



Università
di Catania

UNIVERSITÀ DEGLI STUDI DI CATANIA
Dipartimento di Scienze Umanistiche

Genocide and War: A Corpus-Based Analysis of Register and Lexical Distribution

Across Journalistic and Legal Discourse

Autori	Andrea Modica · Cristina Raciti
Materia	Statistics in Corpus Linguistics
Docente	Marco Venuti
Università	Università degli Studi di Catania
Anno Accademico	2025-2026
Dimensione corpus	250,099 tokens (giornalistico: 239,516 · giuridico: 10,583)
Keywords analizzate	genocide · war (Gaza_press_pov), genocide-Gaza (Onu_documentation corpus)
Tools	Sketch Engine · Microsoft Excel

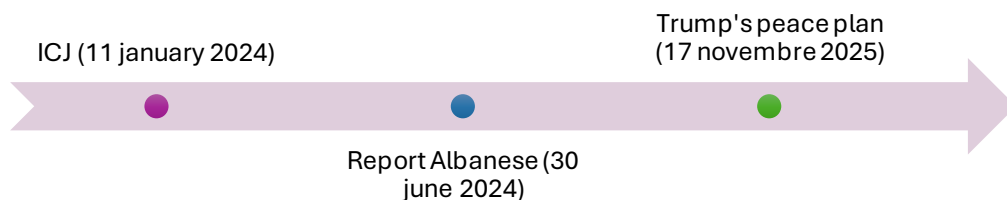
1. Introduzione	2
1.1 Presentazione dell'argomento	2
1.2 CORPUS: Preparazione e struttura	2
1.3 Piattaforma e tool utilizzati	5
2. Analisi	7
2.1 Premessa	7
2.2 Gli strumenti statistici di partenza	7
2.2.1 Frequenze osservate	8
2.2.2 Frequenze attese	9
2.2.3 Scarti quadratici	9
2.2.4 Chi-Quadrato (χ^2), significatività e Cramér's V	10
2.3 Analisi dei collocati	11
2.3.1 Nodo war (gaza press pov)	13
2.3.2 Nodo "genocide" (Gaza_press_pov)	16
2.3.3 Nodo "Genocide" (Onu_documentation corpus)	20
2.3.4 Nodo "Gaza" (Onu_documentation corpus)	24
2.4 Wordsketch Difference	27
2.5 Neighbouring Cluster Analysis (NCA) of "genocide"	29
3. Conclusione	32
3.1 Riepilogo	32
3.2 Distribuzione statistica: chi-quadrato e Cramér's V	32
3.3 Analisi dei collocati: due grammatiche del conflitto	33
3.3.1 Il nodo war nel corpus stampa: registro temporale-diplomatico	33
3.3.2 Il nodo genocide nel corpus stampa: registro accusatorio-epistemico	33
3.3.3 I nodi genocide e Gaza nel Corpus ONU: registro definitorio-preventivo	33
3.4 Word Sketch Difference: L'asimmetria documentata	34
3.5 Neighbouring Cluster Analysis: la dimensione diacronica	35
3.6 Conclusioni finali	35
BIBLIOGRAFIA	37
Formulario	38

1. Introduzione

1.1 Presentazione dell'argomento

Il seguente elaborato si propone di realizzare un'analisi testuale di testi legati alle vicende della guerra Israelo-Palestinese, concentrando il lavoro di ricerca nel rapporto della parola **genocidio** nella documentazione internazionale e la stampa britannica. Il periodo temporale preso in esame spazia dal gennaio 2024 al novembre del 2025, usando come eventi principali - insieme a diversi testi che narrano delle condizioni di vita a Gaza durante i 22 mesi presi in esame:

- Il riconoscimento da parte della Corte Internazionale di Giustizia (ICJ) delle attività di genocidio da parte di Israele ai danni di Gaza (11 gennaio 2024)
- Presentazione del Report sulla situazione in Palestina - *From Economy of Occupation to Economy of Genocide* - redatto dalla Relatrice speciale delle Nazioni Unite sui territori palestinesi occupati Francesca Albanese (30 giugno 2025)
- L'approvazione dal Consiglio di Sicurezza della Nazioni Unite del "Piano di pace di Trump" per Gaza (17 novembre 2025)



Timeline 1

L'ipotesi di ricerca si focalizza sul tipo di comunicazione istituzionale di queste istituzioni sovranazionali, ponendola in comparazione con il tipo di divulgazione attuato dalle maggiori testate giornalistiche in Inghilterra. In particolare, si andrà ad analizzare il ruolo della parola **genocidio** nelle due varianti comunicative e analizzando come viene narrata la vicenda tra i diversi tipi di giornali.

I 3 eventi sono stati scelti per la loro rilevanza internazionale, che permettono di valutare come si è la vicenda nel corso dei ventidue mesi presi in esame.

La selezione dei giornali analizzati in questa ricerca risponde ai requisiti di:

- Ampio pubblico mensile
- Target politicizzato (giornali di "sinistra", giornali di "destra" e giornali "moderati")

1.2 CORPUS: Preparazione e struttura

Una parte importante per la realizzazione del progetto di ricerca è stata la realizzazione di un corpus di testi da poter analizzare. La grandezza complessiva del corpus su cui è stata effettuata l'analisi è di 250.099 parole.

La raccolta dei testi differisce per categoria:

- La documentazione internazionale della Corte Internazionale di Giustizia (ICJ) e delle Nazioni Unite, in questo caso il report di Francesca Albanese (Relatrice Speciale delle Nazioni Unite sui territori palestinesi occupati) è stato reperito dagli archivi digitali online delle stesse istituzionali, accessibili gratuitamente al grande pubblico. La grandezza di suddetto corpus è di **10.583 parole**;
- I testi giornalistici sono stati recuperati attraverso **NexisUni**, strumento messo a disposizione dall'Università degli Studi di Catania per la raccolta di articoli di giornale identificati tramite keywords e data di pubblicazione. Sono state stabilite delle keywords affinché la piattaforma rispettasse il margine contestuale, in ordine dalla più generale alla più specifica: *Gaza, Palestine, Genocide, Genocide Convention, Genocide Allegations, civilians, crimes, UNSpecial Rapporteur, Accusation, Report, UN Security Council, Approval, Trump, Endorsement, Resolution, ceasefire*. Il numero totale di articoli è di 334, e la grandezza complessiva di suddetto corpus è di **239.516 parole**.

Entrando nello specifico, la selezione delle testate risponde a due criteri principali: l'ampio pubblico mensile e la rappresentatività degli orientamenti politici del panorama editoriale britannico. I dati di *readership* utilizzati come riferimento sono quelli forniti dalla *PAMCo* (*Publishers Audience Measurement Company*), la principale fonte di misurazione delle audience delle testate britanniche su base cross-platform (stampa, digitale, mobile), integrati con i dati del *Reuters Institute Digital News Report 2024* (Newman et al. 2024).

Secondo i dati PAMCo più recenti, *The Sun* e *The Daily Mail* registrano la *reach* mensile più alta tra le testate britanniche, con circa 23,2 milioni di lettori ciascuna. *The Daily Telegraph* raggiunge i 13,7 milioni di lettori mensili, mentre il *The Times* si attesta a 10,8 milioni. *The Guardian* conta 22,4 milioni di visitatori unici mensili, mentre *The Mirror* supera i 20 milioni di lettori online. *The Independent*, interamente digitale dal 2016, raggiunge circa 22 milioni di lettori mensili su base cross-platform.

Testata	Schieramento politico	Reach mensile (PAMCo/Ipsos Iris, 2024)
The Guardian	Progressista	22.4 milioni di lettori mensili
The Daily Telegraph	Conservatore	13.7 milioni di lettori mensili
The Times	Moderato-conservatore	10.8 milioni di lettori mensili
The Sun	Populista-conservatore	23.2 milioni di lettori mensili
The Daily Mirror	Progressista	20 milioni di lettori mensili
The Independent	Centrista-liberale	22 milioni di lettori mensili

Le sei testate selezionate coprono complessivamente l'intero spettro politico del giornalismo britannico e rappresentano alcune delle fonti di informazione più lette nel paese, garantendo

così una base empirica solida e rappresentativa per l'analisi del discorso sulla parola *genocide* nella stampa britannica.

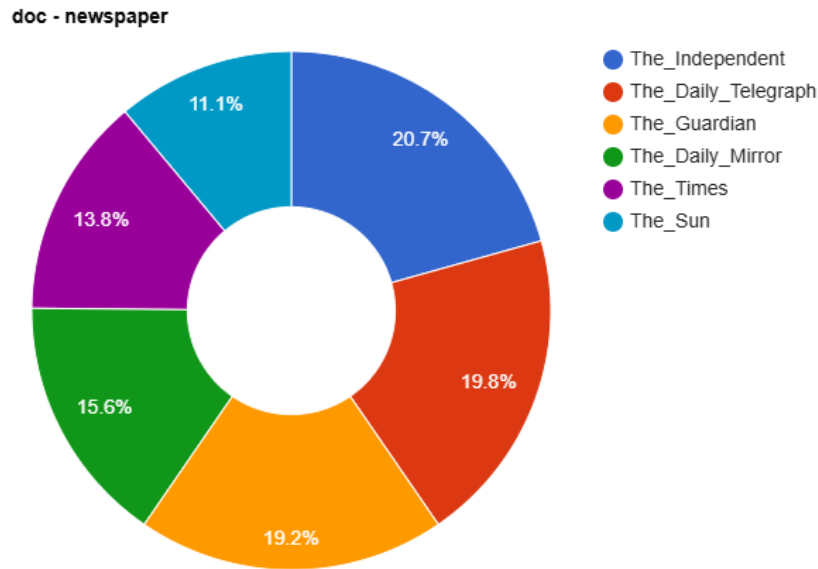


Immagine 1¹

La prima osservazione evidente dalla text type analysis è che il corpus è distribuito in modo relativamente bilanciato tra le sei testate: nessuna singola fonte domina il corpus in modo schiacciante, anche se la differenza tra la testata più rappresentata (*The Independent*, 20.7%) e quella meno rappresentata (*The Sun*, 11.1%) è di circa 10 punti percentuali. *The Independent* è la testata singolarmente più rappresentata con il 20.7%, seguita immediatamente dal *Daily Telegraph* (19.8%) e dal *The Guardian* (19.2%). Queste tre testate da sole rappresentano il 59.7% del corpus - quasi tre articoli su cinque. *The Daily Mirror*, *The Times* e *The Sun* contribuiscono ciascuno meno del 16% e hanno quindi un peso minore nell'aggregato.

¹ Nota di lettura dell'immagine: La seguente immagine è stata elaborata da Sketch Engine, ed evidenzia la suddivisione dei giornali presi in esame.

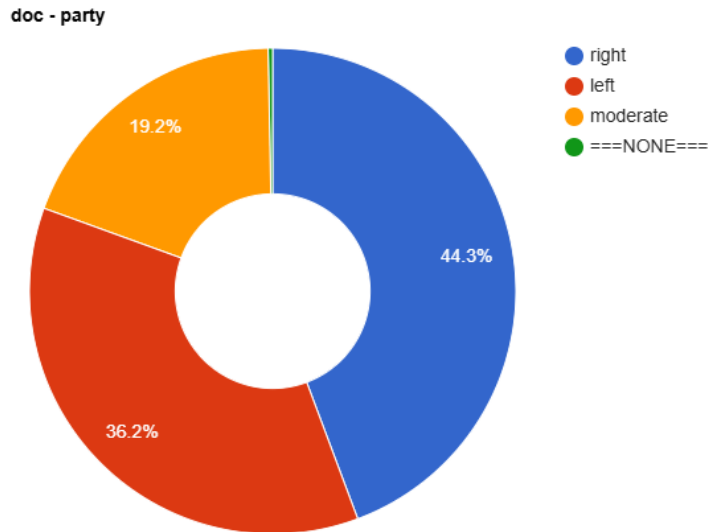


Immagine 2²

Dalla text type sul tag doc.party emerge che il corpus non è bilanciato per orientamento politico. Di fatto, la stampa di destra rappresenta quasi la metà del corpus (44.3%), con un vantaggio di 8 punti percentuali sulla stampa di sinistra (36.2%) e di 25 punti sulla stampa moderata (19.2%).

Il materiale sopra citato è stato raccolto in prima istanza nel formato *Rich Text Format* (conosciuto anche come “.rtf “), successivamente convertiti nel linguaggio di markup *Extensible Markup Language* (conosciuto anche come “XML”). La scelta di questo formato è stata dettata dalla necessità di catalogare gli articoli secondo le varie testate giornalistiche rispettando i margini temporali prefissati e al tempo stesso di rendere le informazioni ricercabili. La ragione principale per cui abbiamo scelto questo linguaggio di codifica è stata quella di evitare di perdere metadati importanti come le date, rivelate cruciali per condurre l’analisi diacronica della parola “genocide“, e gestire le varie informazioni in tag funzionali quali: doc.newspaper per identificare il tipo di giornale; doc.party per identificare gli schieramenti politici delle varie testate giornalistiche; il tag section ha permesso di catalogare i vari eventi in “news“ per il corpus “Gaza_press_pov“, che risulta essere il corpus specialistico della stampa inglese e il suo punto di vista sugli eventi ivi citati.

1.3 Piattaforma e tool utilizzati

Il programma utilizzato per l’analisi è Sketch Engine, software sviluppato nel 2003 dalla Lexical Computing che permette di lavorare grandi corpus di testi attraverso diversi strumenti integrati. La forte compatibilità del software usato con il formato XML ha permesso un’analisi adeguata ai fini della nostra ricerca.

Sketch Engine mette a disposizione vari tool di cui abbiamo usufruito, quali:

² Nota di lettura dell’immagine: La seguente immagine è stata elaborata da Sketch Engine, ed evidenzia la suddivisione percentuale dei giornali in base al loro orientamento politico.

- Text type analysis (doc.newspaper, doc.party): strumento che accedendo ai metadati menzionati ha permesso di visualizzare i metadati delle testate giornalistiche;
- Wordlist: Liste di frequenza, generate per il corpus specialistico degli articoli di stampa Gaza_press_pov e l'Onu_documentation corpus, il corpus generale.
- Word Sketch: Utilizzato per condurre l'analisi collocazionale. Rispettivamente sono stati analizzati i nodi "genocide" e "war" del corpus specialistico, e i nodi "genocide" e "Gaza" del corpus generale;
- Word Sketch difference: Utilizzato per confrontare i nodi "genocide" e "war" dei corpora menzionati;
- Concordance: Utilizzato per verificare le parole in contesto, prenderemo in esame i nodi utilizzati nella collocation analysis per verificarne la frequenza effettiva e la significatività all'interno dei corpora.

Ogni tool sopra citato permette la creazione di grafici attraverso lo stesso software e l'esportazione dei dati in formato .xlsx permettendo, tramite l'uso di software come Excel, la creazione di grafici e calcoli matematici.

2. Analisi

2.1 Premessa

Una volta caricato il corpus unico (250.099 parole) all'interno del software Sketch Engine, suddiviso in due sub-corpora indipendenti (giornalistico di 239.516 parole; giuridico di 10.583 parole) si procede all'analisi seguendo la domanda di ricerca "Qual è il rapporto della parola *genocide* nella documentazione ufficiale e nella stampa?"

Trattando i due sub-corpora come indipendenti l'uno dall'altro, con contesti comunicativi, finalità testuali e registri distinti, è stato possibile ottenere dei dati che permettono di analizzare le differenze d'uso delle keyword prese in esame all'interno di questi due ambiti discorsivi, accomunati dal tema ma caratterizzati da finalità comunicative e convenzioni testuali distinte.

Di seguito, saranno presenti una serie di analisi tecniche commisurate all'obiettivo da raggiungere. Per agevolare la lettura in questa fase iniziale verranno evidenziati i passaggi e i risultati delle singole operazioni senza, però, entrare nello specifico delle formule matematiche ad esse applicate.

Prima di addentrarci nelle varie analisi derivate dalle nostre domande di ricerca è doveroso precisare che la grande differenza di grandezza tra il corpus giornalistico e quello giuridico ha influenzato in maniera rilevante i dati che verranno mostrati. Tale differenza è dovuta al fatto che i testi giuridici presi in esame sono pochi - soltanto 3 -, ma hanno avuto un grande impatto sulla percezione, narrazione e l'effettiva evoluzione di tutta la vicenda.

Le formule e i calcoli precisi di ogni singola operazione possono essere visionati nel formulario allegato al testo.

2.2 Gli strumenti statistici di partenza

Per rispondere alla domanda di ricerca è stato necessario adottare un approccio quantitativo, in grado di misurare in modo oggettivo le differenze distribuzionali delle keyword nei sub-corpora. A tal fine sono stati scelti tre strumenti statistici complementari: il chi-quadrato (χ^2), il p-value e il Cramér's V.

Il **chi-quadrato** è lo strumento di partenza. Permette di verificare se la distribuzione di una parola tra due corpora indipendenti è statisticamente significativa, ovvero se la differenza osservata è reale o riconducibile al caso. È lo strumento standard nella linguistica dei corpora per il confronto di frequenze tra gruppi indipendenti, e si applica in modo appropriato al presente lavoro poiché i due sub-corpora provengono da contesti comunicativi distinti e le frequenze analizzate sono conteggi discreti³.

Il **p-value** accompagna il chi-quadrato e ne quantifica la certezza: indica la probabilità che la differenza osservata sia casuale. Nella presente analisi si adotta la soglia più severa ($p < 0,001$), che corrisponde a una probabilità inferiore allo 0,1% che il risultato sia dovuto al

³ (Brezina 2018 :112-113)

caso. È tuttavia importante non ridurre l'intera analisi al solo p-value: come sottolinea Brezina (2018:21), “significatività statistica”, “importanza pratica” e “significatività linguistica” sono tre concetti distinti che non devono essere confusi

Il **Cramér's V** completa il quadro misurando la forza dell'associazione tra le keyword e il tipo di corpus, indipendentemente dalla dimensione di quest'ultimo. Come osserva Brezina (2018:114), il Cramér's V è una misura standardizzata del chi-quadrato che corregge la crescita del valore di quest'ultimo all'aumentare della dimensione del campione, rendendo così i risultati comparabili tra corpora di dimensioni diverse. Questo strumento è particolarmente rilevante nel caso in esame: la marcata sproporzione tra i due sub-corpora (239.516 parole contro 10.583) rende il chi-quadrato sensibile anche a differenze di piccola entità, rischiando di sovrastimarne la rilevanza linguistica. Il Cramér's V corregge questo limite e restituisce un valore compreso tra 0 e 1. Per l'interpretazione dei valori si fa riferimento alla scala proposta da Cohen (1988) e adottata da Brezina (2018:115): per tabelle 2x2 un valore inferiore a 0,10 indica un effetto piccolo, tra 0,10 e 0,30 un effetto medio e superiore a 0,50 un effetto grande.

Utilizzati congiuntamente, i tre strumenti permettono di rispondere non solo alla domanda “la differenza esiste?”, ma anche “quanto sono sicuro che non sia casuale?” e “quanto è grande?” - fornendo un quadro analitico completo e metodologicamente solido.

2.2.1 Frequenze osservate

Il primo passo dell'analisi consiste nel rilevare quante volte le keyword *genocide* e *war* appaiono effettivamente in ciascun sub-corpus. Questi valori, detti **frequenze osservate**, costituiscono la base di tutti i calcoli successivi e rappresentano il dato grezzo estratto direttamente dal corpus tramite Sketch Engine.

La tabella seguente riporta le occorrenze delle due keyword nei due sub-corpora, con i relativi totali di riga e di colonna:

Frequenze osservate	Corpus giornalistico	Corpus giuridico	Totale
<i>“genocide”</i>	314	77	391
<i>“war”</i>	729	6	735

Come si evince dalla tabella, *genocide* appare 314 volte nel corpus giornalistico e 77 volte in quello giuridico, per un totale di 391 occorrenze complessive. *War* appare invece 729 volte nel corpus giornalistico e sole 6 volte in quello giuridico, per un totale di 735 occorrenze.

A una prima lettura questi dati sembrano indicare una presenza maggiore di entrambe le keyword nel corpus giornalistico. Tuttavia, questo confronto diretto è fuorviante: il corpus giornalistico è circa 22 volte più grande di quello giuridico. Per poter confrontare le frequenze in modo corretto è necessario calcolare quante occorrenze ci aspetteremmo in ciascun sub-corpus se la distribuzione fosse puramente casuale - ovvero le **frequenze attese**, oggetto del paragrafo successivo.

2.2.2 Frequenze attese

Una volta rilevate le frequenze osservate si procede a calcolare le **frequenze attese**, ovvero i valori che ci aspetteremmo di trovare in ciascuna cella della tabella se la distribuzione delle keyword fosse completamente casuale e indipendente dal tipo di corpus. Come sottolinea Brezina (2018:113), le frequenze attese rappresentano la baseline che utilizziamo per stabilire se esiste una relazione reale tra le variabili analizzate.

Il calcolo si basa sulla seguente formula (Brezina 2018:113):

$$\text{Expected frequency} = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

Per ogni cella si moltiplica il totale della riga corrispondente per il totale della colonna corrispondente, dividendo il risultato per il numero totale di parole all'interno del corpus (N=250.099). I valori ottenuti sono stati arrotondati all'unità.

La tabella seguente riporta le frequenze attese per ciascuna keyword nei due sub-corpora:

Frequenza attesa	Corpus giornalistico	Corpus giuridico
“genocide”	374	17
“war”	704	31

Il confronto tra le frequenze osservate e attese rileva dei pattern interessanti. Per *genocide*, nel corpus giuridico ci si aspetterebbe solo 17 occorrenze, mentre ne sono state osservate 77 (più di quattro volte il valore atteso). Nel corpus giornalistico la situazione è inversa: le occorrenze osservate (314) sono inferiori alle attese (374).

Per *war* il pattern è opposto: nel corpus giuridico ci si aspetterebbe 31 occorrenze, ma ne appaiono solo 6 (circa un quinto del valore atteso). Nel corpus giornalistico le occorrenze osservate (729) sono inferiori alle attese (704).

Questi scostamenti tra valori osservati e attesi sono il segnale che la distribuzione delle due keyword non è casuale e che esiste una relazione tra la parola e il tipo di corpus. La misurazione precisa di questi scostamenti è oggetto del paragrafo successivo.

2.2.3 Scarti quadratici

Una volta ottenute le frequenze osservate e quelle attese, il passo successivo consiste nel misurare di quanto ciascuna cella si discosta dal valore atteso. Questo scarto viene calcolato attraverso la formula (Brezina 2018:113):

$$\text{Chi - squared contribution} = \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

Il quadrato dello scarto serve a rendere sempre positivo il risultato, indipendentemente dal verso della differenza - che la parola sia sovrarappresentata o sottorappresentata. Dividere per

la frequenza attesa normalizza il contributo di ciascuna cella rispetto alla sua dimensione attesa: una cella con frequenza attesa bassa e scarto elevato contribuisce in modo sproporzionato al risultato finale.

La tabella seguente riporta il contributo di ciascuna cella al chi-quadrato complessivo:

Scarto quadratico	Corpus giornalistico	Corpus giuridico
“genocide”	9,625668449	211,7677059
Altre parole - “genocide”	0,01	0,34
“war”	0,887784091	20,16129032
Altre parole - “war”	0,002	0,06

I valori mostrano con chiarezza quale cella pesa maggiormente in ciascuna analisi. Per *genocide*, è la cella relativa al corpus giuridico a dominare il risultato, con uno scarto di 211,76 rispetto al 9,62 del corpus giornalistico. Questo riflette la forte sovrarappresentazione delle keyword in quel contesto: 77 occorrenze osservate contro le sole 17 aspettate. Anche per *war* il corpus giuridico (20,16) produce uno scarto nettamente superiore rispetto al corrispettivo giornalistico (0,88): le 6 occorrenze osservate contro le 31 attese generano uno scarto elevato, amplificato dal fatto che la frequenza attesa è bassa.

Le celle relative al corpus giornalistico e alla riga “altre parole” contribuiscono in maniera trascurabile al risultato finale, confermando che è sempre il corpus giuridico il luogo in cui avviene qualcosa di linguisticamente significativo. Il chi-quadrato complessivo si ottiene sommando tutti questi contributi, come illustrato nel paragrafo successivo.

2.2.4 Chi-Quadrato (χ^2), significatività e Cramér’s V

Il **chi-quadrato** si ottiene sommando i quattro scarti calcolati nel passo precedente. La sua formula è (Brezina 2018:113):

$$\chi^2 = \sum \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

Per *genocide* il valore ottenuto è $\chi^2=221,74$, mentre per *war* è $\chi^2=21,111$. Entrambi i valori superano abbondantemente la soglia critica di 10,828 corrispondente a $p < 0,001$ con 1 grado di libertà - il livello di significatività più severo adottato in statistica. Questo significa che la probabilità che la distribuzione osservata delle due keyword sia dovuta al caso è inferiore allo 0,1%. La differenza tra i due sub-corpora è reale e sistematica per entrambe le parole.

Prima di procedere è opportuno verificare il presupposto fondamentale del test del chi-quadrato: tutte le frequenze attese devono essere uguali o superiori a 5 (Brezina 2018:113). Nel caso in esame il valore atteso più basso è 17 per *genocide* nel corpus giuridico e 31 per *war* nello stesso corpus - entrambi ampiamente superiori alla soglia. Il test, quindi, è applicabile senza riserve.

È tuttavia importante non fermarsi alla sola significatività statistica. Come osserva Brezina (2018:21), in corpora di grandi dimensioni anche differenze minime tra gruppi di testi possono risultare statisticamente significative, senza che questo implichi necessariamente una rilevanza linguistica o pratica. Per questo motivo si è scelto di affiancare al chi-quadrato il Cramér's V come misura della forza dell'effetto.

Il **Cramér's V** è stato calcolato a partire dal chi-quadrato, secondo la formula applicabile alle tabelle 2x2 (Brezina 2018:114):

$$V = \sqrt{\frac{\chi^2}{total\ observations \times (no.\ of\ rows\ or\ cols, whichever\ is\ smaller - 1)}}$$

La tabella seguente riporta i risultati complessivi dell'analisi:

Keyword	χ^2	p-value	Cramer's V
Genocide	221.7	p<0,001	0,029776381
War	21.1	p<0,001	0,009187523

I valori ottenuti per *genocide* (V=0,029) e per *war* (V=0,009) collocano entrambe le keyword al di sotto della soglia di 0,10, che secondo la scala di Cohen (1988), adottata da Brezina (2018:115), corrisponde alla fascia del piccolo effetto per tabelle 2x2. Questo risultato non contraddice la significatività, ma la contestualizza: la marcata sproporzione tra i due sub-corpora rende il chi-quadrato sensibile anche a differenze di entità limitata, producendo valori elevati che potrebbero indurre a sovrastimare la rilevanza linguistica del fenomeno. Il Cramér's V corregge questo effetto e indica che, nonostante la significatività statistica, la forza dell'associazione tra le keyword e il tipo di corpus è contenuta.

2.3 Analisi dei collocati

Con lo scopo di indagare il significato contestuale di alcune parole, sono state condotte analisi collocazionali condotte sui due corpora: i nodi "genocide" e "war" del corpus della stampa e i nodi "genocide" e "Gaza" del corpus della documentazione istituzionale. Le analisi collocazionali sono state condotte attraverso la funzione Word Sketch della piattaforma Sketch Engine, che fornisce il profilo grammaticale-collocazionale di un nodo e organizzato per:

- relazioni sintattiche (Gramrel): verbs with "X" as object, modifiers of "X", "... of X", nouns modified by "X", "X" in;
- frequenza relativa (RF);

- ogDice (⁴indicato come “score”): indica la forza e l’esclusività del nodo dell’associazione tra nodo e collocato. Abbiamo tre gradi diversi per identificare lo status di una collocazione: lo score > 11 identifica una collocazione molto forte; lo score 9-11 certifica una collocazione solida; lo score < 9 invece visualizza una collocazione debole, probabilmente dovuta a frequenza generica della parola.

Inoltre, al fine di verificare la direzione semantica effettiva dei nodi è stata verificata la concordanza dei collocati più forti sulla base di uno span di 5 in modo da avere una finestra ampia del contesto in cui le parole co-occorrono al nodo principale. La funzione fondamentale della concordanza è quella di rendere visibile la distribuzione di una parola attraverso il corpus ma allo stesso tempo essa permette di disambiguare. I dati appariranno strutturati con i seguenti valori: AF (Absolute Frequency) è la frequenza assoluta del collocato nel corpus; Coll.Freq. (Collocation Frequency, O_{11}) è la frequenza osservata della co-occorrenza: quante volte il collocato appare effettivamente nella finestra intorno al nodo; T-Score⁵ premia i collocati che co-occorrono molte volte in termini assoluti, anche se il legame non è esclusivo; MI (Mutual Information⁶) misura l'esclusività del legame, ovvero quanto la co-occorrenza osservata supera quella attesa per caso; e infine il LogDice.

⁴ Il LogDice viene calcolato automaticamente da Sketch Engine, questa misura statistica dipende dalle frequenze assolute delle parole esaminate.

Nello specifico, la formula utilizzata dalla piattaforma è:

$$\logDice = 14 + \log_2 \frac{2f_{xy}}{f_x + f_y}$$

Dove:

$f_{\{xy\}}$ = la frequenza con cui le due parole compaiono insieme (co-occorrono).

$f_{\{x\}}$ = la frequenza complessiva della prima parola.

$f_{\{y\}}$ = la frequenza complessiva della seconda parola.

14= si pone come una costante che viene aggiunta arbitrariamente per convertire decimali tipicamente minuscoli e di difficile lettura in numeri interi positivi e leggibili dall'uomo.

$$^5 T = (O_{11} - E_{11}) / \sqrt{O_{11}}$$

dove $E_{11} = (\text{freq_nodo} \times \text{freq_collocato} \times \text{window}) / N$ è la frequenza attesa. La soglia di significatività convenzionale è $T > 1.65$ ($p < 0.05$).

$$^6 MI = \log_2(O_{11} / E_{11})$$

2.3.1 Nodo war (gaza press pov)

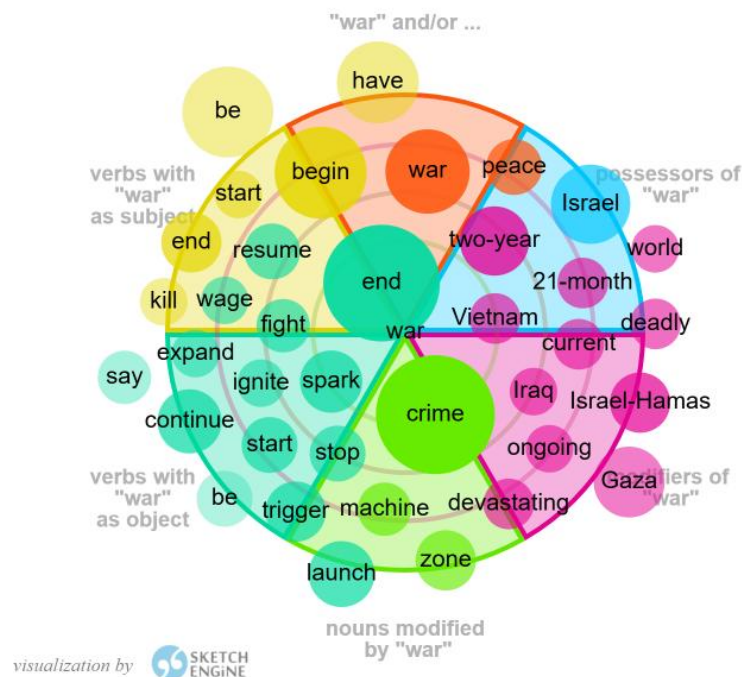


Immagine 37

Il nodo *war* è il termine narrativamente dominante del corpus stampa, dall'analisi delle componenti grammaticali:

Modifiers of "war"

Collocate	AF	Score
<i>Two-year</i>	13	11.37
<i>Israel-Hamas</i>	8	10.76
<i>21-month</i>	4	9.88
<i>Gaza</i>	14	8.93
<i>devastating</i>	4	9.72
<i>deadly</i>	4	9.43
<i>ongoing</i>	3	9.27

La prevalenza dei modificatori temporali segnala che la stampa costruisce la guerra come evento misurabile nel tempo (Two-year, 21-month, ongoing). Il modificatore "Israel-Hamas" cristallizza una visione diadica del conflitto, riducendolo a due attori. I modificatori valutativi (*devastating*, *deadly*) confermano il frame emotivo del corpus.

⁷ Grafico costruito da Wordsketch che mostra i collocati di war organizzati per relazioni grammaticali (creato con indici di frequenza)

Nouns modified by “war”, nominal compound

Compound	AF	Score
<i>War crime</i>	62	13.36
<i>War machine</i>	5	10.41
<i>War zone</i>	8	10.14

War è un termine usato come modificatore nominale: *war crime* è il compound più forte dell'intero Word Sketch: la categorizzazione giuridica della violenza nel corpus stampa passa per il termine meno marcato politicamente. *War zone* identifica il frame geografico della pericolosità del conflitto, *war machine* è una metafora meccanicistica, tipica del discorso critico verso l'apparato militare.

Verbs with “war” as object

Verb	AF	Score
<i>end</i>	59	12.64
<i>launch</i>	11	10.42
<i>continue</i>	10	10.53
<i>spark</i>	9	10.52
<i>stop</i>	6	9.65
<i>wage</i>	3	9.06

“End the war” è la costruzione diplomatica, indica un’azione che si vuole concludere. I verbi di innesco (*spark*, *launch*) costruiscono la narrativa causale. “Wage”, pur raro, è un verbo marcato stilisticamente, quasi arcaico, che attribuisce la massima intenzionalità bellica a un soggetto.

Prepositional constructions and geological frame

“war” in	AF	Score
<i>Gaza</i>	73	11.2
<i>Strip</i>	3	9.3

Possessor of “war”	AF	Score
<i>Israel</i>	20	10.3

War in Gaza domina con 73 occorrenze su 92 totali nella sezione X in - preposizione di localizzazione geografica neutra. Il possessore principale è Israel (freq 20, Score 10.29): Israel's war è la costruzione possessiva dominante.

Concordance

Collocate	AF	Coll.Freq.	T-Score	MI	LogDice
<i>End</i>	91	258	9.46643	7.03066	11.53338
<i>Crimes</i>	49	80	6.96917	7.82688	10.92122
<i>Gaza</i>	159	2898	11.98954	4.34614	10.48078
<i>Israel</i>	89	2030	8.85351	4.02256	10.0359
<i>Hamas</i>	61	1392	7.32946	4.02189	9.86734

1. End

La Coll.Freq. (258) è quasi tre volte l'AF (91). Questo indica che *end* compare 258 volte nella finestra di *war* ma solo 91 volte come lemma autonomo nel corpus, questo significa che la tabella conta tutte le forme flesse. Il T-Score indica che la collocazione è statisticamente molto solida in termini di frequenza assoluta. Il MI è il più basso della tabella, il che significa che *end* non è esclusivo di *war*: compare frequentemente anche in altri contesti del corpus (*end of*, *end the*, ecc.). Il Log Dice è il più alto della tabella: bilanciando frequenza ed esclusività, *end* risulta il collocato complessivamente più robusto. Questo conferma quanto emerso dal Word Sketch: *end the war / end of war* è la costruzione diplomatica dominante del corpus, con 59 occorrenze come verbo oggetto e 27 come costruzione nominale.

2. Crimes

Il MI segnala che *crimes* è più esclusivo di *end* nel suo legame con *war*, quando *crimes* compare nel corpus, tende a farlo in prossimità di *war* più di quanto faccia *end*. *War crimes* è una costruzione più cristallizzata rispetto a *end the war*, che è una costruzione sintattica variabile. Il T-Score è più basso degli altri perché la frequenza assoluta di *crimes* (49) è la più bassa della tabella, e il T-score penalizza le frequenze assolute basse.

3. Gaza

Gaza è il collocato geografico dominante di *war* - la guerra è quasi sempre *war in Gaza* o *war on Gaza* nel corpus. Ma non è un collocato *esclusivo*: la parola è così pervasiva nel corpus che il suo legame con *war* riflette in parte la struttura generale del corpus piuttosto che un legame semantico specifico. Il Log Dice 10.48, che bilancia i due effetti, è la misura più informativa: indica una collocazione solida ma non cristallizzata.

4. Israel

Il profilo di *Israel* è simile a quello di *Gaza* per struttura, ma con valori inferiori. *War* è sistematicamente ancorato all'agente *Israel* nel corpus, coerente con il possessivo *Israel's war* (Score 10.29) rilevato nel Word Sketch.

5. Hamas

Hamas ha lo stesso MI di *Israel* (4.02) e Log Dice inferiore. War è costruito come conflitto tra due attori nominati con pari esclusività collocazionale. Né *Israel* né *Hamas* ha un legame semantico più specifico con *war* dell'altro - entrambi sono attori genericamente presenti nel corpus.

2.3.2 Nodo “genocide” (Gaza_press_pov)

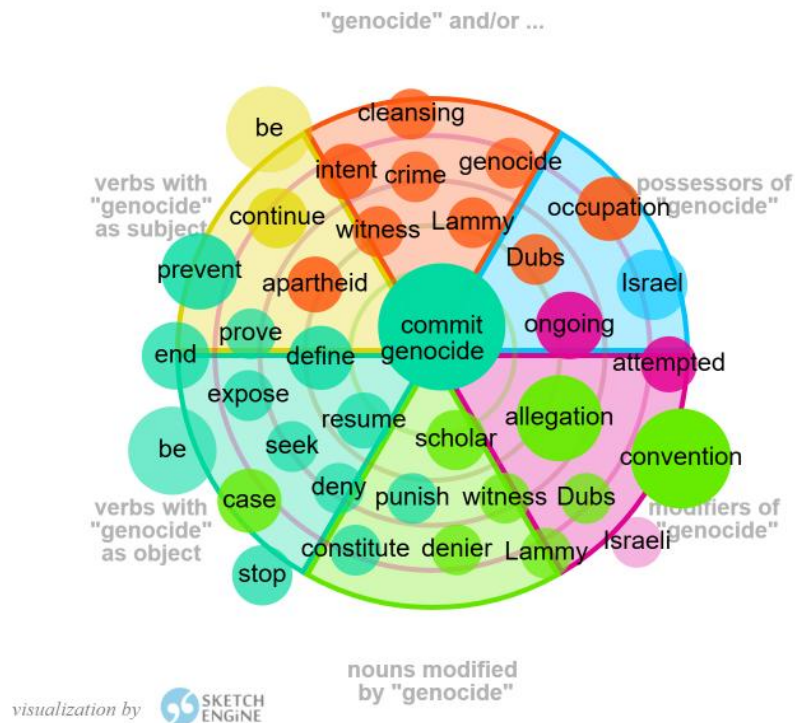


Immagine 3 ⁸

Genocide è collocazionalmente più selettivo di war: frequenze globali inferiori, ma collocati più esclusivi e score più alti.

Modifiers of “genocide”

Collocate	AF	Score
<i>ongoing</i>	6	11.77
<i>attempted</i>	3	11.41
<i>Israeli</i>	2	5.48

Tre soli aggettivi, in minoranza rispetto ai numerosi modificatori di “war”. Da questo possiamo inferire che genocide resiste alla qualificazione aggettivale: nel discorso giornalistico, la parola viene usata tendenzialmente senza modificatori perché porta con sé un peso semantico sufficiente. Quelli che compaiono, però, sono densi di significato. *Ongoing* costruisce il genocidio come evento in corso, di fatto l’evento viene riconosciuto come evento che sta

⁸ Grafico costruito da Wordsketch che mostra i collocati di genocide organizzati per relazioni grammaticali (creato con indici di frequenza)

accadendo nel presente. *Attempted* introduce una sfumatura giuridica precisa: il tentato genocidio è una categoria del diritto internazionale (Convenzione ONU 1948). *Israeli* è molto frequente nel corpus in contesti diversi da “genocide”; quindi, il LogDice la penalizza; tuttavia, la parola ricopre un ruolo importante nell’attribuire le azioni ad un attore ben preciso.

Nouns modified by “genocide”, nominal compounds

Compound	AF	Score
<i>Genocide convention</i>	20	12.83
<i>Genocide allegations</i>	13	12.25
<i>Genocide case</i>	5	10.93

Genocide si risulta essere un modificatore che produce compound nominali. Questi compound sono la mappa del campo semantico che il discorso costruisce attorno alla parola. *Genocide convention* è il collocato più forte della sezione in quanto il discorso aggancia sistematicamente *genocide* alla sua definizione giuridica internazionale, ovvero la Convenzione ONU del 1948. Non si parla di genocidio in astratto, anzi si parla di genocidio *in quanto categoria giuridica codificata*. *Genocide allegations* è il secondo collocato il più forte. *Allegation* introduce una sorta di distanza, poichè probabilmente il giornale riporta l'accusa senza assertarla come fatto. *Genocide case* si riferisce al procedimento ICJ (Corte Internazionale di Giustizia) aperto dal Sudafrica contro Israele nel 2024.

Verbs with “genocide” as object:

Verb	AF	Score
<i>commit</i>	38	12.76
<i>prevent</i>	9	10.79
<i>End</i>	6	9.72
<i>Define</i>	5	10.4

Commit genocide è la collocazione dominante: "commettere un genocidio" è il frame narrativo principale: c'è un agente, c'è un'azione, c'è una vittima implicita. Questo verbo è anche quello giuridicamente codificato (la Convenzione ONU usa esattamente "committed"). I verbi preventivi (*prevent, stop, end*) costituiscono il frame dell’azione politica urgente.

Prepositional constructions

... of “genocide”

Collocate	AF	Score
<i>Israel</i>	9	11.98
<i>Accusation</i>	7	11.54
<i>Economy</i>	6	11.45
<i>Act</i>	5	11.17
<i>definition</i>	4	10.98
<i>Crime</i>	4	10.89

Economy of genocide è un sintagma che rimanda alla definizione fornita dalla segretaria Francesca Albanese che nel suo report del 30 giugno 2025 condanna il genocidio come un atto collegato all'economia perchè finanziato da grandi multinazionali e Stati. *Accusation of genocide*, *act of genocide* e *crime of genocide* consolidano il frame giuridico attribuito alla parola *genocidio*. *Risk of genocide* introduce una dimensione preventiva poiché si parla di *rischio* di genocidio, non solo di genocidio in atto.

"genocide" against	AF	Score
Palestinians	5	12.5

Genocide against Palestinians è la costruzione che identifica le vittime. Dallo score molto alto possiamo inferire che questa costruzione è quasi esclusiva: quando *genocide against* compare, il completamento è quasi sempre *Palestinians*.

In "genocide"	AF	Score
Complicity	3	12.7
Complicit	2	12.1

Complicity in genocide e *complicit in genocide* indicano che il corpus include un discorso sulla corresponsabilità: chi è complice del genocidio, chi lo consente senza agire. Costituiscono un frame morale e giuridico della responsabilità indiretta.

Possessor of "genocide"	AF	Score
Israel	7	8.82

Come nel Word Sketch di war, Israele è il possessore grammaticale. Ma lo Score qui (8.82) è più basso di quello di war (10.29), segno che il possessivo con *genocide* è costruzione meno solida, forse più contestata nel corpus.

Concordance

Collocate	AF	Coll.Freq.	T-Score	MI	LogDice
<i>Committing</i>	27	35	5.18843	9.39371	11.29136
<i>Convention</i>	1	18	4.57807	9.9905	11
<i>Against</i>	33	288	5.68707	6.64258	10.80122
<i>Allegations</i>	15	56	3.8564	7.86764	10.36
<i>Israel</i>	67	2030	7.90093	4.84693	9.86887

Committing

Committing compare 27 volte nel corpus e 35 volte nella finestra di *genocide*. Questo significa che la stragrande maggioranza delle occorrenze di *committing* nel corpus *gaza_press_pov* si trovano in prossimità di *genocide*. Il MI è il secondo più alto della tabella, quasi due punti sopra *against* e quasi cinque punti sopra *Israel*. Insieme, questi valori dicono che *committing genocide* è la collocazione più esclusiva e più robusta del profilo di *genocide* nel corpus stampa.

1. *Convention*

La Coll.Freq. (18) è inferiore all'AF (21), significa che *convention* compare 21 volte nel corpus ma solo 18 di queste occorrenze cadono nella finestra di *genocide*. Il MI 9.99 è il più alto in assoluto: significa che *convention* è il collocato più esclusivo di *genocide* nel corpus stampa, quando *convention* compare nel corpus lo fa quasi esclusivamente in prossimità di *genocide*. La costruzione *Genocide Convention* è talmente cristallizzata da essere quasi un'unità lessicale fissa, un termine tecnico-giuridico che il corpus usa sempre nella stessa forma. Il T-Score è il più basso della tabella: il T-score dice che *convention* non è un collocato frequente di *genocide*; il MI dice che è il collocato più esclusivo.

2. *Against*

La Coll.Freq. (288) è quasi nove volte l'AF (33). *Against* è il collocato con la Coll.Freq. più alta, ma il suo MI basso indica che è un collocato ad alta frequenza non esclusivo. La sua presenza riflette la ricchezza delle costruzioni preposizionali con *genocide*: il termine è sistematicamente costruito come qualcosa che si commette contro qualcuno (*against Palestinians, against people*) e qualcosa contro cui si va contro (*stand against*).

3. *Allegations*

Genocide allegations è uno dei costrutti del sistema di distanza epistemica già identificato nell'analisi dei Word Sketch. Il Log Dice 10.36 conferma che *allegations of genocide / genocide allegations* è una costruzione abbastanza stabile nel corpus stampa ma sufficientemente solida da essere sistematica. La stampa produce *genocide allegations* come formula di hedging: riporta l'accusa senza assertarla come fatto. Il fatto che *allegations* abbia MI quasi uguale a *against* (7.87 vs 6.64) e Log Dice superiore (10.36 vs 10.80) mostra che è

<i>Cultural</i>	1	11.41
<i>Settler-colonial</i>	1	11.30
<i>Ongoing</i>	1	11.30

Live-streamed genocide è un aggettivo radicalmente contemporaneo. Riflette la dimensione mediale del conflitto: il genocidio trasmesso in tempo reale sui social media. *Medicide* termine tecnico che indica la distruzione deliberata del sistema sanitario di una popolazione come forma di genocidio. È un termine medico-legale usato in rapporti specialistici. *Settler-colonial genocide* inquadra il genocidio all'interno di una struttura coloniale. *Cultural genocide* categoria distinta dal genocidio fisico, si riferisce alla distruzione della cultura, della lingua, delle istituzioni di un gruppo.

Nouns modified by “genocide”, nominal compound

Collocate	AF	Score
<i>Convention</i>	14	13.76
<i>Conspiracy</i>	2	11.61

Genocide Convention è il compound più forte della sezione, e coerente con il corpus stampa (dove aveva Score 12.83). Nei documenti ONU la Convenzione del 1948 è il riferimento giuridico fondamentale e il suo abbinamento con *genocide* è quasi automatico. *Genocide conspiracy* è un collocato assente nel corpus stampa. La cospirazione per commettere genocidio è un reato autonomo descritto all'interno della Convenzione ONU. La sua presenza qui indica che il corpus include atti giuridici (atti d'accusa, pareri legali, deliberazioni) che utilizzano il linguaggio tecnico del diritto internazionale penale.

Verbs with “genocide” as object

Verb	AF	Score
<i>Commit</i>	10	12.96
<i>prevent</i>	7	12.45
<i>punish</i>	3	11.63

La struttura verbale del corpus ONU è significativamente diversa da quella del corpus stampa: *Commit* rimane primo, si riscontra coerenza con il corpus stampa, perché "commettere genocidio" è il termine tecnico della Convenzione ONU, presente in entrambi i discorsi;

Prevent nel discorso ONU la prevenzione coincide con l'attribuzione. Questo riflette il mandato istituzionale: la Convenzione del 1948 obbliga gli Stati non solo a punire ma a prevenire il genocidio. *Punish*: la punizione è l'altro obbligo esplicito della Convenzione. Il binomio *prevent and punish* è la formula della Convenzione stessa. Questi verbi si categorizzano come verbi istituzionali che inquadrano il genocidio in un'ottica giuridica.

Collocate+ preposition	AF	Score
<i>Crime of genocide</i>	10	13.46
<i>Act of genocide</i>	4	12.41
<i>Complicity in genocide</i>	2	13.99
<i>Genocide in Strip</i>	5	12.92
<i>Accountability for genocide</i>	1	13.41

Crime of genocide è il collocato più forte dell'intera sezione, la costruzione *the crime of genocide* è la formula giuridica canonica della Convenzione ONU e del diritto internazionale penale. La sua forza collocazionale nel corpus ONU riflette la precisione tecnica del linguaggio istituzionale. *Act of genocide* è un'altra formula tecnica: è la formulazione usata nell'art. II della Convenzione per elencare le condotte vietate. *Genocide in the Strip*, indica che il corpus ONU usa preferibilmente la formulazione tecnica "Gaza Strip" rispetto al nome generico "Gaza", situa il fenomeno "genocide" in un luogo preciso. *Complicity in genocide* è la costruzione giuridica della responsabilità indiretta, disciplinata dall'articolo della Convenzione ONU. La sua posizione come collocato con score massimo nel corpus ONU è un risultato di assoluta rilevanza; similmente *Accountability* ha lo stesso peso all'interno del corpus, seppur con uno score più basso.

Concordance

Collocate	AF	Coll.Freq.	T-Score	MI	LogDice
<i>Convention</i>	16	34	3.94523	6.19043	12.17982
<i>Prevent</i>	13	21	3.56802	6.58601	12.05658
<i>Commit</i>	9	12	2.97423	6.86285	11.66213
<i>Gaza</i>	10	103	2.9524	3.91332	10.81413
<i>Israel</i>	8	111	2.57555	3.48347	10.43014

1. *Convention*

Convention compare 16 volte nel corpus ONU ma 34 volte nella finestra di *genocide*. Il Log Dice 12.18 è il valore più alto di tutta la tabella e il più alto dell'intera serie di analisi su corpora comparabili. Questo indica che *Genocide Convention* è la collocazione nominale più cristallizzata dell'intero corpus ONU: una formula quasi fissa, usata ogni volta che il termine

genocide compare in un documento istituzionale. Il MI è invece il più basso della tabella, inferiore a quello che *convention* aveva nel corpus stampa.

2. *Prevent*

Prevent compare 13 volte nel corpus ONU e 21 volte nella finestra di *genocide*. Anche qui la Coll.Freq. supera l'AF, segnale che la costruzione *prevent genocide* cattura la quasi totalità delle occorrenze di *prevent* nel corpus. Il Log Dice 12.06 e il MI 6.59 dicono che *prevent genocide* è la seconda collocazione più cristallizzata del corpus ONU dopo *Genocide Convention*. Questo risultato è di primaria importanza teorica. *Prevent* non compare tra i collocati di *genocide* nella tabella del corpus stampa, era presente nel Word Sketch (Score 10.79, freq 9) ma non tra i top 5 della concordanza. Nel corpus ONU è il secondo collocato più forte. La differenza riflette il mandato istituzionale: l'art. I della Convenzione ONU del 1948 obbliga gli Stati a *prevent and to punish the genocide*.

3. *Commit*

Commit compare 9 volte nel corpus ONU e 12 nella finestra di *genocide*. Il MI segnala che *commit* è il collocato più *esclusivo* di *genocide* nel corpus ONU: quando compare, lo fa quasi solo in prossimità di *genocide*.

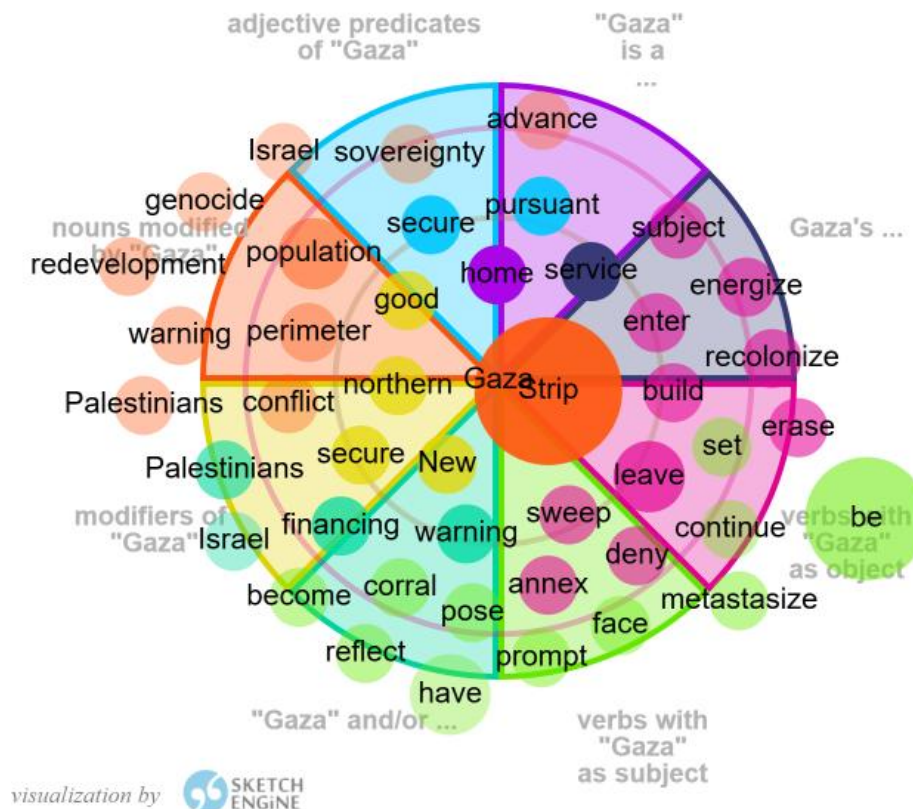
4. *Gaza*

Gaza compare 10 volte nel corpus ONU ma 103 volte nella finestra di *genocide*. Come già visto per *Gaza* e *Israel* nelle tabelle precedenti, questo riflette la pervasività del termine nel corpus: *Gaza* è così centrale nella documentazione che la sua frequenza attesa vicino a *genocide* è già molto alta. Il MI attesta che *Gaza* non è un collocato esclusivo di *genocide* nel corpus ONU, compare con molti altri nodi.

5. *Israel*

Israel compare solo 8 volte nel corpus ONU ma 111 volte nella finestra di *genocide*. Questo rapporto è il più alto dell'intera serie di analisi. Indica che quasi tutte le occorrenze di *Israel* nel corpus ONU si trovano in prossimità di *genocide* - ma il MI dice che il legame non è esclusivo, perché la frequenza attesa di *Israel* vicino a *genocide* è già molto alta dato quanto è centrale nel corpus.

2.3.4 Nodo “Gaza” (Onu_documentation corpus)

Immagine 5¹⁰

Analizzare Gaza come nodo collocazionale significa chiedersi cosa il discorso ONU fa con quel nome: come il lessico e la grammatica che lo circondano costruiscono Gaza come entità nel discorso istituzionale e internazionale. Gaza è il nodo con il profilo collocazionale più ricco del corpus ONU (112 collocati, AF 16), molto più ricco di war (9) e genocide (~40). Per un nome geografico, i collocati rivelano il frame del territorio: se Gaza è costruita come territorio da governare, da occupare, da proteggere, da distruggere, da ricostruire.

Verbs with “Gaza” as object

Verb	AF	Score
Leave	2	12.09
Recolonize	1	11.41
Erase	1	11.30
Deny	1	11.30

¹⁰ Grafico costruito da Wordsketch che mostra i collocati di Gaza (ONU) organizzati per relazioni grammaticali (creato con indici di frequenza)

È la sezione semanticamente più densa: *Recolonize* ed *erase Gaza* sono i due collocati semanticamente più marcati dell'intero corpus ONU. Il corpus istituzionale incorpora il linguaggio della violazione strutturale del diritto internazionale, presumibilmente per descriverlo criticamente o riportarlo come posizione documentata. All'interno del collocato *Recolonize Gaza* è interessante notare che il prefisso *re-* presuppone una colonizzazione precedente, inserendo il conflitto in un frame storico-strutturale coloniale. *Erase Gaza* probabilmente si riferisce alla cancellazione totale del territorio come azione. Il verbo *erase* costruisce Gaza come qualcosa che può essere eliminato dalla mappa. *Deny* costituisce il frame della deprivazione sistematica.

...in “Gaza”, Gaza as place

Collocate	AF	Score
<i>Palestinians</i>	9	12.92
<i>Genocide</i>	1	9.67
<i>Ceasefire</i>	1	10.09
<i>Destruction</i>	1	10.00

Palestinians in Gaza è il collocato più frequente della sezione e il più forte con frequenza > 1. Gaza è costruita nel corpus ONU prima di tutto come il luogo dei palestinesi; non un territorio astratto, non uno spazio militare, ma il luogo abitato da una popolazione specifica. Questo è il collocato con la maggiore solidità statistica nell'intera sezione preposizionale. La presenza di *genocide in Gaza* accanto a *ceasefire in Gaza* e *destruction in Gaza* costruisce un campo semantico della condizione umanitaria e giuridica: Gaza è il luogo dove avvengono distruzione, genocidio, e dove si chiede il cessate il fuoco.

... of “Gaza”

Collocate	AF	Score
<i>Control</i>	2	11.83
<i>Redevelopment</i>	2	11.83
<i>Destruction</i>	2	11.30
<i>Demilitarization</i>	1	10.91
<i>Invasion</i>	1	10.91

La sezione rivela il repertorio di questioni politiche e giuridiche che ruotano attorno a Gaza come oggetto di governo internazionale. *Control of Gaza* e *governance of Gaza* indicano la questione dell'amministrazione del territorio. *Demilitarization* e *reconstruction* sono le due proposte politiche principali del dopo-conflitto. *Invasion of Gaza* usa un termine giuridicamente preciso per caratterizzare l'operazione militare. *Destruction of Gaza* è la descrizione della condizione materiale.

Dall'analisi dei collocati risulta che Gaza è il nodo con il profilo collocazionale più ricco del corpus ONU; molto più ricco di *war* e significativamente più ricco di *genocide*. Il corpus ONU tace su *war* come termine, costruisce *genocide* come categoria giuridica, ma costruisce *Gaza* come entità geografica complessa attorno alla quale si organizza il discorso politico, umanitario, giuridico e morale del sistema internazionale.

Concordance

Collocate	AF	Coll.Freq.	T-Score	MI	LogDice
<i>Strip</i>	16	17	3.96395	6.79377	12.08114
<i>Palestinians</i>	13	60	3.46439	4.67478	11.34289
<i>Genocide</i>	5	25	2.14123	4.5593	10.3107
<i>operations</i>	4	10	1.95759	5.5593	10.16711
<i>destruction</i>	4	28	1.88124	4.07388	9.95561

1. *Strip*

Il collocato *Strip* è quello con l'associazione più forte nel corpus, come mostrano i valori molto alti di MI e LogDice. Questo significa che la parola compare vicino alla node word con una frequenza molto superiore a quella che ci si aspetterebbe casualmente. La concordanza suggerisce quindi che il corpus collega in maniera sistematica il tema principale a uno spazio geografico preciso.

2. *Palestinians*

Il T-score elevato indica che questa associazione è: frequente; stabile; e ricorrente nel tempo. Questo significa che il discorso non ruota soltanto intorno a eventi astratti o politici, ma mette continuamente al centro una popolazione specifica. La ripetizione sistematica del collocato suggerisce inoltre che il gruppo viene presentato come protagonista o vittima degli eventi descritti, e che la narrazione assume una forte componente identitaria e collettiva.

3. *Genocide*

Il collocato *Genocide* è particolarmente significativo perché possiede valori alti di associazione semantica (MI e LogDice). Ne deduciamo che il corpus non usa un linguaggio neutrale ma adotta un lessico estremamente forte e valutativo.

4. *operations*

Il collocato *operations* introduce invece una componente tecnico-militare nel corpus. Anche se la frequenza assoluta è più bassa, l'associazione rimane significativa. Questo suggerisce che il tema viene descritto non solo in termini emotivi o umanitari, ma anche attraverso il linguaggio strategico e istituzionale.

5. *destruction*

Il collocato *destruction* rafforza il tono negativo dell'intero corpus. La sua presenza ricorrente vicino alla node word suggerisce: devastazione; danno materiale; perdita; annientamento.

2.4 Wordsketch Difference

Il Word Sketch Difference confronta automaticamente i profili di due nodi nello stesso corpus, assegnando a ogni collocato un valore Color: +3 (esclusivo del nodo 1), -3 (esclusivo del nodo 2), 0 (condiviso), ± 1 (preferenziale). La misura di associazione è il Log Dice (scala 0-14): valori > 10 indicano collocazioni solide, $> 12-13$ collocazioni cristallizzate. I collocati con freq = 1 vengono citati come segnali qualitativi, non come misure robuste.

Wordsketch difference: genocide/war (gaza_press_pov)



Il WSD del corpus stampa è il più ricco e complesso dell'intera serie: 177 collocati totali. L'asimmetria è invertita rispetto al WSD ONU: nel corpus stampa è *war* a dominare. La stampa è il terreno lessicale della guerra, non del genocidio. *War* ha 113 collocati esclusivi contro 52 di *genocide*. L'inversione rispetto al corpus ONU (dove genocide dominava 65:7) documenta che i due discorsi costruiscono la realtà del conflitto attraverso sistemi terminologici radicalmente diversi.

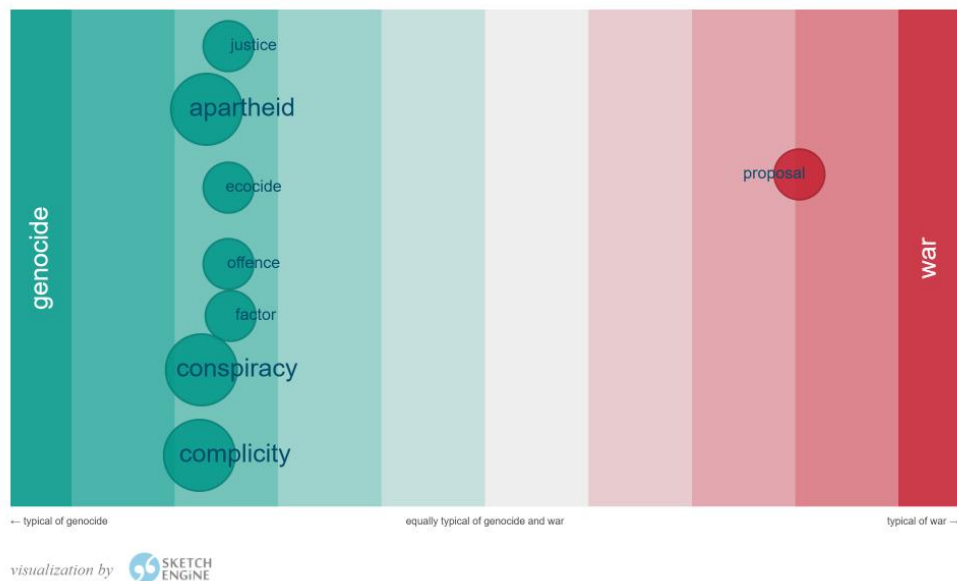
Il profilo semantico di *genocide* nel corpus stampa si organizza attorno a tre cluster:

1. Responsabilità e attribuzione: *commit* (38, Score 12.76), *Israel* (9, Score 11.98), *stand against* (3): il genocidio viene attribuito e si resiste ad esso.
2. Distanza epistemica: *allegation* (13, Score 12.25), *accusation* (7, Score 11.54), *describe* (2, Score 12.09), *amount to* (Score 13.00). La stampa non afferma il genocidio, anzi lo riporta come accusa, come descrizione, come soglia da valutare. Queste costruzioni sono il sistema di hedging giornalistico attorno al termine più giuridicamente e politicamente carico del corpus.
3. Responsabilità indiretta: *complicity* (3, Score 12.68), *complicit* (2, Score 12.09) costituisce il frame morale della corresponsabilità.

Il profilo semantico di *war* nel corpus stampa si organizza in tre cluster distinti:

1. Temporale-cronachistico; *month into the war* (3, Score 12.88), *year* (15, Score 11.67), *start* (20, Score 12.06), *beginning* (6), *end* (7): la stampa misura la guerra in unità di tempo e ne tracciano i confini narrativi.
2. Diplomatico; *end the war* (59, Score 12.64), *end to war* (27, Score 13.39): le due costruzioni più frequenti della soluzione politica; la guerra è il problema che la diplomazia deve risolvere.
3. Agentivo; *Hamas* come avversario (*war against Hamas*, *war with Hamas*), *Israel* come parte (*war between Israel*): la stampa nomina i belligeranti. *Genocide* non ha cluster agentivo equivalente con attori nominati nelle colonne di *war*.

Wordsketch difference: genocide/war (ONU)



I due file WSD del corpus ONU sono stati verificati come identici. Rappresentano un unico dataset - due esportazioni successive della stessa query da Sketch Engine.

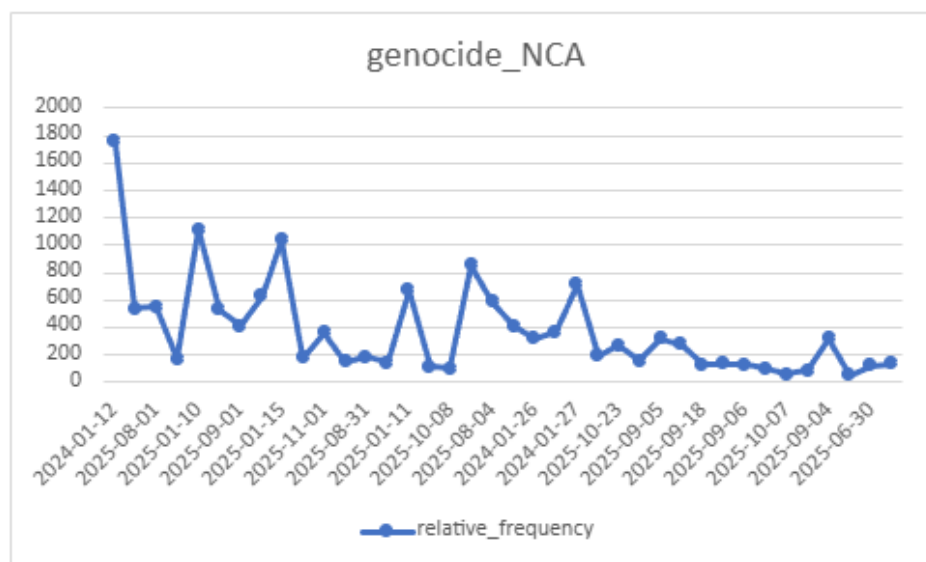
Le tre conclusioni principali che il WSD rende visibili con precisione numerica:

- La distanza epistemica evidenziata da *allegation* (Score 12.25), *accusation* (Score 11.54), è una strategia esclusiva del discorso giornalistico applicata a *genocide* e completamente assente per *war*. Il corpus ONU non ha collocati di hedging per *genocide*: si definisce, si commette, si previene. Nel corpus stampa il *genocidio* si allega, si accusa, si descrive come, ammonta a. Il termine più carico giuridicamente riceve nel discorso mediatico un sistema di attenuatori epistemici che ne riduce la forza assertiva. La guerra non riceve lo stesso trattamento: è un fatto registrato, non un'accusa da valutare.

- Il corpus ONU attiva obblighi giuridici solo per genocidio I verbi *commit*, *prevent*, *punish*, ovvero i tre verbi degli obblighi della Convenzione ONU del 1948, appartengono esclusivamente a genocidio. Il diritto internazionale ha costruito per genocidio un sistema di responsabilità e obblighi codificati che non si applica alla guerra in quanto tale. Questa asimmetria è incorporata nel lessico e nella grammatica del discorso istituzionale.
- Gaza è costruita diversamente dalla guerra e dal genocidio Gaza nel corpus ONU ha il profilo collocazionale più ricco della serie (112 collocati). È costruita simultaneamente come territorio conteso (recolonize, control, governance), come luogo di una popolazione (*Palestinians in Gaza*, freq 9, Score 12.92), come oggetto di violenza strutturale (*erase, invasion, destruction*).

2.5 Neighbouring Cluster Analysis (NCA) of “genocide”

La Neighbouring Cluster Analysis (NCA), è una variante del clustering gerarchico agglomerativo che, a differenza del clustering standard, rispetta il vincolo dell'**adiacenza temporale**: possono essere uniti solo punti che sono vicini nel tempo. Il risultato non è un dendrogramma libero, ma una **segmentazione della linea temporale** in periodi omogenei per frequenza d'uso del termine.



L'asse x è il tempo (gennaio 2024 - novembre 2025); l'asse y è la frequenza relativa (RF) di *genocide* per milione di token in ciascun documento. I cluster sono identificati dai punti di rottura: momenti in cui cambia bruscamente il livello di RF o si apre un gap temporale significativo tra documenti.

È possibile identificare 5 cluster differenti:

- **Cluster A (Gennaio - Marzo 2024):** Il primo cluster certifica il picco assoluto, di fatto il valore di RF più alto (1753.6) si registra il 12 gennaio 2024, seguito da 856.5 il giorno

successivo. Questo picco coincide cronologicamente con il deposito del ricorso del Sudafrica contro Israele alla Corte Internazionale di Giustizia (ICJ), avvenuto il 29 dicembre 2023, e con le prime udienze del gennaio 2024. Il termine *genocide* esplode nella documentazione, presumibilmente nei testi ICJ, nelle dichiarazioni degli Stati e nei rapporti dei Relatori Speciali perché per la prima volta entra ufficialmente nel registro giuridico processuale internazionale. La variabilità interna al cluster è elevata (RF da 135.7 a 1753.6): i documenti dello stesso periodo hanno densità lessicale del termine molto diverse, suggerendo che il corpus mescola testi ad alta carica giuridica (atti processuali) con testi di contorno, ad esempio comunicati e aggiornamenti.

- **Cluster B (Dicembre 2024 - Gennaio 2025):** Il secondo picco è separato dal Cluster A da un gap temporale di 269 giorni. Il secondo cluster segna un picco di intensità ancora alta (RF media 861.7, massimo 1112.4). Il divario è metodologicamente rilevante in quanto gap temporali ampi tra documenti consecutivi indicano cambiamenti di periodo discorsivo. In questo caso coincide probabilmente con nuovi sviluppi nel procedimento ICJ o con nuovi rapporti del Sistema ONU tra fine 2024 e inizio 2025. La variabilità interna è più ridotta rispetto al Cluster A, ne deduciamo che i documenti sono più omogenei nel livello di densità del termine.
- **Cluster C (Maggio - Luglio 2025):** Il cluster C attesta una fase di transizione poiché, dopo un secondo gap di 112 giorni, la RF crolla sotto i 100 punti (96.1 il 7 maggio 2025) prima di risalire parzialmente. Questo cluster è una fase di transizione discorsiva: il termine *genocide* non è scomparso, ma la sua frequenza si è normalizzata. I sei documenti di questo periodo mostrano oscillazioni moderate (96.1 - 411.7). È possibile inferire che il discorso si sia spostato da atti processuali con alta densità del termine a testi di diversa natura (risoluzioni, aggiornamenti, dichiarazioni) che usano *genocide* in modo più sporadico.
- **Cluster D (Agosto - Settembre 2025):** Il cluster attesta una fase densa e oscillante, con 14 documenti è il cluster più denso del corpus, di fatto quasi il 37% dell'intera serie si concentra in due mesi. La variabilità interna è fortissima: in settembre la RF oscilla tra 85.9 (8 settembre) e 533.9 (17 settembre) nello spazio di pochi giorni, con un'alternanza rapida di documenti ad alta e bassa densità del termine. Questo pattern di oscillazione ravvicinata suggerisce che in questo periodo circolano simultaneamente documenti di natura molto diversa: alcuni ad altissima densità giuridica (ad esempio il report della ONU Special Rapporteur Francesca Albanese e altri con uso sporadico del termine, ad esempio notizie del giorno, aggiornamenti e comunicati umanitari).
- **Cluster E (Ottobre - Novembre 2025):** L'ultimo cluster registra le RF più basse del corpus, il minimo assoluto si registra il 6 ottobre 2025 a 53.1. La tendenza è discendente, con l'eccezione di un picco anomalo a 361.3 il 1° novembre che interrompe la discesa prima di ridiscendere a 142.1. Questo picco finale isolato potrebbe corrispondere a un singolo documento ad alta densità, ed effettivamente corrisponde ai

rumours che circolavano sul trattato di pace stilato da Trump (e poi approvato dall'Onu qualche settimana dopo). La tendenza generale del Cluster E è verso la normalizzazione, in quanto il termine *genocide* viene usato con frequenza decrescente nella documentazione ONU, suggerendo o un cambio di frame discorsivo, o uno spostamento verso termini alternativi, come il termine *ceasefire*, o una fase di attesa procedurale.

L'andamento rilevato dai cluster è coerente con quanto rilevato nell'analisi collocazionale: nel corpus ONU *genocide* è un termine tecnico-giuridico, il cui uso è correlato a momenti specifici della vita processuale e istituzionale del conflitto, non a un flusso continuo di produzione documentale.

3. Conclusione

3.1 Riepilogo

La presente ricerca ha analizzato il rapporto tra le keyword *genocide* e *war* in un corpus di 250.099 token, articolato in due sub-corpora: quello giornalistico (239.516 token, composto da 334 articoli di giornale provenienti da sei testate britanniche) e quello della documentazione istituzionale internazionale (10.583 token, composto dai testi dell'ICJ, ONU e dal rapporto della Relatrice Speciale Francesca Albanese). L'arco temporale coperto, compreso tra gennaio 2024 e novembre 2025, include tre eventi di rilevanza internazionale: le misure cautelari ICJ dell'11 gennaio 2024, il rapporto Albanese del 30 giugno 2024 e l'approvazione del Piano di pace al Consiglio di Sicurezza ONU del 17 novembre 2025.

Seguendo il quadro teorico di Brezina (2018), l'analisi ha articolato cinque strumenti metodologici:

- Frequenze osservate e attese con chi-quadrato e Cramér's V
- Word Sketch
- Concordanza
- Word Sketch Difference
- Neighbouring Cluster Analysis

3.2 Distribuzione statistica: chi-quadrato e Cramér's V

Il primo strato dell'analisi ha verificato se la distribuzione di *genocide* e *war* nei due sub-corpora fosse casuale o sistematica. I risultati del test chi-quadrato (χ^2) sono inequivocabili:

Keyword	χ^2	p-value	Cramer's V	Effetto
<i>Genocide</i>	221,746	<0,001	0,029	<piccolo
<i>war</i>	21,111	<0,001	0,009	<piccolo

Entrambe le keyword sono statisticamente significative al livello più severo ($p < 0,001$): la loro distribuzione tra i due sub-corpora non è casuale. Le frequenze osservate rilevano pattern opposti e speculari: *genocide* compare 77 volte nel corpus giuridico contro le 17 attese (sovrarappresentata di 4,5 volte) mentre *war* compare solo 6 volte contro le 31 attese (sottorappresentata di 5 volte). I due termini gravitano, nel senso più letterale, verso registri diversi.

Il Cramér's V contestualizza questa significatività: $V = 0,029$ per *genocide* e $V = 0,009$ per *war* collocano entrambe le keyword al di sotto della soglia del piccolo effetto ($V < 0,10$ secondo Cohen, riportato in Brezina 2018:115). Questo risultato non contraddice la significatività statistica, ma la ridimensiona: la sproporzione dimensionale tra i due sub-corpora (239.516 token vs 10.583, con un rapporto 22:1) amplifica la sensibilità del chi-quadrato, producendo valori significativi anche in presenza di effettivi contenuti. La forza dell'associazione tra le

keyword e il loro corpus è reale ma non intensa. Il valore analitico dei risultati, quindi, risiede nella direzione dei pattern.

3.3 Analisi dei collocati: due grammatiche del conflitto

L'analisi collocazionale ha prodotto il risultato più ricco e più interpretabile dell'intera ricerca. I Word Sketch dei quattro nodi analizzati - *war* e *genocide* nel corpus giornalistico, *genocide* e Gaza nel corpus ONU - rilevano che i due discorsi non differiscono soltanto per frequenza: differiscono strutturalmente per il tipo di relazioni grammaticali che attivano intorno agli stessi termini.

3.3.1 Il nodo *war* nel corpus stampa: registro temporale-diplomatico

War è il termine narrativamente dominante nel corpus giornalistico. Con circa 50 collocati e frequenza massima di 73 occorrenze (*war* in Gaza, Score 11.21), costruisce il conflitto attraverso tre frame principali. Il frame temporale-cronachistico - *two year* (Score 11.37), *21-month*, *year of war*, *month of war* - misura la guerra in unità di tempo: la stampa la costruisce come evento con una durata, un inizio e una fine. Il frame diplomatico - *end of war* (freq 59, Score 12.64), *end to war* (freq 27, Score 13.39) - ne fa l'obiettivo politico centrale, qualcosa che deve finire. Il frame giuridico si concentra nel compound nominale *war crime* (freq 62, Score 13.36): è il collocato più forte dell'intero Word Sketch di *war*, e il fatto che la categorizzazione giuridica della violenza passi attraverso *war* - e non attraverso *genocide* - è uno dei risultati più rilevanti dell'analisi.

3.3.2 Il nodo *genocide* nel corpus stampa: registro accusatorio-epistemico

Genocide è collocazionalmente più selettivo di *war*: frequenze globali inferiori, ma collocati più esclusivi e valori di Log Dice più alti. Il profilo semantico si organizza attorno a tre cluster: responsabilità e attribuzione (*commit*, freq 38, Score 12.76; *Israel's genocide*, Score 11.98), sistema di distanza epistemica (*genocide allegations*, Score 12.25; *accusation of genocide*, Score 11.54; *described as genocide*, Score 12.09; *amounts to genocide*, Score 13.00) e responsabilità indiretta (*complicity in genocide*, Score 12.68; *complicit in genocide*, Score 12.09).

Il cluster della distanza epistemica è il risultato più rilevante per la domanda di ricerca. La stampa non afferma il genocidio: lo allega, lo descrive come, ne valuta la soglia. Questo sistema di hedging - costruzioni che riportano il termine come accusa altrui senza assertarlo come fatto proprio - è sistematico e statisticamente documentato. Lo stesso sistema è assente nel profilo di *war*: la guerra è un fatto registrato, non un'accusa da pesare.

3.3.3 I nodi *genocide* e Gaza nel Corpus ONU: registro definitorio-preventivo

Nel corpus ONU il profilo collocazionale di *genocide* è radicalmente diverso. Con circa 40 collocati esclusivi e valori di Log Dice tra i più alti della serie - *complicity in genocide* Score 13.99, *crime of genocide* Score 13.46, *genocide convention* Score 13.76 - il termine è costruito come fattispecie giuridica con una propria logica interna. I modificatori esclusivi del corpus ONU (*live-streamed*, *medicide*, *settler-colonial*, *cultural*) sono completamente assenti dalla stampa e riflettono la specializzazione tecnica del discorso istituzionale. I verbi *commit* (Score

12.96), prevent (Score 12.45) e punish (Score 11.63) riproducono il testo della Convenzione ONU del 1948 (art. I: "undertake to prevent and to punish"): nel corpus ONU il genocidio attiva obblighi giuridici codificati, non semplicemente narrazioni.

Gaza è il nodo con il profilo collocazionale più ricco del corpus ONU - 112 collocati, superiore sia a *war* (9) che a *genocide* (~40). È costruita simultaneamente come luogo di una popolazione (*Palestinians in Gaza*, freq 9, Score 12.92, il collocato più frequente e stabile dell'intera sezione), come oggetto di violenza strutturale (*recolonize, erase, annex, invasion*), come problema di governance (*control, demilitarization, reconstruction*) e come casa (*Gaza is a home*, Score 13.99). Il corpus istituzionale incorpora anche il frame della minaccia biologica (*metastasize, corral*) - probabilmente per riportarlo criticamente - accanto al frame umanizzante.

3.4 Word Sketch Difference: L'asimmetria documentata

Il Word Sketch Difference ha fornito la prova comparativa più diretta. Nel corpus stampa, *war* domina con 113 collocati esclusivi contro 52 di *genocide* (rapporto 2.2:1): la stampa è il terreno lessicale della guerra, non del genocidio. Nel corpus ONU il rapporto si inverte con violenza: *genocide* ha 65 collocati esclusivi contro 7 di *war* (rapporto 9.3:1).

Dimensione	War (stampa)	Genocide (stampa)	Genocide (ONU)	War (ONU)
Collocati esclusivi	113	52	65	7
Score max affidabile	13.39	13.00	13.99	13.99
Verbi oggetto esclusivi	13	13	8	1 (declare)
Frame dominante	Temporale-diplomatico	Accusatorio	Definitorio-preventivo	Giuridico-elusivo
Gaza con preposizione	in (73 occ.)	in (29 occ.)	in Strip (5 occ.)	On (1 occ.)

Il confronto preposizionale condensa in un unico dato la differenza discorsiva tra i due corpora: la stampa dice *war in Gaza* (preposizione di localizzazione geografica neutra, 73 occorrenze), il corpus ONU dice *war on Gaza* (preposizione di direzione e intenzionalità, 1 occorrenza). Una sola preposizione diversa, ma la seconda costruisce la guerra come azione condotta contro un territorio, non semplicemente localizzata al suo interno.

Il secondo dato rilevante del WSD ONU riguarda i verbi della Convenzione: *commit, prevent* e *punish* hanno Freq2 = 0 nel WSD ONU (non si applicano mai a *war*). Solo *genocide* attiva obblighi giuridici nel discorso istituzionale. *War* nel corpus ONU genera giustificazioni (*self-defence*) e incertezze epistemologiche (*fog of war, narrative of war*) - ma non obblighi.

3.5 Neighbouring Cluster Analysis: la dimensione diacronica

La NCA ha aggiunto alla ricerca una dimensione che nessun altro strumento poteva fornire: la correlazione tra l'intensità d'uso di *genocide* nel corpus ONU e i momenti specifici della vita processuale e istituzionale del conflitto. L'analisi ha identificato cinque cluster su 38 documenti (gennaio 2024 - novembre 2025):

Cluster	Periodo	RF media	Evento corrispondente
A	Gen - Mar 2024	629.5	Misure cautelari ICJ (11 gen 2024) - picco assoluto RF 1753.6
B	Dic 2024 - Gen 2025	861.7	Nuovi sviluppi procedimento ICJ - secondo picco (gap 269 gg da A)
C	Mag - Lug 2025	210.4	Transizione discorsiva - RF normalizzata dopo gap 112 gg
D	Ago - Set 2025	318.8	Fase densa (14 doc., 37% del corpus) - rapporto Albanese e altri
E	Ott - Nov 2025	151.3	Coda discendente - Piano di pace Trump, normalizzazione

Il pattern è coerente con l'analisi collocazionale: *genocide* nel corpus ONU è un termine tecnico-giuridico il cui uso è determinato da eventi istituzionali specifici - atti processuali, rapporti dei Relatori Speciali, sviluppi diplomatici. Non è un termine di uso corrente nel discorso istituzionale: si concentra nei momenti in cui il quadro normativo della Convenzione ONU viene formalmente attivato. La disaggregazione per doc.date delle concordanze conferma questo pattern: la RF di tutti i collocati chiave (*convention, commit, prevent, Israel*) è significativamente più alta nel documento dell'11 gennaio 2024 (atto processuale ICJ) che nel documento del 30 giugno 2025 (rapporto analitico), segnalando che il tipo di documento e la data determinano la densità del termine.

3.6 Conclusioni finali

Per rispondere alla domanda di ricerca, i due discorsi non differiscono soltanto per frequenza: differiscono strutturalmente per il tipo di relazioni grammaticali che attivano. Nel corpus ONU *genocide* è un termine tecnico-giuridico con obblighi codificati (*commit, prevent, punish*), formule canoniche (*crime of genocide, act of genocide, genocide convention*), e una logica interna propria (*genocide arises, occurs, requires*). Nel corpus stampa *genocide* è un termine marcato, usato con selettività, circondato da un sistema sistematico di distanza epistemica (*allegations, accusation, described as, amounts to*) che la guerra non riceve.

Il termine più carico giuridicamente (*genocide*) riceve nel discorso mediatico il sistema di attenuatori più intenso. *War*, inteso come termine narrativo, non tecnico-giuridico, viene assertito come fatto; mentre *genocide* viene riportato come accusa da valutare. Questa asimmetria è incorporata nel profilo collocazionale dei due termini e ha una forma quantitativamente verificabile attraverso i valori di Log Dice e MI.

Il discorso istituzionale ha un lessico della definizione, della prevenzione e della punizione dell'atto di genocidio, non della resistenza morale. Il discorso giornalistico include entrambi: il sistema legale e il sistema morale-politico della risposta. Il corpus stampa è più plurale nelle voci che incorpora, ma simultaneamente più opaco nella categorizzazione.

Infine, la NCA ha dimostrato che *genocide* nel corpus ONU non è un termine di uso continuo: si intensifica in corrispondenza di eventi istituzionali specifici (udienze ICJ, rapporti dei Relatori Speciali) e si normalizza nei periodi intermedi. Questo pattern diacronico conferma il carattere tecnico-giuridico del termine nel discorso istituzionale: la sua densità è determinata dall'attivazione del quadro normativo della Convenzione ONU, non da una narrazione corrente del conflitto.

BIBLIOGRAFIA

- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical guide*. Cambridge University Press.
- International Court of Justice. (2024, January 11). *Application of the Convention on the Prevention of the Crime of Genocide in the Gaza Strip (South Africa v. Israel)*. International Court of Justice.
- Albanese, F. (2025, June 30). *From Economy of Occupation to Economy of Genocide: Report on the Situation in Palestine*. United Nations.
- United Nations Security Council. (2025, November 17). *Resolution approving the "Trump Peace Plan" for Gaza*. United Nations Security Council.
- Newman, N., Fletcher, R., Robertson, C.T., Eddy, K. & Nielsen, R.K. (2024). *Reuters Institute Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024> (ultima visualizzazione: 24/05/2026).
- PAMCo / Ipsos Iris (2024). *UK National Newsbrands Monthly Reach Data*. Publishers Audience Measurement Company. <https://pamco.co.uk/> (ultima visualizzazione: 24/05/2026).

Formulario

Analisi del Corpus

Parametri del corpus

I seguenti dati sono la base per tutti i calcoli a seguire.

Parametro	Dati	Simbolo
Grandezza corpus giornalistico	239,516 tokens	N ₁
Grandezza corpus giuridico	10,583 tokens	N ₂
Dimensione totale corpus	250,099 tokens	N
Tipo di tabella	2 × 2 tabella di contingenza	-

Frequenze osservate (O)

Le frequenze osservate sono i conteggi grezzi estratti direttamente dal corpus tramite Sketch Engine. Esse costituiscono la base di tutti i calcoli successivi.

2.1 - "genocide"

	Giornalistico	Giuridico	Totale riga
"genocide"	314	77	391
Altre parole	239.202	10.506	249.708
Totale colonna	239.516	10.583	250.099

2.2 - "war"

	Giornalistico	Giuridico	Totale riga
"war"	729	6	735
Altre parole	238.787	10.577	249.364
Totale colonna	239.516	10.583	250.099

Frequenze attese (E)

Le frequenze attese rappresentano i valori di riferimento che ci aspetteremmo se la distribuzione della parola chiave fosse del tutto casuale e indipendente dal tipo di corpus

$$Expected\ frequency = \frac{row\ total \times column\ total}{grand\ total}$$

3.1 - "genocide"

Keyword	Corpus giornalistico	Corpus giuridico
"genocide"	$(391 \times 239.516) / 250.099 =$ 374	$(391 \times 10.583) / 250.099 =$ 17
Altre parole	$(249.708 \times 239.516) /$ $250.099 = 239,142$	$(249.708 \times 10.583) /$ 250.099 = 10,566

3.2 - "war"

Keyword	Corpus giornalistico	Corpus giuridico
"war"	$(735 \times 239.516) / 250.099 =$ 704	$(735 \times 10.583) / 250.099 =$ 31
Altre parole	$(249.364 \times 239.516) /$ $250.099 = \mathbf{238,812}$	$(249.364 \times 10.583) /$ $250.099 = \mathbf{10,552}$

4. Scarto quadratico

Il contributo di ciascuna cella al valore complessivo del chi-quadrato viene calcolato utilizzando la formula di seguito. La differenza al quadrato garantisce che tutti i valori siano positivi, e la divisione per **E** normalizza ciascun contributo rispetto al valore di riferimento atteso

$$\text{Chi - squared contribution} = \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

4.1 - "genocide"

Keyword	Corpus giornalistico	Corpus giuridico
genocide	$(314 - 374)^2 / 374 = \mathbf{9.63}$	$(77 - 17)^2 / 17 = \mathbf{211.76}$
Altre parole	$(239,202 - 239,142)^2 /$ $239,142 = \mathbf{0.01}$	$(10,506 - 10,566)^2 / 10,566$ $= \mathbf{0.34}$

4.2 - "war"

Keyword	Corpus giornalistico	Corpus giuridico
"war"	$(729 - 704)^2 / 704 = \mathbf{0,89}$	$(6 - 31)^2 / 31 = \mathbf{20,16}$
Altre parole	$(238.787 - 238.812)^2 /$ $238.812 = \mathbf{0,002}$	$(10.577 - 10.552)^2 / 10.552$ $= \mathbf{0,06}$

5. Chi-quadrato (χ^2)

Il chi-quadrato si ottiene sommando i contributi di tutte e quattro le celle. Il valore risultante viene confrontato con la soglia critica per $df = 1$

$$\chi^2 = \sum \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

5.1 - "genocide"

$$\chi^2 = 9,63 + 211,76 + 0,01 + 0,34 = \mathbf{221,74}$$

χ^2	Soglia critica ($p < 0.001$)	Risultato
221,74	10.828	$221,746 > 10.828 \rightarrow p < 0.001$

5.2 - "war"

$$\chi^2 = 0.89 + 20.16 + 0.002 + 0.06 = \mathbf{21,111}$$

χ^2	Soglia critica ($p < 0.001$)	Risultato
21,111	10.828	$21.111 > 10.828 \rightarrow p < 0.001$

Cramér's V

Il Cramér's V è una misura standardizzata dell'effetto derivata dal chi-quadrato. Essa tiene conto delle dimensioni del campione, rendendo i risultati comparabili tra corpora di dimensioni diverse. Per una tabella 2x2, (min - 1) la formula si semplifica in:

$$V = \sqrt{\frac{\chi^2}{\text{total observations} \times (\text{no. of rows or cols, whichever is smaller} - 1)}}$$

6.1 - "genocide"

$$V = \sqrt{(221,746 / 250,099)} = \sqrt{0.000847} = \mathbf{0.029}$$

6.2 - "war"

$$V = \sqrt{(21.111 / 250,099)} = \sqrt{0.0000844} = \mathbf{0.009}$$

6.3 – scala di interpretazione ($df = 1$, tabella 2x2)

Grande	Piccolo	Medio	Grande
Soglia di Cramér's V	< 0.10	0.10 - 0.30	> 0.50
genocide (V = 0.029)	← Al di sotto della soglia di effetto minimo		
war (V = 0.009)	← Al di sotto della soglia minimo d'effetto		
