

IRENE FIORAVANTI

ADOTTARE UNA MASCHERINA: ANALISI COMPUTAZIONALE E PSICOLINGUISTICA DELLE NUOVE COLLOCAZIONI IN TEMPO DI COVID

1. LE COLLOCAZIONI: UN FENOMENO TESTUALE E PSICOLINGUISTICO

Le collocazioni sono spesso definite come combinazioni ricorrenti di due o più parole, caratterizzate da un'alta frequenza e da una forte associazione reciproca (cfr. Firth 1957; Sinclair 1991; Evert 2007). La loro formazione può essere interpretata come il risultato di un uso ripetuto e frequente: più due parole ricorrono insieme ripetitivamente e frequentemente più queste saranno legate da un'alta associazione reciproca (cfr. Evert 2008). La ripetizione delle parole e delle collocazioni nel linguaggio è condizionata da diversi fattori, quali il contesto che ci circonda e le convenzioni sociolinguistiche (cfr. Sinclair 1991). In questo senso, si potrebbe affermare che nella pandemia da Covid-19 anche il lessico e le parole sono stati contagiati. Si pensi all'uso metaforico di parole come *trincea* e *guerra*, o ai tecnicismi medici entrati nel lessico quotidiano (es., *sintomo*, *tampone*, *vaccino*, *incubazione*, ecc.).

Secondo l'approccio frequentista, le collocazioni sono combinazioni di due o più parole (tecnicamente *co-occorrenze*) che tendono a ricorrere e a presentarsi insieme frequentemente (cfr. Firth 1957; Sinclair 1991; Evert 2008). Inoltre, gli elementi di una collocazione sono legati fra di loro da una sorta di "associazione reciproca" (cfr. Firth 1957), attraendosi a vicenda. La collocazione rappresenta, quindi, la "compagnia abituale" (cfr. Firth 1957) di una data parola (*nodo*), ovvero le parole (*collocati*)

con cui un nodo tende a ricorrere con maggiore probabilità e frequenza.

Le collocazioni sono definite, soprattutto, come un fenomeno testuale (cfr. Sinclair 1991). Più precisamente Sinclair (1991: 170) definisce una collocazione come «the occurrence of two or more words within a short span of each other's in a text». Lo span, definito da Sinclair, consisterebbe in un massimo di 4 parole a destra e a sinistra del nodo ed è in questo span che ricorrerebbero i collocati più probabili del nodo.¹

Più recentemente Evert (2008) ha proposto un metodo empirico per identificare le collocazioni. Definendo le collocazioni come combinazioni di due o più parole che tendono a ricorrere l'una accanto all'altra nel linguaggio naturale, ovvero a co-occorrere, perché legate da un'associazione reciproca, Evert (2008) propone di quantificare questa attrazione in termini statistici attraverso misure di associazione statistica (es., Mutual Information (MI); *t*-score; logDice...; cfr. Gablasova/Brezina/McEnery 2017). Queste misure di associazione calcolano la probabilità che due parole ricorrano insieme contro la probabilità che ricorrano l'una indipendentemente dall'altra in un testo.

Un'altra recente prospettiva è quella che vede le collocazioni come un fenomeno psicolinguistico. Alla base di tale prospettiva vi sarebbe l'idea che il nostro cervello è sensibile all'informazione distribuzionale del linguaggio (cfr. Bybee 2006). Più precisamente, la ripetizione e la frequenza di co-occorrenza di unità lessicali più grandi di due parole lascerebbe traccia nel nostro lessico mentale (cfr. Ellis 2002 e 2012). Il primo che ha identificato le collocazioni come fenomeno psicolinguistico è stato Hoey, il quale definisce una collocazione come una «psychological association between words» e «a psycholinguistic phenomenon the evidence for which can be found statistically in computer corpora» (Hoey 2005: 5). Secondo Hoey (2005: 8), la produzione di collocazioni dipenderebbe da meccanismi di *priming*, assumendo che «every word is mentally primed for collocational use». La definizione di Hoey rappresenta un collegamento fra le collocazioni nel testo e le collocazioni nella mente.

Per gli scopi del presente studio, le collocazioni sono state definite sia come un fenomeno testuale sia psicolinguistico. Nei paragrafi che seguiranno sarà proposta quindi un'analisi del lessico della pandemia,² focalizzandosi su un insieme ristretto di parole, con lo scopo di individuare le nuove collocazioni del Covid, adottando un approccio corpus-based, e conducendo un esperimento di *priming* con lo scopo di determinare la realtà psicolinguistica delle nuove collocazioni.

1 Il limite massimo per lo span di 4 funziona meglio per l'inglese; per altre lingue, come l'italiano, il limite massimo dello span è generalmente fissato a 5. Ciò dipende dal fatto che l'italiano presenta una morfologia e una flessione più articolata dell'inglese; per questo, fissando uno span a 5, è più probabile trovare collocazioni salienti del testo.

2 Si menzionano due ulteriori corpora rappresentativi della lingua post-Covid: Italc-Covid19 (Gagliardi/Suozzi 2020) e il corpus Twitter Covid-19 (Gagliardi/Gregori/Suozzi 2021).

2. LO STUDIO

Il presente studio analizza il lessico giornalistico per individuare le nuove collocazioni della pandemia. È stato preso in considerazione un insieme di parole della pandemia e per ciascuna di esse è stato ricostruito il suo profilo collocazionale pre-Covid (prima della pandemia) e post-Covid (dopo l'inizio della pandemia). È stata, infatti, realizzata un'indagine lessicale corpus-based, che consente un'esplorazione approfondita del lessico e dei profili collocazionali delle parole (cfr. Church/Hanks 1990; Lenci/Montemagni/Pirrelli 2005). Inoltre, per indagare la natura psicolinguistica delle collocazioni, è stato condotto un esperimento di *priming*.

2.1 Materiali

Le parole più frequenti della pandemia (*nodi*) sono state selezionate dalla lista delle "100 parole del coronavirus", pubblicata da Treccani insieme all'Istituto Superiore di Sanità.³ La lista è stata randomizzata e sono state scelte le prime dodici parole come oggetto d'analisi corpus-based e psicolinguistica di questo studio: *mascherina*, *affetto*, *zona*, *distanziamento*, *didattica*, *crisi*, *curva*, *congiunto*, *tampone*, *immunità*, *contagio* e *vaccino*.

Per poter investigare i profili collocazionali pre-Covid delle dodici parole della pandemia, è stato usato come corpus di riferimento *ItTenTen20*, un corpus di italiano costituito da testi raccolti dal web, contenente una vastissima quantità di dati (circa 5 miliardi di parole), che permette un'indagine approfondita dei profili collocazionali delle dodici parole della pandemia, ricavando l'informazione sulla frequenza di ricorrenza dei collocati con i nodi e sul loro grado di associazione.

I profili collocazionali post-Covid, invece, sono stati indagati in un corpus costruito *ad hoc* di linguaggio giornalistico (*ItLesCo – Corpus di Italiano del Lessico del Covid*), contenente articoli pubblicati tra febbraio e dicembre 2020, selezionati da tre testate giornalistiche: *La Repubblica*, *La Stampa* e *Il Post*. La selezione degli articoli è avvenuta attraverso la ricerca avanzata per parole chiave di Google. In tutto sono stati selezionati 378 testi,⁴ annotati in linguaggio XML per etichettarli in base alla testata giornalistica, alla sezione del giornale,⁵ all'anno, e al mese (cfr. tab. 1).

3 La lista contiene i vocaboli diventati di uso comune durante la pandemia da Covid-19, consultabile al sito www.treccani.it.

4 I testi sono così suddivisi fra le tre testate giornalistiche: 144 testi *La Repubblica*; 123 testi *La Stampa*; e 111 testi *Il Post*.

5 Le sezioni degli articoli sono: cronaca, politica, salute, esteri, scuola, economia e scienze.

ItLesCo	
Lingua	Italiano
Codifica	UTF-8
Testi	378
Token	332.304
Parole (type)	285.893
Fraasi	11.361
Periodo	1/02/2020 – 31/12/2020

tab. 1. *Composizione di ItLesCo.*

3. ANALISI CORPUS-BASED DEI PROFILI COLLOCAZIONALI PRE- E POST-COVID

L'analisi dei profili collocazionali pre- e post-Covid dei dodici nodi è stata condotta sia su *ItTenTen20* sia su *ItLesCo* con il software online *SketchEngine* (cfr. Kilgariff *et al.* 2004) applicando la funzione *Word Sketch*, che permette di indagare i collocati di una parola in diverse relazioni sintattiche (verbo e oggetto diretto, modificatori, soggetto e verbo, ecc.); sono riportati per ogni collocato la frequenza di ricorrenza e il grado di associazione con il nodo. Per il grado di associazione è riportato il valore del *logDice*, una misura di associazione che non dipende dalla grandezza del corpus⁶ e può essere, quindi, usata per comparare valori di associazione fra corpora di dimensioni diverse (cfr. Gablasova/Brezina/McEnery 2017).

Applicando la funzione *Word Sketch* di *SketchEngine*, sono stati inseriti come parole chiave della ricerca i dodici nodi della pandemia (es. *mascherina*); successivamente, sono stati analizzati i loro profili collocazionali in tre diverse relazioni sintattiche: la relazione Verbo+Nome (in cui il nodo configurava come oggetto diretto del collocato); la relazione Modificatore+Nome (in cui il collocato ricorreva come modificatore del nodo); e la relazione Nome+Nome (in cui il nodo e il collocato ricorrevano uniti da una preposizione o da una congiunzione). Per ogni relazione sintattica, i collocati sono stati ordinati in termini di *logDice* e, in seguito, sono stati estratti i dieci collocati più associati al nome sia pre- (es. *antismog*) sia post-covid (es. *chirurgica*). Tra i collocati pre-Covid e post-Covid è stata confrontata la media del grado di associazione con i nodi oggetto d'analisi (cfr. tab. 2)⁷ in ciascuna relazione

⁶ Il *logDice* ha il valore massimo standardizzato di 14 (cfr. Gablasova *et al.* 2017); questo significa che l'associazione fra due parole non è influenzata dalle dimensioni del corpus.

⁷ La media riporta il valore medio di *logDice* tra il nodo e il collocato nel profilo collocazionale pre- e post-Covid; la Deviazione Standard indica quanto il valore di *logDice* differisce dalla media; il Range riporta il valore minimo e il valore massimo di *logDice*.

sintattica (cfr. fig. 1 e 2).

Pre-Covid			
	Media	Deviazione Standard	Range
LogDice	5,87	2,65	0,06-11,5
Post-Covid			
	Media	Deviazione Standard	Range
LogDice	10,24	1,31	5,8-13,5

tab. 2. Riassunto dei valori di *logDice* nei profili collocazionali pre- e post-Covid.

mascherina



fig. 1. Word Sketch dei collocati pre-Covid del nodo “mascherina”.

mascherina

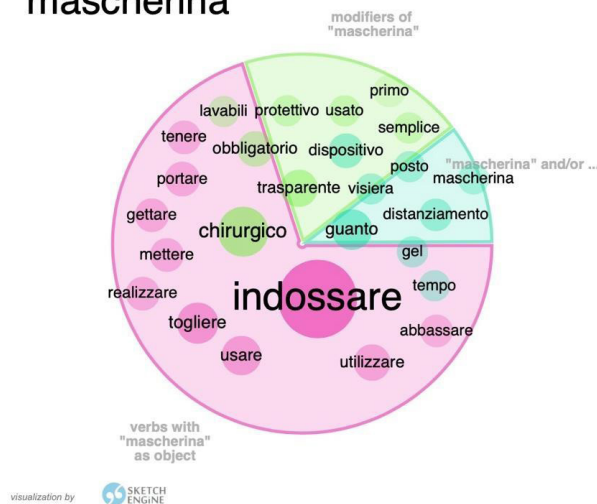


fig. 2. Word Sketch dei collocati post-Covid del nodo “mascherina”.

Dal confronto tra i profili collocazionali pre- e post-Covid dei dodici nodi sono effettivamente emersi nuovi collocati in termini di valore associativo ($\log\text{Dice} > 8$), come *mascherina obbligatoria*, *abbattere la curva*, *affetto stabile*. È interessante notare che sono presenti anche casi in cui lo stesso collocato è associato al nodo sia nel profilo pre-Covid sia nel profilo post-Covid, ma il legame associativo risulta più alto nel profilo post-Covid, come nel caso di *indossare la mascherina* ($\log\text{Dice}$ pre-Covid: 6,81 – $\log\text{Dice}$ post-Covid: 13,5) e di *mascherina chirurgica* ($\log\text{Dice}$ pre-Covid: 5,8 – $\log\text{Dice}$ post-Covid: 13,2). Infine, è emerso che in alcuni casi le collocazioni post-Covid assumono significati in contesto diversi e più specifici rispetto alle collocazioni pre-Covid, per esempio nel caso di *congiunto*, i collocati pre-Covid rimandano a un significato di “aiuto” e “assistenza”, diversamente i collocati post-Covid assumono il significato di “incontrare” e “identificare”.

4. INDAGINE PSICOLINGUISTICA DELLE COLLOCAZIONI POST-COVID

Per investigare la realtà psicologica di queste nuove collocazioni, è stato indagato il *priming* (cfr. McNamara 2005), che può essere definito come quel fenomeno per cui una parola (es. *dottore*) viene elaborata più velocemente da un parlante se preceduta da una parola ad essa correlata (es. *infermiera*) rispetto a una parola ad essa non correlata (es. *burro*). La parola su cui ricade l'elaborazione del parlante è chiamata *target*, mentre il contesto che la precede è chiamato *prime*. Se tra il *prime* e il *target* è presente una relazione semantica e/o associativa (cfr. Lucas 2000), allora l'elaborazione sarà più veloce; al contrario, se non sussiste una relazione semantica e/o associativa fra il *prime* e il *target*, l'elaborazione sarà più lenta.

Studi recenti hanno dimostrato che il fenomeno del *priming* è riscontrabile anche tra gli elementi di una collocazione, suggerendo come non solo le relazioni semantiche ma anche associative diano forma al lessico mentale dei parlanti (cfr. Lucas 2000). In particolare, si parla di *priming* collocazionale (cfr. Hoey 2005) quando si riscontra questo fenomeno tra gli elementi di una collocazione, che dipenderebbe non tanto da relazioni semantiche e/o associative, ma dal fatto che due parole ricorrono insieme frequentemente e ripetutamente.

Nel presente studio è stato, quindi, realizzato un task di decisione lessicale per indagare se ci fosse evidenza di *priming* collocazionale fra i nodi e i loro collocati post-Covid, ovvero se le nuove collocazioni del Covid avessero una realtà psicologica per i parlanti nativi di italiano, confrontandole con le collocazioni pre-Covid e con coppie di parole non identificabili come collocazioni. È stato scelto come metodo di indagine il task di decisione lessicale (cfr. Neely 1991) (ad oggi la tipologia di esperimento più usata per investigare il fenomeno del *priming*) in cui il soggetto deve decidere il più velocemente possibile se una parola esiste o non esiste nel lessico oggetto d'indagine. In questo esperimento hanno partecipato allo studio 48 parlanti nativi di italiano (media età: 28.3), che erano studenti, dottorandi e ricercatori presso

l'Università per Stranieri di Siena e l'Università per Stranieri di Perugia.

4.1 Design dell'esperimento

Ogni esperimento di *priming* consiste nel testare diverse condizioni sperimentali tra il *prime* e il *target*. Generalmente, le condizioni sperimentali sono due: una in cui sussiste un qualche tipo di relazione tra il *prime* e il *target* (condizione di relazione), e una in cui il *prime* e il *target* non sono correlati fra loro (condizione di controllo). Nel costruire il design dell'esperimento si possono seguire due opzioni: presentare lo stesso *prime* seguito da diversi *target*, o mantenere costante il *target* preceduto da diversi *prime*. Considerando che ogni possibile effetto di *priming* dipende dalla velocità di risposta dei partecipanti al *target* e non al *prime*, nel primo caso la differenza nei tempi di risposta tra la condizione di relazione e la condizione di controllo può dipendere non solo dalla presenza/assenza di relazione tra il *prime* e il *target*, ma anche dalle proprietà lessicali di entrambi. Questo limite può essere superato usando la seconda opzione; ciò significa che lo stesso *target* è usato in entrambe le condizioni sperimentali e che qualsiasi differenza nei tempi di reazione dipende esclusivamente da differenze nella relazione fra *prime* e *target* (cfr. Jiang 2012).

Per questo motivo, nel presente studio è stato scelto di mantenere costante il *target*. Più precisamente, nell'esperimento ciascuno dei dodici nodi è stato usato come *target* e inserito in tre condizioni sperimentali diverse: pre-Covid, post-Covid, e non-collocato. Nella condizione pre-Covid, il nodo-*target* (es. *mascherina*) era anticipato da un collocato pre-Covid (es. *antismog*), usato come *prime*; nella condizione post-Covid da un collocato post-Covid (es. *chirurgica*), sempre usato come *prime*; e nella condizione non-collocato, il *prime* era una parola non presente tra i collocati del nodo nel corpus di riferimento *ItTenTen20* o ricorrente tra i collocati del nodo ma con basso valore (< 2) di logDice (es. *elegante*). Di seguito sono mostrati per ciascun nodo-*target* i suoi collocati-*prime* nelle tre condizioni sperimentali (cfr. tab. 3).

Si può notare che non in tutti i casi la coppia *prime-target* è caratterizzata dallo stesso ordine (es. *mascherina antismog* ha ordine 1-2, mentre *didattica attività* ha ordine 2-1) e dalla stessa relazione sintattica (es. in *crisi economica* il collocato è aggettivo modificatore del nome, mentre in *vaccino iniettare*, il nodo è oggetto diretto del verbo). Questi parametri non sono stati mantenuti costanti in quanto è stato scelto di presentare il nodo come *target* e non come *prime*; ciò esclude, come ribadito precedentemente, l'interferenza di altre proprietà lessicali o sintattiche nell'elaborazione dello stimolo.

Nodo	Collocato pre-Covid	Collocato post-Covid	Non-Collocato
mascherina	antismog	chirurgica	elegante
affetto	stretto	stabile	sincero
zona	umida	rossa	piena
distanziamento	emotivo	sociale	ironico
didattica	attività	distanza	supporto
crisi	economica	sanitaria	frequente
curva	percorrere	abbattere	completare
congiunto	assistere	incontrare	difendere
tampone	effettuare	eseguire	rimuovere
immunità	conferire	ottenere	revocare
contagio	propagare	contenere	terminare
vaccino	iniettare	scoprire	bloccare

tab. 3. *I nodi e i loro collocati nelle tre diverse condizioni sperimentali.*

Riassumendo, sono stati usati per l'esperimento 36 item sperimentali (coppia nodo-collocato): 12 collocazioni pre-Covid, 12 collocazioni post-Covid, e 12 non collocazioni.⁸ I 36 item sperimentali sono stati suddivisi in tre liste sperimentali, per evitare che i soggetti incontrassero più di una volta la stessa parola, ma vedessero gli stimoli in tutte e tre le condizioni sperimentali. Ad esempio, il nodo *mascherina* è stato combinato con il collocato *antismog* nella prima lista, con il collocato *chirurgica* nella seconda lista, e con il non-collocato *elegante* nella terza lista. Inoltre, in ciascuna lista sono state inserite 12 coppie in cui il *target* era una pseudoparola (es. camera – *orpilata*). In totale, ciascuna lista sperimentale comprendeva 24 item (12 coppie sperimentali e 12 coppie non-parola).

4.2 Procedura

Il task di decisione lessicale è stato creato e distribuito ai partecipanti attraverso PsyToolkit (www.psychtoolkit.org), una piattaforma online che permette di creare e condurre esperimenti psicolinguistici attraverso un semplice linguaggio di programmazione e un'interfaccia grafica (cfr. Stoet 2010 e 2017). L'esperimento è stato distribuito online, fornendo ad ogni partecipante, assegnato in modo casuale a una delle tre liste, un link che lo/la rimandava direttamente al task di decisione lessicale. Inizialmente, apparivano ai partecipanti le istruzioni dell'esperimento, che spiegavano loro che dovevano decidere il più velocemente e il più accuratamente possibile se le

⁸ Le dodici non-collocazioni sono state usate come condizioni di controllo, in cui non dovrebbe verificarsi il fenomeno del *priming*.

parole che leggevano sullo schermo fossero reali o non reali in italiano. L'esperimento iniziava con un punto di fissazione (“+”) che restava sullo schermo per 500ms; il punto di fissazione veniva sostituito dal *prime* (collocato pre-Covid; collocato post-Covid; non-collocato), che restava brevemente sullo schermo (500ms); e, infine, il *prime* veniva rimpiazzato dal *target* (il nodo o la pseudoparola) che restava sullo schermo per 3000ms (cfr. fig. 3). I soggetti dovevano, quindi, decidere in tre secondi se il *target* era una parola esistente o non esistente in italiano, schiacciando il tasto sulla tastiera del proprio computer per la risposta “Sì”, se ritenevano che il *target* fosse una parola reale in italiano, o per la risposta “No”, se ritenevano che il *target* non fosse una parola esistente in italiano. I tempi di reazione sono stati misurati da quando il *target* appariva sullo schermo a quando il soggetto dava la sua risposta.

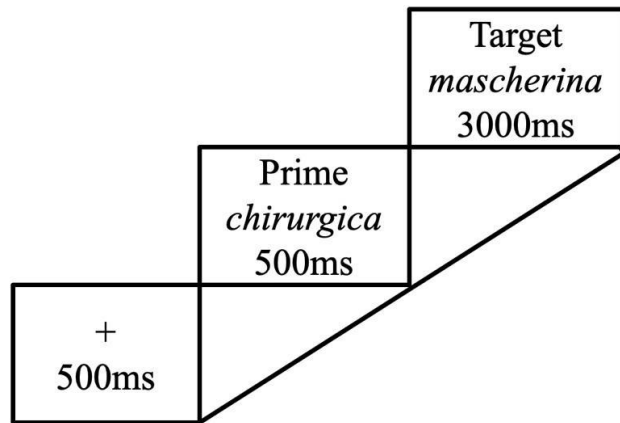


fig. 3. Sequenza degli item nel task di decisione lessicale.

4.3 Analisi e risultati

I tempi di reazione dei partecipanti sono stati analizzati prima con un'analisi statistica descrittiva per indagare la media dei tempi di reazione nelle tre condizioni sperimentali, e in seguito con un'analisi statistica inferenziale per trovare possibili differenze significative nei tempi di reazione tra le tre diverse condizioni sperimentali.

L'analisi descrittiva dei dati (cfr. tab. 4) ha mostrato che la media dei tempi di reazione dei partecipanti è più bassa nella condizione post-Covid (728.8ms), rispetto alla condizione pre-Covid (+14.3ms) e non-collocato (+93.7ms), suggerendo che i tempi di reazione nella condizione sperimentale post-Covid sono più veloci rispetto alle altre due condizioni sperimentali. Inoltre, le medie dei tempi di reazione nelle due condizioni sperimentali pre-Covid e post-Covid sono più basse rispetto alla media dei tempi di reazione nella condizione non-collocato, lasciando intendere un possibile effetto *priming* sia fra gli elementi delle collocazioni pre-Covid sia delle collocazioni post-Covid (cfr. tab. 5).

NUOVE COLLOCAZIONI IN TEMPO DI COVID

Condizione	Media	Mediana	Deviazione Standard
Pre-Covid	743.1	700.0	232.5
Post-Covid	728.8	688.5	236.9
Non-Collocato	822.5	778.5	270.2

tab. 4. *Analisi descrittiva dei tempi di reazione nelle tre condizioni sperimentali.*

Condizione	Media dei tempi di reazione delle collocazioni (ms)	Media dei tempi di reazione delle non-collocazioni (ms)	Differenza nei tempi di reazione (ms)
Pre-Covid	743.1	822.5	79.4
Post-Covid	728.8	822.5	93.7

tab. 5. *Differenza nelle medie dei tempi di reazione tra collocazioni e coppie di controllo.*

In seguito, è stata condotta un'analisi statistica inferenziale per indagare da che cosa dipendesse la differenza nei tempi di reazione tra le tre diverse condizioni sperimentali. Per statistica inferenziale si intende quella parte della statistica che analizza il comportamento di un campione (es., un gruppo di parlanti) per poter poi generalizzare i dati ottenuti alla popolazione stessa (es., l'intero insieme dei parlanti). La statistica inferenziale studia quali fattori influiscono sul comportamento del campione. I fattori costituiscono le variabili indipendenti, mentre il comportamento dei parlanti è indicizzato con valori matematici (es., la media o i tempi di reazione). Per sapere se un dato fattore agisce significativamente sul comportamento del campione, si osserva il *p*-value, la cui soglia è posta a .05; se il *p*-value risulta inferiore a .05, si può assumere che l'ipotesi di partenza sia corretta e che il fattore indagato influisca significativamente sul comportamento dei parlanti.

Una delle tecniche più usate di statistica inferenziale sono i modelli misti (cfr. Baayen/Davidson/Bates 2008). Come riportano Siyanova-Chantura/Spina (2019), i modelli misti sono un metodo statistico raffinato ed elegante, che presentano diversi vantaggi rispetto ad altre analisi statistiche inferenziali, come l'ANOVA, il t-test, le analisi di correlazione, ecc. Per esempio, nei modelli misti possono essere inserite in un unico modello sia variabili continue sia variabili categoriche. Allo stesso modo la variabile dipendente di un modello misto può essere sia continua (giudizi di accettabilità, tempi di reazione) sia categorica (Sì/No). In un modello misto è possibile includere non solo predittori singoli, ma anche interazioni fra due o più variabili indipendenti. Inoltre, i modelli misti, come fa intuire il loro stesso nome, permettono di integrare e di incrociare in un unico modello i partecipanti e gli item.

Nella presente analisi i modelli sono stati costruiti utilizzando come variabile dipendente i tempi di reazione dei partecipanti e come variabili indipendenti i seguenti fattori relativi sia agli item sia ai parlanti (cfr. tab. 6): la condizione sperimentale (pre-Covid; post-Covid; non-collocato); la frequenza frasale della coppia (ricavata

dai due corpora di riferimento e trasformata in logaritmo per normalizzare la distribuzione di frequenza); il valore di logDice (ricavato dai due corpora di riferimento); il sesso dei partecipanti; l'età dei partecipanti; la lista sperimentale a cui apparteneva la coppia *prime-target*,⁹ e l'ordine di presentazione degli stimoli nel task (trial).

Variabile	Parlanti/Item	Caratteristiche
Condizione sperimentale	Item	Categorica a tre livelli (pre-Covid; post-Covid; non-collocato)
Frequenza frasale della coppia	Item	Continua
LogDice	Item	Continua
Sesso	Parlanti	Categorica a tre livelli (M; F; Altro)
Età	Parlanti	Continua
Lista	Item	Categorica a tre livelli (1; 2; 3)
Trial	Item	Numerica

tab. 6. *Le variabili indipendenti e le loro caratteristiche.*

I risultati hanno mostrato che la differenza nei tempi di reazione dipende significativamente dalla condizione sperimentale degli item ($p < .01$). Nessuno degli altri fattori è risultato influenzare significativamente i tempi di risposta. Inoltre, per osservare se ci fossero differenze significative fra le tre condizioni sperimentali, è stata condotta un'analisi *posthoc*, che ha confrontato le medie dei tempi di reazione delle tre condizioni sperimentali fra di loro. L'analisi ha evidenziato che la differenza nei tempi di reazione fra la condizione post-Covid e la condizione non-collocato è significativa ($p < .01$), ma non è significativa fra la condizione pre-Covid e la condizione non-collocato ($p > .05$), e non fra la condizione pre-Covid e post-Covid ($p > .05$).

5. CONCLUSIONI

Il presente studio ha proposto un'analisi sia corpus-based sia psicolinguistica per individuare le nuove collocazioni in tempo di Covid. A tal fine, sono state selezionate dodici parole della pandemia e sono stati esplorati i loro profili collocazionali pre- e post-Covid in un corpus di riferimento dell'italiano e in un corpus costruito *ad hoc*, costituito da articoli di giornale trattanti il tema della pandemia. L'analisi corpus-based ha rilevato nuovi collocati post-Covid delle dodici parole in termini di legame associativo.

Successivamente, per indagare la realtà psicolinguistica delle nuove collocazioni in parlanti nativi di italiano, è stato condotto un esperimento di *priming*. I risultati

⁹ Ciò permette di osservare se le differenze dei tempi di reazione dipendono da come gli stimoli sono stati distribuiti fra le diverse liste.

hanno mostrato un'evidenza del *priming* fra gli elementi delle collocazioni post-Covid, ma non fra gli elementi delle collocazioni pre-Covid. La realtà psicolinguistica delle nuove collocazioni può dipendere dalla loro ripetizione e dalla frequenza di ricorrenza nel linguaggio quotidiano e, quindi, dalla nostra esposizione, come parlanti, al lessico della pandemia. È interessante notare come l'evidenza del *priming* non sia stata trovata nelle collocazioni pre-Covid. Ciò potrebbe dipendere dal fatto che la continua esposizione al contesto della pandemia abbia rafforzato i legami associativi nel lessico mentale dei parlanti fra i nodi e i loro nuovi collocati, e indebolito i legami fra i nodi e i loro collocati pre-Covid.

In conclusione, un'analisi corpus-based e psicolinguistica rappresenta un valido metodo per identificare nuove collocazioni ed esplorare i profili collocazionali delle parole sia dal punto di vista lessicale sia cognitivo.

BIBLIOGRAFIA

- Baayen/Davidson/Bates 2008 = Harald R. Baayen / Donald J. Davidson / Douglas M. Bates, *Mixed-effects modeling with crossed random effects for subjects and items*, in «Journal of Memory and Language», 59, pp. 390-412.
- Bybee 2006 = Joan Bybee, *From usage to grammar: The mind's response to repetition*, in «Language», 82, pp. 711-733.
- Church/Hanks 1990 = Kenneth Ward Church / Patrick Hanks, *Word association norms, mutual information, and lexicography*, in «Computational Linguistics», 16 (1), pp. 22-29.
- Ellis 2002 = Nick C. Ellis, *Reflections on frequency effects in language processing*, in «Studies in Second Language Acquisition», 24, pp. 297-339.
- Ellis 2012 = Nick C. Ellis, *Frequency effects*, in Peter Robinson (a cura di), *The Routledge Encyclopedia of Second Language Acquisition*, New York, Routledge, pp. 260-265.
- Evert 2008 = Stefan Evert, *Corpora and collocations*, in Anke Lüdeling / Merja Kytö (a cura di), *Corpus linguistics. An international handbook*, Berlino, Mouton de Gruyter, pp. 1212-1248.
- Firth 1957 = John Rupert Firth, *A synopsis of linguistic theory, 1930-55*, in *Studies in Linguistic Analysis*, Oxford, Philological Society, pp. 1-32.
- Gablasova/Brezina/McEnery 2017 = Dana Gablasova / Vaclav Brezina / Tony McEnery, *Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence: Collocations in Corpus-Based Language Learning Research*, in «Language Learning», 67 (1), pp. 155-179.
- Gagliardi/Suoizzi 2020 = Gloria Gagliardi / Alice Suoizzi, *ItalC-Covid19: Institutional Communication in Times of COVID-19*, <http://gloriagagliardi.altervista.org/psychological-reactions-to-the-pandemic/>.
- Gagliardi/Gregori/Suoizzi 2021 = Gloria Gagliardi / Lorenzo Gregori / Alice Suoizzi, *L'impatto emotivo della comunicazione istituzionale durante la pandemia di Covid-19: uno studio di Twitter Sentiment Analysis*, in Johanna Monti / Felice Dell'Orletta / Fabio Tamburini (a cura di), *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020*, Torino, Accademia University Press, pp. 205-210.

- Hoey 2005 = Michael Hoey, *Lexical Priming: A new theory of words and language*, Londra e New York, Routledge.
- Jiang 2012 = Nan Jiang, *Conducting reaction time research in second language studies*, Londra e New York, Routledge.
- Kilgariff *et al.* 2004 = Adam Kilgariff / Paver Rychlý / Paverl Smrž / David Tugwell, *The Sketch Engine*, in «Information Technology», 105, pp. 1-116.
- Lenci/Montemagni/Pirrelli 2005 = Alessandro Lenci / Simonetta Montemagni / Vito Pirrelli, *Testo e computer. Elementi di linguistica computazionale*, Roma, Carocci.
- Lucas 2000 = Margery Lucas, *Semantic priming without association: a meta-analytic review*, in «Psychonomic Bulletin & Review», 7 (4), pp. 618-630.
- McNamara 2005 = Timothy P. McNamara, *Semantic priming: perspectives from memory and word recognition*, New York e Hove, Psychology Press.
- Neely 1991 = James H. Neely, *Semantic priming effects in visual word recognition: A selective review of current findings and theories*, in Besner Derek / Humphreys Glyn W. (a cura di), *Basic processes in reading: Visual word recognition*, Hillsdale, NJ, Lawrence Erlbaum Associates, Inc., pp. 264-336.
- Sinclair 1991 = John Sinclair, *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Siyanova-Chanturia/Spina 2019 = Anna Siyanova-Chanturia / Stefania Spina, *Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study*, in «Language Learning», 70 (2), pp. 420-463.
- Stoet 2010 = Gijsbert Stoet, *PsyToolkit - A software package for programming psychological experiments using Linux*, in «Behavior Research Methods», 42 (4), pp. 1096-1104.
- Stoet 2017 = Gijsbert Stoet, *A novel web-based method for running online questionnaires and reaction-time experiments*, in «Teaching of Psychology», 44 (1), pp. 24-31.