 1974 non presenta mai il termine *amore* ma, come evidenziato in Tabella 4.2<sup>9</sup>, usa *grazia* nei libri storici o profetici e *misericordia* o *bontà* (Sal 138, 8) nel libro dei Salmi.

Tabella 4.2: Amore — confronto tra le traduzioni dei versetti in cui occorrono le parole in  $\mathcal{E}$

	.....1Cr 16, 34.....
CEI08	Rendete grazie al Signore perché è buono, perché il suo <b>amore</b> è per sempre.
CEI74	Lodate il Signore, perché è buono, perché la sua <b>grazia</b> dura sempre.
	.....1Cr 16, 41.....
CEI08	... perché lodassero il Signore, perché il suo <b>amore</b> è per sempre.
CEI74	... perché lodassero il Signore, perché la sua <b>grazia</b> dura sempre.
	.....2Cr 5, 13.....
CEI08	... per lodare il Signore perché è buono, perché il suo <b>amore</b> è per sempre, ...
CEI74	... per lodare il Signore perché è buono, perché la sua <b>grazia</b> dura sempre, ...
	.....2Cr 7, 3.....
CEI08	... adorarono e celebrarono il Signore perché è buono, perché il suo <b>amore</b> è per sempre.
CEI74	... adorarono e celebrarono il Signore perché è buono, perché la sua <b>grazia</b> dura sempre.

<sup>9</sup>Per le citazioni dal libro dei Salmi è stato necessario riallineare i versetti per presentare la traduzione dello stesso testo.

- ..... *2Cr 7, 6* .....
- CEI08 ... con tutti gli strumenti musicali che il re Davide aveva fatto per celebrare il Signore, perché il suo **amore** è per sempre, ...
- CEI74 ... con tutti gli strumenti musicali, fatti dal re Davide, celebravano il Signore, perché la sua **grazia** dura sempre, ...
- ..... *2Cr 20, 21* .....
- CEI08 ... perché lodassero il Signore dicendo: " Rendete grazie al Signore, perché il suo **amore** è per sempre ".
- CEI74 ... perché lodassero il Signore dicendo: Lodate il Signore, perché la sua **grazia** dura sempre.
- ..... *Esd 3, 11* .....
- CEI08 Essi cantavano lodando e rendendo grazie al Signore, ripetendo: " Perché è buono, perché il suo **amore** è per sempre verso Israele " ...
- CEI74 Essi cantavano a cori alterni lodi e ringraziamenti al Signore perché è buono, perché la sua **grazia** dura sempre verso Israele ...
- ..... *1Mac 4, 24* .....
- CEI08 Di ritorno cantavano e benedicevano il Cielo perché è buono, perché il suo **amore** è per sempre.
- CEI74 Di ritorno cantavano e innalzavano benedizioni al cielo " perché egli è buono e la sua **grazia** dura sempre ".
- ..... *Sal 100, 5* .....
- CEI08 perché buono è il Signore, il suo **amore** è per sempre, ...
- CEI74 poiché buono è il Signore, eterna la sua **misericordia**, ...
- ..... *Sal 106, 1; 107, 1; 108, 1. 29* .....
- CEI08 Alleluia. Rendete grazie al Signore, perché è buono, perché il suo **amore** è per sempre.
- CEI74 Alleluia. Celebrate il Signore, perché è buono, perché eterna è la sua **misericordia**.
- ..... *Sal 118, 3* .....
- CEI08 Dica la casa di Aronne: " Il suo **amore** è per sempre ".
- CEI74 Lo dica la casa di Aronne: eterna è la sua **misericordia**.
- ..... *Sal 118, 4* .....
- CEI08 Dicano quelli che temono il Signore: " Il suo **amore** è per sempre ".
- CEI74 Lo dica chi teme Dio: eterna è la sua **misericordia**.

- .....*Sal 136, 1–26*.....
- CEI08 ... perché il suo **amore** è per sempre.
- CEI74 ... perché eterna è la sua **misericordia**.
- .....*Sal 138, 8*.....
- CEI08 ... Signore, il tuo **amore** è per sempre: non abbandonare ...
- CEI74 ... Signore, la tua **bontà** dura per sempre: non abbandonare ...
- .....*Ger 33, 11*.....
- CEI08 ... la voce di coloro che cantano: ” Rendete grazie al Signore degli eserciti, perché il suo **amore** è per sempre ” ...
- CEI74 ... il canto di coloro che dicono: Lodate il Signore degli eserciti, perché è buono, perché la sua **grazia** dura sempre ...
- .....*Dn 3, 89*.....
- CEI08 Lodate il Signore, perché egli è buono, perché il suo **amore** è per sempre.
- CEI74 Lodate il Signore, perché egli è buono, perché la sua **grazia** dura sempre.
- .....*Dn 3, 90*.....
- CEI08 ...lodatelo e celebratelo, perché il suo **amore** è per sempre ”.
- CEI74 ...lodatelo e celebratelo, perché la sua **grazia** dura sempre ”.
- 

Guardando ai versetti si nota immediatamente la regolarità della formula (ripetuta 26 volte a concludere ogni versetto nel Salmo 136) con i riferimenti alla lode, alla bontà e all’eternità dell’amore. Per risolvere il dubbio legato alla differenza riscontrata tra i due gruppi di sezioni precedenti si può risalire al testo originale ebraico e alla LXX (riferimento originale per Daniele e il primo libro dei Maccabei) per notare che si tratta della stessa parola **ἔλεος**<sup>10</sup> (la LXX usa invece ἔλεος). Potrebbe essere interessante fare un’analisi simile a questa anche per *amore*, *grazia*, *misericordia*, *bontà* ed eventuali altre traduzioni del termine.

L’altro dato interessante riguarda la parola *affetto*, per la quale il modello identifica due *cluster*, più delle sfumature di significato riconosciute dalla *Concordanza pastorale della Bibbia*, nonostante la tendenza fin qui riscontrata a in-

<sup>10</sup>Tradotto da Alonso Schökel 2013 con: «Favore, beneficio, grazia, servizio, aiuto; misericordia, clemenza, bontà, benevolenza, pietà, compassione, commiserazione, pena; affetto, tenerezza, carità; lealtà, fedeltà; patto, accordo, intesa, obbligazione, promessa, impegno; gradimento, gusto, amabilità, affabilità, simpatia, attrattiva, fascino; compiacimento, attrazione».

dividuarne un numero inferiore. Inoltre, abbiamo già notato che il significato  $\mathcal{M}$  è mediamente molto distante da tutti gli altri (0.8612).

Cercando i significati più simili nel dizionario, si trovano *ammalato, lebbroso, cronico, malattia*: esaminando le occorrenze incluse in questo *cluster* si nota che sono dell'aggettivo (sinonimo di ammalato, colpito, contagiato, infettato, sofferente) e non del sostantivo (sinonimo di sentimento, affezione, amore, amicizia, simpatia, ...). La distanza si nota anche graficamente come in Figura 4.6a in cui il *cluster*  $\mathcal{M}$  si trova etichettato dalla parola *affetta*<sup>11</sup> ed  $\mathcal{L}$  dalla parola *affetto*. Controllando la provenienza di queste occorrenze si nota la quasi coincidenza con quelle provenienti dal Levitico.

I due *cluster* di *carità* sono ben distinti dal gruppo centrale ma non è facile distinguerli, i loro centri, infatti, hanno una distanza del coseno di solo 0.05 e di conseguenza selezionano le stesse parole e definizioni dal confronto con il vocabolario. Anche esaminando le parole chiave e la distribuzione delle occorrenze tra le otto sezioni del testo biblico (sono quasi tutte concentrate nel Nuovo Testamento) non si trova molto per distinguerli.

A livello di libro, però, i due *cluster* sono quasi alternativi, segno che il modello abbia notato qualche differenza tra i contesti che potrebbe essere un utile spunto per un lavoro sul testo.

Confrontando le occorrenze si nota che non tutte siano presenti in entrambe le versioni prese in esame: la differenza riguarda nella maggior parte dei casi il cambiamento di traduzione tra *amore* e *carità*. Le tre occorrenze dell'Antico Testamento nella traduzione CEI del 1974 scompaiono, sostituite da *amore* (Sir 11, 15), *elemosina* (Sir 40, 24) e *pietà* (Pr 19, 17). Curioso il caso di Efesini 1, 4 in cui l'unica variazione nella traduzione è il passaggio da «al suo cospetto» a «di fronte a lui» ma che il sistema ha separato fra i due *cluster*.

Esplorare queste differenze ha permesso anche di notare come in alcuni di questi versetti, l'unica differenza di traduzione è stata la sostituzione della parola *amore* con *carità* (e viceversa) ma neanche in questi casi l'identità del contesto ha prevalso sulla differenza della singola parola ed il sistema ha diviso le due parole

<sup>11</sup>Nella frase: «Nel caso di una persona affetta da un singhiozzo, lo starnuto che sopravviene rimuove il singhiozzo».



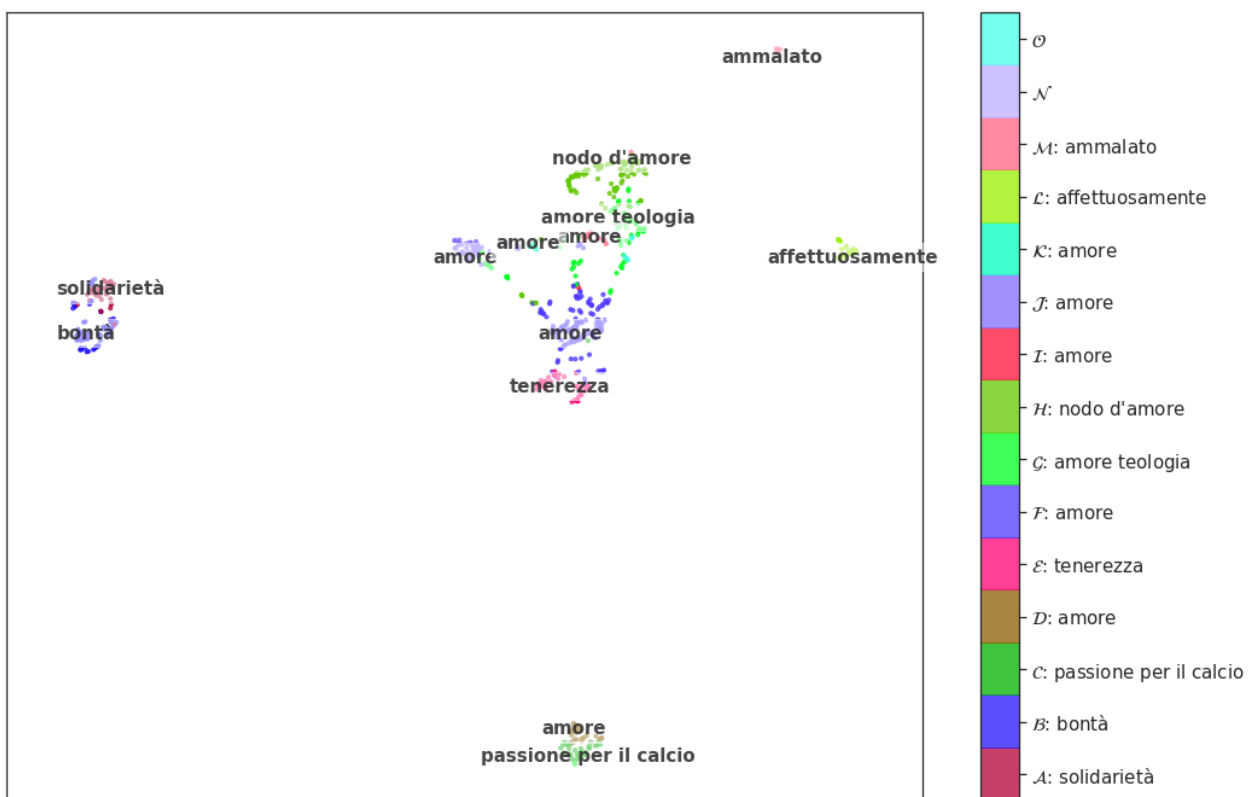
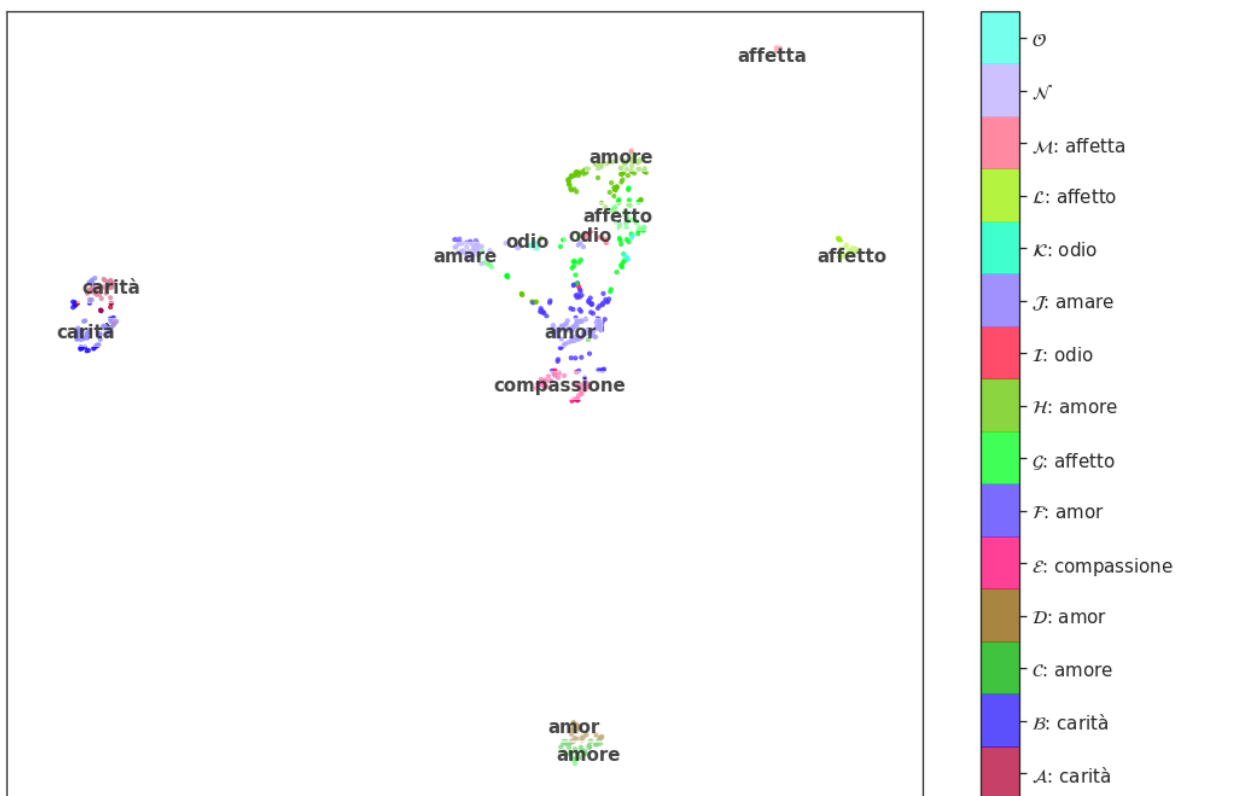


Figura 4.6: Amore — visualizzazione in uno spazio bidimensionale

in *cluster* diversi. Le 10 occorrenze parallele di *amore* non sono state raggruppate ma sono sparse in 5 *cluster* differenti.

Passando infine alla parola *amore*, sono stati già evidenziati i *cluster*  $\mathcal{C}$  e  $\mathcal{D}$  come molto vicini tra loro e lontani dagli altri (come si vede anche dalle Figure 4.6). Raccolgono riferimenti soprattutto dall'Antico Testamento (in particolare da libri storici e Salmi) e si distinguono dalla presenza di *Davide*, *servo*, *Gerusalemme* tra le parole co-occorrenti.

I *cluster*  $\mathcal{I}$  e  $\mathcal{J}$  ospitano *Cristo* e *Gesù* tra le parole più rilevanti nei versetti inclusi<sup>12</sup>. Non stupisce che siano caratteristici del Nuovo Testamento ma in essi sono ospitate anche alcune (poche) occorrenze dall'Antico: *amore della sapienza* (Sir 40, 20) e *amore del Signore* (Sal 89, 2; 103, 17 e Sir 1, 2). Il *cluster*  $\mathcal{I}$ , invece, seleziona tra le definizioni più simili del vocabolario anche *amore fraterno*, e *fraterno* è la più rilevante tra le co-occorrenti nei versetti. Include solo citazioni dall'epistolario del Nuovo Testamento ed in particolare tutte le occorrenze di *amore* nella seconda lettera di Pietro: possiamo prendere questo dato come conferma della centralità del tema dell'amore fraterno nelle lettere (in particolare nella seconda di Pietro)<sup>13</sup>.

A livello del singolo libro, il Cantico è caratterizzato da usi di *amore* del *cluster*  $\mathcal{H}$ , il secondo libro dei Maccabei da *affetto* ( $\mathcal{L}$ ), la prima lettera di Giovanni è quasi equamente diviso tra  $\mathcal{H}$  e  $\mathcal{J}$ . Colpisce la completa assenza dei vangeli dalle occorrenze di *amore*. Bisogna fare però attenzione a non concludere che il tema sia assente dai vangeli. Anche se il sostantivo è assente, in questo esempio non sono state considerate le forme flesse del verbo *amare*, che sono invece molto numerose nel testo evangelico.

## Risultati principali

Le modifiche apportate nella traduzione più recente della CEI non hanno provocato un cambiamento rilevante nel significato delle parole usate per tradurre lo stesso termine greco. Dal punto di vista delle distribuzioni di frequenza, hanno lasciato maggior traccia i cambiamenti apportati all'Antico Testamento.

<sup>12</sup>Sempre calcolando il Tf-Idf considerando i versetti nel *cluster* come documento e tutti i versetti che contengono una delle parole da esaminare come *corpus*.

<sup>13</sup>Cfr. Wiéner 1976, pp. 52–54.

Per quanto riguarda l'applicazione dell'algoritmo di *clustering* a più parole contemporaneamente, nonostante sia stato in grado di riconoscere che il sostantivo *affetto* sia più simile ad *amore* che all'aggettivo, le differenze tra le parole sono rimaste troppo forti per creare *cluster* misti. Rimane interessante la possibilità di avere uno sguardo d'insieme a parole dal significato simile.

### 4.3 Spirito

In tutte le lingue classiche e bibliche « spirito » è una parola suscettibile di sensi diversissimi. Tra lo spirito di vino e l'uomo di spirito, tra « rendere lo spirito » e « vivere secondo lo Spirito », ci sono molte differenze, ma ci sono pure analogie reali. Spirito tende sempre a designare in un essere l'elemento essenziale ed inafferrabile, ciò che lo fa vivere e ciò che emana da esso senza che egli voglia, ciò che è più esso stesso e ciò di cui egli non può rendersi padrone.<sup>14</sup>

La *Concordanza pastorale della Bibbia* individua quattro significati per *spirito*: *di Dio*,  *dono*, *dell'uomo* e *spiriti*, ciascuno con almeno sette sfumature. Questa poliedricità può rendere interessante l'analisi della parola.

In questo caso, all'analisi diacronica tra le diverse sezioni di testo, proviamo ad aggiungere l'analisi sincronica tra tutte e cinque le traduzioni in esame. L'interpretazione teologica dello Spirito dei testimoni di Geova è profondamente diversa da quella cristiana, di conseguenza, nella traduzione del nuovo mondo quasi non occorre la forma *Spirito* con la iniziale maiuscola, in questa indagine vengono considerate entrambe le forme, maiuscola e minuscola.

Come per *legge* (cfr. sezione 4.1) i *cluster* ospitano soprattutto una delle due forme, maiuscola ( $\mathcal{B} - \mathcal{F}$ ) o minuscola ( $\mathcal{A}$  e  $\mathcal{G} - \mathcal{Q}$ ) ma nella metà dei *cluster* è inclusa almeno qualche occorrenza dell'altra forma. In particolare, in quelli contraddistinti dalla forma con iniziale maiuscola, le occorrenze minuscole provengono tutte dalla traduzione del nuovo mondo.

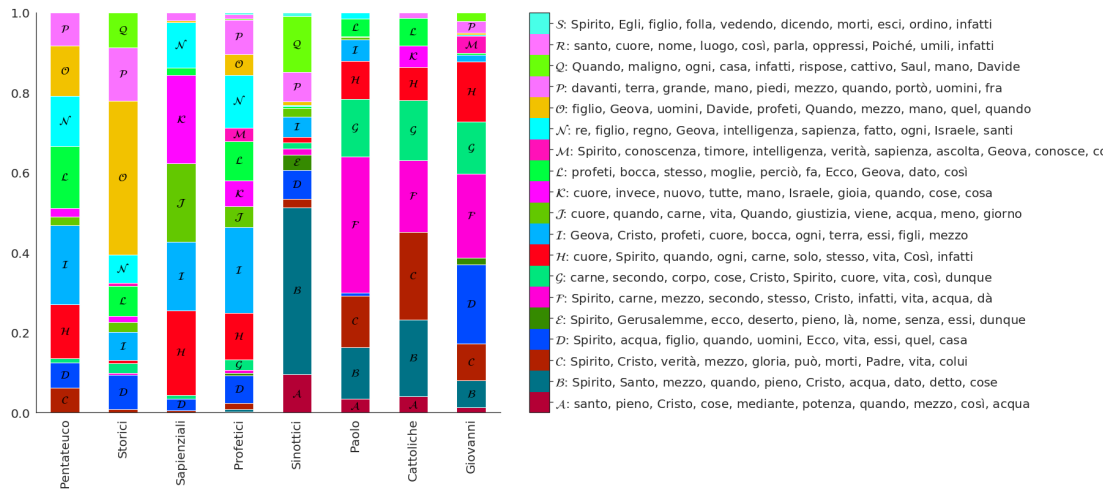
Rispetto a quanto visto per *legge*, nel problema del trasporto ottimo le distanze tra la forma con l'iniziale maiuscola e minuscola o tra le principali sezioni di testo

---

<sup>14</sup>Guillet 1976, p. 1229.

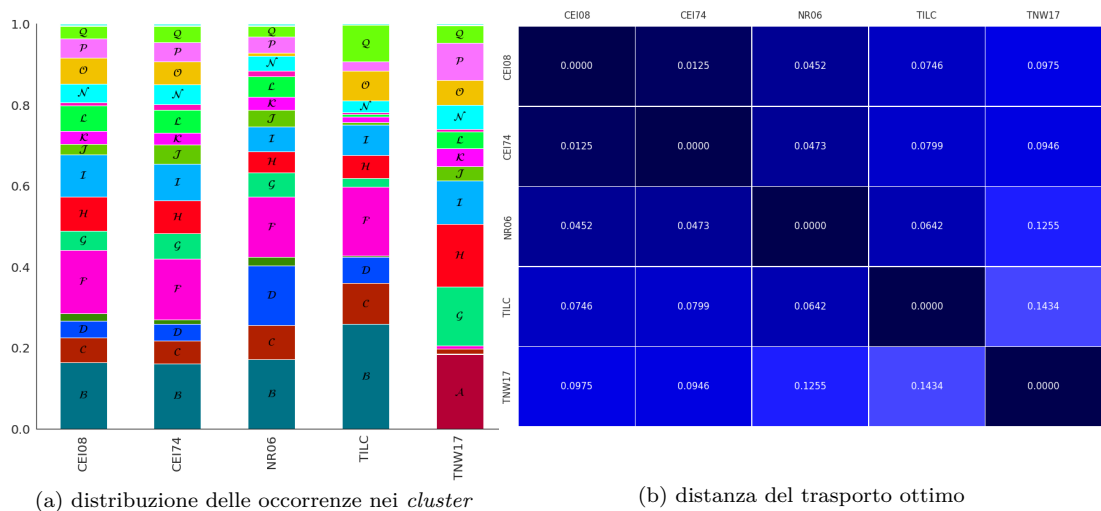
	Pentateuco	Storici	Sapenziali	Profetici	Sinottici	Paolo	Cattoliche	Giovanni
Pentateuco	0.0000	0.1438	0.1762	0.1119	0.2291	0.2435	0.2574	0.2523
Storici	0.1438	0.0000	0.2193	0.1466	0.2235	0.3028	0.3073	0.2900
Sapenziali	0.1762	0.2193	0.0000	0.1301	0.2838	0.2463	0.2759	0.2800
Profetici	0.1119	0.1466	0.1301	0.0000	0.2385	0.2402	0.2615	0.2468
Sinottici	0.2291	0.2235	0.2838	0.2385	0.0000	0.1386	0.1523	0.1464
Paolo	0.2435	0.3028	0.2463	0.2402	0.1386	0.0000	0.0748	0.0970
Cattoliche	0.2574	0.3073	0.2759	0.2615	0.1523	0.0748	0.0000	0.1309
Giovanni	0.2523	0.2900	0.2800	0.2468	0.1464	0.0970	0.1309	0.0000

(a) distanza del trasporto ottimo



(b) distribuzione delle occorrenze nei cluster

Figura 4.7: Spirito — confronto tra le principali sezioni della Bibbia



(a) distribuzione delle occorrenze nei cluster

(b) distanza del trasporto ottimo

Figura 4.8: Spirito — confronto tra le traduzioni

sono mediamente minori ma, come si nota anche in Figura 4.7 resta la chiara contrapposizione tra Antico e Nuovo Testamento.

Tra le traduzioni (cfr. Figura 4.8b) le uniche distanze superiori allo 0.1 sono quelle della traduzione dei Testimoni di Geova rispetto a quelle con influenza di area Riformata (Nuova Riveduta e traduzione interconfessionale, rispettivamente distanti 0.1255 e 0.1434). Per quanto siano molto basse sono comunque più di dieci volte quella tra le due versioni CEI, ed è curioso che la distanza sia così forte da due traduzioni che seguono approcci traduttivi opposti.

Il *cluster*  $\mathcal{A}$  ospita solo occorrenze provenienti dalla traduzione del nuovo mondo che è facile supporre corrispondano ad occorrenze con la forma maiuscola (e dal Nuovo Testamento) nelle altre traduzioni. La dimensione del *cluster*, però, non è sufficiente a coprirle tutte, buona parte delle occorrenze parallele saranno distribuite negli altri *cluster*.

Guardando la Figura 4.8a si nota subito la quasi identità tra le versioni CEI ma anche la similarità tra le tutte le traduzioni cristiane. Le altre differenze che si notano a colpo d'occhio sono il *cluster*  $\mathcal{O}$  che quasi scompare nella Nuova Riveduta rispetto ad un  $\mathcal{D}$  molto più numeroso, ed i *cluster*  $\mathcal{G}$ ,  $\mathcal{J} - \mathcal{L}$  e  $\mathcal{P}$  molto ridotti nella traduzione interconfessionale rispetto alle altre.

Tra le sezioni di testo non sorprende ancora una volta la similarità tra epistolario paolino e le lettere cattoliche (quello che si dirà del *corpus* paolino nel resto di questa analisi si potrà intendere anche delle altre lettere). Buona parte dei *cluster* ospitano occorrenze da tutte o quasi le sezioni di testo ma ci sono comunque differenze macroscopiche.

Il *cluster*  $\mathcal{D}$  è attestato in quasi tutte le sezioni di testo (trascurabile in Paolo) e arriva al 20% delle occorrenze di Giovanni. Al contrario  $\mathcal{F}$ , a parte qualche trascurabile occorrenza in libri storici, profetici e sinottici, si concentra in Paolo e Giovanni. Anche  $\mathcal{C}$  e  $\mathcal{G}$  sono trasversali e si fanno notare in Paolo, Giovanni e,  $\mathcal{C}$  anche nel Pentateuco.  $\mathcal{B}$  è importante (42%) per i sinottici e quasi assente nell'Antico Testamento e sembra riprodotto in scala ridotta da  $\mathcal{A}$ .

$\mathcal{H}$  è un po' meno attestato nei libri storici e nei Sinottici.  $\mathcal{I}$  ed  $\mathcal{N}$  sono concentrati nell'Antico Testamento,  $\mathcal{J}$  e  $\mathcal{K}$  sono concentrati nei Sapienziali (ed in misura minore nei Profetici) ma assenti in Giovanni,  $\mathcal{O}$  è assente solo in Paolo

ma forte soprattutto in libri storici (38%) e Pentateuco, mentre  $\mathcal{Q}$  unisce libri storici e Sinottici.

Come si nota in Figura 4.9, rispetto ad *amore* (cfr. Figura 4.5) i *cluster* sono mediamente più vicini tra loro ed è più difficile individuare gruppi di *cluster* nettamente distinti. Il più facile da notare, quello a prevalenza di occorrenze maiuscole  $\mathcal{B} - \mathcal{F}$ , è comunque più vicino agli altri di quanto lo fossero i due significati distinti di *amore*. Un altro gruppo che si può distinguere può essere  $\mathcal{N} - \mathcal{Q}$  con il dubbio se inserire  $\mathcal{N}$  in questo o in quello centrale. Ma, stando alle distanze, anche il gruppo centrale si potrebbe ulteriormente spezzare.

Guardando la distribuzione in due dimensioni ottenuta con UMAP (in Figura 4.10) si notano i *cluster*  $\mathcal{A} - \mathcal{B}$  separati dagli altri con  $\mathcal{B}$  che sembra avvolgere  $\mathcal{A}$ , a confermare quanto era stato notato rispetto ai due insiemi di occorrenze e alla traduzione dei Testimoni di Geova.

Tra le parole del vocabolario più simili ai centri dei *cluster*, i riferimenti ai doni dello Spirito Santo (*carisma, sapienza, scienza*) sono molto diffusi, specialmente ma non solo nei *cluster*  $\mathcal{A} - \mathcal{F}$ . Altri significati riconducibili al linguaggio cristiano uniscono i *cluster*  $\mathcal{A} - \mathcal{F}$  (quelli con occorrenze con l'iniziale maiuscola insieme a quello caratteristico della traduzione del nuovo mondo): *colomba dello Spirito Santo, paraclito* ed alcuni riferimenti alla Trinità.

Tra i *cluster*  $\mathcal{H} - \mathcal{K}$  (in particolare  $\mathcal{I}$ ) ricorrono i riferimenti al *cuore* ed allo *spirito vitale*.  $\mathcal{M}$  ha numerosi riferimenti a *stato d'animo, forza di volontà*,  $\mathcal{K}$  allo spirito combattivo,  $\mathcal{O}$  ha dei riferimenti al vento: *scirocco e bonaccia*<sup>15</sup>. Per concludere,  $\mathcal{P}$  ed in particolare  $\mathcal{Q}$  hanno riferimenti allo *spirito di defunto*, ed in generale a spiriti maligni: *ridda*<sup>16</sup>, *energumeno*<sup>17</sup>, *posseduto, demonio* ma anche alla *salute mentale*.

Quest'ultimo *cluster*, etichettato con *demonio* in Figura 4.10 sembra un po' distanziato dal gruppo centrale, effettivamente risulta quello con una distanza media dagli altri maggiore.

Esaminando i versetti con occorrenze incluse nel *cluster*  $\mathcal{Q}$  in almeno una traduzione (in Appendice A) si parla nella maggior parte di casi di spiriti scacciati

<sup>15</sup>Ma con il significato figurato di «stato di serenità dello spirito».

<sup>16</sup>Come «ballo degli spiriti infernali».

<sup>17</sup>Come «persona posseduta dal demonio».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	0.000	0.149	0.193	0.241	0.264	0.247	0.179	0.133	0.183	0.250	0.180	0.198	0.262	0.162	0.216	0.218	0.221	-0.2061
B	0.149	0.000	0.093	0.110	0.128	0.111	0.299	0.286	0.341	0.421	0.358	0.363	0.389	0.319	0.345	0.365	0.377	-0.2784
C	0.193	0.093	0.000	0.079	0.125	0.068	0.243	0.192	0.268	0.351	0.328	0.275	0.252	0.204	0.280	0.308	0.365	-0.2266
D	0.241	0.110	0.079	0.000	0.071	0.097	0.289	0.232	0.267	0.353	0.382	0.324	0.348	0.247	0.231	0.236	0.339	-0.2403
E	0.264	0.128	0.125	0.071	0.000	0.087	0.246	0.257	0.304	0.392	0.399	0.357	0.385	0.295	0.331	0.271	0.376	-0.2680
F	0.247	0.111	0.068	0.097	0.087	0.000	0.180	0.210	0.295	0.365	0.365	0.368	0.351	0.305	0.401	0.338	0.420	-0.2630
G	0.179	0.299	0.243	0.289	0.246	0.180	0.000	0.081	0.149	0.176	0.209	0.251	0.262	0.187	0.320	0.222	0.318	-0.2258
H	0.133	0.286	0.192	0.232	0.257	0.210	0.081	0.000	0.089	0.130	0.147	0.147	0.163	0.078	0.201	0.116	0.200	-0.1664
I	0.183	0.341	0.268	0.267	0.304	0.295	0.149	0.089	0.000	0.098	0.181	0.181	0.229	0.105	0.193	0.166	0.271	-0.2075
J	0.250	0.421	0.351	0.353	0.392	0.365	0.176	0.130	0.098	0.000	0.120	0.233	0.272	0.164	0.252	0.208	0.288	-0.2546
K	0.180	0.358	0.328	0.382	0.399	0.365	0.209	0.147	0.181	0.120	0.000	0.177	0.232	0.188	0.298	0.253	0.251	-0.2543
L	0.198	0.363	0.275	0.324	0.357	0.368	0.251	0.147	0.181	0.233	0.177	0.000	0.170	0.135	0.198	0.179	0.204	-0.2349
M	0.262	0.389	0.252	0.348	0.385	0.351	0.262	0.163	0.229	0.272	0.232	0.170	0.000	0.151	0.285	0.293	0.347	-0.2744
N	0.162	0.319	0.204	0.247	0.295	0.305	0.187	0.078	0.105	0.164	0.188	0.135	0.151	0.000	0.107	0.157	0.236	-0.1899
O	0.216	0.345	0.280	0.231	0.331	0.401	0.320	0.201	0.193	0.252	0.298	0.198	0.285	0.107	0.000	0.132	0.216	-0.2504
P	0.218	0.365	0.308	0.236	0.271	0.338	0.222	0.116	0.166	0.208	0.253	0.179	0.293	0.157	0.132	0.000	0.090	-0.2221
Q	0.221	0.377	0.365	0.339	0.376	0.420	0.318	0.200	0.271	0.288	0.251	0.204	0.347	0.236	0.216	0.090	0.000	-0.2824

Distanza media

Figura 4.9: Spirito — distanza del coseno tra i centri dei cluster

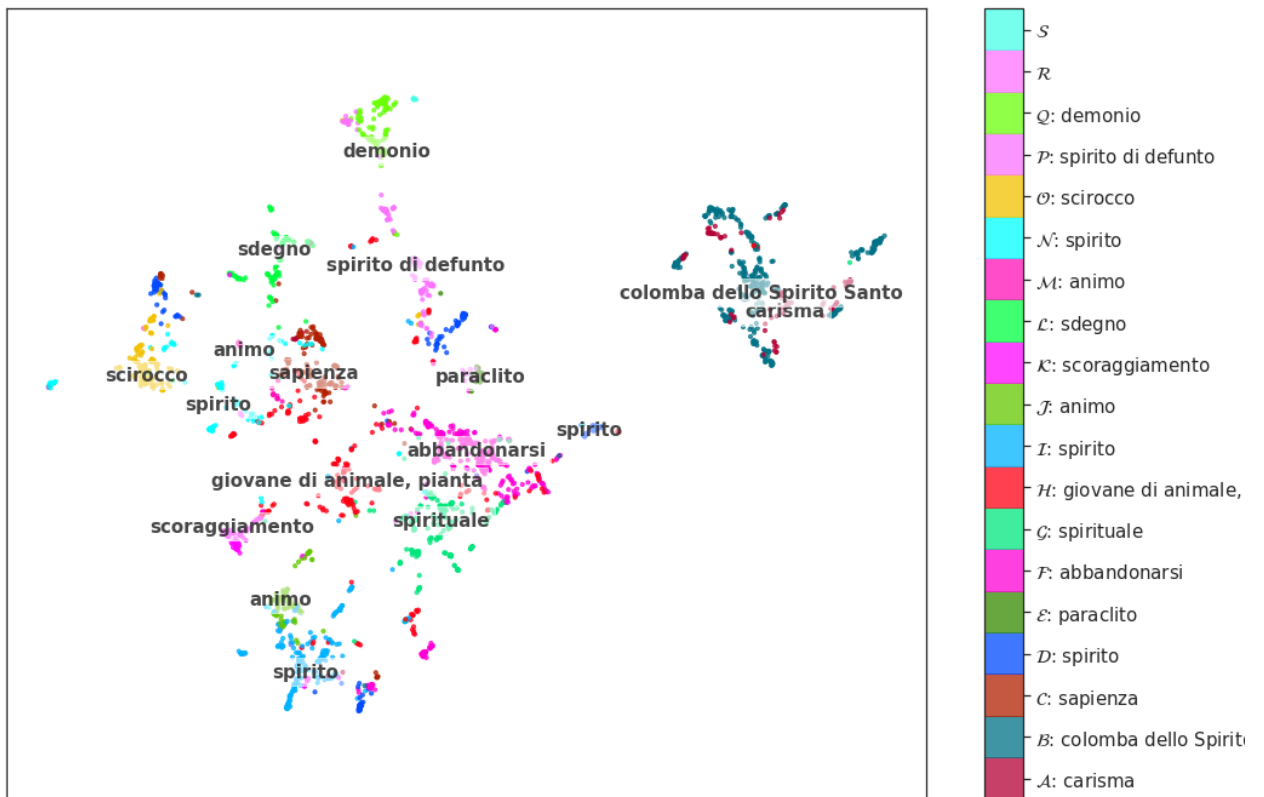


Figura 4.10: Spirito — visualizzazione in uno spazio bidimensionale, cluster etichettati con le definizioni del vocabolario

nel Nuovo Testamento e sono seguiti dagli aggettivi *cattivo*, *impuro*, *maligno*, *immondo* o preceduti dal verbo *scacciare*. Nella maggior parte dei casi il modello ha inserito nello stesso *cluster* le occorrenze parallele, fanno eccezione i versetti dove il termine è reso invece con *demonio*, non inserito nell'analisi, alcune occorrenze di *spirito cattivo* e tutte quelle di *cattivo spirito* che sono state invece inserite nel *cluster*  $\mathcal{P}$ , sei occorrenze di *Spirito impuro*, con la maiuscola perché all'inizio del discorso diretto, tutti inseriti in  $\mathcal{T}$  (uno dei tre *cluster* di dimensioni trascurabili).

Quando *spirito* occorreva due volte nella stessa frase ma solo una volta accompagnato dall'aggettivo, il sistema ha avuto difficoltà ad inserire in  $\mathcal{Q}$  anche l'occorrenza senza aggettivo.

Una ulteriore applicazione di questo modello può essere quella di rispondere alla domanda su quanto un'espressione sia biblicamente fondata. Avendo a disposizione il modello di partenza e quello adattato sulla Bibbia, sottoponendo la stessa frase mascherata ad entrambi i modelli e confrontandone le risposte possiamo scoprire se un'espressione è più o meno probabile nel modello adattato.

In particolare è stata introdotta nella traduzione italiana della terza edizione del Messale romano l'immagine della «rugiada dello Spirito» di cui alcuni non riconoscono la provenienza biblica.

Sottoponendo sia la frase completa che l'espressione ridotta ad entrambi i modelli si ottengono i risultati in Tabella 4.3. L'intenzione originale sarebbe stata quella di provare mascherando alternativamente sia *rugiada* che *Spirito*, ma *rugiada* non appartiene al vocabolario del modello base e mettere tre *token* maschera moltiplicando le probabilità avrebbe portato a numeri troppo bassi per poterli confrontare in modo significativo.

Nessuno dei due modelli inserisce *spirito* tra i candidati più probabili a prendere il posto di *[MASK]* nella frase intera, ma è il modello base a stimare come più probabile (con probabilità comunque molto basse) la parola che realmente prende quel posto nella frase.

Provando invece soltanto con l'immagine della «rugiada dello Spirito», il modello adattato sulla Bibbia suggerisce come prima risposta proprio *spirito*, anche se con la lettera minuscola e, seppure in misura molto minore, fa crescere la probabilità di comparire anche della parola con l'iniziale maiuscola.



Tabella 4.3: Spirito – risultati del riempimento del *token* mascherato

bert-base-it		BibliBERT	
parola	probabilità	parola	probabilità
«Ti preghiamo: santifica questi doni con la rugiada del tuo [MASK].»			
1 cuore	14.80 %	volto	38.64 %
2 amore	11.59 %	cielo	17.20 %
3 sangue	10.95 %	corpo	7.87 %
4 Signore	4.19 %	Dio	6.32 %
5 sorriso	3.52 %	capo	3.69 %
...		...	
spirito	2.67 %	spirito	1.05 %
Spirito	1.80 %	Spirito	0.02 %
«La rugiada dello [MASK].»			
1 zucchero	10.07 %	spirito	15.02 %
2 zolfo	7.16 %	protegge	6.78 %
3 stagno	5.58 %	rugiada	5.13 %
4 champagne	5.14 %	sposo	4.57 %
5 sguardo	4.61 %	splendore	2.97 %
...		...	
spirito	1.16 %		
Spirito	0.50 %	Spirito	0.69%

Una curiosità che non si può evitare di notare è che in quest'ultimo caso la seconda e la terza proposta sono in realtà scorrette in italiano: *protegge* è un verbo e *rugiada* è un sostantivo ma femminile, difficile immaginare uno dei due che segue la preposizione articolata *dello*. Questo non è comunque un problema così raro nemmeno con il modello base, quando gli si presentano poche parole.

Guardando questi risultati si può dire che, per quanto ci siano immagini più ricorrenti nella Bibbia per indicare la presenza del Signore (il suo *volto* è sicuramente una di queste), non si può non riconoscere che l'immagine della *rugiada dello spirito* sia in qualche modo biblicamente fondata e che la lettura della Bibbia la possa far sorgere più facilmente.

## Risultati principali

Il numero di *cluster* e le distanze del coseno tra i centroidi confermano la poliedricità dei significati di *spirito* come lo fa la varietà dei risultati nel confronto con il dizionario italiano. L'esperimento con il *masked language modeling* ha dimostrato che con il *fine-tuning* effettuato il modello è stato capace di apprendere anche immagini bibliche generate da accostamenti di parole che nell'uso *standard* della lingua difficilmente si troverebbero insieme.

## 4.4 Segno

Si chiama segno ciò che, per rapporto naturale o per convenzione, fa conoscere il pensiero o la volontà di una persona, l'esistenza o la veridicità di una cosa. (...)

Adattandosi alla nostra natura, anche Dio, per salvare gli uomini, fa loro dei segni ( **אֵימֹתָא** ; *σημεῖα* ). Spesso vengono chiamati prodigi simbolici ( **מִפְתֻּיִם** ; *τέρατα* ) e meraviglie ( **נִפְלְאוֹת** ; *θαυμάσια* ), perché Dio « significa » la propria potenza e il proprio amore innanzitutto attraverso la trascendenza della sua azione salvifica. Per questo i miracoli, per la loro efficacia e il loro carattere straordinario, occupano un posto privilegiato tra i segni divini relativi alla storia della salvezza.<sup>18</sup>

Questo uso di *segno* come miracolo comune nella Bibbia non è tra i significati contemplati dal dizionario scaricato, e può essere quindi interessante osservare il comportamento del modello in questo caso.

Partendo dall'uso della parola nelle cinque traduzioni, vengono individuati 17 *cluster*, di cui ben 6 (più di un terzo) hanno dimensioni trascurabili. Il numero totale è in linea con quanto ottenuto con le altre parole, ma i piccoli gruppi sparsi sono molti di più.

Anche la *Concordanza pastorale della Bibbia* distingue *segno* in due significati: «marchio o gesto» e «segni di Dio», dividendo ciascuno dei due in sei sfumature o contesti distinti. Questa volta il numero di *cluster* considerati sembra abbastanza

<sup>18</sup>Ternant 1976, pp. 1173–1174.

in linea con quello di questa concordanza. Rimane da chiedersi se ci sia una qualche corrispondenza tra questi significati.

Guardando ai sei *cluster* piccoli, si nota che rappresentano usi particolari di *segno* in locuzioni o espressioni idiomatiche come si può vedere negli esempi in Tabella 4.4. Le altre occorrenze incluse in ciascuno di questi *cluster* ripropongono la stessa espressione non usata mai da tutti i traduttori nello stesso versetto ma, in alcuni casi, usata da traduttori diversi in versetti diversi. In nessuno di questi casi *segno* è usato con il significato di miracolo che ci aspettiamo essere il principale nel testo biblico.

Tabella 4.4: Segno — esempi dai *cluster* di dimensioni ridotte ( $\mathcal{L} - \mathcal{Q}$ )

$\mathcal{L}$	2Mac 14, 43	Non avendo però portato a <b>segno</b> il colpo per la fretta ...
$\mathcal{M}$	Zc 2, 4	... Le corna hanno disperso Giuda a tal <b>segno</b> che nessuno osa più alzare la testa ...
$\mathcal{N}$	Gal 3, 3	... dopo aver cominciato nel <b>segno</b> dello Spirito, ora volete finire nel <b>segno</b> della carne?
$\mathcal{O}$	Is 13, 2	... chiamateli a gran voce, fate <b>segno</b> con la mano, ed entrino nelle porte dei principi!
$\mathcal{P}$	Pr 22, 28	Non spostare un antico <b>segno</b> di confine, stabilito ...
$\mathcal{Q}$	Gdc 18, 4	Egli raccontò la sua storia per filo e per <b>segno</b> e concluse: ...



Figura 4.11: Segno — distanza del coseno tra i centri dei *cluster*

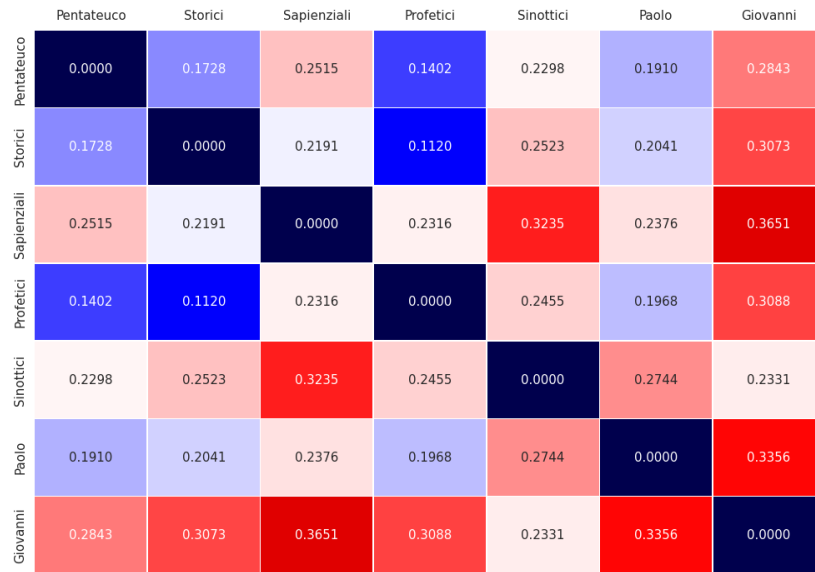
Confrontando le distanze del coseno tra i centri degli 11 più grandi (in Figura 4.11), si nota subito il  $\mathcal{K}$  con una distanza media superiore allo 0.5. Questa differenza è confermata dal confronto con il vocabolario: è l'unico a non avere *segno* nel senso di «indicazione evidente di qualcosa» come parola più vicina al centro e questa definizione non appare nemmeno tra le più simili. Il baricentro si trova invece vicino ad esempio a *baldoria* «falò in segno di festa» ed *umiliare* «prostrarsi in segno di riverenza». Tra gli esempi tratti dal vocabolario seleziona *segno* nella frase «quel pugile, prima della gara, mostrò al suo avversario, il dorso della mano in **segno** di sfida». Guardando ai versetti inclusi nel *cluster* (in Tabella 4.5) si nota la ricorrenza della locuzione *in segno di*. I versetti non hanno *segno* in più di una o due traduzioni e sono sparsi per tutto il testo. Spesso è un'aggiunta al testo originale per rendere esplicito il significato di un gesto o di un'usanza.

Tabella 4.5: Segno — versetti in cui occorrono le parole del *cluster*  $\mathcal{K}$ 

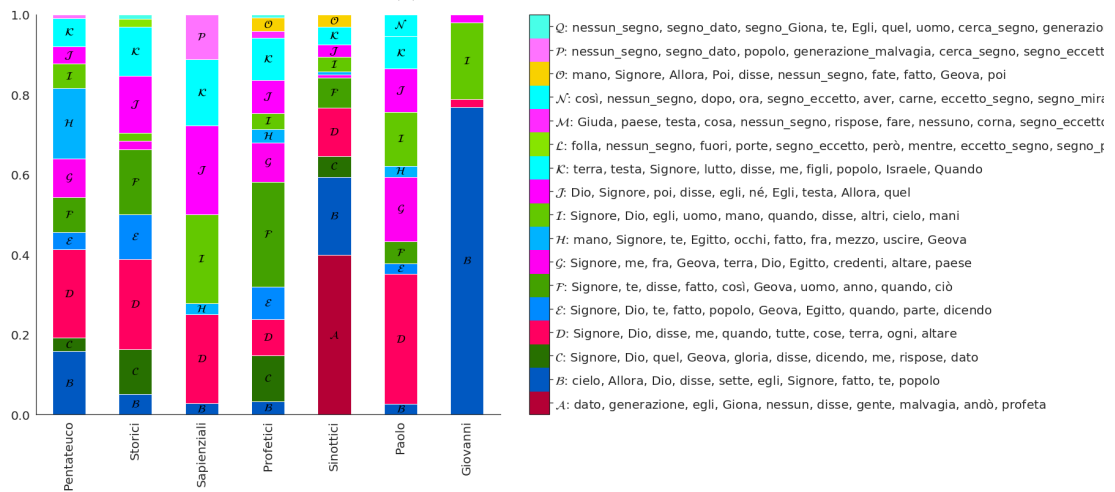
Es	28, 2	<i>NR06</i>	... farai dei paramenti sacri, in <b>segno</b> di dignità e di gloria.
Es	28, 40	<i>NR06</i>	... farai delle mitre in <b>segno</b> di dignità e come ornamento.
Lv	7, 12	<i>TNW17</i>	Se lo presenta in <b>segno</b> di ringraziamento, ...
Lv	7, 13	<i>TNW17</i>	... sacrificio ... offerto in <b>segno</b> di ringraziamento.
Lv	7, 15	<i>TNW17</i>	... sacrifici ... offerti in <b>segno</b> di ringraziamento ...
Lv	10, 6	<i>TILC</i>	... non vi stracciate le vesti in <b>segno</b> di lutto: ...
Lv	19, 28	<i>TILC</i>	Non fatevi incisioni sul corpo in <b>segno</b> di lutto; ...
Nm	14, 5	<i>TILC</i>	... si gettarono a terra ... in <b>segno</b> di disperazione.
2Sam	1, 2	<i>TILC</i>	... la testa coperta di polvere in <b>segno</b> di lutto. ...
2Sam	15, 32	<i>TILC</i>	... il capo coperto di polvere in <b>segno</b> di dolore.
1Cr	10, 12	<i>TILC</i>	... in <b>segno</b> di lutto, digiunarono per sette giorni.
1Cr	15, 16	<i>NR06</i>	... per cantare in <b>segno</b> di gioia.
		<i>TILC</i>	... dovevano cantare e suonare forte in <b>segno</b> di festa ...
Gdt	4, 10	<i>TILC</i>	... si vestirono di panno ruvido in <b>segno</b> di penitenza.
Gdt	10, 23	<i>TILC</i>	... si gettò in ginocchio ... in <b>segno</b> di sottomissione, ...
Gdt	14, 16	<i>TILC</i>	... si strappò i vestiti in <b>segno</b> di disperazione.
1Mac	2, 14	<i>TILC</i>	... si coprirono di sacco in <b>segno</b> di lutto e piansero a lungo.
1Mac	3, 47	<i>TILC</i>	Quel giorno in <b>segno</b> di lutto digiunarono ...
1Mac	4, 39	<i>TILC</i>	Allora in <b>segno</b> di lutto si strapparono le vesti, ...

2Mac	10, 25	TILC	... In <b>segno</b> di tristezza, si cosparsero il capo di cenere ...
Gb	1, 20	TILC	... si rase i capelli in <b>segno</b> di lutto. ...
Gb	37, 13	CEI08	Egli le manda o per castigo del mondo o in <b>segno</b> di bontà.
		CEI74	Le manda o per castigo della terra o in <b>segno</b> di bontà.
Sal	26, 6	TILC	In <b>segno</b> d'innocenza, lavo le mie mani. ...
Sal	73, 13	TILC	... e in <b>segno</b> di innocenza ho lavato le mie mani,
Sir	13, 7	TILC	... scuoterà la testa in <b>segno</b> di disprezzo.
Is	15, 2	TILC	... si sono rasati le teste e le barbe in <b>segno</b> di lutto.
Is	32, 12	TNW17	Battetevi il petto in <b>segno</b> di dolore ...
Ger	16, 6	TILC	... si taglierà i capelli in <b>segno</b> di lutto.
Ger	47, 5	TILC	Gli abitanti di Gaza si sono rasati i capelli in <b>segno</b> di lutto, ... si faranno ferite in <b>segno</b> di lutto?
Ger	48, 37	TILC	In <b>segno</b> di lutto si sono tagliati tutti i capelli ...
Lam	2, 15	TNW17	... ti battono le mani in <b>segno</b> di disprezzo. ...
Ez	21, 12	TNW17	... Pertanto battiti la coscia in <b>segno</b> di dolore.
Ez	26, 16	TILC	... in <b>segno</b> di lutto, si leveranno i loro mantelli, ...
Mi	1, 10	TILC	... rotolatevi nella polvere in <b>segno</b> di dolore.
Mi	1, 16	TILC	... in <b>segno</b> di lutto per i vostri figli, ...tagliatevi i capelli, ...
Mi	4, 14	TILC	Ora fatevi pure dei tagli in <b>segno</b> di lutto, ...
Sof	2, 15	TILC	... in <b>segno</b> di disprezzo fischierà e farà gesti con le mani.
Mt	11, 17	TNW17	... ma non vi siete battuti il petto in <b>segno</b> di dolore ! '
Mt	27, 39	TILC	... scuotevano la testa in <b>segno</b> di disprezzo, lo insultavano
Mc	15, 29	TILC	... scuotevano la testa in <b>segno</b> di disprezzo, ...
Lc	8, 52	TNW17	... piangevano e si battevano il petto in <b>segno</b> di dolore, ...
Lc	23, 27	TNW17	... si battevano il petto in <b>segno</b> di dolore e piangevano ...
At	18, 6	TILC	... Allora Paolo si stracciò le vesti in <b>segno</b> di sdegno ...
Gal	2, 9	CEI08	... diedero ... la destra in <b>segno</b> di comunione, ...
		CEI74	... diedero ... la loro destra in <b>segno</b> di comunione, ...
		NR06	... diedero ... la mano destra in <b>segno</b> di comunione ...

Confrontando le sezioni di testo (in Figura 4.12) si nota che il *cluster*  $\mathcal{A}$  è presente solo nei sinottici, dove è molto rilevante, con più di un terzo delle occorrenze totali. L'opera giovannea si distingue dal significato  $\mathcal{B}$  in tre quarti dei casi ed il resto assorbito quasi tutto da  $\mathcal{I}$ . Quest'ultimo significato è presente in ma-



(a) distanza del trasporto ottimo



(b) distribuzione delle occorrenze nei cluster

Figura 4.12: Segno — confronto tra le principali sezioni della Bibbia

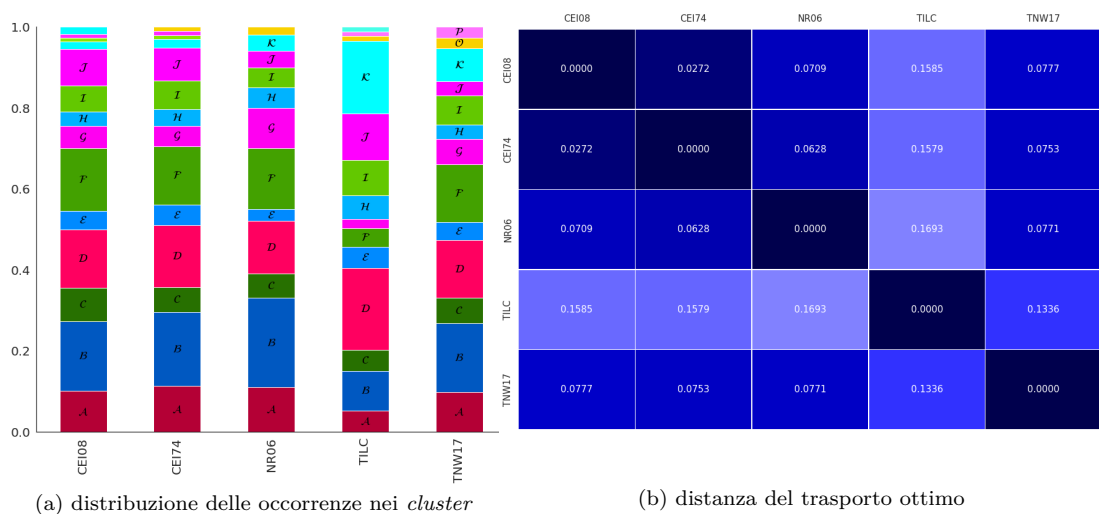


Figura 4.13: Segno — confronto tra le traduzioni

niera significativa anche in Paolo e nei Sapienziali.  $\mathcal{H}$ , invece, include soprattutto parole provenienti dal Pentateuco.

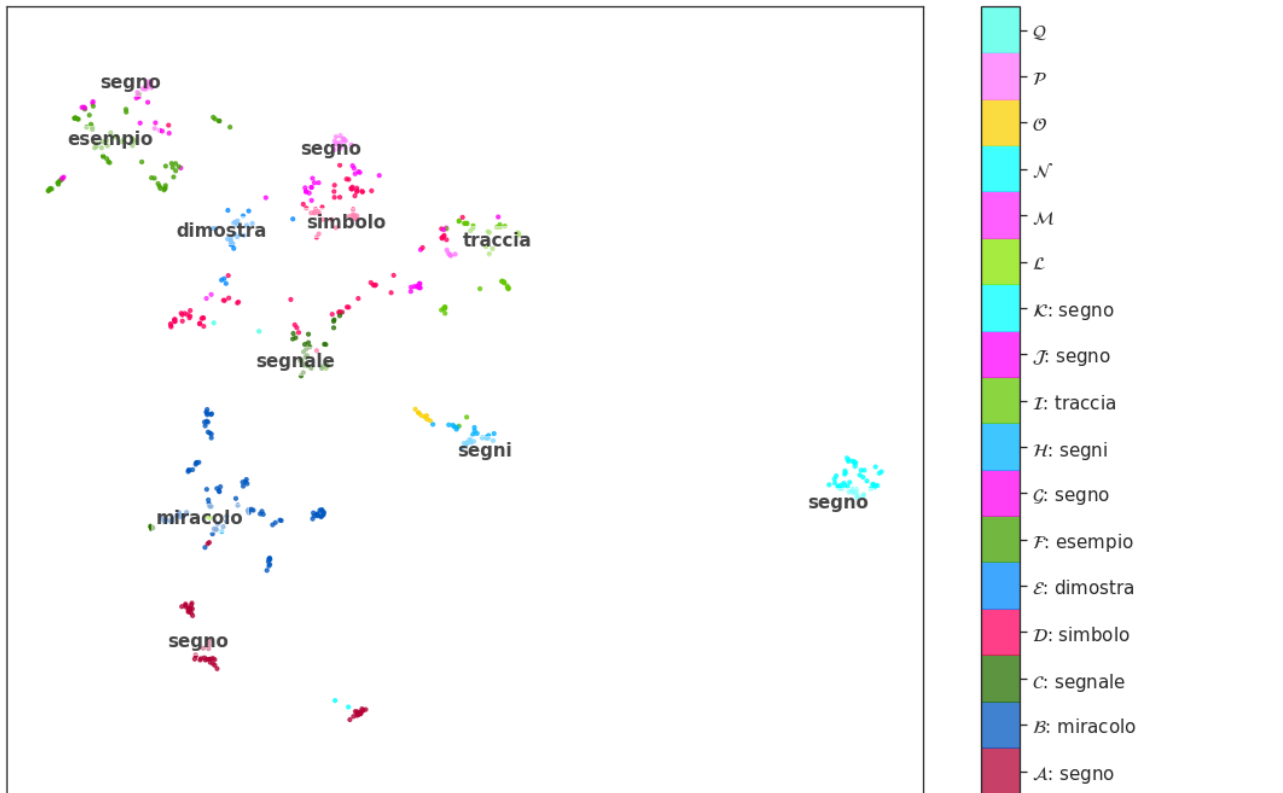
Lo stesso emerge anche dalle distanze del trasporto ottimo in cui opera giovannea, libri sapienziali e Sinottici sono più distanti tra loro e dalle altre sezioni (che invece sono relativamente ravvicinate).

Guardando i versetti di provenienza delle occorrenze di *segno* incluse nel *cluster*  $\mathcal{A}$  si nota che provengono da Marco 8, 12 e i passi paralleli nei vangeli sinottici (Mt 12, 39; 16, 4; Lc 11, 29), in tutte le traduzioni e con due o tre occorrenze a versetto. Il versetto nella versione lucana è reso nella traduzione CEI del 2008: «Mentre le folle si accalcavano, Gesù cominciò a dire: Questa generazione è una generazione malvagia; essa cerca un **segno**, ma non le sarà dato alcun **segno**, se non il **segno** di Giona.». Come esplicita la traduzione interconfessionale nella prima occorrenza di ognuno di questi versetti, si parla di un «segno miracoloso» anche se questa esplicitazione non ha aiutato il sistema, che nella metà dei versetti ha incluso la parola in  $\mathcal{B}$ . Se  $\mathcal{K}$  fornisce informazioni utili dal punto di vista stilistico,  $\mathcal{A}$  è un ottimo esempio di successo del sistema come sussidio all'esegesi.

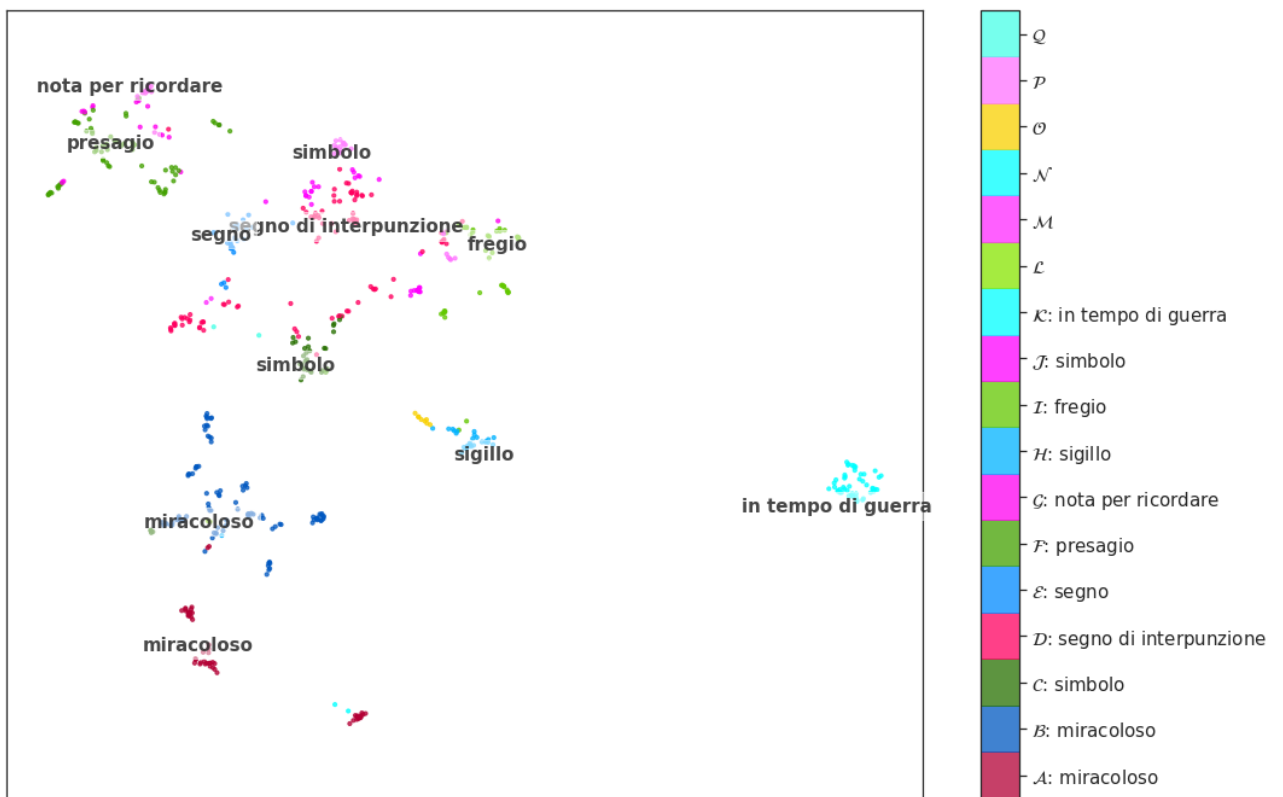
Questo versetto spiega infatti la differenza tra vangeli sinottici ed opera giovannea, se nei primi c'è un'insistenza minore sui miracoli per riconoscere il Cristo, il *segno* in Giovanni è una categoria teologica fondamentale. La distanza interna all'opera lucana (0.3437 nel trasporto ottimo tra Vangelo e Atti) conferma invece la maggiore insistenza nel secondo libro sui *segni* come indicazione di continuità tra l'opera di Gesù e quella della prima comunità cristiana.

Confrontando le traduzioni, è quella interconfessionale che si distingue per una distanza del trasporto ottimo da ciascuna delle altre traduzioni più che doppia rispetto ai confronti che non la includono (in Figura 4.13b).

Anche tra le numerosità nei *cluster* (in Figura 4.13a) si vede immediatamente la differenza, con alcuni *cluster* molto meno numerosi ( $\mathcal{A} - \mathcal{B}$ ,  $\mathcal{F} - \mathcal{G}$ ) ed altri più grandi ( $\mathcal{D}$ ,  $\mathcal{H} - \mathcal{K}$ ) tra cui risalta in modo particolare il  $\mathcal{K}$  (cosa che non stupisce dato lo stile traduttivo e quanto è stato già evidenziato sopra su questo *cluster*). Le altre traduzioni sono invece molto simili, con solo il  $\mathcal{G}$  un po' più numeroso nella Nuova Riveduta, il  $\mathcal{J}$  nelle traduzioni CEI e di nuovo il  $\mathcal{K}$  in quella del nuovo mondo.



(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 4.14: Segno — visualizzazione in uno spazio bidimensionale



Il significato di miracolo, associato a  $\mathcal{B}$  in Figura 4.14a, è collegato anche ai cluster  $\mathcal{A} - \mathcal{I}$  con *miracoloso*, *auspicio*, *presagio*, *segno premonitore* tra le definizioni del vocabolario. Questo fornisce una qualche indicazione alla domanda rimasta in sospeso sul confronto con i contesti individuati dalla *Concordanza pastorale della Bibbia*: il modello sembra aver riunito in molti meno cluster gli usi di *segno* corrispondenti all'uso dell'italiano *standard*, mentre ha ulteriormente suddiviso quelli relativi all'uso prettamente biblico dei segni divini.

### Risultati principali

Oltre a confermare il significato particolare che la parola *segno* assume nella Bibbia, l'analisi ha permesso di individuare i testi in cui questo uso particolare è più rilevante. Questo anche in presenza un discreto numero di usi diversi, anche in espressioni idiomatiche. Inoltre, la distanza del trasporto ottimo tra le traduzioni, nonostante nessun valore particolarmente elevato, si è dimostrata capace di essere un aiuto nell'individuare parole che vengono utilizzate in modo non *standard* dal linguaggio biblico: trovarla sensibilmente più alta solo nel confronto con la traduzione in lingua corrente può essere un segno rivelatore.

## 4.5 Davide

La figura di David, come uomo e come re, ha un rilievo tale da rimanere per sempre per Israele come tipo del Messia che deve nascere dalla sua stirpe. A partire da David, l'alleanza con il popolo passa ormai attraverso il re, come ricorda Ben Sira al termine del ritratto che ne delinea (Sir 47, 2-11). Quindi il trono di Israele è il trono di David (Is 9, 6; Lc 1, 32); le sue vittorie annunziano quella che il Messia, pieno dello spirito che riposò sul figlio di Jesse (1 Sam 16, 13; Is 11, 1-9), riporterà sull'ingiustizia. Con la vittoria della sua risurrezione, Gesù adempirà le promesse fatte a David (At 13, 32-37) e darà alla storia il suo senso (Ap 5, 5). Come ha fatto il personaggio di David ad assumere questo posto di elezione nella storia della salvezza?<sup>19</sup>

---

<sup>19</sup>Motte 1976, pp. 249-251.

Una buona parte della Bibbia è composta di testi narrativi che raccontano gli avvenimenti attorno a luoghi e personaggi di cui si riportano i nomi. Il nome, nella cultura del vicino oriente antico, esprime l'identità di chi lo porta in modo molto più profondo di una semplice etichetta, descrive la sua essenza, il suo scopo nel mondo.<sup>20</sup>

Difficilmente un modello computazionale come questo, specie lavorando su testi tradotti, potrà arrivare a cogliere così tanto ma, vista la rilevanza del tema ed il numero di nomi propri nel testo, può essere interessante vedere cosa è in grado di fare. Molti personaggi scompaiono dal testo velocemente come sono comparsi ed altri hanno nomi difficili da rendere nelle lingue occidentali e danno origine a forme diverse a volte anche all'interno della stessa traduzione. Un nome che mantiene la medesima forma tra le cinque traduzioni in esame e che è citato in diverse sezioni del testo biblico è quello di Davide, su cui è ricaduta la scelta in questo esperimento.

Guardando alle cinque traduzioni prese in esame, salta immediatamente agli occhi l'assenza di caratteristiche specifiche o differenze sensibili, come si vede in Figura 4.15a e confermano questa impressione le distanze del trasporto ottimo comprese tra 0.006 e 0.027.

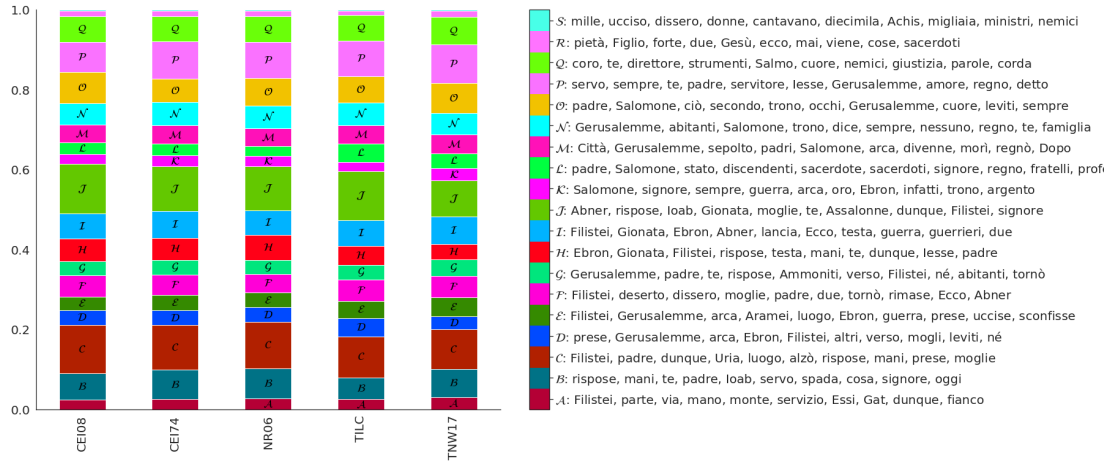
Dal confronto tra le sezioni di testo e tra i libri si ottiene una situazione molto differente. Ovviamente non ci sono occorrenze nel Pentateuco, Davide compare per la prima volta nel primo libro di Samuele ed è citato per la prima volta nell'ottavo libro della Bibbia che racconta la storia della sua antenata Rut.

Nei libri storici, in cui è raccontata la vicenda di Davide e dei suoi successori sul trono di Giuda e Israele appaiono una grande varietà di significati (in Figura 4.15c), nelle altre sezioni di testo, invece, sono al massimo tre i significati principali chiaramente individuabili:  $\mathcal{N}$  e  $\mathcal{P}$  presenti in misura trascurabile anche nei libri storici,  $\mathcal{Q}$  ed  $\mathcal{R}$  di appannaggio esclusivo o quasi del libro dei Salmi e dei Vangeli sinottici rispettivamente.

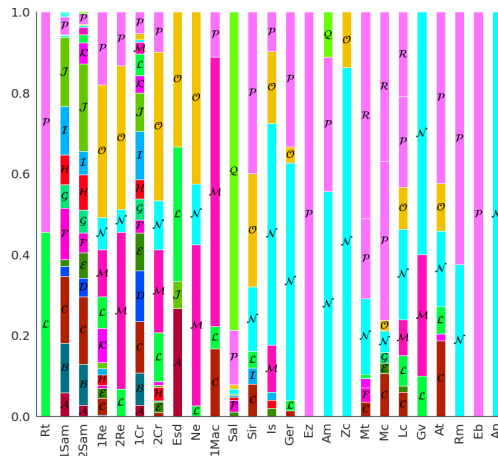
Il *cluster*  $\mathcal{R}$ , presente solo nei Vangeli sinottici, ha tra le parole più rilevanti quelle co-occorrenti nei versetti *Gesù* e *Figlio*, facile intuire che si parli delle profezie messianiche sul figlio di Davide associate a Gesù. Il confronto tra le

---

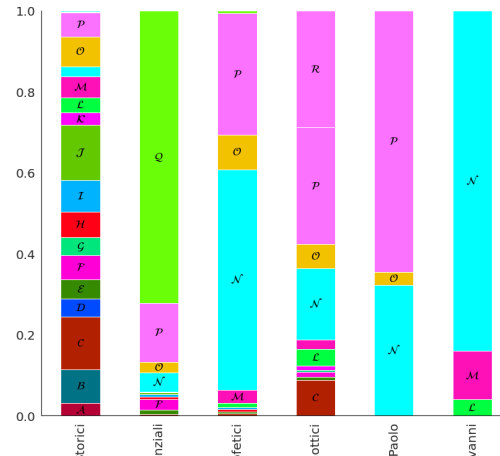
<sup>20</sup>Cfr. Cazelles e Dupont 1976, p. 764.



(a) per traduzione



(b) per libro



(c) per sezione di testo

Figura 4.15: Davide — distribuzione delle occorrenze nei cluster

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ø	P	Q	R	
A	0.000	0.122	0.101	0.126	0.113	0.102	0.125	0.133	0.118	0.127	0.190	0.203	0.323	0.267	0.213	0.180	0.323	0.381	-0.1851
B	-0.122	0.000	0.028	0.064	0.075	0.090	0.101	0.155	0.125	0.094	0.146	0.211	0.333	0.292	0.244	0.185	0.294	0.387	-0.1733
C	-0.101	0.028	0.000	0.036	0.046	0.056	0.085	0.130	0.096	0.092	0.136	0.198	0.315	0.267	0.224	0.151	0.293	0.371	-0.1544
D	-0.126	0.064	0.036	0.000	0.049	0.105	0.113	0.182	0.148	0.145	0.164	0.214	0.326	0.294	0.260	0.205	0.341	0.420	-0.1878
E	-0.113	0.075	0.046	0.049	0.000	0.056	0.086	0.130	0.107	0.107	0.128	0.168	0.288	0.246	0.203	0.156	0.304	0.374	-0.1550
F	-0.102	0.090	0.056	0.105	0.056	0.000	0.056	0.077	0.088	0.097	0.178	0.197	0.325	0.279	0.233	0.159	0.317	0.355	-0.1628
G	-0.125	0.101	0.085	0.113	0.086	0.056	0.000	0.089	0.140	0.126	0.184	0.175	0.326	0.296	0.246	0.206	0.334	0.387	-0.1809
H	-0.133	0.155	0.130	0.182	0.130	0.077	0.089	0.000	0.046	0.057	0.163	0.109	0.235	0.186	0.162	0.115	0.272	0.280	-0.1482
I	-0.118	0.125	0.096	0.148	0.107	0.088	0.140	0.046	0.000	0.033	0.156	0.156	0.253	0.209	0.197	0.129	0.275	0.331	-0.1533
J	-0.127	0.094	0.092	0.145	0.107	0.097	0.126	0.057	0.033	0.000	0.115	0.112	0.212	0.175	0.172	0.129	0.237	0.287	-0.1364
K	-0.190	0.146	0.136	0.164	0.128	0.178	0.184	0.163	0.156	0.115	0.000	0.086	0.208	0.183	0.168	0.142	0.292	0.299	-0.1728
L	-0.203	0.211	0.198	0.214	0.168	0.197	0.175	0.109	0.156	0.112	0.086	0.000	0.111	0.074	0.076	0.106	0.227	0.202	-0.1545
M	-0.323	0.333	0.315	0.326	0.288	0.325	0.326	0.235	0.253	0.212	0.208	0.111	0.000	0.080	0.167	0.202	0.275	0.210	-0.2463
N	-0.267	0.292	0.267	0.294	0.246	0.279	0.296	0.186	0.209	0.175	0.183	0.074	0.080	0.000	0.081	0.099	0.207	0.140	-0.1985
Ø	-0.213	0.244	0.224	0.260	0.203	0.233	0.246	0.162	0.197	0.172	0.168	0.076	0.167	0.081	0.000	0.054	0.224	0.232	-0.1856
P	-0.180	0.185	0.151	0.205	0.156	0.159	0.206	0.115	0.129	0.129	0.142	0.106	0.202	0.099	0.054	0.000	0.190	0.210	-0.1539
Q	-0.323	0.294	0.293	0.341	0.304	0.317	0.334	0.272	0.275	0.237	0.292	0.227	0.275	0.207	0.224	0.190	0.000	0.242	-0.2733
R	-0.381	0.387	0.371	0.420	0.374	0.355	0.387	0.280	0.331	0.287	0.299	0.202	0.210	0.140	0.232	0.210	0.242	0.000	-0.3005

Distanza media

Figura 4.16: Davide — distanza del coseno tra i centri dei cluster

frequenze di occorrenza confermano la maggiore rilevanza del tema in Matteo, rivolto principalmente ad ascoltatori di origine giudaica, rispetto a Marco e Luca.

Dalla distanza del trasporto ottimo tra le sezioni non emerge granché di significativo, è guardando ai libri che escono fuori alcune indicazioni più chiare. I libri di Rut, Esdra, Salmi, Amos e la lettera ai Romani sono piuttosto distanti tra loro e da tutti gli altri.

Nel libro dei Salmi, più di tre quarti delle occorrenze sono in  $\mathcal{Q}$ , che racchiude quasi tutte le intestazioni dei Salmi che riportano nel primo versetto l'attribuzione a Davide come in «Al maestro del coro. Per strumenti a corda. Salmo. Di Davide.» (Sal 4, 1 CEI08). In questo *cluster* sono state inserite occorrenze da altri due versetti in cui si parla di Davide come musicista e cantore (2Sam 23, 1; Am 6, 5). Alcune delle occorrenze simili sono state inserite però in  $\mathcal{P}$ , spesso quando vicino all'espressione «servo del Signore» o in  $\mathcal{D}$  (in alcune traduzioni di Sal 7, 1). Guardando la Figura 4.17, si nota un gruppo di parole inserite in  $\mathcal{P}$  ma circondate dalle occorrenze inserite in  $\mathcal{Q}$ .

In Figura 4.15b si nota come Ezechiele ed Ebrei citino Davide soltanto con il significato  $\mathcal{P}$ , mentre il problema del trasporto ottimo li ponga ad una distanza non trascurabile di 0.23. Guardando a questi versetti in Tabella 4.6, Ebrei ed Ezechiele non sembrano avere molto in comune.

Tabella 4.6: Davide — versetti in cui occorrono le parole del *cluster*  $\mathcal{P}$   
(Ezechiele ed Ebrei, traduzione CEI08)

Ez 34, 23	Susciterò per loro un pastore che le pascerà, il mio servo <b>Davide</b> . Egli le condurrà al pascolo, sarà il loro pastore.
Ez 34, 24	Io, il Signore, sarò il loro Dio, e il mio servo <b>Davide</b> sarà principe in mezzo a loro: io, il Signore, ho parlato.
Ez 37, 24	Il mio servo <b>Davide</b> regnerà su di loro e vi sarà un unico pastore per tutti; seguiranno le mie norme, osserveranno le mie leggi e le metteranno in pratica.
Ez 37, 25	Abiteranno nella terra che ho dato al mio servo Giacobbe. In quella terra su cui abitarono i loro padri, abiteranno essi, i loro figli e i figli dei loro figli, per sempre; il mio servo <b>Davide</b> sarà loro re per sempre.

---

---

Eb 4, 7	Dio fissa di nuovo un giorno, oggi, dicendo mediante <b>Davide</b> , dopo tanto tempo: Oggi, se udite la sua voce, non indurite i vostri cuori!
Eb 11, 32	E che dirò ancora? Mi mancherebbe il tempo se volessi narrare di Ge- deone, di Barak, di Sansone, di Iefte, di <b>Davide</b> , di Samuele e dei profeti;

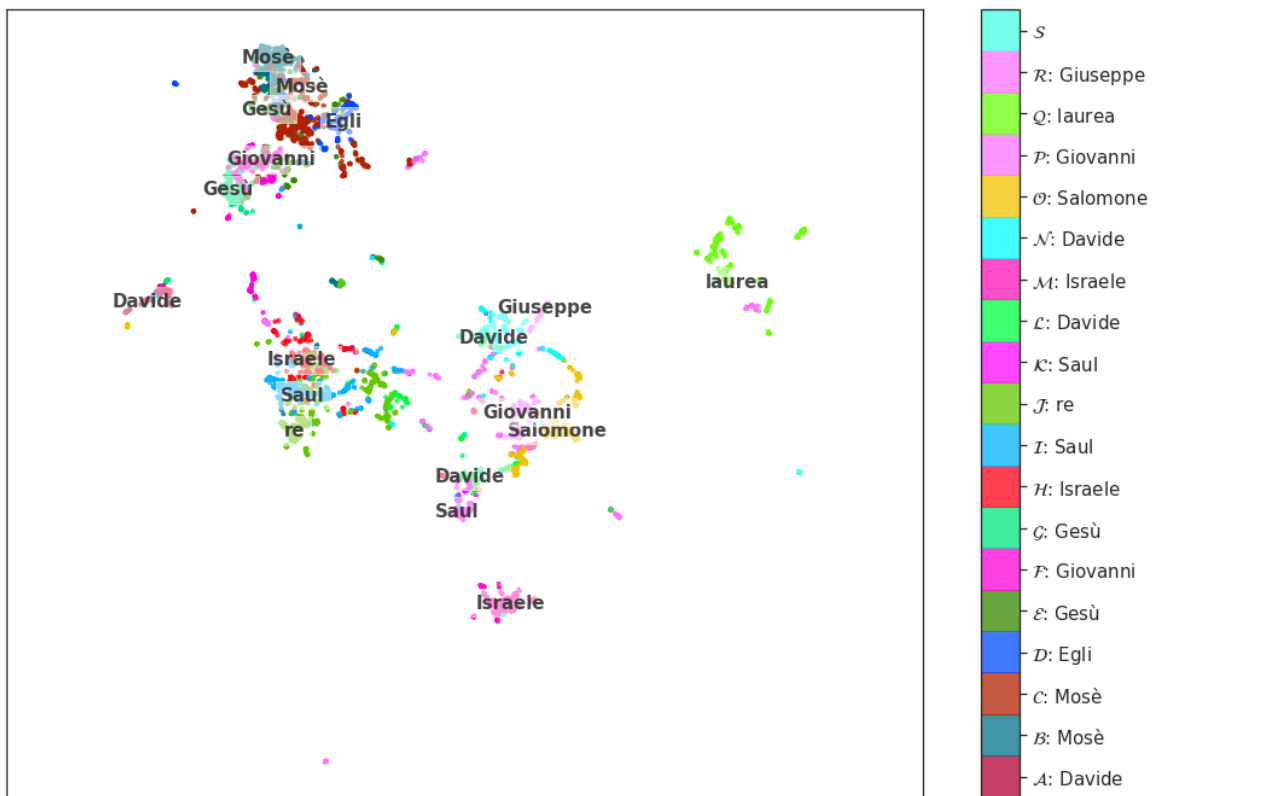
---

Tra le somiglianze, infine, non stupisce la vicinanza del primo libro delle Cronache con i due libri di Samuele, trattano infatti dell'ascesa e del regno di Davide, né quella del secondo con i due libri dei Re, i quali trattano del regno dei suoi successori. Questi ultimi sono particolarmente vicini ai profeti Isaia e Geremia, la cui azione si svolge al tempo del regno di Israele. Questo gruppo di libri storici e profetici è quello a cui si avvicinano di più gli altri libri. Un'altra conferma arriva dal libro del Siracide, che ospita una sezione che riassume le vicende di alcuni personaggi chiave dell'Antico Testamento: il significato di Davide in questo libro si rivela simile ai libri che raccontano la sua storia.

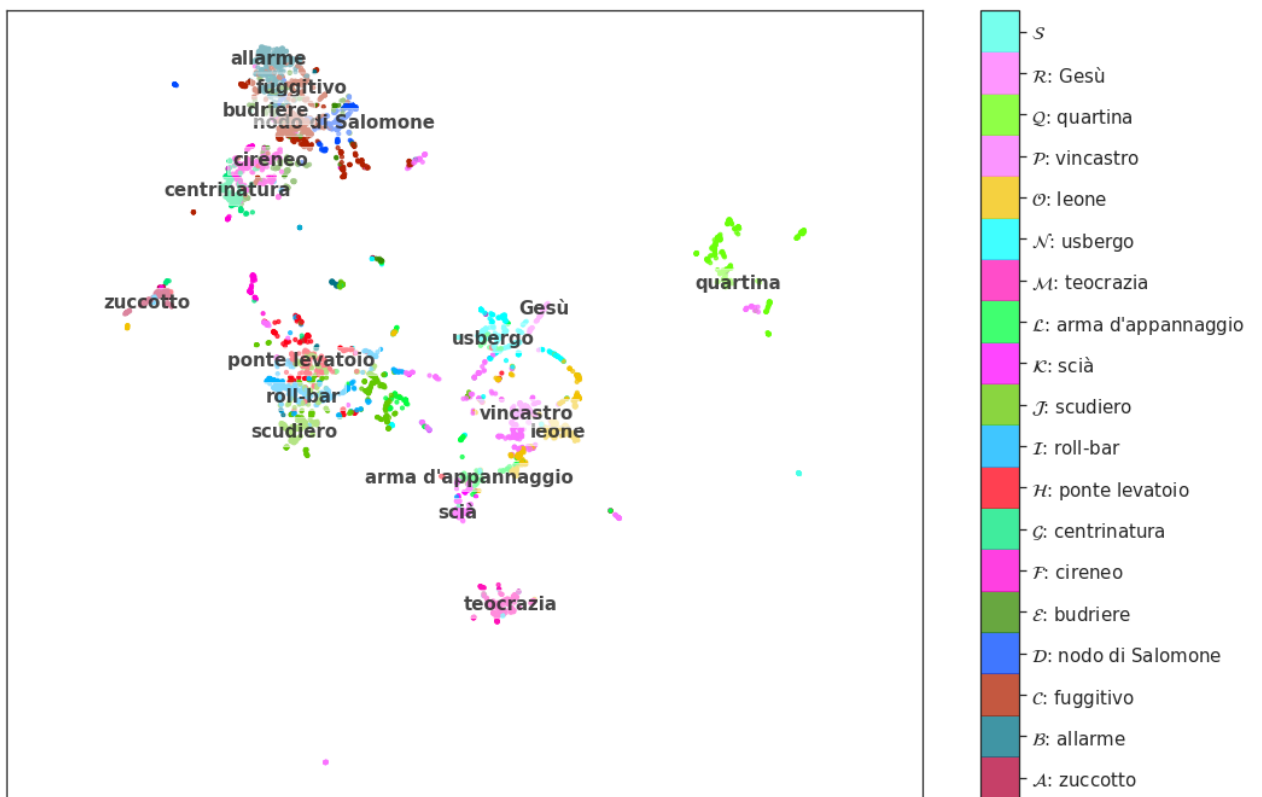
Guardando le distanze del coseno tra i centri dei *cluster* (in Figura 4.16), si distinguono i già visti  $\mathcal{Q}$  ed  $\mathcal{R}$ , insieme ad  $\mathcal{M}$ , come mediamente più distanti dagli altri, in particolare da  $\mathcal{A} - \mathcal{G}$ . Si notano anche  $\mathcal{P}$ , ed in misura minore  $\mathcal{L}$  ed  $\mathcal{O}$  come non particolarmente distanti da alcun altro *cluster*. Inoltre si notano diversi gruppetti di *cluster* particolarmente ravvicinati tra loro, come conferma la visualizzazione in due dimensioni (in Figura 4.17).

La *Concordanza pastorale della Bibbia* individua un solo significato per Davide con otto sfumature: *storia di Davide*, *promessa fatta a Davide*, *fare il bene come Davide*, *Davide cantore e profeta*, *casa*, *stirpe*, *famiglia di Davide*, *figlio di Davide*, *per amore di Davide* e *Davide, il Messia*. In questo caso i *cluster* individuati sono molti di più delle sfumature di significato, ma la maggior parte di questi *cluster* ne ha più di un altro con una similarità del coseno maggiore di 0.9, probabilmente la soglia impostata si è rivelata troppo alta per permetterne la fusione.

Il confronto con il vocabolario si distingue per similarità del coseno (tra centri dei *cluster* e definizioni) dimezzate rispetto agli altri esempi visti fin qui. Nonostante questo, si ottengono dei risultati significativi: sono numerosi i riferimenti al mondo militare (il racconto delle imprese militari di Davide occupa buona parte



(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 4.17: Davide — visualizzazione in uno spazio bidimensionale

della sua storia), a parole collegate a re, e a personaggi biblici. In  $\mathcal{P}$  si aggiunge un riferimento al mondo pastorale (già evidenziato in Tabella 4.6), Davide stesso era un pastore prima di essere scelto come re e, nella Bibbia, la metafora pastorale è spesso applicata ai capi del popolo.

Il *cluster*  $\mathcal{Q}$  si distingue invece per i riferimenti al mondo musicale e della preghiera, mentre da  $\mathcal{R}$  escono riferimenti a Gesù ed al *figlio di David*. Infine,  $\mathcal{M}$  seleziona alcuni riferimenti a luoghi distanti da Davide nel tempo e nello spazio; andando a guardare le occorrenze incluse in questo *cluster* si nota che in tutti si parla della *Città di Davide*, cioè Gerusalemme. Il confronto è stato in grado di cogliere questo riferimento e, probabilmente, non ha trovato tra le definizioni a disposizione esempi di luogo più calzanti in riferimento a Davide.

Dal confronto con gli esempi del vocabolario (in Figura 4.17a), è sempre  *Davide* la parola più vicina da tre diverse citazioni bibliche, la maggior parte dei successivi più vicini sono altri personaggi biblici (*Mosè, Gesù, Giovanni, Saul, Israele*).

### Risultati principali

Forse non è stata catturata l'essenza di Davide, ma ne è stata ricostruita la storia di pastore, soldato, re e cantore, insieme alla relazione con Gesù ed altri personaggi e luoghi chiave della Bibbia. Questa analisi ha dimostrato le possibilità del sistema realizzato di cogliere gli aspetti chiave legati a personaggi o luoghi nell'applicazione ai nomi propri.

## 4.6 Signore

In un testo tardivo [dell'AT] (Lv 24, 11-16), «il nome» designa Jahve senz'altra precisazione, come farà più tardi il linguaggio rabbinico. Per un rispetto sempre più accentuato, il giudaismo tenderà di fatto a non più osare di pronunciare il nome rivelato all'Horeb. Nella lettura esso sarà sostituito da Dio (*Elohim*) o più spesso da *Adonai*, «mio Signore». Perciò i Giudei che tradurranno i libri sacri dall'ebraico in greco non trasriveranno mai il nome di Jahve ma lo renderanno con Κύριος, Signore.<sup>21</sup>

---

<sup>21</sup>Cazelles e Dupont 1976, p. 766.

Quando il testo ebraico verrà vocalizzato, i masoreti segnaleranno di non pronunciare il nome e trovando le consonanti יהוה (JHWH), inseriranno le vocali della parola con cui viene sostituito nella lettura, facendo così nascere la forma artificiosa *Jehowah* che ha portato al nome di Dio usato dai testimoni di Geova.

I primi cristiani trasferiscono il titolo di Signore al Risorto, e questo viene riflesso dai testi del Nuovo Testamento in cui il titolo è usato dai discepoli per riferirsi a Gesù anche prima della resurrezione.

Nonostante i testi originali per il Nuovo Testamento siano tutti in greco e quindi non si presenti mai il tetragramma sacro JHWH, le traduzioni dei testimoni di Geova<sup>22</sup> usano questo nome anche per tradurre κύριος.

Oltre alla rilevanza dell'uso del termine nella Scrittura, è interessante guardare a questa decisione singolare nella traduzione del nuovo mondo, e per fare un confronto è ragionevole scegliere la Nuova Riveduta per due motivi. Prima di tutto è l'unica altra, tra le traduzioni scelte, che adotta il canone ristretto, inoltre entrambe le traduzioni dichiarano di seguire un paradigma traduttivo più vicino alla letteralità del testo.

Considerando che nella quasi totalità dei casi il tetragramma sacro viene reso con Signore nelle traduzioni cristiane, lasciamo da parte la parola Dio e confrontiamo, nelle due traduzioni indicate, soltanto *Signore* e *Geova*.

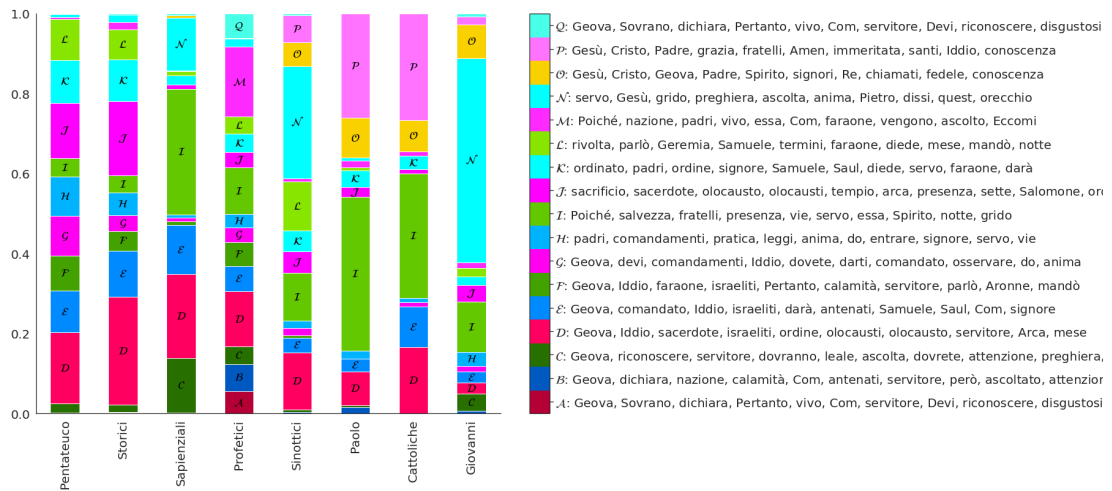
Tra i risultati, la prima curiosità viene dalle parole più rilevanti tra le co-occorrenti:  $\mathcal{A}$  e  $\mathcal{Q}$  hanno selezionato le stesse. Entrambi accolgono termini solo della traduzione del nuovo mondo ma  $\mathcal{A}$  include solo *Signore* e  $\mathcal{Q}$  solo *Geova*. Nonostante i termini vicini più rilevanti siano gli stessi, la distanza del coseno tra i centri dei *cluster* è piuttosto alta (0.545). Guardando alla provenienza delle occorrenze, si nota che portano agli stessi versetti in cui si usa l'espressione *Signore Geova*, tradotta generalmente con Signore Dio nelle altre traduzioni.

Confrontando le Figure 4.18b e 4.18c si nota che non c'è quasi sovrapposizione tra le traduzioni e non ci sono *cluster* che ospitano entrambi i termini. Neanche nel caso appena affrontato, in cui ricorrono nella stessa frase in posizioni adiacenti, il contesto della frase riesce ad essere più forte della differenza tra le due parole appresa dal sistema.

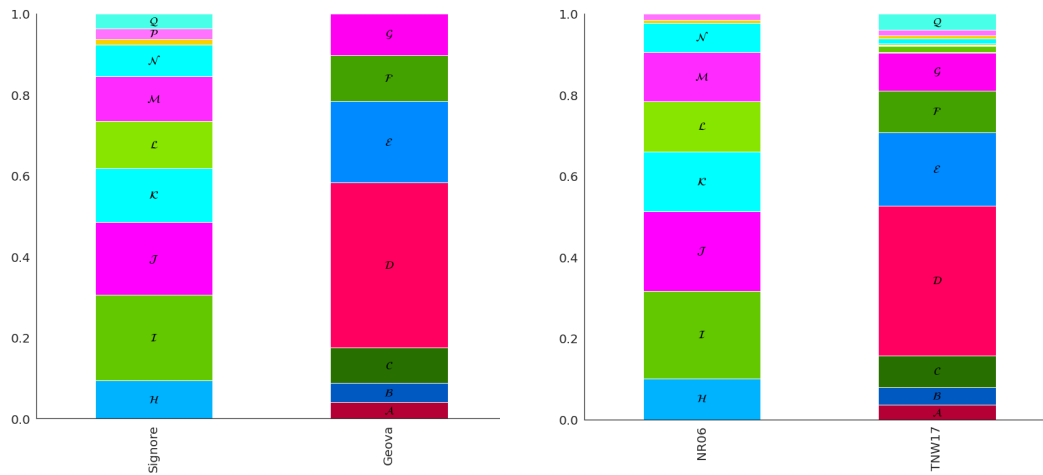
---

<sup>22</sup>Come la traduzione del nuovo mondo presa in esame in questo lavoro.





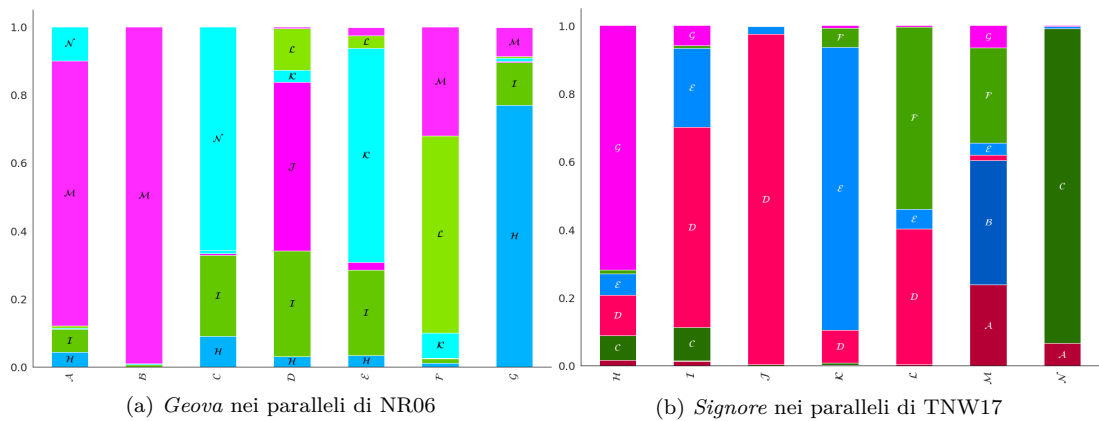
(a) per sezione di testo



(b) per parola

(c) per traduzione

Figura 4.18: Geova — distribuzione delle occorrenze nei cluster



(a) Geova nei paralleli di NR06

(b) Signore nei paralleli di TNW17

Figura 4.19: Confronto tra i cluster di Geova e quelli del termine parallelo in NR06

In Figura 4.19 è riportato il confronto tra la suddivisione in *cluster* di *Geova* nella traduzione del nuovo mondo e delle occorrenze negli stessi versetti di *Signore* nella Nuova Riveduta. Il sistema non ha diviso le parole allo stesso modo. Sia i paralleli di  $\mathcal{A}$  che quelli di  $\mathcal{B}$  confluiscono in  $\mathcal{M}$  che ospita una pari quantità di paralleli di  $\mathcal{F}$ . C'è una buona sovrapposizione (ma non completa) soltanto tra  $\mathcal{C}$  e  $\mathcal{N}$  e tra  $\mathcal{G}$  e  $\mathcal{H}$ .

Guardando alla distanza del trasporto ottimo tra le sezioni di testo (in Figura 4.20), si ottiene l'attesa differenza tra Antico e Nuovo Testamento ma c'è distanza anche tra le sezioni all'interno del Nuovo, le uniche vicine tra loro sono il *corpus* paolino e le lettere cattoliche.

All'interno delle sezioni dell'Antico Testamento non ci sono grandi differenze tra i libri, con l'unica eccezione di Ezechiele che si distingue dagli altri libri profetici.

All'interno del Nuovo Testamento, invece, ci sono maggiori differenze. Colpisce in particolare che il vangelo secondo Marco sia più vicino ad Atti degli Apostoli che agli altri sinottici (e comunque più a Luca che a Matteo). Per lungo tempo è stata sostenuta la dipendenza di Marco da Matteo proprio per la loro similarità ma non stupisce che, essendo più antico, usi meno *Signore* per riferirsi a Gesù, cosa che invece Matteo fa molto spesso.

I *cluster*  $\mathcal{A}$  e  $\mathcal{B}$ , ed ovviamente anche  $\mathcal{M}$  che ospita la maggior parte delle occorrenze di *Signore* corrispondenti nella Nuova Riveduta e  $\mathcal{Q}$  che ospita parole adiacenti a quelle in  $\mathcal{A}$ , sono caratteristici dei libri profetici. Guardando alle distanze del coseno tra i centri (in Figura 4.21), sia  $\mathcal{A} - \mathcal{B}$  che  $\mathcal{M}$  sono ai margini di due gruppi di *cluster* piuttosto vicini tra loro che accolgono occorrenze di *Geova* ( $\mathcal{A} - \mathcal{G}$ ) o di *Signore* ( $\mathcal{H} - \mathcal{M}$ ).

$\mathcal{C}$ , oltre che nel libro dei Salmi, è molto presente anche in Daniele ed Apocalisse che condividono il genere apocalittico. Il *cluster* corrispondente  $\mathcal{N}$  è molto presente anche nei quattro vangeli ed ospita anche un numero non trascurabile di occorrenze di *Signore* nella traduzione del nuovo mondo che dovrebbero essere concentrate in questi libri. Al contrario di  $\mathcal{C}$ , che è ben inserito all'interno del gruppo  $\mathcal{A} - \mathcal{G}$ ,  $\mathcal{N}$  si allontana da quel gruppo di *cluster*, rimanendo più vicino ai *cluster* di *Signore* che a quelli di *Geova*.

	Pentateuco	Storici	Sapientziali	Profetici	Sinottici	Paolo	Cattoliche	Giovanni
Pentateuco	0.0000	0.0496	0.1597	0.1438	0.2524	0.3430	0.2879	0.4007
Storici	0.0496	0.0000	0.1405	0.1284	0.2358	0.3342	0.2782	0.3836
Sapientziali	0.1597	0.1405	0.0000	0.1526	0.2356	0.3218	0.2586	0.3392
Profetici	0.1438	0.1284	0.1526	0.0000	0.2539	0.3204	0.2598	0.3829
Sinottici	0.2524	0.2358	0.2356	0.2539	0.0000	0.2305	0.2289	0.1738
Paolo	0.3430	0.3342	0.3218	0.3204	0.2305	0.0000	0.1035	0.3012
Cattoliche	0.2879	0.2782	0.2586	0.2598	0.2289	0.1035	0.0000	0.3526
Giovanni	0.4007	0.3836	0.3392	0.3829	0.1738	0.3012	0.3526	0.0000

Figura 4.20: Geova — distanza del trasporto ottimo tra le sezioni della Bibbia

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	0.000	0.209	0.256	0.290	0.330	0.269	0.352	0.622	0.609	0.612	0.614	0.586	0.503	0.669	0.658	0.739	0.545	-0.4913
B	0.209	0.000	0.170	0.212	0.242	0.185	0.255	0.521	0.505	0.513	0.519	0.494	0.386	0.550	0.624	0.709	0.586	-0.4175
C	0.256	0.170	0.000	0.134	0.160	0.190	0.193	0.474	0.443	0.498	0.491	0.497	0.488	0.418	0.583	0.713	0.657	-0.3977
D	0.290	0.212	0.134	0.000	0.103	0.155	0.128	0.417	0.383	0.341	0.411	0.418	0.493	0.558	0.618	0.692	0.695	-0.3780
E	0.330	0.242	0.160	0.103	0.000	0.066	0.107	0.408	0.411	0.443	0.351	0.398	0.511	0.537	0.624	0.707	0.711	-0.3817
F	0.269	0.185	0.190	0.155	0.066	0.000	0.150	0.457	0.471	0.483	0.397	0.355	0.455	0.563	0.651	0.716	0.650	-0.3883
G	0.352	0.255	0.193	0.128	0.107	0.150	0.000	0.358	0.436	0.458	0.431	0.471	0.483	0.574	0.629	0.642	0.695	-0.3976
H	0.622	0.521	0.474	0.417	0.408	0.457	0.358	0.000	0.066	0.126	0.078	0.146	0.135	0.298	0.288	0.379	0.474	-0.3280
I	0.609	0.505	0.443	0.383	0.411	0.471	0.436	0.066	0.000	0.087	0.081	0.131	0.138	0.242	0.217	0.365	0.476	-0.3163
J	0.612	0.513	0.498	0.341	0.443	0.483	0.458	0.126	0.087	0.000	0.104	0.121	0.194	0.341	0.327	0.445	0.519	-0.3508
K	0.614	0.519	0.491	0.411	0.351	0.397	0.431	0.078	0.081	0.104	0.000	0.057	0.163	0.310	0.326	0.461	0.515	-0.3318
L	0.586	0.494	0.497	0.418	0.398	0.355	0.471	0.146	0.131	0.121	0.057	0.000	0.155	0.319	0.363	0.488	0.501	-0.3438
M	0.503	0.386	0.488	0.493	0.511	0.455	0.483	0.135	0.138	0.194	0.163	0.155	0.000	0.297	0.296	0.405	0.365	-0.3416
N	0.669	0.550	0.418	0.558	0.537	0.563	0.574	0.298	0.242	0.341	0.310	0.319	0.297	0.000	0.307	0.478	0.540	-0.4375
O	0.658	0.624	0.583	0.618	0.624	0.651	0.629	0.288	0.217	0.327	0.326	0.363	0.296	0.307	0.000	0.228	0.430	-0.4481
P	0.739	0.709	0.713	0.692	0.707	0.716	0.642	0.379	0.365	0.445	0.461	0.488	0.405	0.478	0.228	0.000	0.382	-0.5342
Q	0.545	0.586	0.657	0.695	0.711	0.650	0.695	0.474	0.476	0.519	0.515	0.501	0.365	0.540	0.430	0.382	0.000	-0.5463

Figura 4.21: Geova — distanza del coseno tra i centri dei cluster

Oltre a  $Q$ , che ospita le espressioni *Signore Geova*, anche  $O$  e  $P$  occupano, come  $N$ , una posizione defilata, e ospitano occorrenze di *Signore* corrispondenti in entrambe le traduzioni e sono caratteristici del Nuovo Testamento, a confermare un certo cambiamento nel significato dato alla parola tra Antico e Nuovo, che hanno una distanza del trasporto ottimo di 0.2614. Un indizio della rilevanza del Nuovo Testamento in questi cluster l'avevano data già la presenza di *Gesù* e *Cristo* tra le parole più rilevanti selezionate tra le co-occorrenti nei versetti.

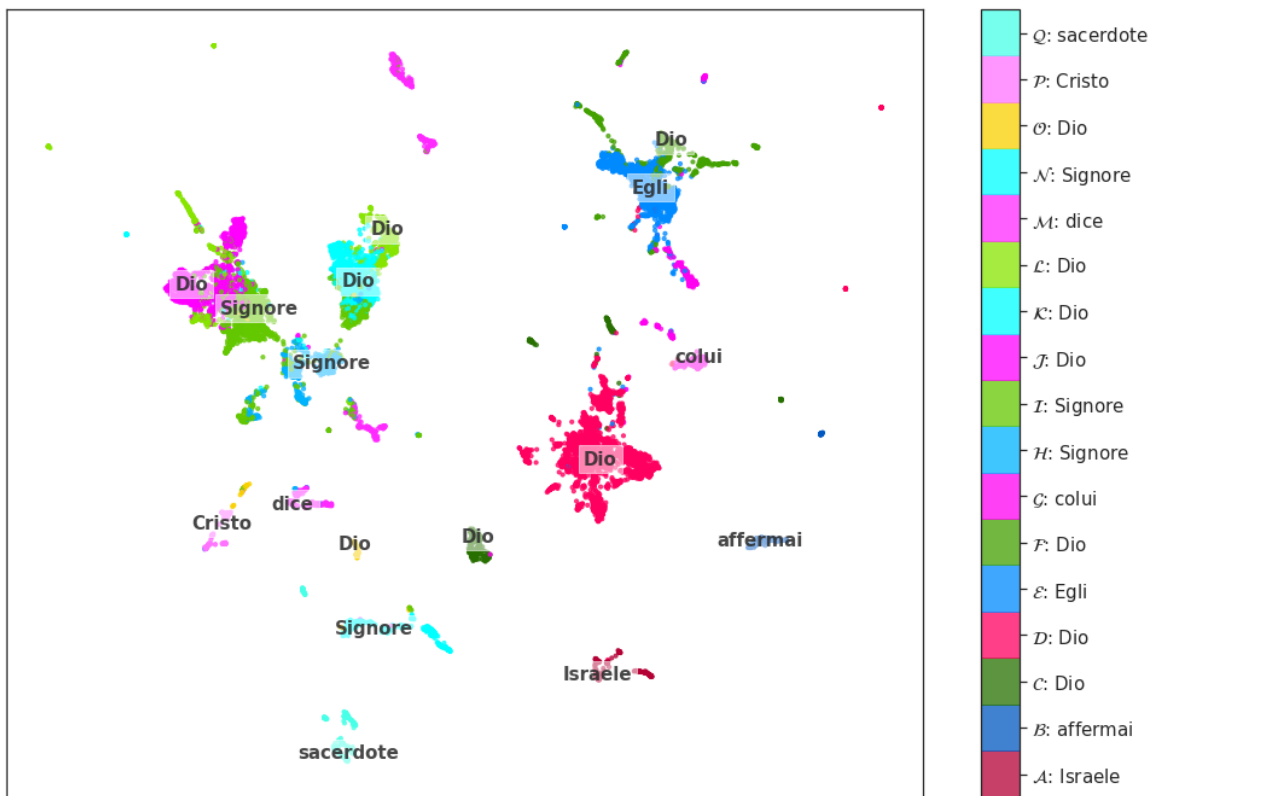
Guardando al vocabolario, questa divisione tra Antico e Nuovo Testamento si trova anche nelle definizioni. «Geova: di Dio che si trova nella Bibbia» è presente tra le più vicine ai *cluster*  $\mathcal{A} - \mathcal{M}$  (quindi anche in quelli che riportano le occorrenze di *Signore* nella Nuova Riveduta) ma non nei successivi  $\mathcal{N} - \mathcal{Q}$ .

La maggior parte delle parole selezionate inquadrano perfettamente il tema religioso sia tra le definizioni che tra gli esempi. I più difficili da inquadrare, guardando sia ai risultati che alle distanze delle definizioni trovate, sono stati  $\mathcal{M}$  e  $\mathcal{Q}$ .  $\mathcal{M}$  seleziona diversi nomi di luogo: Città del Messico, Lussemburgo, labronici (come in Figura 4.22b). Una spiegazione potrebbe essere come con Davide il riferimento alla città del Signore, ma, controllando le occorrenze, in questo caso non sono presenti espressioni simili. Per  $\mathcal{Q}$  il problema sono anche le distanze maggiori dalle definizioni selezionate, fino a scegliere ad esempio *Gesù* senza che siano presenti occorrenze dal Nuovo Testamento o la definizione dei buoni ordinari del tesoro.

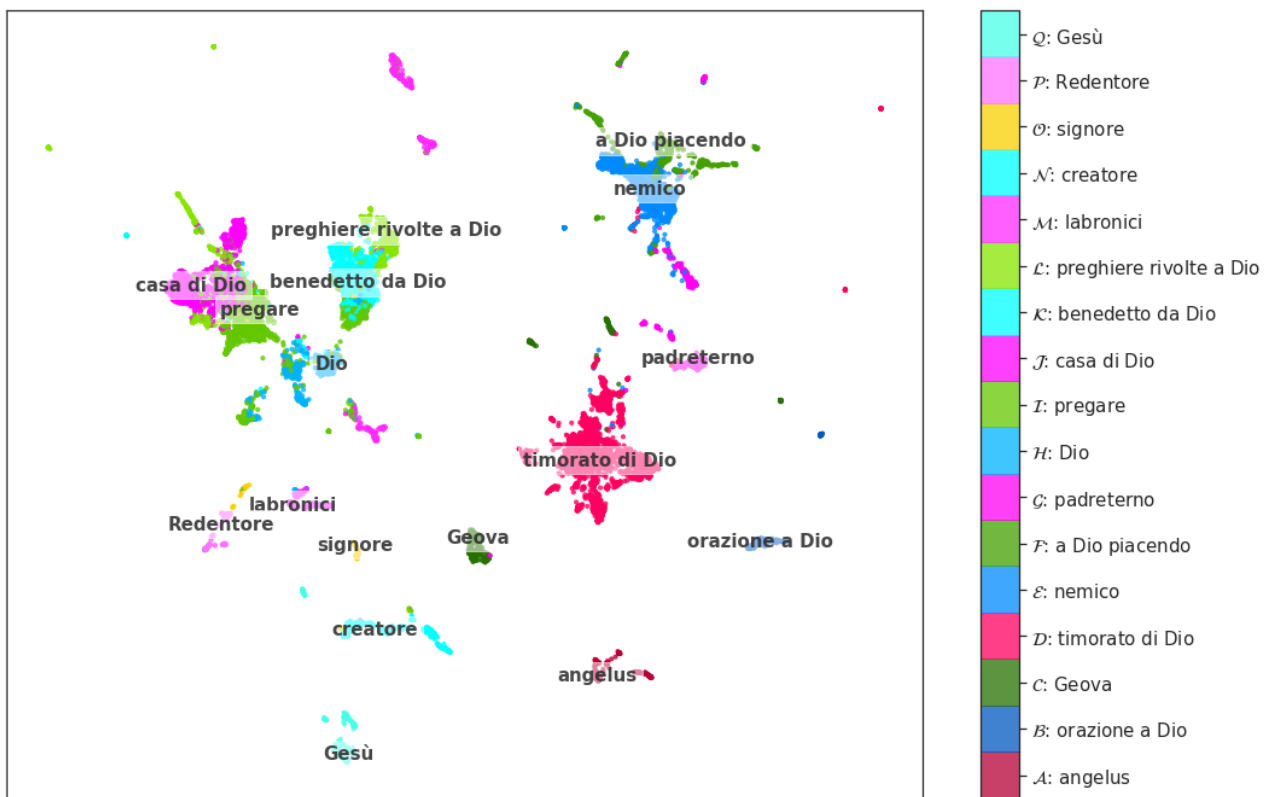
### Risultati principali

Dalle Figure 4.22 si nota la visibile differenza tra le occorrenze di *Geova* (tutte a destra), quelle di *Signore* e quelle del Nuovo Testamento (in basso a sinistra). Nonostante *Geova* sia una delle parole che il modello ha imparato solo dal testo fornito nel *fine-tuning* e ricorra nelle stesse frasi di *Signore*, il significato appreso è rimasto distante da quello della parola che sostituisce, fino ad una distanza del coseno superiore allo 0.5 tra  $\mathcal{A}$  e  $\mathcal{Q}$  che ospitano i vettori di parole la cui unica differenza è la forma, occorrendo sempre adiacenti negli stessi versetti.

L'algoritmo di *clustering* è riuscito ad individuare questa corrispondenza ma non quella tra gli usi delle due parole. Il fatto che non sia stato possibile individuare chiaramente una relazione tra i *cluster* delle due parole può essere considerato un fallimento dell'algoritmo o un'ulteriore prova della differenza tra le due parole nelle traduzioni considerate. Anche UMAP, d'altra parte, ha riprodotto la distanza ma non la specularità dei significati.



(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 4.22: Geova — visualizzazione in uno spazio bidimensionale

## 4.7 Dodici

Quando nei libri sacri si incontrano indicazioni numeriche, bisogna (...) chiedersi se, nell'intenzione dell'autore, il numero in questione si doveva intendere nel suo valore aritmetico esatto, oppure come una approssimazione od ancora nel suo significato simbolico. (...)

Il 12 è il numero delle lunazioni dell'anno e suggerisce quindi l'idea di un ciclo annuale completo (...)

Il 12, in quanto cifra delle dodici tribù, è anch'esso una cifra perfetta, che si applica simbolicamente al popolo di Dio. Di qui il suo uso significativo per i dodici apostoli di Gesù, che governeranno le dodici tribù del nuovo Israele (Mt 19, 28 par.). Anche la nuova Gerusalemme dell'Apocalisse ha dodici porte su cui sono incisi i nomi delle dodici tribù (Ap 21, 12), e dodici basamenti che portano i nomi dei dodici apostoli (21, 14). Così pure il popolo salvato è un numero di 144.000, dodici migliaia per ogni tribù di Israele (7, 4-8). Ma le dodici stelle che coronano la donna (altro simbolo della nuova umanità) potrebbero fare allusione alle dodici costellazioni zodiacali (12, 1).<sup>23</sup>

La *Concordanza pastorale della Bibbia* divide le occorrenze del numero 12 in quattro contesti: l'uso come numero cardinale, come riferimento alle tribù d'Israele, come riferimento agli apostoli e come numero cardinale (di dodicesimo).

Le versioni CEI usano spesso *Dodici* con la maiuscola per riferirsi agli apostoli, mentre la traduzione del nuovo mondo usa spesso il numero in cifre. Per vedere quanto queste scelte influenzino la comprensione del modello proviamo ad esaminare le tre forme (iniziale maiuscola, tutto minuscole e in cifre) in tutte le cinque traduzioni.

Il modello ha individuato inizialmente 12 *cluster* con almeno 25 occorrenze e ben 17 di dimensioni trascurabili, ma che insieme riguardavano una porzione rilevante delle occorrenze. Per provare a reinserirli nell'analisi è stata tentata una ulteriore fase di fusione dei *cluster* piccoli diminuendo la soglia della distanza ad una deviazione standard dalla media. Le occorrenze non hanno dato origine

---

<sup>23</sup>Fraine e Grelot 1976, p. 775.777.779.

a nessun nuovo *cluster* ma sono state tutte riassorbite dai 12 più grandi già individuati.

La distribuzione delle parole nei *cluster* (in Figura 4.23b) riporta una prima novità rispetto ai casi già esaminati: ci sono tre *cluster* ( $\mathcal{C}$ ,  $\mathcal{F}$  e  $\mathcal{H}$ ) con un numero non trascurabile di tutte e tre le forme prese in considerazione e quattro ( $\mathcal{B}$ ,  $\mathcal{E}$ ,  $\mathcal{G}$  e  $\mathcal{I}$ ) con un numero non trascurabile di due forme (minuscola e in cifre). Questo avveniva già prima della fase ulteriore di fusione dei *cluster* ridotti, non è quindi dovuto a questa operazione.

In ogni caso, la maggior parte delle occorrenze con l’iniziale maiuscola è inserita in  $\mathcal{I}$  e delle occorrenze in cifre in  $\mathcal{A}$ . Calcolando la distanza tra le tre forme con il problema del trasporto ottimo, si nota che la forma con l’iniziale maiuscola è quella più distante dalle altre due che invece sono relativamente più simili.

Confrontando le traduzioni (in Figura 4.23c), si nota la solita similarità tra le due della CEI aggiunta a quella con la Nuova Riveduta, la cui unica differenza visibile è la presenza di  $\mathcal{K}$  a sostituire  $\mathcal{L}$ . Questo fa capire che la Nuova Riveduta non usa la forma maiuscola ed il sistema ha raccolto le occorrenze corrispondenti con la forma minuscola tutte nello stesso *cluster*. La similarità tra i due è ben rappresentata dall’immagine in Figura 4.25.

Per motivi diversi si distinguono dalle altre sia la traduzione interconfessionale che quella del nuovo mondo. La prima, comunque più vicina alle altre secondo la distanza del trasporto ottimo, ha alcuni *cluster* di dimensioni minori (in particolare  $\mathcal{B}$  e  $\mathcal{L}$ ) mentre accoglie molte più occorrenze in  $\mathcal{D}$  e  $\mathcal{J}$ . La traduzione del nuovo mondo, comunque con una distanza del trasporto ottimo sotto la soglia di attenzione, si differenzia soprattutto per l’uso del numero in cifre (che riunisce in  $\mathcal{A}$  ciò che è diviso tra più *cluster* con il numero in lettere) a parte alcune occorrenze con la maiuscola (in  $\mathcal{L}$ ).

Più che confrontare le grandi sezioni di testo, visto che il numero ricorre in un numero gestibile di libri, vale la pena fare il confronto a questo livello di granularità. Tra i sapienziali è presente solo in Siracide, tra i profeti solo in Geremia ed Ezechiele ed è assente in tutte le lettere del Nuovo Testamento.

In Genesi potrebbe esserci un *cluster* per accogliere le occorrenze del numero riferite ai dodici figli di Giacobbe che daranno origine alle tribù. La parola *figli*

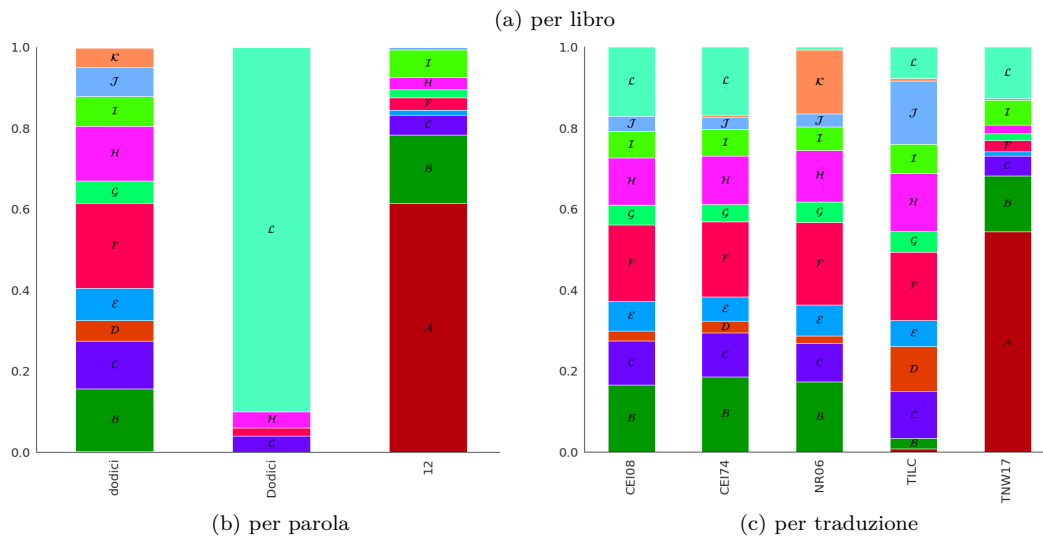
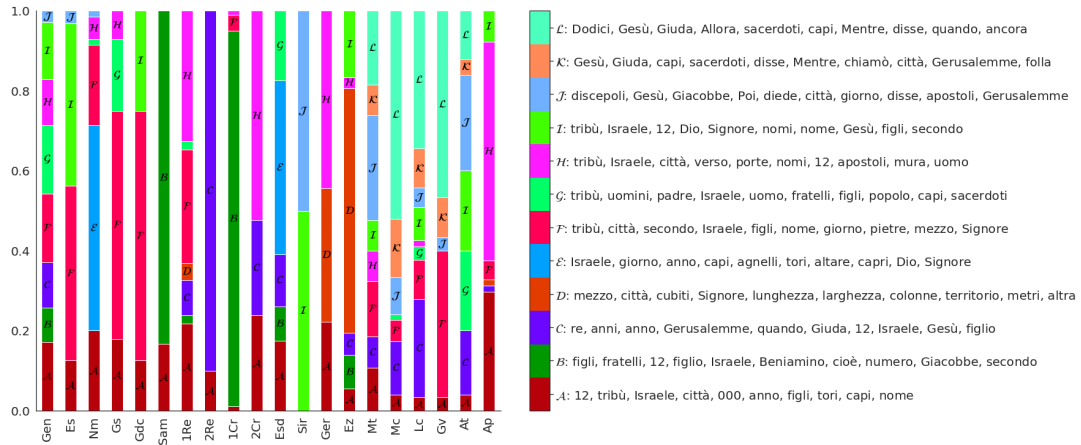


Figura 4.23: 12 — distribuzione delle occorrenze nei *cluster*

	A	B	C	D	E	F	G	H	I	J	K	L	
A	0.0000	0.3098	0.3353	0.3585	0.2927	0.2251	0.2483	0.1981	0.2780	0.3201	0.4096	0.5464	-0.3202
B	0.3098	0.0000	0.3155	0.3320	0.3017	0.2557	0.2604	0.3203	0.4328	0.4184	0.3540	0.5188	-0.3472
C	0.3353	0.3155	0.0000	0.1661	0.3013	0.2121	0.2499	0.2535	0.3805	0.3477	0.3841	0.5251	-0.3156
D	0.3585	0.3320	0.1661	0.0000	0.2425	0.1909	0.2749	0.2391	0.3882	0.3849	0.4518	0.6043	-0.3303
E	0.2927	0.3017	0.3013	0.2425	0.0000	0.1650	0.2297	0.1726	0.3496	0.3365	0.4321	0.5654	-0.3081
F	0.2251	0.2557	0.2121	0.1909	0.1650	0.0000	0.1205	0.0849	0.2230	0.2483	0.3218	0.4745	-0.2292
G	0.2483	0.2604	0.2499	0.2749	0.2297	0.1205	0.0000	0.1325	0.2972	0.2701	0.3280	0.4901	-0.2638
H	0.1981	0.3203	0.2535	0.2391	0.1726	0.0849	0.1325	0.0000	0.1624	0.1635	0.2843	0.4286	-0.2218
I	0.2780	0.4328	0.3805	0.3882	0.3496	0.2230	0.2972	0.1624	0.0000	0.2019	0.3288	0.4680	-0.3191
J	0.3201	0.4184	0.3477	0.3849	0.3365	0.2483	0.2701	0.1635	0.2019	0.0000	0.1350	0.2836	-0.2827
K	0.4096	0.3540	0.3841	0.4518	0.4321	0.3218	0.3280	0.2843	0.3288	0.1350	0.0000	0.1500	-0.3254
L	0.5464	0.5188	0.5251	0.6043	0.5654	0.4745	0.4901	0.4286	0.4680	0.2836	0.1500	0.0000	-0.4595

Distanza media

Figura 4.24: 12 — distanza del coseno tra i centri dei *cluster*



appare però tra le rilevanti in diversi *cluster* e nessuno è caratteristico del primo libro della Bibbia.

Il *cluster*  $\mathcal{E}$ , presente solo in Numeri ed Esdra, insieme alle parole co-occorrenti più rilevanti<sup>24</sup> fa pensare all'uso del numero per indicare le tribù in relazione alle offerte nel tempio. Almeno in parte, le occorrenze corrispondenti nella traduzione del nuovo mondo sembrano essere state inserite invece in  $\mathcal{A}$ . Esplorare i versetti (in Tabella 4.7) conferma questa ipotesi. Come dimostra Esdra 6, 17 la ricorrenza del numero 12 (e dei suoi multipli) in queste offerte dipende dal numero delle tribù di Israele.

Tabella 4.7: Dodici — versetti in cui occorrono le parole del *cluster*  $\mathcal{E}$  (traduzione CEI08)

- Nm 7, 3 Portarono la loro offerta davanti al Signore: sei carri coperti e **dodici** capi di bestiame grosso, ... (TNW17: in  $\mathcal{A}$ )
- Nm 7, 84 Questi furono i doni per la dedicazione dell'altare da parte dei capi d'Israele, il giorno in cui esso fu unto: **dodici** piatti d'argento, **dodici** vassoi d'argento, **dodici** coppe d'oro; (TILC e TNW17: il primo in  $\mathcal{A}$ )
- Nm 7, 87 Totale del bestiame per l'olocausto: **dodici** giovenchi, **dodici** arieti, **dodici** agnelli di un anno con la loro oblazione, e **dodici** capri per il sacrificio per il peccato. (TNW17: tutti in  $\mathcal{A}$ )
- Nm 29, 17 Il secondo giorno offrirete **dodici** giovenchi, ... (TNW17: in  $\mathcal{A}$ )
- Esd 6, 17 offrirono per la dedicazione di questo tempio di Dio ... **dodici** capri come sacrifici espiatori per tutto Israele, secondo il numero delle tribù d'Israele. (TNW17: in  $\mathcal{A}$ )
- Esd 8, 35 Quelli che venivano dall'esilio, i deportati, offrirono olocausti al Dio d'Israele: **dodici** tori per tutto Israele, ... **dodici** capri per il peccato, tutto come olocausto al Signore. (CEI74: in  $\mathcal{B}$ ; TNW17: in  $\mathcal{A}$ )
- 

Il secondo libro di Samuele ed il primo delle Cronache, che parlano entrambi del regno di Davide, sono simili e caratterizzati dalla presenza di molte occorrenze raccolte in  $\mathcal{B}$ . Una similarità, anche guardando alla distanza del trasporto ottimo, si nota probabilmente è presente per lo stesso motivo, anche tra il primo libro dei Re ed il secondo delle Cronache, mentre le occorrenze del secondo libro dei Re sono raccolte quasi tutte in  $\mathcal{C}$ .

<sup>24</sup>«Israele, giorno, anno, capi, agnelli, tori, altare, capri, Dio, Signore» come in Figura 4.23a.

I due libri profetici di Geremia ed Ezechiele accolgono quasi tutte le occorrenze raccolte in  $\mathcal{D}$  dove il numero è usato per misurare una lunghezza. Il modello ha messo insieme le lunghezze di dodici cubiti delle traduzioni più letterali con altri versetti che riportavano dodici metri o chilometri dove la traduzione interconfessionale ha convertito le misure.

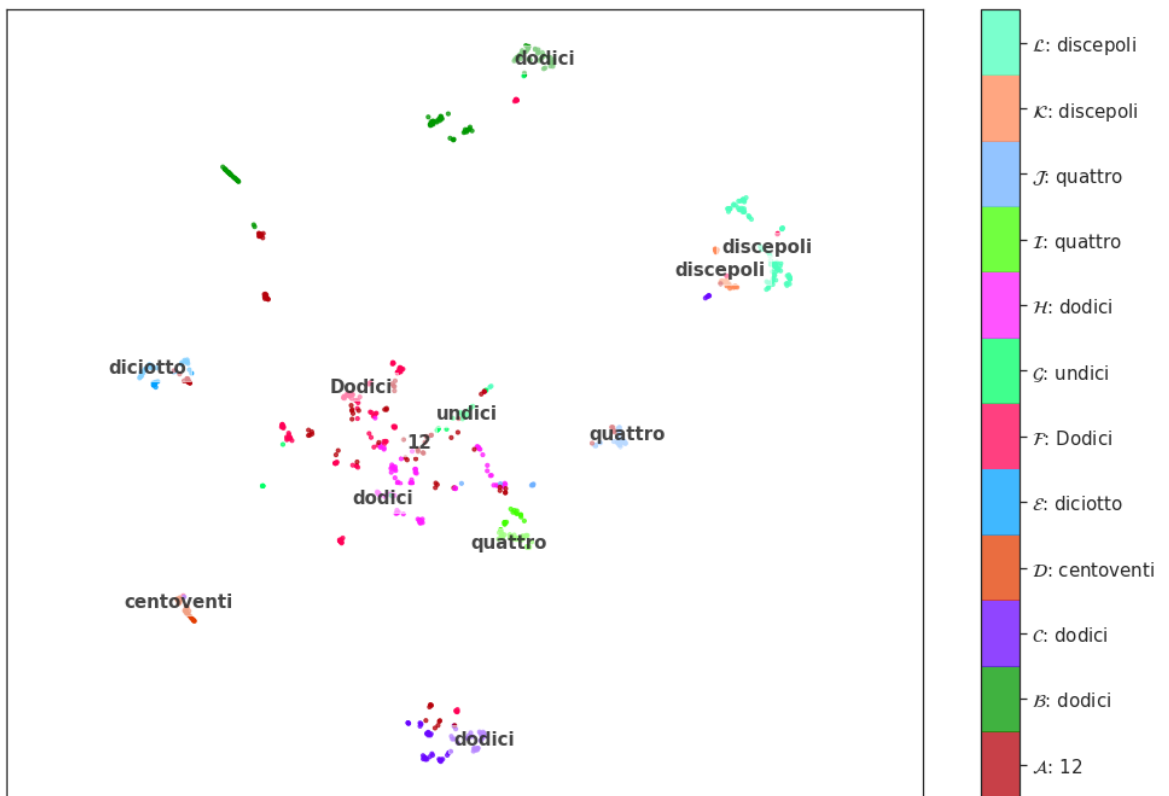
Mentre come era atteso  $\mathcal{K} - \mathcal{L}$  sono caratteristici del Nuovo Testamento, ad eccezione dell'Apocalisse,  $\mathcal{J}$ , nonostante sia descritto da parole come *discepoli*, *Gesù*, *apostoli*, non è esclusivo del Nuovo Testamento ma accoglie occorrenze da Genesi, Esodo e Siracide (la metà delle poche presenti).

Guardando alle distanze tra i centri dei *cluster* (in Figura 4.24) si nota che siano più sparsi rispetto agli esempi precedenti. Il gruppo  $\mathcal{J} - \mathcal{L}$  che abbiamo già individuato come caratteristico del Nuovo Testamento, si nota più per la distanza dagli altri (in particolare di  $\mathcal{L}$ ), che per la vicinanza all'interno. Anche  $\mathcal{A}$  ed ancora di più  $\mathcal{B}$  si distinguono per non essere particolarmente vicini a nessun altro.

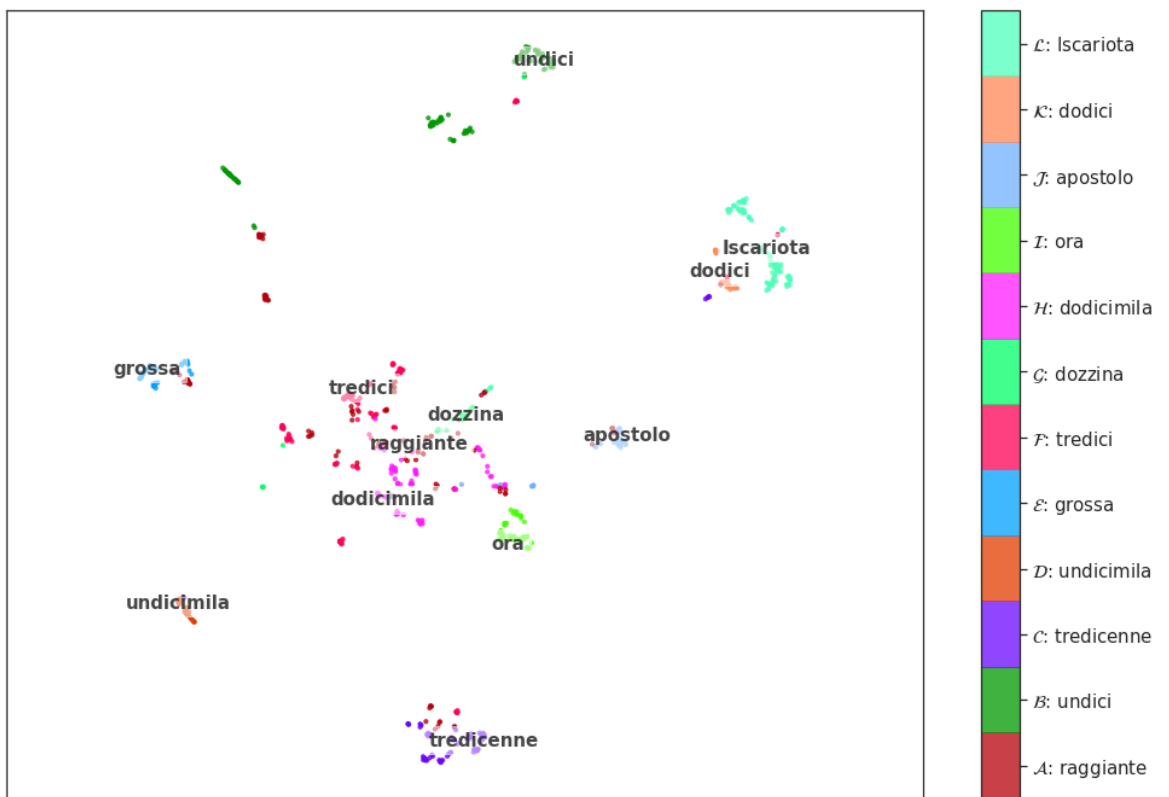
Sia dalle definizioni che dagli esempi del vocabolario, i risultati principali sono semplicemente numeri, come si vede in Figura 4.25. Lo strumento fornisce comunque risultati confortanti, tra le poche parole selezionate spuntano *discepoli*, *apostolo*, *Iscariota* e solo per i *cluster* caratteristici del Nuovo Testamento. In alcuni di questi ultimi è assente ogni riferimento numerico nella definizione.

### Risultati principali

I significati simbolici come quello di numero che indica completezza non sono stati colti. Il riferimento quantitativamente più rilevante, cioè quello anticotestamentario alle dodici tribù sembra avere influito meno sul significato colto dal sistema, che ha riconosciuto semplicemente il numero come una quantità. Nel Nuovo Testamento, invece, l'uso del numero da solo, con la maiuscola come se fosse un nome proprio ma anche con la minuscola nella Nuova Riveduta, ha permesso almeno di riconoscerlo come sinonimo di apostoli o discepoli.



(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 4.25: 12 — visualizzazione in uno spazio bidimensionale

## Capitolo 5

# I risultati – parole suggerite dal modello

Nel capitolo precedente le parole sono state scelte *a priori* basandosi su curiosità o conoscenze già acquisite. Uno dei vantaggi dell’algoritmo di *clustering* adottato è la scalabilità. Inoltre, continuando ad usare i *cluster* e non le singole occorrenze, anche la valutazione del cambiamento semantico con il problema del trasporto ottimo ha tempi di elaborazione contenuti.

L’impegno maggiore, in termini di tempo di elaborazione, rimane quello del *clustering*, una volta preparati i dati per misurare ad esempio il cambiamento semantico tra le traduzioni, è possibile e conveniente usare gli stessi dati per misurarlo anche tra porzioni di testo.

È quindi praticabile chiedere direttamente ai dati quali possano essere le parole più interessanti da analizzare. In questo capitolo, si cercheranno le parole che si distinguono per il cambiamento semantico per poi cercare *a posteriori* quali possano essere le cause del cambiamento evidenziato e individuare eventuali aspetti particolari del significato con cui vengono usate.

Escludendo le *stopwords*, il vocabolario complessivo delle cinque traduzioni comprende 8523 parole. Di ciascuna parola che occorresse almeno 25 volte sono state estratte le rappresentazioni vettoriali che sono state sottoposte al processo di *clustering*. Solo 1903 parole hanno dato origine ad almeno un *cluster* di dimensioni rilevanti. Considerando la legge di Zipf, questa differenza tra il numero di parole nel vocabolario e quelle che danno luogo a *cluster* di dimensioni rilevanti non deve

Tabella 5.1: Parole con distanza del trasporto ottimo maggiore tra le traduzioni CEI

1	Parola	0.4187	11	Satana	0.1818
2	Va	0.3218	12	continua	0.1807
3	Mare	0.2797	13	età	0.1688
4	Re	0.2558	14	dati	0.1688
5	impuro	0.2198	15	completamente	0.1671
6	amato	0.1969	16	rifiutato	0.1651
7	va	0.1950	17	termini	0.1582
8	sì	0.1894	18	sud	0.1573
9	Sur	0.1851	19	Legge	0.1536
10	registrati	0.1829	20	bruciare	0.1499

Tabella 5.2: Parole con distanza del trasporto ottimo maggiore tra due traduzioni

1	rivolta	0.5939	CEI74 – TILC	
2	viola	0.5839	NR06 – TILC	<i>assente in TNW17</i>
3	dichiara	0.5535	NR06 – TNW17	
4	riportato	0.5098	CEI08 – TNW17	<i>solo in CEI08, TILC, TNW17</i>
5	completo	0.4905	CEI74 – TILC	
6	compresi	0.4889	NR06 – TILC	<i>assente in CEI08</i>
7	Porta	0.4802	TILC – TNW17	<i>assente in CEI08</i>
8	venti	0.4693	TILC – TNW17	
9	pranzo	0.4684	CEI74 – TNW17	<i>assente in NR06</i>
10	termini	0.4636	NR06 – TILC	<i>assente in TNW17</i>
11	parla	0.4520	NR06 – TNW17	
12	schiera	0.4484	TILC – TNW17	
13	tratti	0.4457	TILC – TNW17	
14	ufficiale	0.4442	NR06 – TNW17	<i>assente in CEI08</i>
15	base	0.4408	NR06 – TNW17	
16	Grande	0.4257	TILC – TNW17	
17	maestro	0.4234	CEI08 – TNW17	
18	Parola	0.4203	CEI74 – NR06	<i>assente in TNW17</i>
19	motivo	0.4198	TILC – TNW17	
20	Sovrano	0.4186	CEI08 – TNW17	<i>solo in CEI08, TNW17</i>

sorprendere, sempre considerando la soglia stabilita in 25 occorrenze perché un *cluster* possa ritenersi tale.

Calcolando le distanze del problema del trasporto ottimo è possibile trovare le parole che si distinguono per un uso diverso tra le traduzioni o tra le sezioni di testo. Confrontando le due traduzioni della CEI, escludendo *stopword* e parole che non danno luogo a *cluster* di dimensioni rilevanti, in Tabella 5.1 sono riportate le venti per cui il sistema rileva un cambiamento maggiore, accompagnate ciascuna dalla distanza del trasporto ottimo misurata.

Solo due parole superano una distanza di 0.3, indicata da Montariol, Martinc e Pivovarova 2021 come soglia per un cambiamento rilevante: *Parola* e *Va*. Di questa seconda può essere interessante sottolineare come gli imperativi siano molto più frequenti nella Bibbia che nel linguaggio *standard* e quindi, volendo approfondire, si potrebbe indagare anche su quanto l'adattamento al dominio l'adattamento al dominio abbia modificato la comprensione del modello. Diverse altre parole della lista possono dare luogo ad un'analisi interessante anche dal punto di vista teologico biblico, ma l'analisi di questo capitolo si soffermerà sulla prima suggerita dai dati: *Parola*.

Considerando tutte le traduzioni e prendendo la distanza maggiore per ogni parola (in Tabella 5.2), bisogna arrivare alla diciottesima posizione per trovarne una che non provenga dalla traduzione interconfessionale o da quella del nuovo mondo ed è proprio *Parola*. Questo è un altro buon indizio del fatto che siano le due traduzioni che si differenziano maggiormente non solo dalle altre, ma anche tra loro.

In questa seconda lista entrano un numero (*venti*) e diverse parole comuni di non grande interesse teologico. Il modello in questo caso sembra suggerire *rivolta*, in particolare per il diverso significato con cui è utilizzata nella traduzione CEI del 1974 e in quella interconfessionale.

Oltre che tra le traduzioni, un simile confronto si può fare tra sezioni di testo, considerando le più estese, sono stati confrontati prima di tutto il Nuovo Testamento rispetto ai libri cronologicamente più vicini, cioè i deuterocanonici dell'Antico Testamento, che spesso condividono anche la lingua del testo originale. Le parole il cui significato varia maggiormente sono riportate in Tabella 5.3.

Tabella 5.3: Parole con distanza del trasporto ottimo maggiore tra NT e deuterocononici

<b>5 traduzioni</b>		<b>Solo CEI08</b>		
1	Figlio	0.5146	proposito	0.2458
2	piano	0.4983	posti	0.2174
3	successo	0.4974	insegna	0.1955
4	comandi	0.2665	segunte	0.1881
5	lascero	0.4799	pratica	0.1873
6	pezzi	0.4782	passa	0.1862
7	passo	0.4672	memoria	0.1819
8	Felice	0.4450	profondo	0.1784
9	segunte	0.4401	superiore	0.1756
10	quindici	0.4400	abiti	0.1730
11	sì	0.4334	uso	0.1712
12	completo	0.4284	teneva	0.1706
13	sud	0.4258	peso	0.1682
14	accese	0.4249	scendere	0.1661

Tabella 5.4: Parole con distanza del trasporto ottimo maggiore tra NT e AT ebraico

<b>5 traduzioni</b>		<b>Solo CEI08</b>		
1	bordo	0.5893	senso	0.2198
2	Ben	0.5261	detta	0.2125
3	fallo	0.5103	pensiero	0.2030
4	Parola	0.5035	quasi	0.1985
5	coro	0.5017	gettati	0.1882
6	passo	0.4885	libera	0.1756
7	discende	0.4838	insegna	0.1742
8	successo	0.4838	uso	0.1663
9	dichiara	0.4801	lasciate	0.1638
10	comandi	0.4752	spalle	0.1569
11	Figlio	0.4724	finire	0.1540
12	strumenti	0.4675	nati	0.1512
13	maestro	0.4646	chiede	0.1484
14	completo	0.4594	posti	0.1465

Le distanze sono paragonabili a quelle tra le traduzioni, riducendo il campo d'indagine alla sola traduzione più recente della CEI e sottoponendo le stesse sezioni di testo a confronto se ne ottengono di molto minori. Questo potrebbe essere dovuto al sommarsi dell'effetto della distanza delle traduzioni, ma le parole ricavate non compaiono nella lista di quelle che variano di più tra le traduzioni. Una spiegazione alternativa potrebbe essere legata al fatto che questi testi siano riportati solo in tre delle traduzioni prese in esame.

Questa seconda ipotesi è confutata dalle distanze ottenute nel confronto del Nuovo Testamento con i libri che appartengono al canone ristretto dell'Antico Testamento (in Tabella 5.4). I valori delle distanze ottenute sono comparabili ai precedenti, ma sono poche le parole che compaiono in entrambe le liste.

La parola che cambia di più tra deutero-canoniche e Nuovo Testamento è anche una di quelle che cambia maggiormente nel confronto con l'Antico Testamento ebraico ed è su questa, cioè *Figlio*, che si soffermerà l'ultima delle analisi delle parole.

## 5.1 Parola

Il sistema identifica 3 *cluster* di dimensioni rilevanti e 10 di dimensioni ridotte che, insieme, rappresentano più di metà delle occorrenze totali.

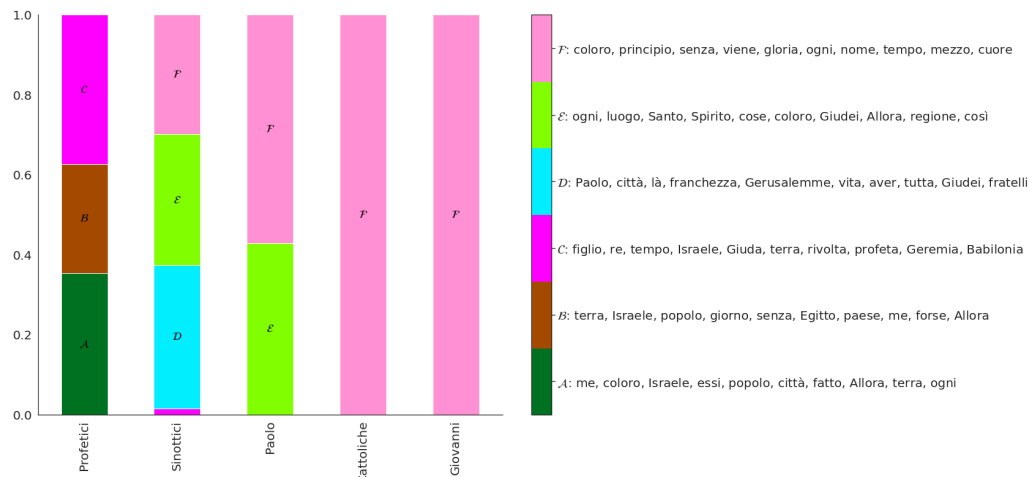
Considerando anche che non ci sono quasi occorrenze di *Parola* (con l'iniziale maiuscola) nella traduzione del nuovo mondo, si può provare ad effettuare una nuova fase di fusione dei *cluster* abbassando la soglia del numero di occorrenze per considerarli rilevanti (fissata in base al numero di traduzioni in esame) ed abbassando la soglia della distanza ad una deviazione standard dalla media come fatto sopra (4.7). In questa ulteriore fase, i 10 *cluster* di dimensioni ridotte sono stati raggruppati dando origine ad altri tre di dimensioni rilevanti.

Calcolando la distanza del trasporto ottimo, tra le sezioni di testo (in Figura 5.1a) e tra le traduzioni (in Figura 5.2b) si nota una netta differenza tra due coppie di traduzioni e tra Antico e Nuovo Testamento. La stessa polarizzazione si vede dalle distanze tra i centri dei *cluster* (in Figura 5.3) che sono chiaramente divisi in due gruppi:  $\mathcal{A} - \mathcal{C}$  (con  $\mathcal{A}$  e  $\mathcal{B}$  particolarmente vicini) e  $\mathcal{D} - \mathcal{F}$ .





(a) distanza del trasporto ottimo



(b) distribuzione delle occorrenze nei cluster

Figura 5.1: Parola — confronto tra le principali sezioni della Bibbia

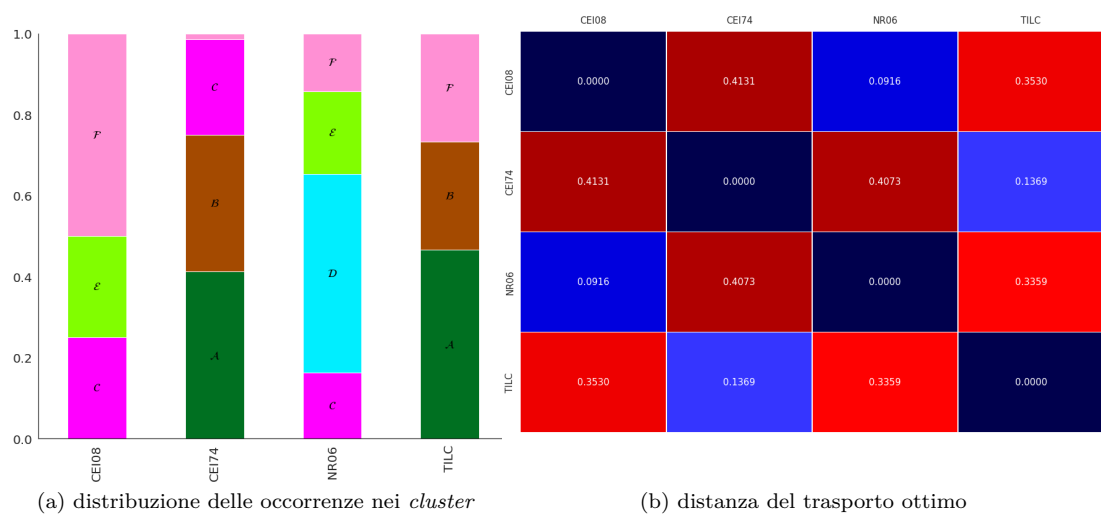


Figura 5.2: Parola — confronto tra le traduzioni

Le due coppie di traduzioni sono inedite: la CEI del 2008 è piuttosto vicina alla Nuova Riveduta e distante dalla CEI del 1974 e da quella interconfessionale.

Cambiando il numero di *cluster* c'è una leggera modifica delle distanze del trasporto ottimo, ad esempio la distanza tra le due CEI diminuisce circa dell'1%. La portata della variazione conferma che l'ulteriore fase di fusione non cambia molto in queste misure ma aiuta molto l'interpretazione.

Confrontando i *cluster* con la provenienza delle parole, si nota una chiara distinzione.  $\mathcal{A} - \mathcal{C}$  provengono quasi esclusivamente dai libri profetici e  $\mathcal{D} - \mathcal{F}$  dal Nuovo Testamento. La stessa divisione, un po' meno netta, c'è tra la traduzione CEI del 1974 e quella interconfessionale (incluse nei primi *cluster*) e le altre due traduzioni.

In particolare, la traduzione interconfessionale non ha occorrenze in  $\mathcal{C}$  (che invece ospita le poche parole dell'Antico Testamento della traduzione CEI del 2008 e Nuova Riveduta), ma alcune in  $\mathcal{F}$ .  $\mathcal{D}$  è presente solo in Atti e nella Nuova Riveduta.

Guardando ai versetti da cui provengono le parole, non si trovano differenze sensibili tra  $\mathcal{A}$  e  $\mathcal{B}$  in cui alcuni oracoli profetici sono intervallati o conclusi dalla formula «Parola del Signore», dove altre traduzioni usano «Oracolo del Signore» o «dice il Signore» e formule simili. La distanza del coseno veramente bassa (0.0205) è confermata da questa similarità riscontrata.

$\mathcal{C}$  è leggermente più eterogeneo, ed è infatti il risultato della fase ulteriore di fusione dei *cluster* di dimensioni ridotte. Come si vede in Tabella 5.5 sono messe insieme forme del tipo «Parola che il Signore a rivolto a...» con introduzioni di discorsi o simili attribuiti ad un uomo. Come si nota, sono espressioni quasi identiche, sono infatti formule che si sono cristallizzate nel tempo che i tre traduttori hanno deciso di riprodurre seguendo piuttosto letteralmente il testo originale. La traduzione interconfessionale e quella del nuovo mondo qui usano invece *messaggio* per tradurre  $\text{דְבַר}$ , termine chiave della teologia dell'Antico Testamento, la cui radice ricorre 1441 volte nel testo e che la LXX traduce letteralmente  $\Lambda\acute{o}\gamma\omicron\varsigma$ . La citazione dagli Atti è chiaramente un intruso nell'elenco, è l'unico caso in questo *cluster* in cui l'occorrenza è presente solo nella traduzione CEI del 1974, quella del 2008 usa infatti *voce* che traduce più letteralmente il greco  $\phi\omega\nu\acute{\eta}$ .

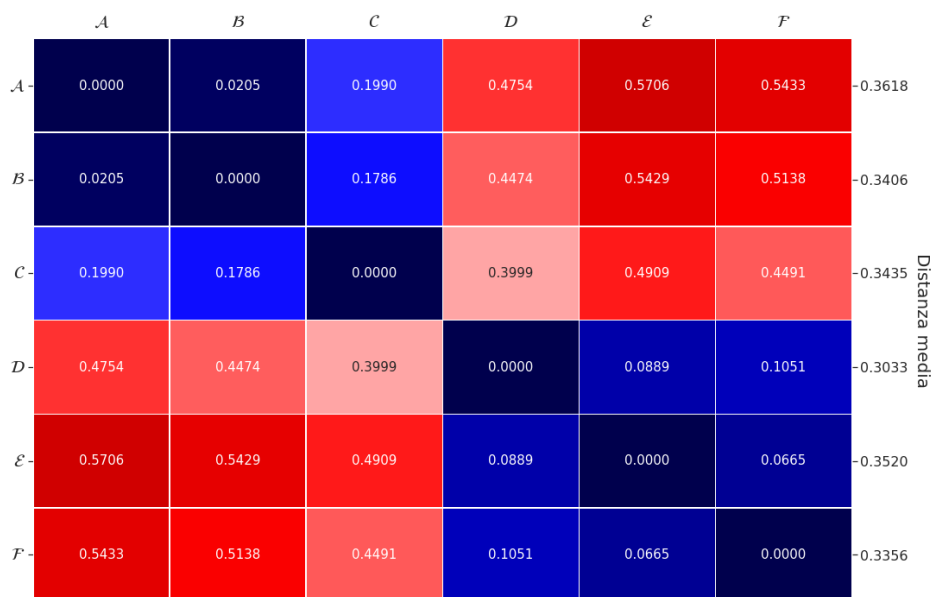


Figura 5.3: Parola — distanza del coseno tra i centri dei *cluster*

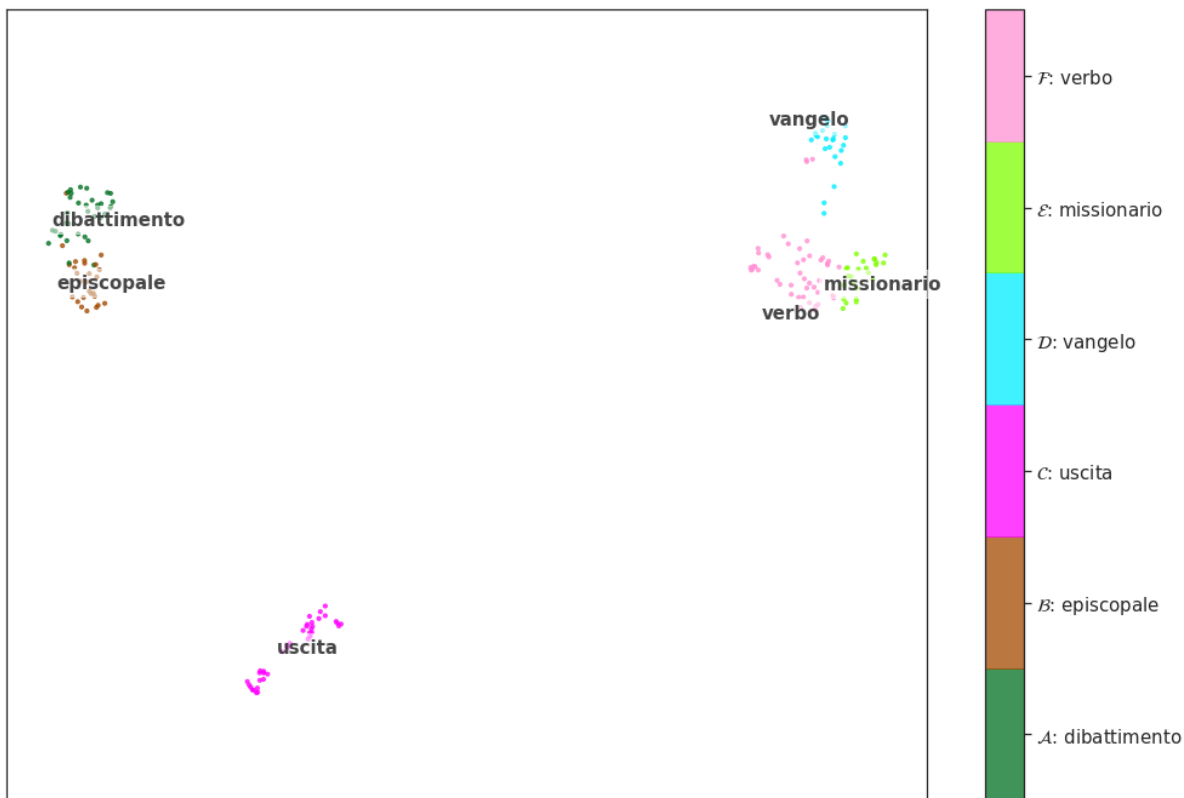


Figura 5.4: Parola — visualizzazione in uno spazio bidimensionale, *cluster* etichettati con le definizioni del vocabolario

Tabella 5.5: Parola — versetti in cui occorrono le parole del *cluster C*

2Sam	23, 1	Queste sono le ultime parole di Davide: « <b>Parola</b> di Davide, ... (NR06)
Is	2, 1	<b>Parola</b> che Isaia, figlio di Amots, ebbe in visione ... (NR06)
Ger	14, 1	<b>Parola</b> rivolta dal Signore a Geremia in occasione ... (CEI08) <b>Parola</b> che il Signore rivolse a Geremia in occasione ... (CEI74)
Ger	30, 1	<b>Parola</b> rivolta a Geremia da parte del Signore: (CEI08) <b>Parola</b> che fu rivolta a Geremia da parte del Signore: (CEI74)
Ger	32, 1	<b>Parola</b> rivolta a Geremia dal Signore nell'anno ... (CEI08) <b>Parola</b> che fu rivolta a Geremia dal Signore nell'anno ... (CEI74)
Ger	34, 1	<b>Parola</b> che fu rivolta dal Signore a Geremia, quando ... (CEI08) <b>Parola</b> che fu rivolta a Geremia dal Signore, quando ... (CEI74)
Ger	46, 1	<b>Parola</b> del Signore che fu rivolta al profeta ... (CEI08, CEI74) <b>Parola</b> del Signore che fu rivolta a Geremia ... (NR06)
Ger	46, 13	<b>Parola</b> che il Signore comunicò al profeta Geremia ... (CEI08, CEI74)
Ger	47, 1	<b>Parola</b> del Signore che fu rivolta al profeta Geremia ... (CEI08, CEI74)
Ger	49, 34	<b>Parola</b> che il Signore rivolse al profeta Geremia ... (CEI08, CEI74)
Ger	50, 1	<b>Parola</b> che il Signore pronunciò contro Babilonia, ... (CEI08, CEI74)
Os	1, 1	<b>Parola</b> del Signore rivolta a Osea, ... (CEI08, CEI74, NR06)
Gl	1, 1	<b>Parola</b> del Signore, rivolta a Gioele, ... (CEI08, CEI74, NR06)
Mi	1, 1	<b>Parola</b> del Signore, rivolta a Michea ... (CEI08, CEI74, NR06)
Sof	1, 1	<b>Parola</b> del Signore che fu rivolta a Sofonia, ... (CEI08, CEI74, NR06)
Zc	12, 1	Oracolo. <b>Parola</b> del Signore su Israele. Oracolo del Signore che ha dispiegato i cieli ... (CEI08) Oracolo. <b>Parola</b> del Signore su Israele. Dice il Signore che ha steso i cieli ... (CEI74) Oracolo, parola del Signore, riguardo a Israele. <b>Parola</b> del Signore che ha disteso i cieli ... (NR06)
Ml	1, 1	Oracolo. <b>Parola</b> del Signore a Israele per mezzo di ... (CEI08, CEI74)
At	12, 22	Il popolo acclamava: ” <b>Parola</b> di un dio e non di un uomo! ”. (CEI74)

Dal testo si nota che questo traduttore ha deciso di usare la maiuscola per *Parola* quando è usato per indicare la diffusione del vangelo nelle prime comunità cristiane. Il modello ha colto bene questo uso, selezionando *vangelo*, *missionario* (cfr. Figura 5.4), *apostolo* e *predicazione* dalle definizioni del vocabolario.

In  $\mathcal{E}$  sono presenti altre occorrenze provenienti da Atti e Nuova Riveduta ma che hanno almeno una differenza che salta agli occhi: solo e in tutti questi casi, anche la traduzione CEI del 2008 ha deciso di rendere il termine con la maiuscola. Il sistema ha notato una qualche differenza in queste occorrenze che deve aver colto anche questo secondo traduttore. Insieme a queste occorrenze, sono incluse in  $\mathcal{E}$  alcune provenienti dalle lettere di Paolo nella traduzione CEI del 2008 con lo stesso significato di Vangelo e dall'*incipit* di Luca per entrambi i traduttori. Completa l'elenco la parabola del seminatore nel Vangelo secondo Marco sempre nella traduzione CEI del 2008.

$\mathcal{F}$  è ancora più eterogeneo di  $\mathcal{C}$ , anche qui conseguenza dell'ultima fase di fusione dei *cluster*. In particolare, ospita il prologo di Giovanni (ed altre occorrenze dall'opera giovannea) nella Nuova Riveduta, nella traduzione interconfessionale e in quella del nuovo mondo (le uniche occorrenze con la maiuscola in questa traduzione), dove invece le traduzioni CEI usano il termine *Verbo* diventato caratteristico della traduzione cattolica. Anche qui, dalle definizioni del vocabolario il sistema è stato capace di estrarre *verbo* (nel suo significato teologico) e *Cristo*.

Aver calcolato i centri dei *cluster* di tutte le parole nelle cinque traduzioni fornisce la possibilità di confrontare quelli in esame con le altre parole presenti nelle traduzioni della Bibbia, permettendo di scoprire a quali significati ogni *cluster* è più vicino. Questa è una indicazione sia del significato della parola in quel gruppo di occorrenze, che delle altre parole simili da indagare (come fa ad esempio la *Concordanza pastorale della Bibbia*). In Tabella 5.6 sono riportate le cinque parole più simili per ogni *cluster* con le relative similarità del coseno. Per leggere i dati occorre ricordare che tra  $\mathcal{A}$  ed  $\mathcal{E}$ , i più lontani tra i *cluster* del termine in esame è stata misurata una similarità del coseno pari a 0.43. Tra queste vale la pena di sottolineare, per  $\mathcal{F}$ , almeno la presenza di *Legge*, parola già analizzata, nel suo uso ad indicare la Bibbia.

### Risultati principali

Partendo dai dati e non da quanto già acquisito è stato possibile scoprire un cambiamento significativo nell'uso di una parola con un importante significato teologico. Nonostante la scelta di continuare ad usare *Verbo* nel prologo di Gio-

Tabella 5.6: Parola — parole della Bibbia più simili per *cluster*

$\mathcal{A}$	Oracolo	parola	Dice	oracolo	Così
	0.8519	0.7980	0.7101	0.6920	0.5897
$\mathcal{B}$	Oracolo	parola	Dice	oracolo	Preghiera
	0.8573	0.8088	0.6965	0.6944	0.5681
$\mathcal{C}$	parola	Oracolo	Preghiera	parole	messaggio
	0.7855	0.7428	0.5906	0.5794	0.5785
$\mathcal{D}$	parola	messaggio	vangelo	notizia	Vangelo
	0.8670	0.6368	0.5977	0.5676	0.5535
$\mathcal{E}$	parola	vangelo	Vangelo	predicazione	messaggio
	0.8146	0.5962	0.5834	0.5679	0.5424
$\mathcal{F}$	parola	Legge	Bibbia	Chiesa	predicazione
	0.8181	0.5424	0.5294	0.5217	0.5214

vanni potesse far pensare che non ci fossero stati grandi cambiamenti in questo campo semantico, nella revisione della traduzione CEI tra il 1974 ed il 2008 *Parola* viene usata in contesti quasi completamente diversi. L'unica costante è rimasta nell'uso a inizio di frase nei versetti introduttivi degli oracoli profetici, dove la scelta di usare la maiuscola è comunque obbligata.

## 5.2 Rivolta

Il sistema identifica 8 *cluster* di dimensioni rilevanti e 13 di dimensioni ridotte che ospitano un totale di 79 occorrenze. Si può provare anche qui ad effettuare una ulteriore fase di fusione dei *cluster* di dimensioni ridotte, sempre abbassando la soglia della distanza ad una deviazione standard dalla media. In questa ulteriore fase non è stato creato nessun nuovo *cluster*, la metà delle occorrenze è stata inserita in  $\mathcal{F}$ , che raddoppia le sue dimensioni, le altre divise tra  $\mathcal{G}$ ,  $\mathcal{H}$  ed una piccola parte in  $\mathcal{A}$ .

Guardando alle distanze tra i centri (in Figura 5.7) si identificano 4 gruppi di *cluster* ravvicinati:  $\mathcal{A} - \mathcal{B}$  e  $\mathcal{C} - \mathcal{F}$  comunque abbastanza vicini tra loro e poi  $\mathcal{G}$  e  $\mathcal{H}$  distanti da tutto il resto (in particolare  $\mathcal{H}$ ). La stessa situazione si verifica

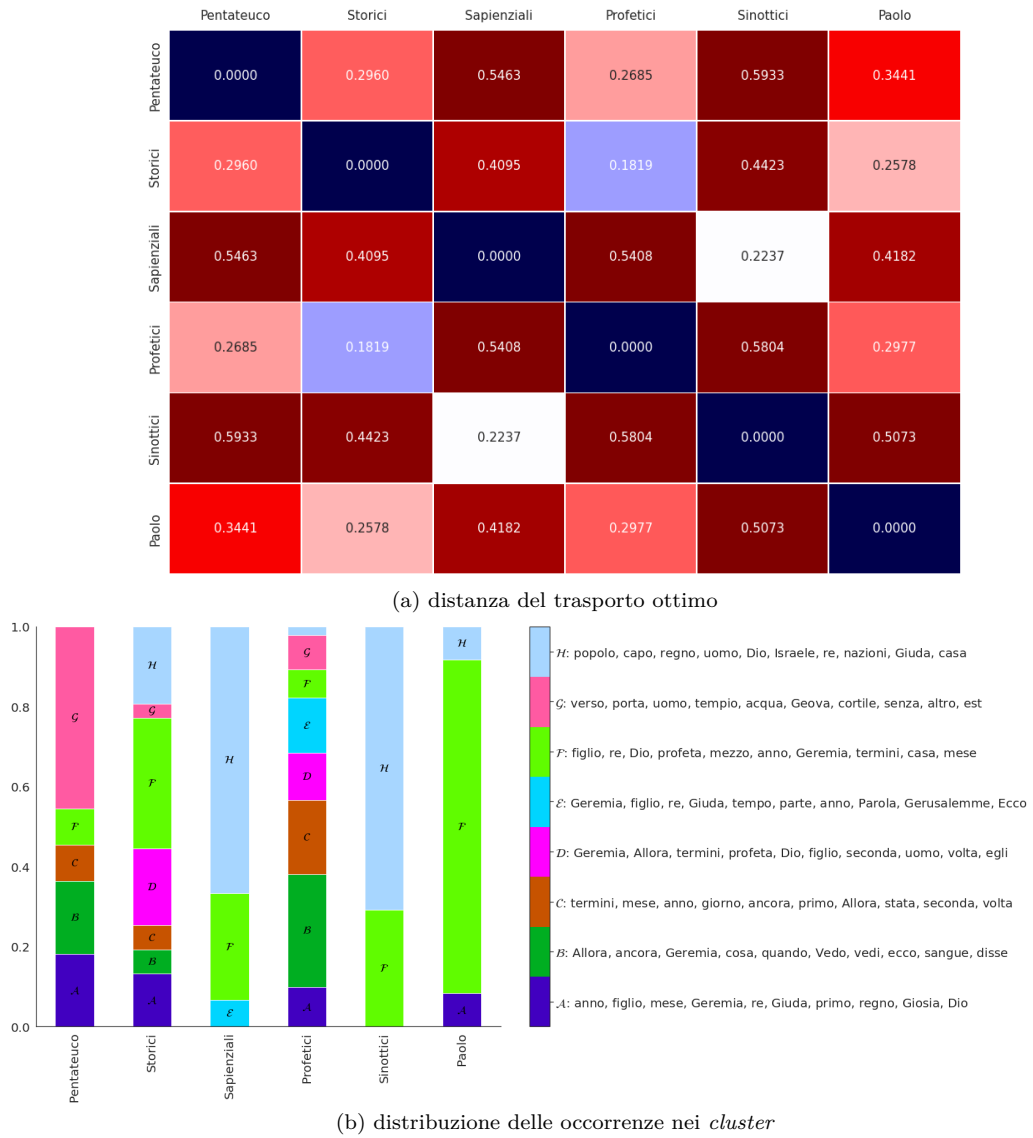


Figura 5.5: Rivolta — confronto tra le principali sezioni della Bibbia

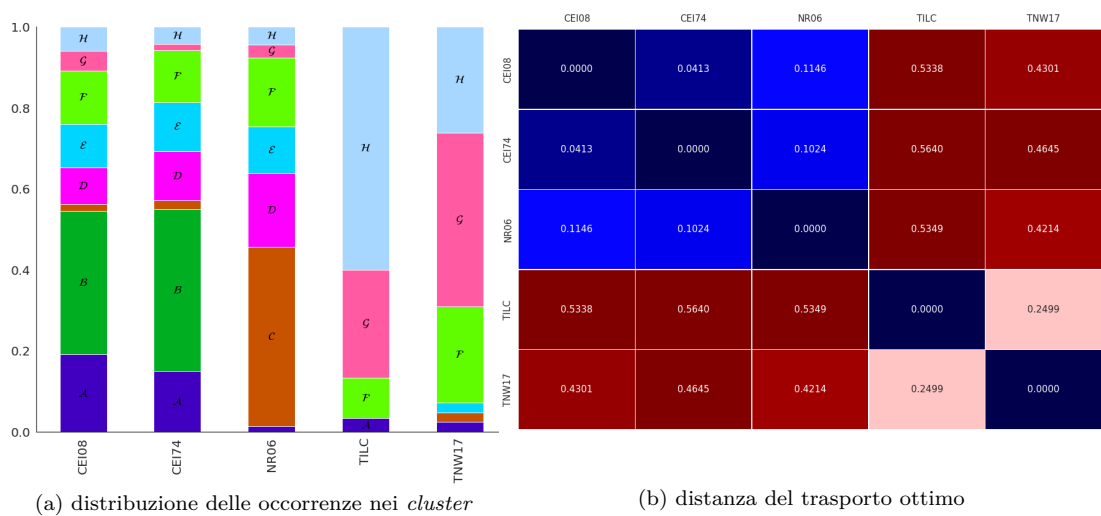


Figura 5.6: Rivolta — confronto tra le traduzioni

calcolando la distanza del trasporto ottimo tra le traduzioni (in Figura 5.6b), con le due CEI e Nuova Riveduta vicine tra loro e distanti dalle altre.

Le sezioni di testo sono tutte piuttosto distanti tra loro, ed anche guardando ai libri al loro interno, solo tra i libri profetici se ne riescono a trovare alcuni con una distanza del trasporto ottimo ridotta. Confrontando le provenienze delle occorrenze nei *cluster*, le distanze si riflettono bene in relazione alle traduzioni, meno in rapporto alle sezioni di testo.

Le due traduzioni della CEI sono molto simili tra loro e alla Nuova Riveduta, a parte  $\mathcal{A} - \mathcal{B}$  che sono sostituiti da  $\mathcal{C}$ . Nella traduzione interconfessionale e in quella del nuovo mondo scompaiono quasi  $\mathcal{A} - \mathcal{E}$  con  $\mathcal{G}$  e  $\mathcal{H}$  molto rilevanti.

Mentre la distribuzione delle occorrenze nei *cluster* rende conto della similarità tra le traduzioni, non sottolinea la differenza tra la traduzione interconfessionale e quella del nuovo mondo che hanno una distanza del trasporto ottimo non trascurabile (0.25). Questo potrebbe però essere causato dalla fase aggiuntiva di fusione dei *cluster* che può aver neutralizzato le differenze dal punto di vista visivo ma non per il problema del trasporto ottimo che considera anche la posizione dei centri.

L'unico *cluster* che ha circa lo stesso peso in tutte le traduzioni è  $\mathcal{F}$ , proprio quello in cui sono confluite la maggior parte delle occorrenze che erano sparse prima dell'intervento aggiuntivo sui *cluster*.  $\mathcal{F}$  è anche l'unico presente, seppur con pesi diversi, in tutte le grandi sezioni di testo. Il parallelo tra  $\mathcal{A} - \mathcal{B}$  e  $\mathcal{C}$  nelle traduzioni si riflette anche nelle sezioni di testo. Soltanto il *corpus* paolino fornisce alcune occorrenze ad  $\mathcal{A}$  ma non a  $\mathcal{C}$ . Nel Pentateuco e nei libri storici e profetici, la grandezza di  $\mathcal{C}$  è proporzionata a quelle di  $\mathcal{A}$  e  $\mathcal{B}$ . Anche  $\mathcal{D}$  ed  $\mathcal{E}$ , sempre provenienti dalle traduzioni CEI e dalla Nuova Riveduta, vengono dall'Antico Testamento, entrambi dei libri profetici,  $\mathcal{D}$  dagli storici ed  $\mathcal{E}$  dai sapienziali.

Dei due *cluster* più rilevanti per la traduzione interconfessionale e per quella del nuovo mondo,  $\mathcal{G}$  è quasi totalmente proveniente dal Pentateuco, mentre  $\mathcal{H}$  ha fonti diverse, ed è particolarmente rilevante nei Sinottici e nei Sapienziali. Anche in questo caso la similarità tra queste due sezioni non è confermata da una distanza del trasporto ottimo più bassa delle altre ma comunque non trascurabile (0.2237).





Figura 5.7: Rivolta — distanza del coseno tra i centri dei *cluster*

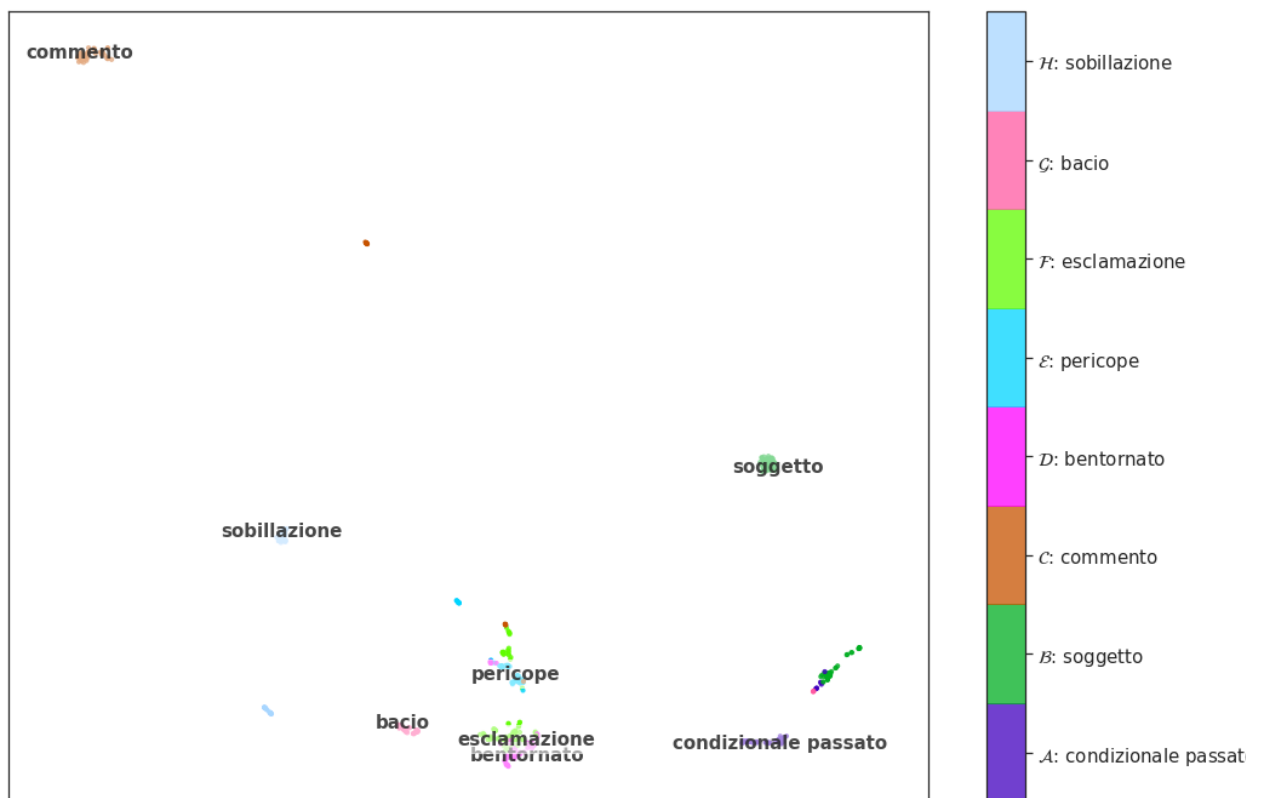


Figura 5.8: Rivolta — visualizzazione in uno spazio bidimensionale, *cluster* etichettati con le definizioni del vocabolario

Guardando ai versetti le cui occorrenze sono incluse in  $\mathcal{F}$  (quello che è cresciuto di più dopo la riduzione del numero di *cluster*), si nota che più della metà degli usi sono simili a «la parola del Signore fu **rivolta** a ...» e provengono dalle traduzioni della CEI o dalla Nuova Riveduta, anche se a volte alcune occorrenze parallele sono inserite nei *cluster*  $\mathcal{A} - \mathcal{E}$ .

In Tabella 5.7 sono raccolti gli altri usi non riconducibili a questo significato. C'è un discreto numero di usi per indicare preghiere rivolte al Signore, specialmente provenienti dalla Nuova Riveduta; di strumenti di violenza rivolti verso qualcuno; di speranza, specie nella traduzione interconfessionale; di mente in quella del nuovo mondo. Tra gli esempi riportati spicca Mi 7, 6 in cui *rivolta* non è il participio passato di rivolgere ma il presente indicativo di rivoltarsi.

Tabella 5.7: Rivolta — versetti in cui occorrono le parole del *cluster*  $\mathcal{F}$  (significati diversi)

Rt	1, 13	... poiché la mano del Signore è <b>rivolta</b> contro di me ». (CEI08, TNW17)
1Sam	1, 17	... il Dio d'Israele esaudisca la preghiera che gli hai <b>rivolta!</b> » (NR06)
1Sam	14, 20	... la spada dell'uno era <b>rivolta</b> contro l'altro ... (NR06)
1Re	8, 38	... ogni supplica che ti sarà <b>rivolta</b> da qualsiasi individuo ... (NR06)
2Re	19, 20	... » Ho udito la preghiera che mi hai <b>rivolta</b> riguardo ... (NR06)
2Cr	6, 29	ogni supplica che ti sarà <b>rivolta</b> da un individuo ... (NR06)
Gb	15, 11	... una parola moderata <b>rivolta</b> a te? (CEI08, CEI74)
Sal	130, 6	L'anima mia è <b>rivolta</b> al Signore più che le sentinelle ... (CEI08)
Pr	8, 4	» Uomini, sto chiamando voi! La mia voce è <b>rivolta</b> a tutti. (TNW17)
Is	37, 21	... Ho udito la preghiera che mi hai <b>rivolta</b> riguardo ... (NR06)
Lam	3, 3	La sua mano è <b>rivolta</b> contro di me di continuo, ... (TNW17)
Dn	5, 10	La regina, ..., <b>rivolta</b> al re, gli disse: ... (CEI08, CEI74)
Mi	7, 6	... la figlia si <b>rivolta</b> contro la madre, ... (CEI08, CEI74)
Sof	2, 5	... La parola del Signore è <b>rivolta</b> contro di te, o Canaan, ... (NR06)
Sof	3, 2	... confidato nel Signore, non si è <b>rivolta</b> al suo Dio. (CEI08, CEI74)
At	25, 24	... tutta la folla dei Giudei si è <b>rivolta</b> a me, ... (CEI08, NR06, TNW17)
At	28, 28	... salvezza di Dio viene ora <b>rivolta</b> ai pagani ... (CEI74, NR06, TILC)
Rm	11, 4	Ma cosa dice la dichiarazione divina a lui <b>rivolta?</b> » ... (TNW17)
2Cor	1, 18	... la parola che vi abbiamo <b>rivolta</b> non è « sì » e « no ». (NR06)
Ef	4, 1	... in modo degno della vocazione che vi è stata <b>rivolta</b> , (NR06)
Fil	3, 19	... e la loro mente è <b>rivolta</b> alle cose terrene. (TNW17)

Col	3, 2	Tenete la mente <b>rivolta</b> alle cose di sopra, ... (TNW17)
1Ts	1, 3	... alla vostra speranza fermamente <b>rivolta</b> verso Gesù Cristo, ... (TILC)
Eb	12, 5	... l'esortazione a voi <b>rivolta</b> come a figli ... (tutte tranne TILC)
Gc	5, 15	E la preghiera <b>rivolta</b> con fede farà star bene il malato, ... (TNW17)
1Pt	1, 13	Tutta la vostra speranza sia <b>rivolta</b> verso quel dono di grazia ... (TILC)
1Pt	3, 21	...ma è invocazione di salvezza <b>rivolta</b> a Dio ... (CEI08, CEI74)

---

Si può concludere che per questo *cluster* la fusione ha funzionato fino ad un certo punto: ha inserito diversi elementi estranei insieme ad alcuni che si sono rivelati simili a quelli iniziali. Ma si è anche verificato un problema alla base con alcune occorrenze parallele e molto simili inserite in *cluster* diversi.

Confrontando i primi sei *cluster* con le parole del vocabolario, alcune parole significative ricorrono in più di uno: *profezia, declamazione, MMS, preghiere rivolte a Dio, pericope, concione*<sup>1</sup>. Oltre a queste è curiosa la presenza di diverse definizioni di termini grammaticali come *soggetto, condizionale passato, esclamazione, frase composta, parola composta*.

La parola più rilevante tra le co-occorrenze nei versetti inclusi in  $\mathcal{G}$  è *verso*, insieme alle altre, fa pensare all'uso di *rivolta* come aggettivo e non seguito da una persona ma da una direzione. Questa intuizione è avvalorata dalle parole selezionate dal vocabolario: *bacio* come in Figura 5.8 con il significato obsoleto di «volto a tramontana, rivolto a nord», *settentrione, nord-ovest, rivolto in basso*.

$\mathcal{H}$ , invece è l'unica che si avvicina ad *insurrezione* tra gli esempi del vocabolario, insieme a *discordia, rivoltare, sedizione, sobillazione* (in Figura 5.8), *sommossa*. Devono essere quindi occorrenze del sostantivo o del verbo rivoltarsi.

Infine, anche in questo caso il confronto con i centri dei *cluster* delle altre parole estratte dalla Bibbia (in Tabella 5.8 le prime cinque per ogni *cluster*) si rivela interessante e capace di spiegare la differenza tra i significati, nonostante delle similarità del coseno leggermente inferiori rispetto a quanto visto sopra per *Parola*.

Escludendo le parole che condividono la stessa radice, per i primi sei *cluster* abbiamo, tra le altre, *pronunciata* e *rivelata*; per  $\mathcal{G}$  abbiamo *guarda* e per  $\mathcal{H}$  *tumulto* e *disordini*: tutte confermano quanto intuito dalle altre analisi.

---

<sup>1</sup>«Discorso solenne tenuto in pubblico».

Tabella 5.8: Rivolta — parole della Bibbia più simili per *cluster*

$\mathcal{A}$	pronunciata	rivelata	rivolte	portata	rivolto
	0.6814	0.6613	0.6487	0.6465	0.6139
$\mathcal{B}$	trasmesso	rivolto	rivolte	pronunciata	mostrata
	0.6106	0.6004	0.5995	0.5972	0.5901
$\mathcal{C}$	rivolto	rivolte	rivolge	rivelata	riferita
	0.7059	0.6870	0.5843	0.5776	0.5767
$\mathcal{D}$	rivolto	rivolte	riferita	pronunciata	rivolse
	0.7718	0.6645	0.6299	0.6054	0.5938
$\mathcal{E}$	rivolto	rivolte	pronunciata	rivolse	rivolge
	0.7575	0.6946	0.6449	0.6267	0.5985
$\mathcal{F}$	rivolto	rivolte	diretta	rivolge	pronunciata
	0.7961	0.7660	0.6512	0.6273	0.6231
$\mathcal{G}$	rivolto	rivolte	rivolti	guardava	guarda
	0.8331	0.7972	0.7184	0.6243	0.6211
$\mathcal{H}$	rivolte	tumulto	disordini	ribellione	lotta
	0.7662	0.6020	0.5358	0.5230	0.5214

### Risultati principali

Grazie a questa analisi è stato possibile intuire che, probabilmente, gli autori della traduzione in lingua corrente avranno ritenuto l'uso del participio passato di rivolgere come tipico del linguaggio biblico e forse anche connotato confessionalmente, viste le differenze tra CEI e Nuova Riveduta e l'assenza dalla traduzione del nuovo mondo.

La presenza del sostantivo e dell'aggettivo in tutte le traduzioni rischiava di considerare la differenza semplicemente in termini di un minore utilizzo della parola se si fossero semplicemente esaminate la frequenza di occorrenza. Anche l'annotazione morfologica, nonostante sia in grado di segnalare qualcosa in più, non avrebbe potuto sottolineare le differenze tra CEI e Nuova Riveduta.

### 5.3 Figlio

*Figlio* viene suddiviso dal sistema in 19 *cluster* di cui solo due sono di dimensioni ridotte. Guardando alle distanze tra i centri dei *cluster* (in Figura 5.9) si notano prima di tutto due gruppi di significati:  $\mathcal{A} - \mathcal{G}$  ed  $\mathcal{H} - \mathcal{Q}$ . Specialmente all'interno del primo gruppo si possono distinguere altre tre suddivisioni:  $\mathcal{A} - \mathcal{C}$ ,  $\mathcal{D} - \mathcal{F}$  e  $\mathcal{G}$ .

La supposta dipendenza della distanza del trasporto ottimo tra le sezioni di testo da quella tra le traduzioni viene smentita da questi dati. Nonostante la grande distanza tra le porzioni di testo a tutti i livelli di granularità, specialmente all'interno dell'Antico Testamento e tra Antico e Nuovo, la distanza del trasporto ottimo massima tra due traduzioni è poco sopra lo 0.1 tra la traduzione interconfessionale e quella del nuovo mondo. Questa similarità si nota anche in Figura 5.10a dalla distribuzione delle occorrenze nei *cluster* delle varie traduzioni.

Del primo gruppo di *cluster*,  $\mathcal{A}$  e  $\mathcal{B}$  provengono soprattutto dalle traduzioni della CEI,  $\mathcal{D}$  proviene quasi esclusivamente dalla Nuova Riveduta ed  $\mathcal{E} - \mathcal{F}$  dalle traduzioni CEI e del nuovo mondo, mentre in quella interconfessionale  $\mathcal{C}$  è molto più esteso che nelle altre. Del secondo gruppo spicca soltanto  $\mathcal{N}$  proveniente solo dalla Nuova Riveduta e dalla traduzione del nuovo mondo.

Le differenze sono ben più visibili tra le sezioni di testo (in Figura 5.10c). La divisione nei due gruppi di *cluster* riproduce quasi esattamente la diversa provenienza delle parole:  $\mathcal{A} - \mathcal{G}$  provengono dall'Antico Testamento e  $\mathcal{H} - \mathcal{Q}$  dal Nuovo.

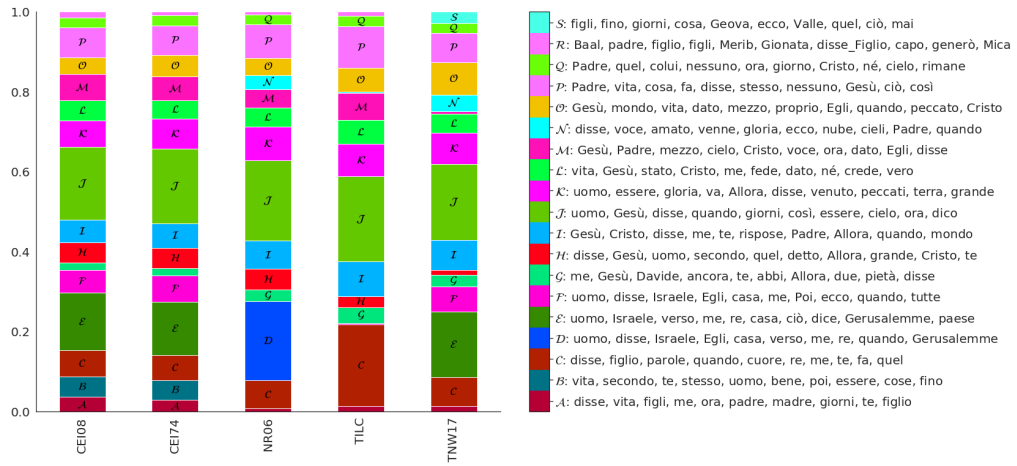
Guardando al confronto con le parole dal vocabolario, tutte quelle selezionate da  $\mathcal{A} - \mathcal{C}$  sono in relazione a figlio come bambino o ragazzo. Da tutti i *cluster* tra  $\mathcal{D}$  e  $\mathcal{Q}$  escono fuori riferimenti a *Gesù*, *Cristo* o *figlio di David*. Quelli provenienti dal Nuovo Testamento hanno in prevalenza parole collegate al mondo religioso cristiano. In  $\mathcal{M} - \mathcal{O}$  sono presenti riferimenti al concepimento e all'Annunciazione.

Dal confronto con i *cluster* delle altre parole della Bibbia la maggior parte dei più simili sono del lemma figlio in varie combinazioni di genere, numero ed iniziale. La presenza di *Ezechiele* in  $\mathcal{E}$  è dovuta al fatto che tutte le occorrenze da quel *cluster* provengono dal suo libro e la traduzione interconfessionale decide

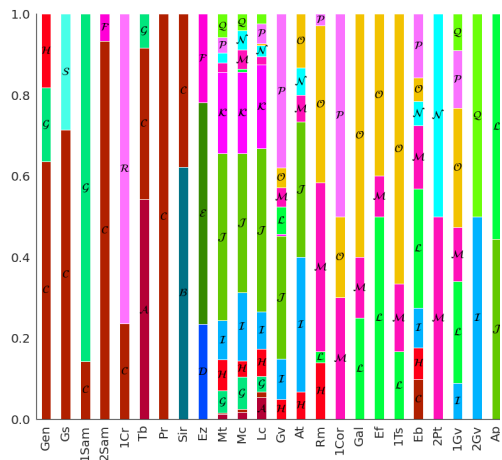
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	0.000	0.092	0.176	0.388	0.381	0.316	0.253	0.437	0.452	0.517	0.524	0.498	0.422	0.376	0.463	0.459	0.482	-0.3897
B	-0.092	0.000	0.157	0.394	0.376	0.358	0.336	0.496	0.526	0.580	0.568	0.553	0.484	0.456	0.515	0.498	0.544	-0.4333
C	0.176	0.157	0.000	0.308	0.314	0.284	0.263	0.500	0.516	0.587	0.581	0.571	0.480	0.491	0.571	0.599	0.618	-0.4385
D	-0.388	0.394	0.308	0.000	0.084	0.087	0.266	0.519	0.545	0.557	0.515	0.609	0.641	0.639	0.695	0.719	0.726	-0.4806
E	-0.381	0.376	0.314	0.084	0.000	0.052	0.273	0.459	0.473	0.437	0.390	0.515	0.557	0.578	0.609	0.628	0.638	-0.4228
F	-0.316	0.358	0.284	0.087	0.052	0.000	0.207	0.408	0.421	0.404	0.359	0.473	0.514	0.532	0.578	0.593	0.604	-0.3868
G	0.253	0.336	0.263	0.266	0.273	0.207	0.000	0.257	0.258	0.355	0.384	0.361	0.384	0.387	0.460	0.509	0.510	-0.3413
H	-0.437	0.496	0.500	0.519	0.459	0.408	0.257	0.000	0.070	0.219	0.245	0.188	0.240	0.261	0.305	0.315	0.355	-0.3296
I	-0.452	0.526	0.516	0.545	0.473	0.421	0.258	0.070	0.000	0.137	0.177	0.098	0.157	0.226	0.238	0.263	0.274	-0.3019
J	-0.517	0.580	0.587	0.557	0.437	0.404	0.355	0.219	0.137	0.000	0.036	0.079	0.136	0.256	0.212	0.214	0.245	-0.3107
K	-0.524	0.568	0.581	0.515	0.390	0.359	0.384	0.245	0.177	0.036	0.000	0.111	0.181	0.309	0.255	0.237	0.301	-0.3232
L	-0.498	0.553	0.571	0.609	0.515	0.473	0.361	0.188	0.098	0.079	0.111	0.000	0.085	0.228	0.147	0.172	0.200	-0.3054
M	-0.422	0.484	0.480	0.641	0.557	0.514	0.384	0.240	0.157	0.136	0.181	0.085	0.000	0.134	0.087	0.120	0.150	-0.2982
N	-0.376	0.456	0.491	0.639	0.578	0.532	0.387	0.261	0.226	0.256	0.309	0.228	0.134	0.000	0.111	0.214	0.243	-0.3402
O	-0.463	0.515	0.571	0.695	0.609	0.578	0.460	0.305	0.238	0.212	0.255	0.147	0.087	0.111	0.000	0.149	0.185	-0.3486
P	-0.459	0.498	0.599	0.719	0.628	0.593	0.509	0.315	0.263	0.214	0.237	0.172	0.120	0.214	0.149	0.000	0.073	-0.3602
Q	-0.482	0.544	0.618	0.726	0.638	0.604	0.510	0.355	0.274	0.245	0.301	0.200	0.150	0.243	0.185	0.073	0.000	-0.3843

Distanza media

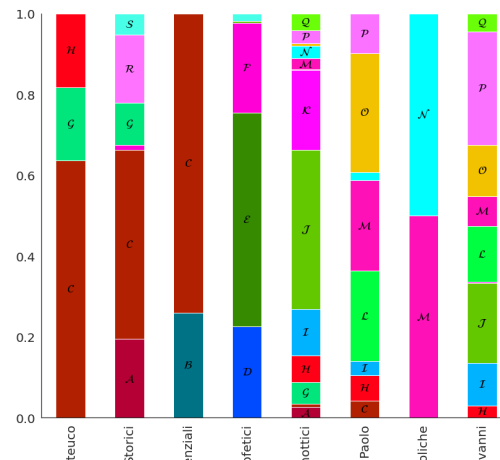
Figura 5.9: Figlio — distanza del coseno tra i centri dei cluster



(a) per traduzione



(b) per libro



(c) per sezione di testo

Figura 5.10: Figlio — distribuzione delle occorrenze nei cluster

Tabella 5.9: Figlio — parole della Bibbia più simili per *cluster*

$A$	figlio	Figli	figli	figlia	Figlia
	0.7175	0.6752	0.6246	0.6046	0.6032
$B$	Figli	figlio	figli	figlia	Figlia
	0.7083	0.6897	0.6338	0.5984	0.5867
$C$	figlio	Figli	figlia	Popolo	Figlia
	0.8387	0.7137	0.6779	0.6760	0.6291
$D$	figlio	Popolo	Figlia	Uomo	Casa
	0.7092	0.5542	0.5408	0.5053	0.5029
$\mathcal{E}$	figlio	Ezechiele	Popolo	Figlia	Figli
	0.6838	0.4983	0.4970	0.4967	0.4917
$\mathcal{F}$	figlio	Uomo	Figli	Popolo	Figlia
	0.7178	0.5413	0.5242	0.5211	0.5119
$\mathcal{G}$	figlio	Uomo	Figli	Figlia	Popolo
	0.7114	0.5739	0.5569	0.5410	0.5253
$\mathcal{H}$	figlio	figli	figlia	Figli	discendenza
	0.7565	0.6673	0.5289	0.4979	0.4852
$\mathcal{I}$	figlio	figli	Figli	Cristo	Signore
	0.6584	0.5938	0.5006	0.4785	0.4670
$\mathcal{J}$	Regno	regno	figlio	Figli	figli
	0.5246	0.5195	0.5107	0.4745	0.4732
$\mathcal{K}$	Regno	regno	figlio	figli	Figli
	0.5096	0.5077	0.4785	0.4562	0.4484
$\mathcal{L}$	figlio	Regno	regno	Cristo	figli
	0.5470	0.5063	0.5004	0.4809	0.4762
$\mathcal{M}$	figlio	figli	Chiesa	Cristo	figlia
	0.5574	0.5011	0.4760	0.4595	0.4568
$\mathcal{N}$	figlio	figlia	figli	bambino	Chiesa
	0.6378	0.5199	0.5118	0.4889	0.4743
$\mathcal{O}$	figlio	Chiesa	figlia	Messia	figli
	0.6153	0.5131	0.4945	0.4762	0.4759
$\mathcal{P}$	figlio	Chiesa	figli	Padre	Regno
	0.5475	0.4953	0.4950	0.4934	0.4690
$\mathcal{Q}$	figlio	figli	Chiesa	Padre	figlia
	0.5338	0.4984	0.4773	0.4703	0.4401

di sostituire all'espressione *Figlio dell'uomo* il nome del profeta, come in Ezechiele 3, 1 in cui « Figlio dell'uomo, mangia ciò che ti sta davanti, mangia questo rotolo, poi va' e parla alla casa d'Israele » (CEI08) diventa « Ezechiele, mangia questo rotolo. Poi va' e parla al popolo d'Israele » (TILC).

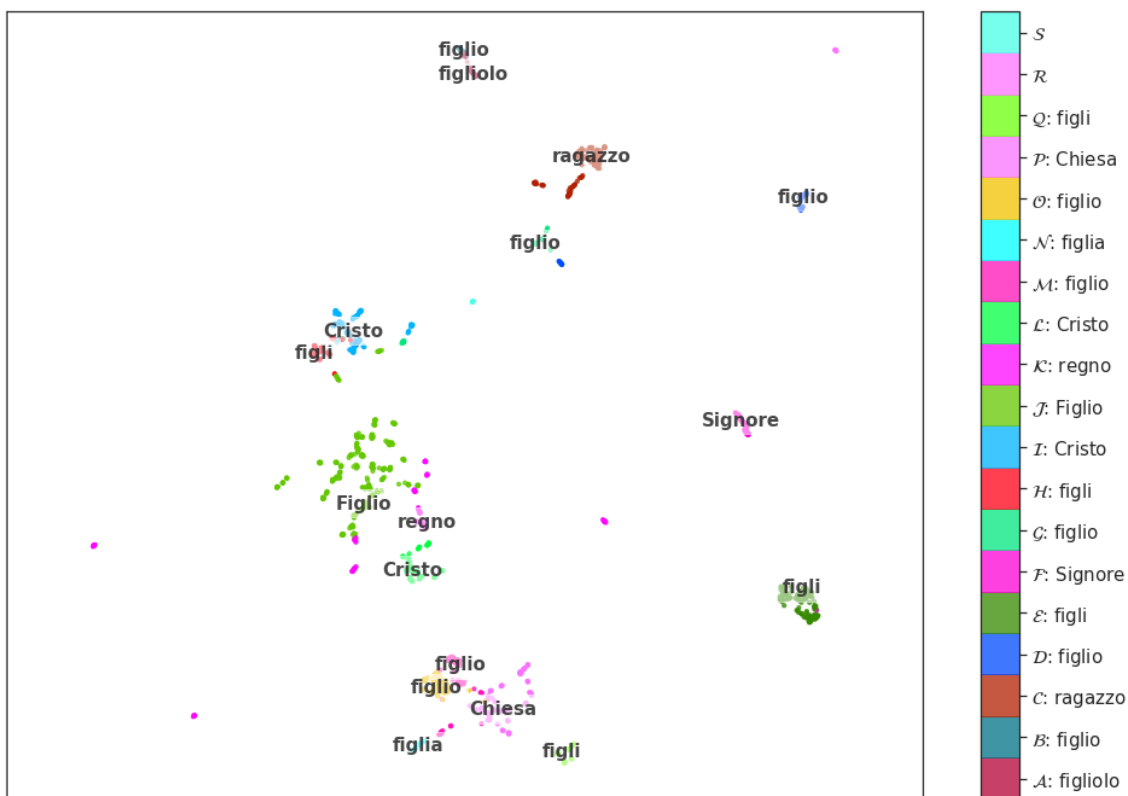
Un altro cambiamento messo in atto dal traduttore è la sostituzione di *casa* con *popolo*. In questo senso di casato, famiglia o popolo è probabile che vada letto il *Casa* che compare tra le più simili a  $\mathcal{D}$ .

Tra  $\mathcal{J}$  ed  $\mathcal{L}$  compare il riferimento al regno, questa parola occorre solo in due dei circa 50 versetti da cui provengono le parole incluse in quei *cluster*, ma includono parabole del regno o annunci messianici.

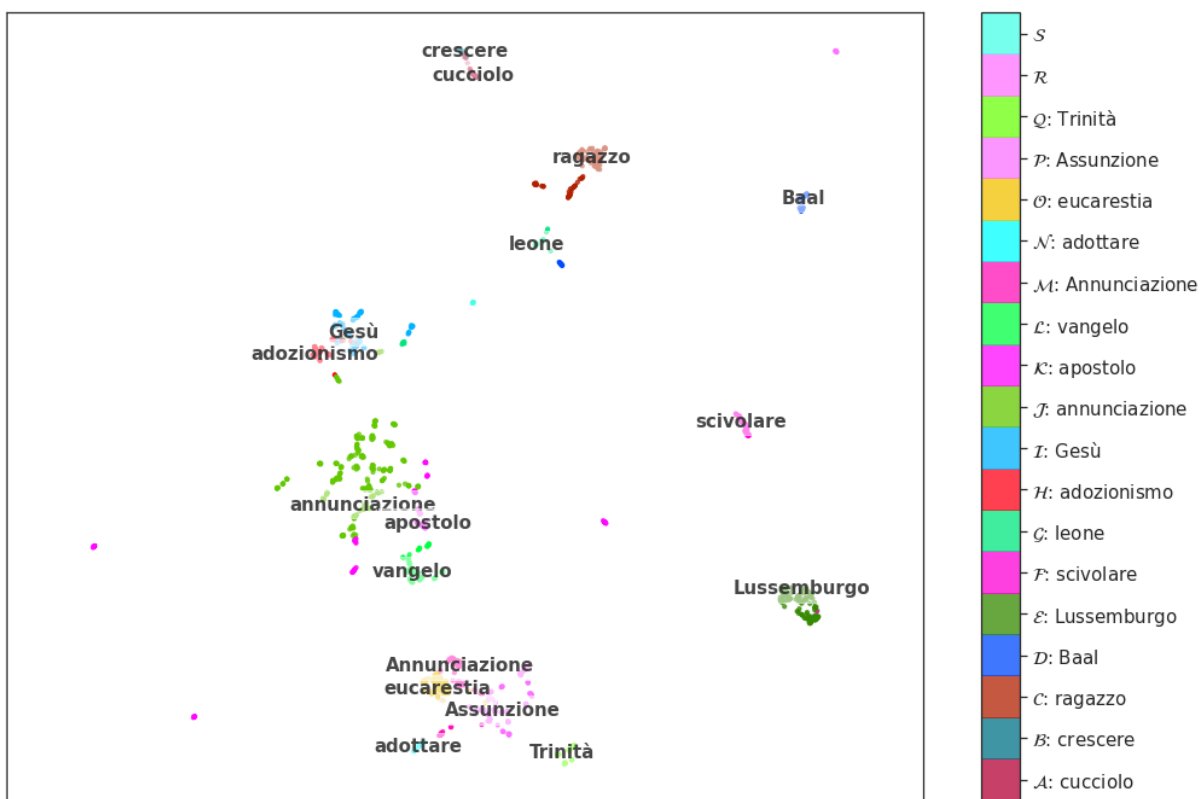
### Risultati principali

In quest'ultima analisi, quello che colpisce sono le potenzialità del confronto con i *cluster* delle altre parole, anche grazie alla presenza di traduzioni diverse. Nonostante la presenza predominante delle forme flesse di figlio, si ricavano indicazioni preziose sul significato dietro ai *cluster* che né il vocabolario né le parole co-occorrenti avevano permesso di cogliere. Guardando alle similarità del coseno ottenute si può notare ad esempio che il centroide di  $\mathcal{E}$  è più vicino ad *Ezechiele* che ai *cluster*  $\mathcal{L} - \mathcal{Q}$  o che  $\mathcal{D}$  è più vicino a *Popolo*, *Uomo* e *Casa* più che a  $\mathcal{H} - \mathcal{Q}$ , ben dieci dei diciassette totali.





(a) cluster etichettati con gli esempi del vocabolario



(b) cluster etichettati con le definizioni del vocabolario

Figura 5.11: Figlio — visualizzazione in uno spazio bidimensionale

## Capitolo 6

### I risultati – ricerca dei versetti

Creando le rappresentazioni vettoriali dei versetti è possibile riprodurre il procedimento già visto, applicando algoritmi di *clustering* ma usando criteri di selezione diversa rispetto a quelli usati per le parole. Non avendo un'etichetta che permetta di classificarli, si possono usare le indicazioni di collocazione del versetto o al più si potrebbe riprodurre lo stesso lavoro fatto con le parole prendendo tutti i versetti che contengono uno o più termini.

Piuttosto che riprodurre lo stesso lavoro già visto, in questo capitolo si cercherà di ottenere altre indicazioni interessanti dal lavoro sui versetti.

Nelle cinque traduzioni prese in esame, visto il diverso approccio alla traduzione, non è possibile stabilire una rapporto biunivoco tra le parole. Un simile confronto sarebbe forse possibile solo tra due traduzioni interlineari. A livello di versetti è invece possibile stabilire una certa corrispondenza anche in relazione al testo di partenza.

Per ottenere gli *embedding* di ogni versetto è stato sfruttato il lavoro fatto, sommando i vettori delle parole che erano già stati estratti dal modello tenendo traccia del versetto e della traduzione di provenienza.

Create le rappresentazioni vettoriali dei versetti, si può scegliere una misura di distanza per quantificare la differenza di significato tra due traduzioni. Scegliendo un indicatore che sintetizzi le distanze tra ogni coppia di traduzioni si può determinare l'accordo tra i traduttori.

Un forte disaccordo può essere causato ad esempio dalle diversità stilistiche, dalla difficoltà che si incontrano nel tradurre il versetto o dalla particolare inter-

pretazione che ne viene data in alcuni contesti confessionali. Su questo tipo di ricerca si soffermeranno le sezioni 6.1 e 6.2.

La ricerca del versetto più simile tra due traduzioni, al contrario, può fornire informazioni sulle differenze di numerazione e diventare uno strumento utile per mappare gli slittamenti. Allo stesso modo, all'interno di una stessa traduzione, è possibile trovare i passi paralleli tra diversi libri o le frasi che si ripetono all'interno dello stesso.

L'applicazione più utile di uno strumento capace di indicare il versetto più simile può essere la ricerca della provenienza di una citazione quando non si hanno indicazioni che possano restringere il campo e quando non si è certi della letteralità della citazione. Nell'ultima parte di questo capitolo (6.3) si proverà a stimare la capacità di fornire la risposta esatta di uno strumento realizzato per quest'ultimo tipo di ricerca.

## 6.1 I versetti del Nuovo Testamento con maggiore disaccordo

Il Nuovo Testamento è un buon punto di partenza, non dovrebbero esserci grandi differenze di numerazione e nel testo di riferimento. Abbiamo preso come riferimento la traduzione interconfessionale in lingua corrente perché è quella che potrebbe causare problemi di allineamento della numerazione per la presenza di versetti raggruppati.

La traduzione interconfessionale divide il Nuovo Testamento in 7869 versetti, di questi, 33 sono assenti in qualche traduzione<sup>1</sup> e 64 sono il frutto della fusione di più versetti dove è stato ritenuto opportuno non rispettare la suddivisione per rendere più semplice la lettura senza essere vincolati a forme estranee all'italiano standard.

---

<sup>1</sup>I motivi possono essere molteplici. La traduzione del nuovo mondo esclude intere pericopi non considerandole autentiche perché incompatibili con il messaggio. Le altre traduzioni, invece, possono aver scelto di escludere qualche singolo versetto perché ritenuto una ripetizione o un inserimento del copiatore in seguito alle ricerche filologiche sulle fonti più antiche effettuate in seguito alla divisione in versetti. Non essendoci in tutti i casi discussi un consenso degli studiosi ed essendo state fatte in tempi diversi, traduttori diversi possono aver preso decisioni diverse.

In questa analisi sono stati esclusi i primi perché, lavorando con solo cinque traduzioni, escluderne anche una rischiava di falsare i risultati. Visto il numero esiguo di versetti raggruppati non è stata dedicata loro una graduatoria a parte perché non ci sarebbero stati campioni sufficienti per un confronto significativo. Comunque sono stati confrontati, come preventivato, con la media dei versetti corrispondenti nelle altre traduzioni.

Come misura del disaccordo nella traduzione di un versetto è stata scelta la media delle distanze del coseno, calcolata per ogni coppia di traduzioni. La mediana, che avrebbe potuto essere una misura più stabile, è stata scartata proprio perché non avrebbe considerato la presenza di una sola traduzione diversa. Usare il massimo, invece, avrebbe indicato soprattutto i versetti in cui anche solo un traduttore non avesse concordato con gli altri. In Figura 6.1 è riportata questa misura del disaccordo medio dei traduttori del Nuovo Testamento. Solo poco più dell'1% dei versetti ha una distanza media del coseno superiore a 0.2, per più del 90% è inferiore a 0.15 e in più della metà dei versetti è inferiore a 0.1.

In Tabella 6.1 sono riportati i dieci versetti con maggiore disaccordo secondo il modello realizzato in questo lavoro, con la distanza media di ciascuna traduzione dalle altre. La traduzione mediamente più differente varia, in alcuni casi sono tutte diverse l'una dall'altra, in altri ce ne è una che si distingue dalle altre. Sono presenti nella lista sia versetti molto brevi che più lunghi, in alcuni la differenza è probabilmente dovuta ad una diversa suddivisione dei versetti (At 15, 18; 2, 21; Rm 15, 10).

Tabella 6.1: Versetti del NT con maggiore disaccordo

..... <i>At 15, 18</i> .....		
CEI08	0.35	note da sempre.
CEI74	0.35	dice il Signore che fa queste cose da lui conosciute dall'eternità.
NR06	0.30	a lui note fin dall'eternità ”.
TILC	0.39	perché le vuole da sempre.
TNW17	0.36	note dall'antichità’.
..... <i>1Cor 13, 9</i> .....		
CEI08	0.32	Infatti, in modo imperfetto noi conosciamo e in modo imperfetto profetizziamo.
CEI74	0.32	La nostra conoscenza è imperfetta e imperfetta la nostra profezia.

- NR06 0.35 poiché noi conosciamo in parte, e in parte profetizziamo;
- TILC 0.37 La scienza è imperfetta, la profezia è limitata,
- TNW17 0.30 Infatti abbiamo una conoscenza incompleta e profetizziamo in modo  
incompleto,  
.....*Rm 5, 4*.....
- CEI08 0.25 la pazienza una virtù provata e la virtù provata la speranza.
- CEI74 0.30 e la virtù provata la speranza.
- NR06 0.30 la pazienza, esperienza, e l'esperienza, speranza.
- TILC 0.38 la perseveranza ci rende forti nella prova, e questa forza ci apre alla  
speranza.
- TNW17 0.39 la perseveranza, a sua volta, una condizione approvata; la  
condizione approvata, a sua volta, speranza;  
.....*1Cor 13, 7*.....
- CEI08 0.27 Tutto scusa, tutto crede, tutto spera, tutto sopporta.
- CEI74 0.26 Tutto copre, tutto crede, tutto spera, tutto sopporta.
- NR06 0.26 soffre ogni cosa, crede ogni cosa, spera ogni cosa, sopporta ogni  
cosa.
- TILC 0.49 Chi ama è sempre comprensivo, sempre fiducioso, sempre paziente,  
sempre aperto alla speranza.
- TNW17 0.26 Copre ogni cosa, crede ogni cosa, spera ogni cosa, sopporta ogni  
cosa.  
.....*Lc 20, 30*.....
- CEI08 0.25 Allora la prese il secondo
- CEI74 0.25 Allora la prese il secondo
- NR06 0.31 Il secondo pure la sposò,
- TILC 0.37 Anche il secondo
- TNW17 0.33 Allora a sposare la donna fu il secondo,  
.....*At 2, 21*.....
- CEI08 0.59 E avverrà:.
- CEI74 0.22 Allora chiunque invocherà il nome del Signore sarà salvato.
- NR06 0.20 E avverrà che chiunque avrà invocato il nome del Signore sarà  
salvato ”.
- TILC 0.22 Allora, chiunque invocherà il nome del Signore sarà salvo.
- TNW17 0.25 E chiunque invocherà il nome di Geova sarà salvato ”’.

		..... <i>Gv 3, 30</i> .....
CEI08	0.23	Lui deve crescere; io, invece, diminuire ”.
CEI74	0.23	Egli deve crescere e io invece diminuire.
NR06	0.37	Bisogna che egli cresca e che io diminuisca.
TILC	0.36	È lui che deve diventare importante. Io invece devo mettermi da parte.
TNW17	0.25	Lui deve continuare a crescere, mentre io devo continuare a diminuire ”.
		..... <i>Rm 15, 10</i> .....
CEI08	0.52	E ancora:.
CEI74	0.22	E ancora: Rallegratevi, o nazioni, insieme al suo popolo.
NR06	0.21	E ancora: « Rallegratevi, o nazioni, con il suo popolo ».
TILC	0.27	E ancora: Nazioni, rallegratevi con il popolo che Dio ha scelto.
TNW17	0.22	E ancora viene detto: ” Gioite, nazioni, con il suo popolo ”.
		..... <i>Mt 3, 8</i> .....
CEI08	0.23	Fate dunque un frutto degno della conversione,
CEI74	0.22	Fate dunque frutti degni di conversione,
NR06	0.24	Fate dunque dei frutti degni del ravvedimento.
TILC	0.46	Dimostrate con i fatti che avete cambiato vita
TNW17	0.28	Producete quindi frutti che dimostrino pentimento.
		..... <i>1Cor 14, 32</i> .....
CEI08	0.23	Le ispirazioni dei profeti sono sottomesse ai profeti,
CEI74	0.22	Ma le ispirazioni dei profeti devono essere sottomesse ai profeti,
NR06	0.27	Gli spiriti dei profeti sono sottoposti ai profeti,
TILC	0.40	Chi profetizza deve controllare il suo dono.
TNW17	0.27	E il dono dello spirito dei profeti dev’essere controllato dai profeti.

---

Per arrivare al versetto più discusso in cui la traduzione interconfessionale ha raggruppato i versetti bisogna raggiungere la cinquantaquattresima posizione, occupata da seconda lettera di Pietro 1, 20-21.

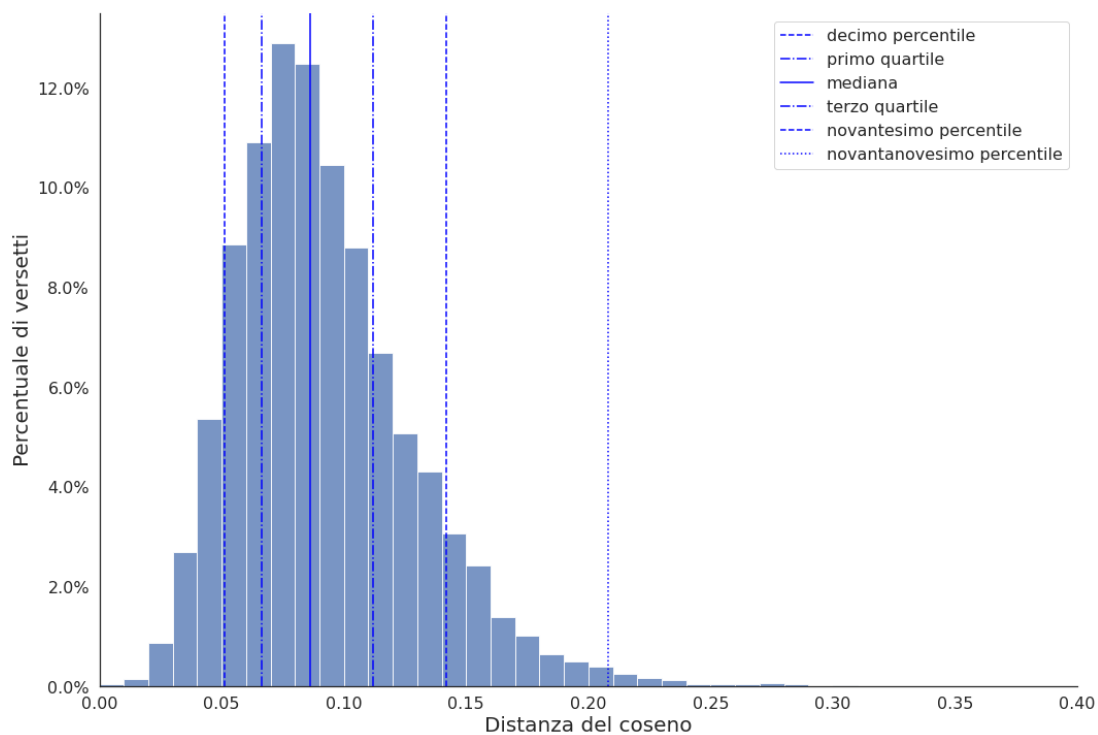


Figura 6.1: Distanza media tra le traduzioni dei versetti del NT

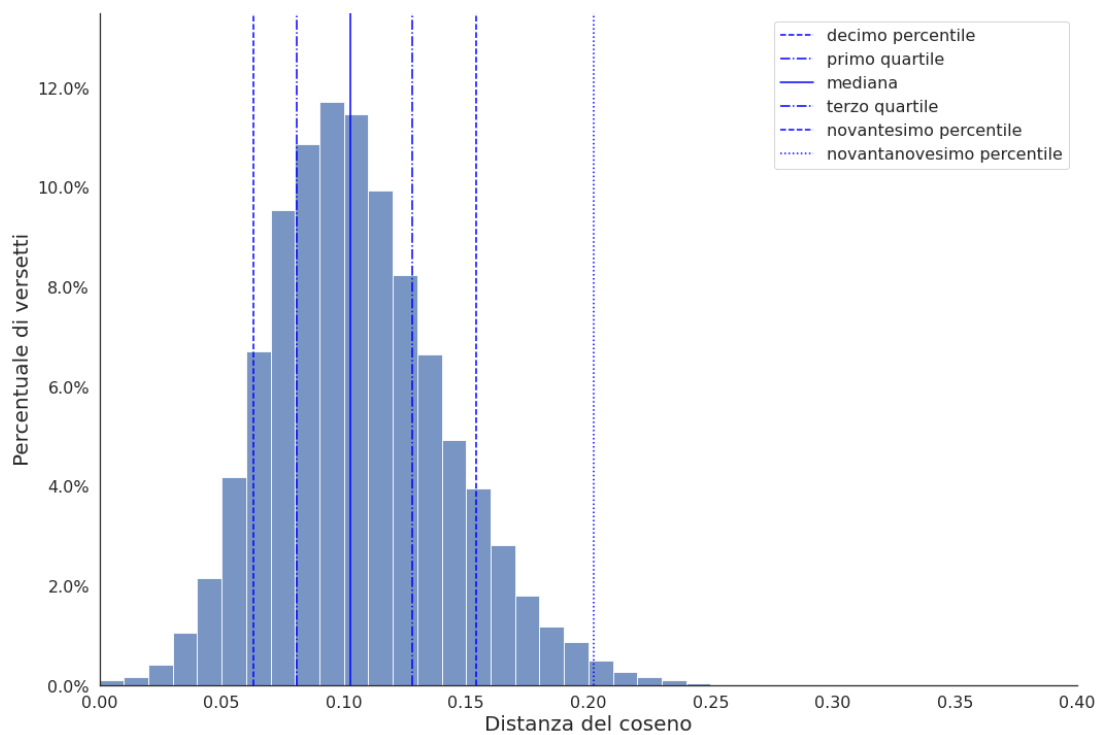


Figura 6.2: Distanza media tra le traduzioni dei versetti dell'AT

## 6.2 I versetti dell'Antico Testamento con maggiore disaccordo

Nel confronto tra i versetti dell'Antico Testamento entrano in gioco i problemi di allineamento del testo, con la diversità di numerazione dei versetti. Alcuni disallineamenti sono noti e sarebbe anche possibile scegliere una numerazione ed allineare tutti a quella. Rimarrebbe comunque disallineato un certo numero di versetti, non tutti chiaramente indicati, e riconoscibili con certezza di fatto solo da un esame diretto del testo. In alcuni casi questi versetti sono spostati tra la fine di un capitolo e l'inizio del successivo, causando lo slittamento della numerazione di tutti i versetti di quel capitolo.

Per calcolare l'accordo medio cercheremo quindi per ogni versetto il più simile nelle altre traduzioni, tra quelli dello stesso libro: avremo così sia una misura della similarità sia una possibile indicazione dei versetti con numerazione disallineata tra due traduzioni. In questo modo si otterrà una stima per difetto del disaccordo nel caso in cui il versetto più simile trovato non sia la traduzione dello stesso testo di partenza.

Come per il Nuovo Testamento erano stati esclusi i versetti che mancavano in una qualche traduzione, sono stati presi in considerazione solo i libri inclusi nel canone ristretto dell'Antico Testamento. Questo non esclude che, come ad esempio per le aggiunte greche del libro di Daniele, ci siano versetti che non abbiano il proprio parallelo in alcune delle traduzioni prese in considerazione.

L'accordo dei traduttori diminuisce leggermente rispetto al Nuovo Testamento. Come si vede in Figura 6.2 la mediana e gli altri quantili sono tutti maggiori di quelli riscontrati sopra. Continua ad esserci solo circa l'1% di versetti con una distanza media superiore a 0.2.

In Tabella 6.2 sono riportati i dieci versetti con maggiore disaccordo, per ciascuna traduzione è riportata anche la distanza media dalle altre. Nel caso in cui la numerazione di uno dei versetti identificati non corrisponda con quella degli altri, sono riportati anche capitolo e versetto di provenienza tra parentesi tonde ed il testo del corrispondente tra quadre.



Tabella 6.2: Versetti dell'AT con maggiore disaccordo

..... <i>Gb 23, 14</i> .....		
CEI08	0.69	Egli esegue il decreto contro di me come pure i molti altri che ha in mente.
CEI74	0.66	Compie, certo, il mio destino e di simili piani ne ha molti.
NR06	0.75	egli eseguirà quel che di me ha decretato; di cose come queste ne ha molte in mente.
TILC	0.67	Che bel destino! Dalla polvere già ne spunta un'altra ». (8, 19) [Farà quel che ha deciso per me, come decide ogni cosa per tutti.]
TNW17	0.71	Eseguirà pienamente ciò che ha decretato riguardo a me, e ha in mente tante altre cose come queste.
..... <i>Gb 8, 19</i> .....		
CEI08	0.76	Ecco la gioia del suo destino e dalla terra altri rispuntano.
CEI74	0.76	Ecco la gioia del suo destino e dalla terra altri rispuntano.
NR06	0.70	Tali sono le delizie che gli procura il suo comportamento! Dalla polvere, altri dopo di lui germoglieranno.
TILC	0.67	Che bel destino! Dalla polvere già ne spunta un'altra ».
TNW17	0.67	È così che scomparirà, mentre altri spunteranno dalla polvere.
..... <i>Gb 28, 8</i> .....		
CEI08	0.73	non lo calpestano le bestie feroci, non passa su di esso il leone.
CEI74	0.67	non battuto da bestie feroci, né mai attraversato dal leopardo.
NR06	0.74	Le fiere superbe non vi hanno messo piede, il leone non c'è mai passato.
TILC	0.73	Laggiù i leoni non sono mai arrivati, né vi sono passati altri animali feroci.
TNW17	0.73	Nessun animale imponente ci cammina sopra, né il giovane leone vi si aggira per cacciare.
..... <i>Lam 3, 20</i> .....		
CEI08	0.77	Ben se ne ricorda la mia anima e si accascia dentro di me.
CEI74	0.76	Ben se ne ricorda e si accascia dentro di me la mia anima.
NR06	0.74	Io me ne ricordo sempre, e ne sono intimamente prostrato.
TILC	0.69	Ricordo, ricordo tutto e mi sento abbattuto.
TNW17	0.67	Sicuramente te ne ricorderai e ti chinerali su di me.

..... *Gb 36, 17* .....

- CEI08 0.73 Ma se di giudizio iniquo sei pieno, giudizio e condanna ti seguiranno.  
 CEI74 0.76 Ma se colmi la misura con giudizi da empio, giudizio e condanna ti seguiranno.  
 NR06 0.74 Ma se giudichi le vie di Dio come fanno gli empi, il suo giudizio e la sua sentenza ti piomberanno addosso.  
 TILC 0.71 ma tu ora sei stato giudicato malvagio, sei sotto il giudizio e la condanna.  
 TNW17 0.69 Sarai soddisfatto del giudizio contro i malvagi, quando verrà emessa la sentenza e la giustizia verrà difesa.

..... *Qo 5, 8* .....

- CEI08 0.75 In ogni caso, la terra è a profitto di tutti, ma è il re a servirsi della campagna.  
 CEI74 0.74 l'interesse del paese in ogni cosa è un re che si occupa dei campi.  
 NR06 0.73 Ma vantaggioso per un paese è, per ogni rispetto, un re che si occupi dei campi. (5, 9)  
 [Chi osserva il comandamento non conosce disgrazia, e il cuore dell'uomo saggio sa che c'è un tempo e un giudizio,]  
 TILC 0.69 Tutti traggono beneficio dalla terra, anche il re dipende dai campi.  
 TNW17 0.74 Inoltre il prodotto della terra viene diviso fra tutti; persino il re si serve di quello che viene dai campi. (5, 9)  
 [Chi osserva i comandi non subirà danno, e il cuore saggio conosce i tempi e i modi giusti.]

..... *2Sam 22, 45* .....

- CEI08 0.71 stranieri cercavano il mio favore, all'udirmi, subito mi obbedivano,  
 CEI74 0.73 I figli degli stranieri mi onorano appena sentono, mi obbediscono.  
 NR06 0.78 I figli degli stranieri mi hanno reso omaggio, al solo udir parlare di me, mi hanno ubbidito.  
 TILC 0.73 A una mia parola, essi ubbidiscono, gente estranea si inchina davanti a me.  
 TNW17 0.69 Stranieri verranno a inchinarsi servilmente; ciò che udranno su di me li spingerà a ubbidirmi.

..... <i>Dn 5, 25</i> .....		
CEI08	0.74	E questo è lo scritto tracciato: Mene, Tekel, Peres,
CEI74	0.69	di cui è questa la lettura: mene, tekkel, peres,
NR06	0.75	Ecco le parole che sono state scritte: " Mené, Mené, Techel, U - Parsin ".
TILC	0.77	Ecco quel che c'è scritto: Mene, Mene, Tekel, Parsin.
TNW17	0.69	Questa è la scritta che è stata tracciata: MENÈ, MENÈ, TECHEL e PARSIN.
..... <i>Dt 2, 16</i> .....		
CEI08	0.75	Quando da mezzo al popolo scomparvero per morte tutti quegli uomini atti alla guerra,
CEI74	0.75	Quando tutti quegli uomini atti alla guerra furono passati nel numero dei morti,
NR06	0.70	Quando la morte ebbe fatto scomparire tutti quei guerrieri,
TILC	0.74	« Dopo che quegli uomini capaci di combattere furono morti,
TNW17	0.72	" Una volta morti tutti gli uomini di guerra del popolo,
..... <i>Ger 6, 2</i> .....		
CEI08	0.76	La bella e incantevole figlia di Sion io riduco al silenzio.
CEI74	0.66	E' forse simile a un tenero prato la figlia di Sion?
NR06	0.76	La bella, la voluttuosa figlia di Sion, io la distruggo!
TILC	0.74	Sion, la figlia, è bella e incantevole, eppure sarà distrutta.
TNW17	0.74	La figlia di Sion assomiglia a una donna bella e raffinata.

---

I versetti delle aggiunte a Daniele non compaiono tra i primi dieci, è presente un versetto dal libro di Daniele ma, come si nota, è tra quelli riportati in tutte le traduzioni.

Il libro di Giobbe, con quattro tra i prime cinque versetti potrebbe essere quello che ha creato più dubbi di traduzione. Soltanto nel primo di questi versetti il modello ha sbagliato ad identificare il parallelo in una traduzione. Invece, come si può vedere dal testo, in *Qohelet* 5, 8 il sistema ha riconosciuto il versetto che presenta una numerazione differente: nella Nuova Riveduta e nella traduzione del nuovo mondo è slittato di un versetto in avanti.

### 6.3 Il confronto con le antifone

Da questi primi risultati il modello sembra essere in grado di riconoscere i versetti corrispondenti superando il rumore dovuto al diverso stile traduttivo con un buon tasso di successo.

In una situazione reale, il rumore inserito nel testo può essere dovuto ad una parafrasi, all'uso di sinonimi o ad una citazione che non corrisponda ai confini di un versetto.

Per avere una stima dell'efficacia del modello in queste situazioni abbiamo usato alcune citazioni della Bibbia usati nella liturgia cattolica. Dai libri liturgici disponibili in rete sono stati estratte le citazioni in cui erano riportati i versetti di provenienza. Per semplicità di esposizione, useremo da qui in avanti il termine «antifone» per riferirci a queste citazioni ma sono state estratte anche citazioni più brevi o più articolate.

In Figura 6.3 sono riportate le distanze del coseno tra il vettore dell'antifona e la somma dei versetti di provenienza indicati. Come traduzione per il confronto abbiamo usato la traduzione della CEI del 1974, dalla quale dovrebbe provenire la maggior parte di esse e alla cui numerazione in genere ci si riferiva. La mediana è poco meno di 0.14, non molto maggiore di quanto visto tra le traduzioni, ma la coda è molto più lunga con il 10% dei dati più distanti di 0.33.

Limitandoci alle antifone in cui era stato riportato il riferimento ad un solo versetto, è stato possibile fare un confronto con tutti gli *embedding* dei versetti per trovare il più simile. In Figura 6.4 sono riportate le distanze tra l'antifona ed il versetto più simile. La forma dell'istogramma cambia sensibilmente nonostante la mediana resti pressoché invariata. Aumentano i casi con distanze molto basse, quasi il 25% delle antifone ha una distanza inferiore allo 0.02, e diminuiscono quelle con distanze maggiori.

Tra le citazioni più problematiche per il sistema (in Tabella 6.3) ci sono chiari errori di numerazione (Ct 4, 9; Sal 21, 1; 59, 5), parafrasi (Sal 8, 8; 17, 3; Mt 19, 12), una citazione in latino (Gv 19, 30), qualche errore dell'OCR (Gv 1, 18; Rm 6, 4; Sal 30, 6) e quasi tutte sono citazioni parziali di un versetto.

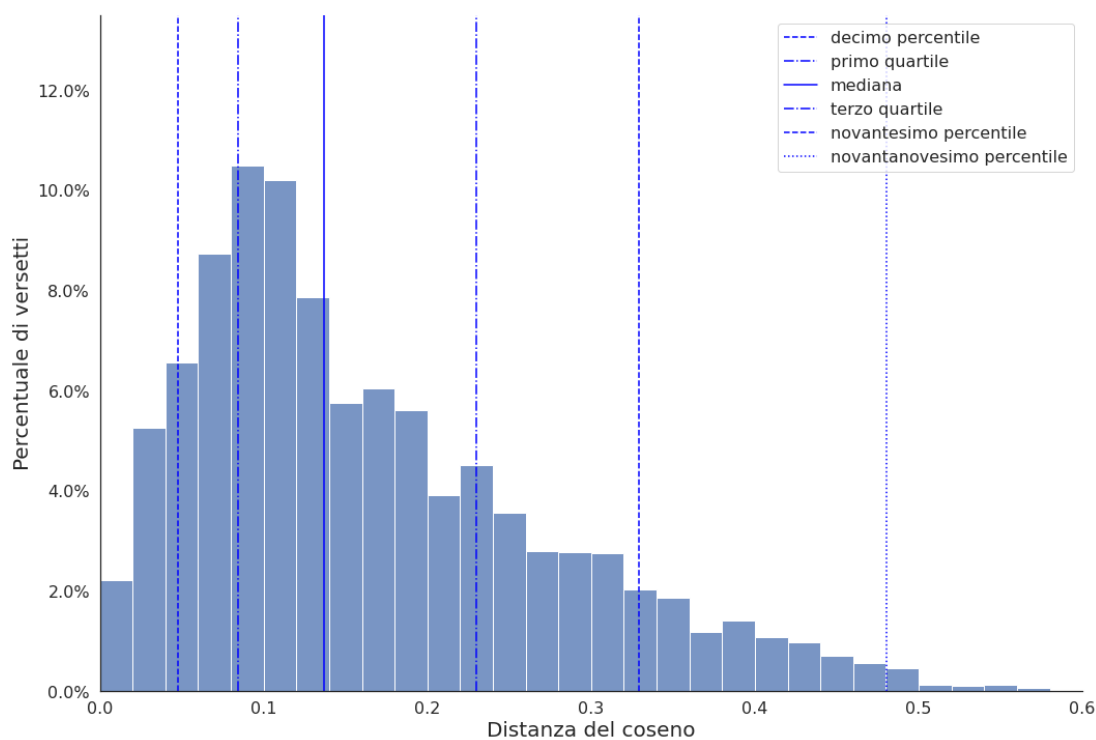


Figura 6.3: Distanza media tra l'antifona ed i versetti corrispondenti (CEI74)

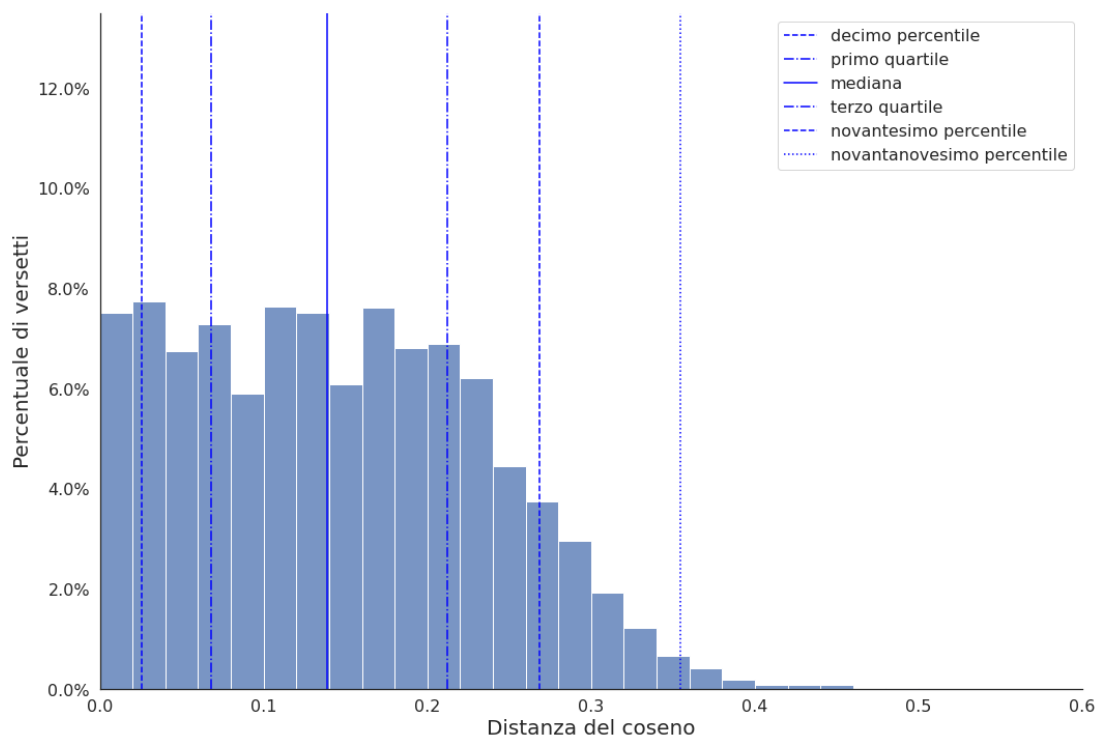


Figura 6.4: Distanza media tra l'antifona ed il versetto più simile

Tabella 6.3: Antifone più distanti dal versetto di partenza

<b>Antifona</b>	<b>Versetto</b>
.....	..... <i>Ct 4, 9 (0.66)</i> .....
Io sono ferita d'amore	Tu mi hai rapito il cuore, sorella mia, sposa, tu mi hai rapito il cuore con un solo tuo sguardo, con una perla sola della tua collana!
.....	..... <i>Gv 1, 18 (0.65)</i> .....
Nes-suno mai vide Dio	Dio nessuno l'ha mai visto: proprio il Figlio unigenito, che è nel seno del Padre, lui lo ha rivelato.
.....	..... <i>Gv 19, 30 (0.65)</i> .....
«Consummatum est»	E dopo aver ricevuto l'aceto, Gesù disse: " Tutto è compiuto! ". E, chinato il capo, spirò.
.....	..... <i>Sal 8, 8 (0.63)</i> .....
Tutto infatti ha sottomesso ai suoi piedi	tutti i greggi e gli armenti, tutte le bestie della campagna;
.....	..... <i>Rm 6, 4 (0.61)</i> .....
«Cam mi nia-mo in una vita nuova»	Per mezzo del battesimo siamo dunque stati sepolti insieme a lui nella morte, perché come Cristo fu risu- scitato dai morti per mezzo della gloria del Padre, così anche noi possiamo camminare in una vita nuova.
.....	..... <i>Sal 17, 3 (0.60)</i> .....
Io metterò la mia fiducia in lui	Signore, mia roccia, mia fortezza, mio liberatore; mio Dio, mia rupe, in cui trovo riparo; mio scudo e baluardo, mia potente salvezza.
.....	..... <i>Mt 19, 12 (0.58)</i> .....
La castità abbracciata in vista del Regno dei cieli	Vi sono infatti eunuchi che sono nati così dal ven- tre della madre; ve ne sono alcuni che sono stati resi eunuchi dagli uomini, e vi sono altri che si sono fatti eunuchi per il regno dei cieli. Chi può capire, capisca ". .....
.....	..... <i>Sal 21, 1 (0.58)</i> .....
«Dio mio, Dio mio, perché mi hai abbandonato?»	Al maestro del coro. Sull'aria: " Cerva dell'aurora ". Salmo. Di Davide.

.....	<i>Sal 30, 6 (0.57)</i> .....
«Mi affido alle tue ma- 1282 6 febbraio ni»	Mi affido alle tue mani; tu mi riscatti, Signore, Dio fedele.
.....	<i>Sal 59, 5 (0.56)</i> .....
«Svegliati, vieni incontro a me e guarda»	Hai inflitto al tuo popolo dure prove, ci hai fatto bere vino da vertigini.

---

Gli errori di numerazione sono gli unici ad essere veramente un problema per la bontà della nostra stima di efficacia, ma la difficoltà di eliminarli è proprio uno dei motivi dell'utilità di un sistema capace di fare questa ricerca. Le altre cause che sembrano aver originato valori alti della distanza possono essere riconducibili a situazioni di un uso reale, per questo è stato deciso di non intervenire correggendoli per effettuare una nuova stima.

Il tasso di successo nell'identificare il versetto riportato (o il corrispondente in una delle altre traduzioni) è stato del 69%, ma solo nel 40% di questi casi (il 27.5% del totale) il versetto individuato è stato quello della traduzione che avevamo considerato come fonte (CEI74).

Escludendo dal confronto i versetti nelle altre traduzioni si ottiene comunque un tasso di successo appena inferiore (il 68%) e differenze minime anche sulle distanze.

Tabella 6.4: Antifone più distanti dal versetto riconosciuto dal sistema

<b>Versetto riconosciuto</b>	<b>Versetto indicato</b>
.....	<i>Nes - suno mai vide Dio (0.49)</i> .....
Quando disperdeva i re l'Onnipotente, nevicava sullo Zalmon (Sal 67, 15 CEI74)	Dio nessuno l'ha mai visto: proprio il Figlio unigenito, che è nel seno del Padre, lui lo ha rivelato. (Gv 1, 18)
.....	<i>«Mi affido alle tue ma- 1282 6 febbraio ni» (0.48)</i> .....
I rimpatriati possedevano 435 cammelli e 6. 720 asini. (Ne 7, 68 TILC)	Mi affido alle tue mani; tu mi riscatti, Signore, Dio fedele. (Sal 30, 6)

..... «Cam mi nia - mo in una vita nuova» (0.46) .....

Zenan, Adasa, Migdal - Gad, (Gs 15, 37 TNW17)	Per mezzo del battesimo siamo dunque stati sepolti insieme a lui nella morte, perché come Cristo fu risuscitato dai morti per mezzo della gloria del Padre, così anche noi possiamo camminare in una vita nuova. (Rm 6, 4)
--	---

..... Magni - ficate il Signore con me (0.45) .....

Magnificate con me il Signore, esaltiamo insieme il suo nome. (Sal 34, 4 CEI08)	GHIMEL. Celebrate con me il Signore, esaltiamo insieme il suo nome. (Sal 33, 4)
--	--

..... andategli incontro (0.44) .....

Andarono quindi da lui per parlargliene. (1Mac 9, 59 TILC)	A mezzanotte si levò un grido: Ecco lo sposo, andategli incontro! (Mt 25, 6)
---	---

..... Co - sa attendo ora se non il Signore? (0.43) .....

Sono io forse Dio solo da vicino - dice il Signore - e non anche Dio da lontano? (Ger 23, 23 CEI74)	Ora, che attendo, Signore? In te la mia speranza. (Sal 38, 8)
---	--

..... «Consummatum est» (0.43) .....

Danna, Kiriath - Sannà ( Debir ), (Gs 15, 49 TILC)	E dopo aver ricevuto l'aceto, Gesù disse: " Tutto è compiuto! ". E, chinato il capo, spirò. (Gv 19, 30)
---	---

..... «e stava loro sottomesso» (0.42) .....

Scese dunque con loro e venne a Nazaret e stava loro sottomesso. Sua madre cu- stodiva tutte queste cose nel suo cuore. (Lc 2, 51 CEI08)	Partì dunque con loro e tornò a Nazaret e stava loro sottomesso. Sua madre serbava tutte queste cose nel suo cuore. (Lc 2, 51)
---	--

..... «Vi mando come agnel - li» (0.41) .....

« Andate! Io vi mando come agnelli in mezzo ai lupi. (Lc 10, 3 TILC)	Andate: ecco io vi mando come agnelli in mezzo a lupi; (Lc 10, 3)
---	--

..... Io sono ferita dall'amore? (0.41) .....

Ma vi conosco: non avete in voi l'amore di Dio. (Gv 5, 42 CEI08)	Sostenetemi con focacce d'uva passa, rin- fratemi con pomi, perché io sono malata d'amore. (Ct 2, 5)
---	--



Tra le dieci citazioni con distanza maggiore dal versetto più simile riconosciuto (in Tabella 6.4) non tutte erano già presenti nella lista precedente.

Nei primi tre casi e nella citazione latina, il versetto riconosciuto come più simile è piuttosto difficile da associare al testo di partenza, ma negli altri la somiglianza che ha portato all'errore di riconoscimento è abbastanza comprensibile.

È piuttosto confortante per le prestazioni del modello che tra queste ultime 10 ci siano anche tre casi (Sal 33, 4; Lc 2, 51; 10, 3) in cui, nonostante la distanza, è stato riconosciuto il versetto giusto.

Inoltre, nel caso in cui si cerchi la provenienza di una frase, si potrebbe essere disposti a controllare più di un solo risultato, specialmente di fronte alla difficoltà di alcune ricerche in cui non si trova una parola chiave che possa aiutare nella ricerca sul testo o su una concordanza.

Considerando i primi cinque versetti più simili nel confronto con la sola traduzione CEI del 1974, il modello trova il versetto indicato nell'82% dei casi, una percentuale sensibilmente maggiore di quella ottenuta sul primo versetto riconosciuto.

# Conclusioni

L'analisi delle parole ha fornito molti spunti interessanti, alcuni di questi costituivano il motivo della scelta, altri erano attesi. Nell'analisi è stato difficile cercare di prescindere dalle conoscenze teologiche e bibliche acquisite ed anche le informazioni non esplicitamente attese sono state spesso analizzate attraverso il filtro di queste conoscenze.

Il modello si è comunque rivelato utile per rispondere a qualche domanda anche in questo lavoro dove la vera questione era più la capacità stessa del modello di aiutare a dare una risposta che il contenuto in sé della risposta. Si è cercato quindi di provare a fare domande diverse piuttosto che andare in profondità su una o poche questioni.

Tra i molti risultati interessanti ottenuti nel corso delle analisi, uno ha destato particolarmente l'attenzione: il successo dell'algoritmo di *clustering* nel creare insiemi di occorrenze di provenienza omogenea nonostante fosse un algoritmo non supervisionato e non addestrato allo scopo.

Che parole o forme diverse confluissero in *cluster* diversi si sarebbe forse potuto dare per scontato, ma il passaggio della traduzione, limando inevitabilmente le differenze sintattiche e riducendo anche più quelle di vocabolario, avrebbe potuto rendere più difficile al modello ottenere qualche risultato.

Non rientrava negli scopi di questo lavoro, ma i risultati analizzati impongono di sottolineare le conclusioni di Montariol, Martinc e Pivovarova 2021 riguardo alla misura del cambiamento semantico. Considerare sia la posizione del centro del *cluster* che le frequenze di occorrenza ha permesso di riconoscere e misurare somiglianze e differenze che sarebbero sfuggite considerando solo le frequenze. Nei casi in cui sembravano fornire indicazioni diverse, l'esame dei testi e le conoscenze acquisite hanno confermato le indicazioni del problema del trasporto ottimo.

L'utilità degli strumenti matematici non erano in discussione, però merita spazio una considerazione che non è stata fatta nel corpo della tesi. La banale applicazione del problema del commesso viaggiatore<sup>2</sup> ai centri dei *cluster* ha permesso di riordinarli semplicemente ed automaticamente in modo da aumentare nettamente l'interpretabilità immediata e visiva dei dati. Questo non è stato utile solamente quando c'era da valutare le distanze tra i baricentri, ma anche nel confronto con il vocabolario e nelle altre osservazioni fatte.

Il lavoro con questo modello ha dato numerosi spunti su prospettive future di un approccio simile al mondo biblico. Inizialmente sono state individuate nei versetti le unità atomiche del testo ed il modello è stato quindi addestrato versetto per versetto, con una notevole semplificazione del lavoro. Per altri versi può essere stato limitante, non sono state fornite informazioni al modello sulla relazione tra i versetti e sono stati forniti contesti molto più piccoli (e quindi molte meno informazioni) rispetto a quelli che sarebbe stato in grado di gestire. D'altra parte quasi ogni suddivisione (e ordinamento) del testo è discutibile, in alcuni casi persino la divisione tra libri. La scelta ragionata di qualsiasi altra suddivisione richiederebbe studi specifici e prese di posizione decise.

Restando a questo come modello di partenza la stessa questione potrebbe essere affrontata con un *fine-tuning* sulla previsione della frase successiva. Altre estensioni di questo modello potrebbero riguardare l'ampliamento del *corpus* di addestramento per includere letteratura apocrifia o intertestamentaria, letteratura cristiana antica o testi del magistero e della liturgia: tutti questi dipendono in qualche modo dal testo biblico e possono essere anche averne influenzato le traduzioni successive.

In alternativa, un modello specialistico dedicato alla ricerca biblica più approfondita non potrebbe non svincolarsi dalle traduzioni. Prima di affrontare una simile impresa andrebbero sicuramente valutate le risorse a disposizione. Richiederebbe infatti una ingente raccolta di altro materiale nelle lingue bibliche e di pensare ad un addestramento da zero.

---

<sup>2</sup>Pappalardo e Passacantando 2004.

# Appendice A

## Spirito: versetti nel *cluster Q*

Per le parole incluse in altri *cluster*, esso è indicato tra parentesi, se *spirito* è reso con altre parole, queste sono indicate in corsivo.

Tabella A.1: Spirito — confronto tra le traduzioni dei versetti in cui occorrono le parole in *Q*

	..... <i>1Sam 16, 14</i> .....
CEI08	Lo <b>spirito</b> ( <i>O</i> ) del Signore si era ritirato da Saul e cominciò a turbarlo un cattivo <b>spirito</b> ( <i>P</i> ), venuto dal Signore.
CEI74	Lo <b>spirito</b> ( <i>O</i> ) del Signore si era ritirato da Saul ed egli veniva atterrito da uno <b>spirito</b> cattivo, da parte del Signore.
NR06	Lo <b>Spirito</b> ( <i>A</i> ) del Signore si era ritirato da Saul; e uno <b>spirito</b> ( <i>P</i> ) cattivo, permesso dal Signore, lo turbava.
TILC	Lo <b>spirito</b> ( <i>O</i> ) del Signore aveva abbandonato Saul. Ora egli era tormentato da uno <b>spirito</b> maligno mandato dal Signore.
TNW17	Ora Saul era stato abbandonato dallo <b>spirito</b> ( <i>N</i> ) di Geova, e uno <b>spirito</b> cattivo da parte di Geova lo terrorizzava.
	..... <i>1Sam 16, 15</i> .....
CEI08	... ” Ecco, un cattivo <b>spirito</b> ( <i>P</i> ) di Dio ti turba.
CEI74	... ” Vedi, un cattivo <b>spirito</b> ( <i>P</i> ) sovrumano ti turba.
NR06	... « Ecco, un cattivo <b>spirito</b> ( <i>P</i> ) permesso da Dio ti turba.
TILC	... — Vedi, Dio ha inviato uno <b>spirito</b> maligno a tormentarti.
TNW17	... ” Hai visto che uno <b>spirito</b> cattivo da parte di Dio ti terrorizza.
	..... <i>1Sam 16, 16</i> .....
CEI08	... Quando il cattivo <b>spirito</b> ( <i>O</i> ) di Dio sarà su di te, ...
CEI74	... Quando il sovrumano <b>spirito</b> cattivo ti investirà, ...

- NR06 ... quando il cattivo **spirito** ( $\mathcal{P}$ ) permesso da Dio verrà su di te, ...
- TILC ... Così, quando lo **spirito** cattivo ti investirà, ...
- TNW17 ...Ogni volta che uno **spirito** cattivo da parte di Dio ti affiggerà, ...  
 ..... *1Sam 16, 23* .....
- CEI08 Quando dunque lo **spirito** ( $\mathcal{O}$ ) di Dio era su Saul, ... si sentiva meglio e lo **spirito** cattivo si ritirava da lui.
- CEI74 Quando dunque lo **spirito** sovrumano investiva Saul, ... si sentiva meglio e lo **spirito** cattivo si ritirava da lui.
- NR06 Or quando il cattivo **spirito** ( $\mathcal{Q}$ ) permesso da Dio veniva su Saul, ...stava meglio e il cattivo **spirito** ( $\mathcal{Q}$ ) andava via da lui.
- TILC Quando lo **spirito** maligno inviato da Dio investiva Saul, ... si calmava e lo **spirito** cattivo si allontanava da lui.
- TNW17 Ogni volta che uno **spirito** cattivo da parte di Dio affliggeva Saul, ... e trovava sollievo e lo **spirito** cattivo si ritirava da lui.  
 ..... *1Sam 18, 10* .....
- CEI08 ... un cattivo **spirito** ( $\mathcal{O}$ ) di Dio irruppe su Saul, ...
- CEI74 ... un cattivo **spirito** ( $\mathcal{P}$ ) sovrumano s'impadronì di Saul, ...
- NR06 ... un cattivo **spirito** ( $\mathcal{P}$ ) permesso da Dio si impadronì di Saul, ...
- TILC ... uno **spirito** maligno mandato da Dio s'impadronì di Saul ...
- TNW17 ... Saul fu afflitto da uno **spirito** cattivo da parte di Dio ...  
 ..... *1Sam 19, 9* .....
- CEI08 Ma un cattivo **spirito** ( $\mathcal{O}$ ) del Signore fu su Saul. ...
- CEI74 Ma un sovrumano spirito **cattivo** ( $\mathcal{P}$ ) si impadronì di Saul. ...
- NR06 Allora uno **spirito** ( $\mathcal{P}$ ) cattivo, permesso dal Signore, s'impadronì di Saul. ...
- TILC Ma uno **spirito** maligno, mandato dal Signore, s'impadronì di Saul. ...
- TNW17 Ora Saul ... fu sopraffatto da uno **spirito** cattivo da parte di Geova.  
 ..... *Tb 6, 8* .....
- CEI08 ... uomo o donna, invasata dal demonio o da uno **spirito** cattivo, ...
- CEI74 ... uomo o donna, invasata dal demonio o da uno **spirito** cattivo ...
- TILC ... un uomo o a una donna tormentati dal demonio o da qualche **spirito** maligno ...

- ..... *Zc 13, 2* .....
- CEI08 ... anche i profeti e lo **spirito** ( $\mathcal{L}$ ) di impurità farò sparire dal paese.
- CEI74 ... anche i profeti e lo **spirito** immondo farò sparire dal paese.
- NR06 ... anche i profeti e gli *spiriti* immondi farò sparire dal paese.
- TILC ... Farò sparire dalla regione quelli che si dicono profeti e il disgustoso desiderio di adorare gli idoli.
- TNW17 ... e farò sparire dal paese i profeti e lo **spirito** ( $\mathcal{L}$ ) di impurità.  
..... *Mt 9, 32* .....
- CEI08 ... un muto *indemoniato*.
- TILC ... un uomo che non poteva parlare a causa di uno **spirito** maligno.  
..... *Mt 9, 33* .....
- CEI08 E dopo che il *demonio* fu scacciato, quel muto cominciò a parlare. ...
- TILC Quando Gesù scacciò questo **spirito**, il muto si mise a parlare. ...  
..... *Mt 12, 22* .....
- CEI08 In quel tempo fu portato a Gesù un *indemoniato*, cieco e muto, ...
- TILC Portarono a Gesù un uomo che era cieco e muto, perché uno **spirito** maligno era dentro di lui. ...  
..... *Mt 12, 43* .....
- CEI08 Quando lo **spirito** impuro esce dall'uomo, ...
- CEI74 Quando lo **spirito** immondo esce da un uomo, ...
- NR06 « Quando lo **spirito** immondo esce da un uomo, ...
- TILC « Quando uno **spirito** maligno è uscito da un uomo ...
- TNW17 » Quando uno **spirito** impuro esce da un uomo, ...  
..... *Mt 15, 22* .....
- CEI08 ... Mia figlia è molto tormentata da un *demonio* ».
- TILC ... Mia figlia sta molto male, uno **spirito** maligno la tormenta.
- TNW17 Mia figlia è *indemoniata* e soffre moltissimo ».  
..... *Mc 1, 23* .....
- CEI08 ... un uomo posseduto da uno **spirito** impuro ...
- CEI74 Allora un uomo ... posseduto da uno **spirito** immondo, ...
- NR06 ... un uomo posseduto da uno **spirito** immondo, ...
- TILC ... un uomo tormentato da uno **spirito** maligno. ...
- TNW17 ... un uomo posseduto da uno **spirito** impuro, ...

- ..... *Mc 1, 26* .....
- CEI08 E lo **spirito** impuro, straziandolo ...
- CEI74 E lo **spirito** immondo, straziandolo ...
- NR06 E lo **spirito** immondo, straziandolo ...
- TILC Allora lo **spirito** maligno scosse con violenza quell'uomo, ...
- TNW17 E lo **spirito** impuro, dopo aver fatto venire all'uomo le convulsioni ...
- ..... *Mc 3, 30* .....
- CEI08 ... " È posseduto da uno **spirito** impuro ".
- CEI74 ... " E' posseduto da uno **spirito** immondo ".
- NR06 ... « Ha uno **spirito** immondo ».
- TILC ... « Uno **spirito** maligno è dentro di lui ».
- TNW17 ... " È posseduto da uno **spirito** impuro ".
- ..... *Mc 5, 2* .....
- CEI08 ... un uomo posseduto da uno **spirito** ( $\mathcal{P}$ ) impuro.
- CEI74 ... un uomo posseduto da uno **spirito** immondo.
- NR06 ... un uomo posseduto da uno **spirito** immondo,
- TILC ... Costui era tormentato da uno **spirito** maligno
- TNW17 ... un uomo posseduto da uno **spirito** impuro ...
- ..... *Mc 5, 8* .....
- CEI08 ... " Esci, **spirito** impuro, da quest'uomo! ".
- CEI74 ... " Esci, **spirito** immondo, da quest'uomo! ".
- NR06 ... « **Spirito** ( $\mathcal{T}$ ) immondo, esci da quest'uomo! »
- TILC ... disse allo **spirito** maligno di uscire da quell'uomo;
- TNW17 ... " Esci da quest'uomo, **spirito** impuro! ".
- ..... *Mc 7, 25* .....
- CEI08 ... la cui figlioletta era posseduta da uno **spirito** impuro, ...
- CEI74 ... che aveva la sua figlioletta posseduta da uno **spirito** immondo, ...
- NR06 ... la cui bambina aveva uno **spirito** immondo, ...
- TILC ... sua figlia era tormentata da uno **spirito** maligno.
- TNW17 ... che aveva una bambina posseduta da uno **spirito** impuro ...

- ..... *Mc 7, 29-30* .....
- CEI08 ... va': il *demonio* è uscito da tua figlia ”.  
... e il *demonio* se n'era andato.
- TILC ... Torna a casa tua: lo **spirito** maligno è uscito da tua figlia.  
... lo **spirito** maligno se n'era andato.  
..... *Mc 9, 17* .....
- CEI08 ... mio figlio, che ha uno **spirito** muto.
- CEI74 ... mio figlio, posseduto da uno **spirito** muto.
- NR06 ... mio figlio che ha uno **spirito** muto;
- TILC ... mio figlio perché è tormentato da uno **spirito** maligno che non lo lascia parlare.
- TNW17 ... mio figlio perché ha uno **spirito** muto.  
..... *Mc 9, 18* .....
- TILC Quando lo prende, dovunque si trovi, lo getta a terra, ... Ho chiesto ai tuoi discepoli di scacciare questo **spirito**, ma non ci sono riusciti.
- TNW17 Ogni volta che lo **spirito** (*P*) lo attacca, lo sbatte a terra, ... Ho chiesto ai tuoi discepoli di espellerlo, ma non ci sono riusciti ”.  
..... *Mc 9, 25* .....
- CEI08 ... minacciò lo **spirito** impuro dicendogli: ” **Spirito** (*T*) muto e sordo, io ti ordino, esci da lui ...
- CEI74 ... minacciò lo **spirito** immondo dicendo: ” **Spirito** (*T*) muto e sordo, io te l'ordino, esci da lui ...
- NR06 ... sgridò lo **spirito** immondo, dicendogli: « **Spirito** (*T*) muto e sordo, io te lo comando, esci da lui ...
- TILC ... minacciò lo **spirito** maligno dicendo: « **Spirito** (*T*) che impedisce di parlare e di ascoltare, esci da questo ragazzo ... Te lo ordino! ».
- TNW17 ... rimproverò lo **spirito** impuro, dicendogli: ” **Spirito** (*T*) muto e sordo, te lo ordino: esci da lui ...  
..... *Mc 9, 28* .....
- CEI08 ... ” Perché noi non siamo riusciti a scacciarlo? ”.
- TILC ... — Perché noi non siamo stati capaci di scacciare quello **spirito**?  
..... *Lc 4, 33* .....
- CEI08 ... c'era un uomo che era posseduto da un *demonio* impuro; ...
- NR06 ... si trovava un uomo che aveva uno **spirito** di *demonio* impuro, ...



- TILC ... c'era un uomo posseduto da uno **spirito** maligno ...
- TNW17 ... c'era un uomo posseduto da uno **spirito**, un demonio impuro, ...  
 ..... *Lc 4, 35* .....
- CEI08 ... E il *demonio* lo gettò a terra in mezzo alla gente ...
- TILC ... Allora lo **spirito** maligno gettò a terra quel pover'uomo davanti a tutti ...  
 ..... *Lc 8, 29* .....
- CEI08 Gesù aveva ordinato allo **spirito** impuro di uscire da quell'uomo. Molte volte infatti si era impossessato di lui; ...
- CEI74 Gesù infatti stava ordinando allo **spirito** immondo di uscire da quell'uomo. Molte volte infatti s'era impossessato di lui; ...
- NR06 Gesù, infatti, aveva comandato allo **spirito** immondo di uscire da quell'uomo, di cui si era impadronito da molto tempo; ...
- TILC Parlava così perché Gesù stava comandando allo **spirito** maligno di uscire da quell'uomo. Molte volte infatti quello **spirito** ( $\mathcal{P}$ ) si era impossessato di lui. ...
- TNW17 ( Gesù, infatti, aveva ordinato allo **spirito** impuro di uscire dall'uomo. Parecchie volte questo si era impossessato di lui, ...  
 ..... *Lc 8, 30* .....
- CEI08 Gesù gli domandò: " Qual è il tuo nome? " ...
- TILC Gesù domandò allo **spirito** maligno: — Come ti chiami? ...  
 ..... *Lc 9, 39* .....
- CEI08 Ecco, uno **spirito** ( $\mathcal{P}$ ) lo afferra ... lo scuote, ...
- CEI74 Ecco, uno **spirito** ( $\mathcal{P}$ ) lo afferra ... lo scuote ...
- NR06 Ecco, uno **spirito** ( $\mathcal{P}$ ) si impadronisce di lui ... lo **spirito** ( $\mathcal{P}$ ) lo contorce, ...
- TILC Talvolta uno **spirito** maligno lo assale, ... Poi gli fa venire le convulsioni ...
- TNW17 Ed ecco, uno **spirito** ( $\mathcal{P}$ ) se ne impadronisce, ... Lo **spirito** ( $\mathcal{P}$ ) gli provoca convulsioni ...  
 ..... *Lc 9, 40* .....
- CEI08 Ho pregato i tuoi discepoli di scacciarlo, ...
- TILC Ho chiesto ai tuoi discepoli di scacciare questo **spirito** maligno, ...  
 ..... *Lc 9, 42* .....
- CEI08 ... il *demonio* lo gettò a terra ... minacciò lo **spirito** impuro, ...
- CEI74 ... il *demonio* lo gettò per terra ...minacciò lo **spirito** immondo, ...
- NR06 ... il *demonio* lo gettò per terra ...sgridò lo **spirito** immondo, ...

- TILC ... lo **spirito** maligno lo buttò a terra ... gridò contro lo **spirito** maligno ...
- TNW17 ... il *demonio* lo gettò a terra ...rimproverò lo **spirito** impuro, ...  
 ..... *Lc 11, 14* .....
- CEI08 Gesù stava scacciando un *demonio* che era muto. ...
- TILC Gesù stava scacciando uno **spirito** maligno che aveva reso muto un uomo. ...  
 ..... *Lc 11, 24* .....
- CEI08 Quando lo **spirito** impuro esce dall'uomo, ...
- CEI74 Quando lo **spirito** immondo esce dall'uomo, ...
- NR06 « Quando lo **spirito** immondo esce da un uomo, ...
- TILC « Quando uno **spirito** maligno è uscito da un uomo, ...
- TNW17 » Quando uno **spirito** impuro esce da un uomo, ...  
 ..... *Lc 13, 11* .....
- CEI08 ... uno **spirito** teneva inferma ...
- CEI74 ... uno **spirito** che la teneva inferma; ...
- NR06 ... uno **spirito** ( $\mathcal{P}$ ) che la rendeva inferma, ...
- TILC ... uno **spirito** maligno la teneva ricurva ...
- TNW17 ... era posseduta da uno **spirito** ( $\mathcal{P}$ ) che la faceva stare male: ...  
 ..... *Gv 10, 21* .....
- CEI08 ... può forse un *demonio* aprire gli occhi ai ciechi? ».
- TILC ... Uno **spirito** maligno non può dare la vista ai ciechi ».  
 ..... *At 16, 16* .....
- CEI08 ... che aveva uno **spirito** ( $\mathcal{L}$ ) di divinazione: ...
- CEI74 ... che aveva uno **spirito** ( $\mathcal{L}$ ) di divinazione ...
- NR06 ... posseduta da uno **spirito** ( $\mathcal{L}$ ) di divinazione. ...
- TILC ... Uno **spirito** maligno si era impossessato di lei ...
- TNW17 ... posseduta da uno **spirito** ( $\mathcal{P}$ ), un *demonio* di divinazione, ...  
 ..... *At 16, 18* .....
- CEI08 ... si rivolse allo **spirito** ( $\mathcal{H}$ ) e disse: ... E all'istante lo **spirito** ( $\mathcal{P}$ ) uscì.
- CEI74 ... si volse e disse allo **spirito** ( $\mathcal{H}$ ): ... E lo **spirito** ( $\mathcal{P}$ ) partì all'istante.
- NR06 ... si voltò e disse allo **spirito** ( $\mathcal{H}$ ): ... Ed *egli* uscì in quell'istante.
- TILC ... Si voltò bruscamente e disse allo **spirito** maligno: ... In quello stesso istante lo **spirito** maligno si allontanò dalla schiava.
- TNW17 ... si voltò e disse allo **spirito** ( $\mathcal{H}$ ): ... E in quell'istante lo **spirito** ( $\mathcal{P}$ ) uscì.

- ..... *At 19, 15* .....
- CEI08 Ma lo **spirito** cattivo rispose loro: ...
- CEI74 Ma lo **spirito** cattivo rispose loro: ...
- NR06 Ma lo **spirito** maligno rispose loro: ...
- TILC Ma una volta lo **spirito** maligno rispose loro: ...
- TNW17 Ma uno degli *spiriti* malvagi rispose loro: ...
- ..... *At 19, 16* .....
- CEI08 E l'uomo che aveva lo **spirito** cattivo si scagliò su di loro, ...
- CEI74 E l'uomo che aveva lo **spirito** cattivo, slanciato su di loro, ...
- NR06 E l'uomo che aveva lo **spirito** maligno si scagliò su tutti loro; ...
- TILC Poi l'uomo posseduto dallo **spirito** maligno si scagliò contro di loro ...
- TNW17 Allora l'uomo posseduto dallo **spirito** malvagio si scagliò su di loro, ...
- ..... *Ap 18, 2* .....
- CEI08 ... ed è diventata covo di demoni, rifugio di ogni **spirito** impuro, ...
- CEI74 ... ed è diventata covo di demoni, carcere di ogni **spirito** immondo, ...
- NR06 ... È diventata ricettacolo di demoni, covo di ogni **spirito** immondo, ...
- TILC ... È diventata dimora di demoni, rifugio di tutti gli *spiriti* immondi, ...
- TNW17 ... ed è diventata dimora di demoni e rifugio di ogni **spirito** impuro ...
-

# Bibliografia

- Alonso Schökel, Luis (1989). *Manuale di poetica ebraica*. Brescia: Queriniana.
- (2013). *Dizionario di Ebraico biblico*. Cinisello Balsamo: Edizioni san Paolo.
- Barbaglia, Silvio (2014). “Bibbia e strumenti informatici”. In: *Introduzione generale alla Bibbia*. A cura di Rinaldo Fabris et al. 2<sup>a</sup> ed. LOGOS Corso di studi biblici 1. Torino: Elledici. Cap. A2, pp. 727–737.
- Boscolo, Gastone (2011). *La Bibbia nella storia. Introduzione generale alla Sacra Scrittura*. 2<sup>a</sup> ed. Padova: Edizioni Messaggero.
- Cazelles, Henri e Jacques Dupont (1976). “Nome”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 764–769.
- Coenen, Andy et al. (2019). *Visualizing and Measuring the Geometry of BERT*.
- Costa, Eugenio (1998). “Come hanno lavorato i revisori. Il contesto e il metodo.” In: *La traduzione della Bibbia nella Chiesa italiana. Il Nuovo Testamento*. A cura di Carlo Buzzetti e Carlo Ghidelli. Cinisello Balsamo: Edizioni san Paolo, pp. 127–138.
- Devlin, Jacob et al. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186.
- Fabris, Rinaldo et al., cur. (2014). *Introduzione generale alla Bibbia*. 2<sup>a</sup> ed. LOGOS Corso di studi biblici 1. Torino: Elledici.
- Flamary, Rémi et al. (2021). “POT: Python Optimal Transport”. In: *Journal of Machine Learning Research* 22.78, pp. 1–8.

- Fraine, Jean De e Pierre Grelot (1976). “Numeri”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 775–781.
- Frermann, Lea e Mirella Lapata (2016). “A Bayesian Model of Diachronic Meaning Change”. In: *Transactions of the Association for Computational Linguistics*. Vol. 4, pp. 31–45.
- Ghidelli, Carlo (1998). “I risultati raggiunti. Qualche esempio.” In: *La traduzione della Bibbia nella Chiesa italiana. Il Nuovo Testamento*. A cura di Carlo Buzzetti e Carlo Ghidelli. Cinisello Balsamo: Edizioni san Paolo, pp. 139–154.
- Giulianelli, Mario, Marco Del Tredici e Raquel Fernández (2020). “Analysing lexical semantic change with contextualised word representations”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 3960–3973.
- Grelot, Pierre (1976). “Legge”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 589–601.
- Guillet, Jacques (1976). “Spirito”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 1229–1232.
- Gulordava, Kristina e Marco Baroni (2011). “A Distributional Similarity Approach to the Detection of Semantic Change in the Google Books Ngram Corpus”. In: *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*, pp. 67–71.
- Harris, Zellig S. (1954). “Distributional structure”. In: *Word* 10.2–3, pp. 146–162.
- Hu, Renfen, Shen Li e Shichen Liang (2019). “Diachronic Sense Modeling with Deep Contextualized Word Embeddings: An Ecological View”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Firenze: Association for Computational Linguistics, pp. 3899–3908.
- Landauer, Thomas K. e Susan T. Dumais (1997). “A solution to Plato’s problem. The latent semantic analysis theory of acquisition, induction, and representation of knowledge”. In: *Psychological Review* 104.2, pp. 211–240.
- Lenci, Alessandro (2018). “Distributional Models of Word Meaning”. In: *Annual Reviews of Linguistics* 4, pp. 151–171.

- Lenci, Alessandro, Simonetta Montemagni e Vito Pirrelli (2005). *Testo e computer. Elementi di linguistica computazionale*. Roma: Carocci.
- Leon-Dufour, Xavier (1976). “Introduzione”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. XIII–XVIII.
- Martinc, Matej et al. (2020). “Capturing evolution in word usage: Just add more clusters?” In: *Companion Proceedings of the Web Conference 2020*. New York: Association for Computing Machinery, pp. 343–349.
- McInnes, Leland, John Healy e James Melville (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*.
- Mikolov, Tomas et al. (2013). *Efficient Estimation of Word Representations in Vector Space*.
- Montariol, Syrielle, Matej Martinc e Lidia Pivovarová (2021). “Scalable and Interpretable Semantic Change Detection”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, pp. 4642–4652.
- Mosbach, Marius, Maksym Andriushchenko e Dietrich Klakow (2021). *On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines*.
- Motte, René (1976). “David”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 249–253.
- Nida, Eugene A. (1995). “Principi di traduzione esemplificati dalla traduzione della Bibbia”. In: *Teorie contemporanee della traduzione*. A cura di Siri Nergard. Milano: Bompiani, pp. 149–180.
- Ortiz Suárez, Pedro Javier, Benoit Sagot e Laurent Romary (2019). “Asynchronous pipelines for processing huge corpora on medium to low resource infrastructures”. In: *Proceedings of the Workshop on Challenges in the Management of Large Corpora*. A cura di Piotr Bański et al. Mannheim: Leibniz-Institut für Deutsche Sprache, pp. 9–16.
- Pappalardo, Massimo e Mauro Passacantando (2004). *Metodi e modelli matematici di ottimizzazione per la gestione*. Pisa: Edizioni Plus.

- Passelecq, Georges e Ferdinand Poswick, cur. (2012). *Concordanza pastorale della Bibbia. Indice analitico e analogico delle voci e dei temi di interesse pastorale nella Bibbia italiana disposti in ordine alfabetico*. 2<sup>a</sup> ed. Bologna: EDB.
- Pražák, Ondřej et al. (2020). “UWB at SemEval-2020 Task 1: Lexical Semantic Change Detection”. In: *Proceedings of the 14th International Workshop on Semantic Evaluation*. Barcellona, pp. 246–254.
- Reimers, Nils e Iryna Gurevych (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*.
- Scaiola, Donatella (1997). “Libro dei Salmi”. In: *Libri sapienziali ed altri scritti*. A cura di Antonio Bonora, Michelangelo Priotto et al. LOGOS Corso di studi biblici 4. Torino: Elledici. Cap. 7, pp. 117–133.
- Schlechtweg, Dominik et al. (2020). “SemEval-2020 task 1: Unsupervised lexical semantic change detection”. In: *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcellona: International Committee for Computational Linguistics, pp. 1–23.
- Schuster, Mike e Kaisuke Nakajima (2012). “Japanese and Korean Voice Search”. In: *International Conference on Acoustics, Speech and Signal Processing*, pp. 5149–5152.
- Schweter, Stefan (2020). *Italian BERT and ELECTRA models*. Ver. 1.0.1.
- Solomon, Justin (2018). *Optimal Transport on Discrete Domains*.
- “Sussidi per la lettura” (2014). In: *La Bibbia in lingua corrente*. Torino: Elledici, pp. 11–12.
- Tang, Xuri (2018). “A state-of-the-art of semantic change computation”. In: *Natural Language Engineering* 24.5, pp. 649–676.
- Tenney, Ian, Dipanjan Das e Ellie Pavlick (2019). “BERT Rediscovered the Classical NLP Pipeline”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Firenze: Association for Computational Linguistics, pp. 4593–4601.
- Ternant, Paul (1976). “Segno”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 1173–1180.

- Tiedemann, Jörg (2012). “Parallel Data, Tools and Interfaces in OPUS”. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation*.
- Van Rossum, Guido e Fred L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, California: CreateSpace.
- Wiedemann, Gregor et al. (2019). *Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings*.
- Wiéner, Claude (1976). “Amore”. In: *Dizionario di Teologia Biblica*. A cura di Xavier Leon-Dufour et al. 2<sup>a</sup> ed. Casale Monferrato: Marietti, pp. 44–54.