
Irene Fioravanti

T RA LE PAROLE
NELLA MENTE

**Studi interdisciplinari
sulle collocazioni lessicali**

ML

Materiali Linguistici. Collana a cura del Dipartimento di Studi Umanistici – Sezione di Linguistica Teorica e Applicata – dell’Università di Pavia

Materiali Linguistici è una collana fondata nel 1990, che pubblica studi su vari aspetti delle lingue naturali e della loro evoluzione storica. La collana riflette la molteplicità di punti di vista teorici e metodologici che la linguistica oggi comprende, con riferimento in particolare ai seguenti temi: descrizione dei sistemi linguistici, tipologia, sociolinguistica, semantica, pragmatica e apprendimento linguistico. *Ogni volume è sottoposto a processo di peer review.*

Materiali Linguistici. Book Series edited by the Department of Humanities – Section of Theoretical and Applied Linguistics – University of Pavia

Materiali Linguistici is a peer-reviewed series founded in 1990. It publishes studies on various aspects of natural languages and their historical evolution. The series reflects the wide range of methodological and theoretical approaches of contemporary linguistics, with particular reference to the following themes: description of linguistic systems, typology, sociolinguistics, semantics, pragmatics and language learning.

Direzione – General Editors

Anna Giacalone Ramat, Elisa Roma (*Università di Pavia*)

Comitato Scientifico di Redazione – Editorial Board

Università di Pavia

Ilaria Fiorentini

Maicol Formentelli

Maria Freddi

Elisabetta Jezek

Silvia Luraghi

Gianguido Manzelli

Maria Pavesi

Chiara Zanchi

Cecilia Andorno, *Università di Torino*

Annalisa Baicchi *Università di Genova*

Giuliano Bernini, *Università di Bergamo*

Sonia Cristofaro, *Sorbonne Université*

Pierluigi Cuzzolin, *Università di Bergamo*

Guglielmo Inglese, *Università di Torino*

Caterina Mauri, *Università di Bologna*

Vito Pirrelli, *CNR Pisa*

Michele Prandi, *Università di Genova*

Irina Prodanof, *CNR Pisa*

Paolo Ramat, *Accademia dei Lincei*

Andrea Sansò, *Università dell’Insubria*

Massimo Vedovelli, *Università di Siena*

Segreteria – Editorial Assistant

Ilaria Fiorentini (*Università di Pavia*)

Dipartimento di Studi umanistici – Sezione di Linguistica

C.so Carlo Alberto 5, I-27100 Pavia; tel. 0382984484.

Per maggiori informazioni: www.lettere.unipv.it/diplinguistica

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “Informatemi” per ricevere via e-mail le segnalazioni delle novità.

Il volume è stato pubblicato con il contributo dell'Università per Stranieri di Siena – Scuola Superiore di Dottorato e di Specializzazione.



1a edizione. Copyright © 2022 by FrancoAngeli s.r.l., Milano, Italy

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

Irene Fioravanti

TRA LE PAROLE NELLA MENTE

Studi interdisciplinari
sulle collocazioni lessicali

FRANCOANGELI

Al mio papà

Indice

Sigle e abbreviazioni	pag.	11
Introduzione	»	13
Parte prima Lo stato dell'arte		
1. La natura intrigante delle collocazioni	»	21
1.1. Che cosa sono le collocazioni?	»	21
1.2. L'approccio frequentista	»	22
1.2.1. I neo-Firthiani	»	23
1.2.2. Le misure di associazione	»	25
1.3. L'approccio fraseologico	»	28
1.3.1. Il <i>continuum</i> fraseologico di Cowie	»	29
1.3.2. Il modello fraseologico di Howarth	»	30
1.3.3. Un approccio formale alle collocazioni lessicali	»	31
1.4. Proposte di classificazione	»	33
1.4.1. Il concetto di 'restricted sense'	»	33
1.4.2. Classificazioni sulla lingua italiana	»	35
1.5. Le collocazioni come sequenze formulari	»	41
1.6. Le collocazioni come fenomeno cognitivo	»	45
1.7. La «migliore» definizione di collocazione	»	47
2. Le collocazioni nell'apprendimento L2	»	48
2.1. Il ruolo chiave delle collocazioni nell'apprendimento L2	»	48

2.2. I modelli sull'apprendimento L2 delle collocazioni	pag.	50
2.3. Produzione e uso delle collocazioni in L2	»	52
2.4. L'influenza della L1 sullo sviluppo dei <i>pattern</i> collocazionali in L2	»	59
2.5. Lo sviluppo della conoscenza collocazionale in L2	»	64
2.6. L'apprendimento esplicito e implicito delle collocazioni	»	70
2.7. Conclusioni	»	75

Parte seconda

La ricerca

3. Strumenti per una ricerca interdisciplinare	»	81
3.1. Linguistica dei corpora e psicolinguistica	»	81
3.2. Metodi quantitativi per identificare le collocazioni	»	83
3.2.1. Sketch Engine: uno strumento multifunzione	»	86
3.2.2. <i>LexIt</i> per l'esplorazione dei profili distribuzionali delle parole	»	88
3.2.3. I modelli distribuzionali	»	90
3.3. Metodologie psicolinguistiche	»	94
3.3.1. Studi offline sulle collocazioni in L1 e L2	»	96
3.3.2. La tecnica del <i>priming</i>	»	98
3.3.2.1. La metodologia del <i>priming</i> nelle collocazioni in L1 e L2	»	102
3.3.3. La metodologia <i>eye-tracking</i>	»	104
3.3.3.1. <i>L'eye-tracking</i> nella ricerca sul processing delle collocazioni	»	108
3.4. Conclusioni	»	110
4. Collocazioni in L1 e L2: uno studio <i>corpus-based</i> e comportamentale	»	111
4.1. Introduzione	»	111
4.2. Studio 1: la metodologia	»	112
4.2.1. Studio 1: partecipanti e procedura	»	116
4.2.2. Studio 1: analisi e risultati	»	117
4.3. Studio 2: la metodologia	»	121
4.3.1. Studio 2: partecipanti e procedura	»	122
4.3.2. Studio 2: analisi e risultati	»	123
4.4. Discussione dei risultati	»	127

5. Il fenomeno del <i>collocational priming</i> in L1 e L2	pag.	130
5.1. La teoria del <i>collocational priming</i>	»	130
5.2. La metodologia	»	131
5.3. Partecipanti e procedura	»	135
5.4. Analisi e risultati	»	137
5.5. Discussione dei risultati	»	140
6. Uno studio <i>eye-tracking</i> in L1 e L2	»	144
6.1. Introduzione	»	144
6.2. La selezione degli stimoli	»	145
6.3. Partecipanti e procedura	»	148
6.4. Le analisi e i risultati	»	149
6.4.1. La prima AOI: l'intera combinazione	»	151
6.4.2. La seconda AOI: il verbo	»	152
6.4.3. La terza AOI: il nome	»	153
6.5. Discussione dei risultati	»	154
Parte terza		
Conclusioni		
7. Il nocciolo della questione	»	161
7.1. Cos'è quindi una collocazione?	»	161
7.2. Le collocazioni nell'apprendimento L2	»	168
7.3. Implicazioni pedagogiche	»	173
7.4. Conclusioni	»	174
Bibliografia	»	177

Sigle e abbreviazioni

AOI	Area of Interest
BNC	British National Corpus
CEA	Computer-aided Analysis
CIA	Contrastive Interlanguage Analysis
Cfr.	Confronta
DS	Deviazione Standard
E	Frequenza attesa
ER	Error Rates
Ecc.	Eccetera
FL	Foreign Language
FLA	Foreign Language Acquisition
Gradit	Grande Dizionario Italiano dell'Uso
ICE	International Corpus of English
ICLE	International Corpus of Learners of English
ISI	Interstimulus Interval
L1	Lingua prima
L2	Lingua seconda
LCR	Learner Corpus Research
LDT	Lexical Decision Task
LMI	Local Mutual Information
LOCCLI	Longitudinal Corpus of Chinese Learners of Italian
LOCNESS	Louvain Corpus of Native English Essays
LS	Lingua Straniera

MDS	Modelli Distribuzionali Semantici
MI	Mutual Information
NR	Non-word Ratio
O	Frequenza osservata
p.e.	Per esempio
PEC	Perugia corpus
PPMI	Positive Pointwise Mutual Information
QCER	Quadro Comune Europeo di Riferimento
RP	Relatedness Proportion
SCF	Subcategorization Frame
SD	Semantica Distribuzionale
SLA	Second Language Acquisition
SOA	Stimulus Onset Asynchrony
TR	Tempo di Reazione

Introduzione

You shall know a word by the company it keeps.

(J.R. Firth)

Mantenere una promessa, cronaca nera, accendere un mutuo, sono combinazioni lessicali conosciute come *collocazioni*, uno dei fenomeni più intriganti e, al tempo stesso, più complessi del lessico. Definite tradizionalmente solo come combinazioni frequenti di due o più parole, queste unità lessicali sono caratterizzate, in realtà, da proprietà distribuzionali e fraseologiche complesse. Inoltre, mostrano un alto grado di convenzionalità, che le rende cristallizzate nell'uso, difficilmente soggette a trasformazioni, tendenzialmente, infatti, non useremmo *conservare una promessa* o *cronaca oscura*.

Proprio per loro complessità, le collocazioni pongono una sfida alla linguistica teorica e a chi si occupa di lessico. Esistono, infatti, diversi approcci adottati per loro definizione e identificazione, tra i quali emergono l'approccio frequentista (Firth, 1957; Sinclair, 1991; Evert, 2008) e l'approccio fraseologico (Cowie, 1994, 1998; Howarth, 1998). Il primo definisce le collocazioni come combinazioni co-occorrenti di due o più parole, che tendono a ricorrere frequentemente insieme nella lingua parlata e scritta; il secondo, invece, distingue le collocazioni da altri tipi di combinazioni lessicali (ad esempio, le combinazioni libere e le espressioni idiomatiche), in termini di specifiche proprietà quali la composizionalità semantica e la fissità lessicale. Nei modelli fraseologici, infatti, le combinazioni lessicali sono di solito poste lungo un *continuum* di fissità lessicale e di composizionalità, che va dalle combinazioni libere alle espressioni idiomatiche, con le collocazioni in posizione intermedia, ed è proprio questa loro natura intermedia che rende la definizione di collocazione complessa.

Non bisogna poi tralasciare che le collocazioni hanno anche una realtà psicolinguistica (Hoey, 2005). Proprio in virtù del fatto che i due elementi di una collocazione tendono a ricorrere insieme frequentemente, tra di essi si crea un'associazione reciproca che fa sì che all'incontrare uno dei due

elementi il richiamo al secondo sia facilitato e veloce. Ciò suggerisce importanti implicazioni sulla struttura del lessico mentale dei parlanti.

Allontanandosi dalla linguistica generale e passando alla linguistica acquisizionale, è ormai provato che le collocazioni rappresentino una sfida anche per gli apprendenti di una lingua seconda (L2) o straniera (LS) (Henriksen, 2013). Le collocazioni, essendo molto frequenti, ricorrono in una buona parte dell'input linguistico a cui un apprendente è esposto. Ne deriva che questi item sono un tassello importante nell'uso di una L2 o di una LS: conoscere e riconoscere le collocazioni permette agli apprendenti di maneggiare una L2 in modo fluente e di elaborare e comprendere l'input linguistico facilmente e velocemente. Nonostante ciò, gli apprendenti tendono a produrre collocazioni devianti, non tipiche della L2/LS che stanno apprendendo, non percependo i legami associativi fra gli elementi di una collocazione. Da cosa dipende questa difficoltà? Tra i fattori che possono influenzare negativamente l'apprendimento delle collocazioni vi è una scarsa esposizione alla L2; al contrario, l'essere sufficientemente esposti alla L2 permetterebbe agli apprendenti di notare e immagazzinare l'informazione sulle co-occorrenze tra le parole nell'input linguistico e di usare espressioni tipiche e convenzionali della L2 (Hoey, 2005).

Come avviene, quindi, l'apprendimento delle collocazioni in L2? I meccanismi alla base dell'apprendimento delle collocazioni in L2 sono simili o differenti da quelli alla base dell'apprendimento delle collocazioni in L1? A queste e tante altre domande sta cercando di rispondere la ricerca sulle collocazioni in L2, che ha preso pieno ritmo negli ultimi anni e indaga questo fenomeno non solo a livello di produzione, ma anche a livello di elaborazione applicando gli strumenti propri della psicolinguistica. Dalla ricerca sulle collocazioni in L2 arrivano risultati interessanti e utili anche per lo sviluppo di nuovi metodi didattici che possono rafforzare l'apprendimento delle collocazioni in L2.

Tenendo conto della complessità di questo fenomeno sia dal punto di vista teorico sia dal punto di vista acquisizionale, questo volume cerca di individuare le collocazioni come un fenomeno non solo lessicale e testuale, ma anche cognitivo, ponendosi tre obiettivi principali.

Il primo scopo di questo libro, infatti, è quello di fornire una trattazione scientifica dei diversi approcci teorici che hanno cercato di definire che cosa sono le collocazioni, offrendo un'ampia rassegna di studi e ricerche sulle collocazioni in L1 e L2, utili per illustrare la complessità del fenomeno, evidenziando le metodologie adottate e i principali risultati ottenuti. L'obiettivo, alla luce di quanto rilevato da tale revisione della letteratura, è quello di proporre una definizione che identifichi una collocazione sia dal punto di vista frequentista e fraseologico, sia dal punto di vista cognitivo.

Il secondo obiettivo è quello di fornire nuovi strumenti di ricerca su questo fenomeno lessicale e cognitivo. Se le collocazioni sono così complesse, come possiamo fare ricerca sperimentale su queste combinazioni? Quali metodologie dobbiamo adottare? Ci si affaccerà, quindi, su due discipline in particolare: la linguistica dei corpora e la psicolinguistica, che adoperano metodi d'analisi diversi e si pongono obiettivi differenti. Infatti, da un lato la linguistica dei corpora ha come suo strumento d'analisi un corpus per identificare *pattern* d'uso del linguaggio; dall'altro, la psicolinguistica studia i processi cognitivi alla base dell'acquisizione, della comprensione e della produzione del linguaggio, presentandosi come una disciplina sperimentale. Nel presente volume si vedrà come integrando le tecniche della linguistica dei corpora a quelle della psicolinguistica sia possibile ottenere un quadro il più completo possibile del fenomeno delle collocazioni.

Il terzo obiettivo, infine, è quello di fornire materiali di studio e di ricerca in lingua italiana. Infatti, la maggior parte della bibliografia su questo tema è in lingua inglese e le ricerche condotte sull'apprendimento delle collocazioni in L2 analizza come lingua target soprattutto l'inglese, mentre altre lingue, tra cui l'italiano, sono ancora poco investigate. Il presente volume cerca di sopperire a questa mancanza, presentando, da un lato, una rassegna in italiano dei principali approcci teorici e degli studi in L2, e dall'altro mostrando come possa essere condotta una ricerca sperimentale sulle collocazioni in italiano L1 e L2.

Venendo ora alla struttura di questo libro, questa si suddivide in tre parti. Nella prima parte del volume, sarà esaminato, come già anticipato, lo stato dell'arte, approfondendo sia i principali approcci teorici sull'identificazione e la definizione delle collocazioni (Capitolo 1), sia i principali studi della ricerca sulle collocazioni in L1 e L2 con una descrizione dettagliata delle metodologie adottate e dei risultati (Capitolo 2).

La seconda parte del libro sarà, invece, dedicata agli strumenti di ricerca. Nel Capitolo 3 saranno illustrate le tecniche d'analisi e gli scopi della linguistica dei corpora e della psicolinguistica. Saranno trattati i metodi quantitativi e computazionali per identificare le collocazioni nei corpora attraverso strumenti d'analisi come *Sketch Engine* e *LexIt*. Successivamente, saranno approfondite le metodologie psicolinguistiche sia *offline*, come la raccolta di intuizioni dei parlanti, sia *online*, con particolare rilievo alla metodologia del *priming* ed *eye-tracking*. Inoltre, per ogni metodologia saranno presentati gli studi più rilevanti per mostrare concretamente come è stata applicata e quali risultati sono emersi.

A seguire, i tre capitoli successivi presenteranno tre studi condotti con lo scopo di indagare le collocazioni in L1 e in L2. Nello specifico, nel

Capitolo 4 si propone un'indagine sulle intuizioni dei parlanti L1 e L2 di italiano sul grado di fissità lessicale delle collocazioni estratte da un corpus di italiano di riferimento. Il Capitolo 5, invece, riporta uno studio *priming* condotto per osservare come le collocazioni derivate da un corpus di italiano siano rappresentate nel lessico mentale dei parlanti L1 e L2 e come l'esposizione alla lingua target rafforzi i legami associativi fra gli elementi delle collocazioni nella mente degli apprendenti. Nel Capitolo 6 sarà mostrato uno studio *eye-tracking* realizzato per indagare l'elaborazione in tempo reale delle collocazioni in parlanti L1 e L2 di italiano. Per ogni studio saranno descritti la raccolta del materiale, il design sperimentale, l'analisi dei dati e i risultati, che verranno discussi e interpretati in base alle domande di ricerca di ciascuno studio.

Infine, nell'ultima parte del volume saranno collegati fra loro i risultati dei tre studi empirici (Capitolo 7), andando più a fondo nell'analisi della natura delle collocazioni e presentando nuovi spunti per possibili ricerche future e suggerimenti pedagogici.

Prima di iniziare questo percorso alla scoperta delle collocazioni, vorrei spendere due parole sul perché ho mosso i primi passi della mia carriera da ricercatrice proprio su questo fenomeno del lessico. Ciò che mi ha sempre affascinato del lessico è il comportamento delle parole. Le parole in modo del tutto arbitrario si legano fra loro e si scelgono, instaurando associazioni che possono rafforzarsi o indebolirsi. Credo che il comportamento associativo delle parole sia una potente metafora dei rapporti umani; anche fra noi i legami si creano a volte in modo del tutto imprevedibile, senza un perché chiaro ed esplicito. Le persone rappresentano per noi le *nostre* collocazioni, con cui ricorriamo spesso o raramente, con cui stringiamo rapporti più o meno forti.

Voglio, quindi, ringraziare le persone che sono *associate* a questo libro, e molte di esse a me e alla mia vita, senza le quali questa monografia non esisterebbe. Questo volume è stato pubblicato con il contributo finanziario della Scuola Superiore di Dottorato e di Specializzazione (SSDS) dell'Università per Stranieri di Siena: ringrazio, pertanto, il Magnifico Rettore dell'Università, Tomaso Montanari, il Direttore della SSDS, Marina Benedetti, e la Coordinatrice del Dottorato, Giovanna Frosini.

Ringrazio la Prof.ssa Marina Benedetti e il Prof. Alessandro Lenci, per essere stati delle guide e dei punti di riferimento e per aver creduto nell'originalità di questo lavoro. Ringrazio la Dr. Anna Siyanova-Chanturia, per tutto ciò che mi ha insegnato e che mi insegna tuttora. Ringrazio i revisori che hanno contribuito a migliorare questo volume. Grazie a

Tommaso Gorni per aver creduto nell'argomento di questo libro e per la professionalità.

Ringrazio la Prof.ssa Stefania Spina e la Prof.ssa Francesca Malagnini, per avermi sostenuta nella ricerca e in molto altro, e i miei colleghi dell'Università per Stranieri di Perugia, Luciana Forti e Fabio Zanda, per i confronti e gli scambi proficui che hanno giovato anche a questo libro.

Un grazie di cuore va a tutti i colleghi e amici della 307, per aver condiviso insieme i tre anni del dottorato all'insegna dell'amicizia. In particolare, grazie a Maria Vittoria, a Martina, e ad Azzurra, per tutto il sostegno che mi hanno dato, in questo ultimo anno più che mai, e per aver essere state felici insieme a me per il raggiungimento di questo obiettivo. Grazie a Cèlia, a Simona, e a Cecilia, per aver brindato a questo libro.

Grazie alle mie amiche di sempre, Elisa, Erica e Elena, per essere state sempre presenti ad ogni mio traguardo, per il sostegno, per essere un porto sicuro.

Grazie alla Nuova Zelanda, per tutta l'energia che quella terra mi ha dato.

Un grazie pieno di amore a Giulio, per l'aiuto con il titolo, e soprattutto per aver scelto di restare.

Infine, un grazie speciale va alle mie collocazioni più forti: a tutta la mia famiglia, a mia sorella Elena, per credere sempre in me, a mia mamma Milvia, per avermi insegnato la determinazione, e, soprattutto, al mio papà Roberto, per tutta la resistenza che mi ha sempre trasmesso. A lui dedico questo libro, so che ne sarebbe fiero.

Parte prima

Lo stato dell'arte

1. La natura intrigante delle collocazioni

1.1. Che cosa sono le collocazioni?

Chiunque abbia a che fare con le collocazioni, non può negare quanto queste rappresentino un fenomeno affascinante e intrigante del linguaggio. La loro pervasività risiede nel fatto che tendono a ricorrere sia nella lingua scritta sia nella lingua parlata più frequentemente di quanto ci si aspetti. Infatti, l'analisi di corpora di vaste dimensioni ormai ci dimostra come certi *pattern* lessicali si ripetano frequentemente nel linguaggio.

Una delle caratteristiche più affascinanti delle collocazioni è l'arbitrarietà del legame fra i loro elementi, caratterizzato da un alto grado di convenzionalità e familiarità e non riconducibile a ragioni grammaticali. Notiamo immediatamente che espressioni come *perdere la concentrazione*, *strisce pedonali*, *sfogliare un giornale*, non sono combinate in modo creativo, ma ci risuonano come fisse, convenzionali, e familiari. Comunemente diciamo *Ti auguro la buonanotte* e non *Ti auspico la buonanotte*, o sappiamo che un *parente* può essere *stretto* o *alla lontana*, mentre un *amico* può essere *stretto* ma non *alla lontana*. Saremmo, però, in grado di spiegare perché un *amico* può essere *stretto* ma non *alla lontana*?

L'arbitrarietà e la convenzionalità delle collocazioni e, più in generale, delle possibilità combinatorie delle parole, rendono altamente difficile la loro definizione e classificazione. Un'ulteriore difficoltà è rappresentata dal fatto che le collocazioni sembrano essere combinazioni né totalmente fisse né totalmente libere, né completamente trasparenti né completamente opache dal punto di vista semantico. Spesso sono, infatti, definite come combinazioni intermedie fra combinazioni libere ed espressioni idiomatiche.

Vista la difficoltà che le collocazioni pongono alla linguistica teorica, sono stati proposti diversi approcci per descrivere e definirne le caratteristiche principali. Nel presente capitolo illustreremo i principali: l'approccio

frequentista; l'approccio fraseologico; l'approccio formulare; e l'approccio cognitivo. Infine, vedremo come ognuno di essi contribuisce a studiare l'intrigante natura delle collocazioni.

1.2. L'approccio frequentista

Nell'approccio frequentista le collocazioni sono concepite come combinazioni di due o più parole, che tendono a ricorrere insieme frequentemente. Come sottolineato da Bartsch (2004), non si sa con certezza chi sia stato il primo linguista a utilizzare il termine *collocazione* in questo senso, ma gli studiosi sono concordi nel designare J.R. Firth (1957) come colui che ha introdotto il termine *collocazione* nella linguistica teorica. Riprendendo l'affermazione di Wittgenstein «the meaning of words lies in their use» (1953: 80), Firth la sviluppa introducendo il concetto di *meaning by collocation* e la nozione di *context of situation*. *Meaning* e *context* sono, infatti, alla base della concezione di linguaggio di Firth, il quale ritiene che l'obiettivo principale dell'analisi linguistica sia l'analisi del significato degli enunciati linguistici. Firth interpreta il linguaggio come un insieme di eventi enunciati dai parlanti in contesti reali. Il linguista britannico afferma, inoltre, che il significato degli enunciati deriva proprio dalla particolare situazione in cui sono pronunciati, o meglio, da quello che chiama appunto *context of situation*.

Riassumendo, Firth interpreta il significato come una funzione che un item linguistico assume in uno specifico contesto. È qui che, a livello lessicale, entra in gioco la nozione di *collocazione*, una co-occorrenza di due o più parole, co-occorrenza che contribuisce al significato stesso delle parole in una collocazione. Come afferma Firth nel suo più celebre enunciato «You shall know a word by the company it keeps!» (1957: 11), è proprio dalle parole che co-occorrono con una certa parola in un dato contesto che possiamo dedurre e analizzare il significato della parola in questione.

Come spiega il linguista, infatti, in un testo possono ricorrere frasi del tipo «Don't be such an ass», «You silly ass!», e «What an ass he is!». Questi esempi, «Don't be such an...», «You silly...!», e «What an... he is!», rappresentano la compagnia familiare di «ass», ovvero le parole con cui «ass» tende a co-occorrere più frequentemente. Secondo Firth, il significato di una parola emerge proprio dalle sue abituali occorrenze con altre parole. Bisogna però non confondere *collocazione* con *contesto*. Il contesto fa riferimento, come abbiamo visto, alla specifica situazione in cui l'enunciato è pronunciato e usato dal parlante, comprendendo anche fattori extralinguistici; invece, la collocazione riguarda la compagnia di una parola, ovvero

le parole che con questa sono più comunemente e frequentemente usate. Alcune parole tendono a ricorrere comunemente con altre perché tra di esse vi è quella che Firth definisce *mutual expectancy*, ovvero un'attrazione reciproca: le parole, infatti, si attraggono e si richiamano a vicenda.

Il lavoro di Firth è di notevole importanza, in quanto introduce per la prima volta il concetto di *collocazione* nella teoria linguistica e sposta l'analisi sul contesto in cui le parole sono usate e sul fatto che le co-occorrenze fra le parole non siano casuali, ma siano piuttosto dettate da una sorta di attrazione reciproca. Tuttavia, Firth non dà una definizione univoca e specifica di collocazione. Ad esempio, non chiarisce il numero di parole che entrano a far parte di una collocazione, ma parla genericamente di uno spazio compreso fra le due e le undici parole ortografiche. Inoltre, sembra affermare che la parola analizzata non appartenga all'entità collocazione, la quale farebbe riferimento, quindi, solo alle parole che co-occorrono con questa.

1.2.1. *I neo-Firthiani*

La nozione di *collocazione* di Firth è sviluppata in modo più preciso da Halliday e Sinclair, ai quali si è soliti riferirsi come «neo-Firthiani». Halliday (1961) definisce una collocazione come un'associazione sintagmatica fra item lessicali, che emerge a livello testuale, dipendente dalla probabilità che a una certa distanza (in termini di item lessicali) da una parola *x* ricorrano gli item lessicali *a*, *b*, *c*, ecc. (1961: 276). La definizione di Halliday appare più precisa della nozione di collocazione di Firth, in quanto descrive la collocazione sia come una relazione sintagmatica fra item lessicali sia come un fenomeno testuale. Inoltre, Halliday identifica come item lessicale il lessema e tutte le sue derivazioni, interpretandolo, al contrario di Firth, come parte integrante della collocazione stessa.

Halliday (1966), come si è appena detto, pone i termini di una collocazione su un piano sintagmatico. Vediamo l'esempio da lui proposto: se prendiamo in considerazione le parole «argument» e «tea», è probabile che entrambe ricorrano con la parola «strong»; più precisamente, «argument» e «tea» rappresentano il *nodo* della collocazione, mentre «strong» il suo *collocato*. Al contrario, è meno probabile che «strong» sia un collocato di «car» e «powerful» sia un collocato di «tea», mentre è più probabile trovare le collocazioni «powerful car» e «strong tea». L'esempio delineato da Halliday (1966) mostra che le relazioni paradigmatiche fra due parole non sono costanti, ovvero non è detto che entrambe possano essere collocati di uno stesso nodo, ma dipendono dalle relazioni sintagmatiche che entrambe hanno con determinati argomenti (p.e., «tea», «car», «argument», ecc.).

Halliday deduce, quindi, che le collocazioni si instaurano a livello sintagmatico, precisando che «lexis seems to require the recognition merely of linear co-occurrence together with some measure of significant proximity» (1966: 152). Halliday, come Firth, non definisce chiaramente questa prossimità, ma introduce la nozione di collocazione come concetto statistico parlando di *probability of occurrence*, affermando che l'occorrenza di un item in una collocazione¹ può essere analizzata o discussa solo in termini di probabilità (1966: 159). In un lavoro precedente lo studioso aveva, inoltre, sostenuto che «any given item thus enters into a range of collocation, the items with which it is collocated being ranged from more to less probable» (1961: 276). Questa probabilità di occorrenza viene definita in termini di frequenza, ovvero deve essere tenuta in considerazione la frequenza di un item in una determinata collocazione relativamente alla sua totale frequenza di occorrenza. Questo permette di determinare quanto è probabile che un item ricorra in una determinata collocazione, come se fosse l'item stesso a predire la sua più probabile collocazione (1961: 156). Ciò suggerisce che l'identificazione di una collocazione deve basarsi su una misura quantitativa che aiuti a stabilire se il fatto che il nodo e il collocato occorrano insieme sia o meno dovuto al caso.

Seguendo i lavori di Halliday (1961, 1966) e Firth (1957), Sinclair (1966, 1987, 1991) definisce *collocazione* l'occorrenza di due o più parole, fra cui intercorre uno *span* ristretto, in un testo² (1991: 170). Anche Sinclair, quindi, sposa la definizione di Halliday, identificando le collocazioni come fenomeno testuale, apportando però un ulteriore contributo alla nostra comprensione delle collocazioni, in quanto precisa e quantifica la prossimità fra il nodo e il suo collocato. Introduce, infatti, il concetto di *span* come lo spazio che intercorre fra il nodo e il collocato di una collocazione, proponendo una dimensione dello *span* di quattro parole a destra e a sinistra del nodo in cui è probabile trovare i collocati più probabili del nodo stesso (Jones & Sinclair, 1974).

Inoltre, Sinclair (1991), seguendo la posizione di Halliday (1966), interpreta la ricorrenza delle collocazioni in un testo in termini statistici. Distingue, infatti, fra collocazioni *significantive* e *casuali*. Le collocazioni *casuali* ricorrerebbero in un testo per caso, mentre i collocati che tendono a ricorrere con un certo nodo più probabilmente di quanto ci si aspetterebbe rappresentano i più comuni collocati del nodo, formando quelle che Sinclair definisce, appunto, collocazioni *significantive*.

Fortemente collegato al concetto di collocazione è il modello proposto da Sinclair dell'*open choice* e dell'*idiom principle* (Sinclair, 1991). Il primo

1. Halliday utilizza il termine *environment*.

2. Sinclair sostituisce il termine *environment* usato da Halliday con *text*.

principio afferma che il testo è il risultato di un vasto numero di scelte complesse ed è conosciuto anche come modello *slot-and-filler*; il testo consisterebbe di una serie di slot che possono essere riempiti con una qualsiasi parola del lessico, a patto che si conservi il criterio di grammaticalità. Invece, il secondo principio afferma che «a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments» (1991: 110). In sostanza, il linguaggio umano è potenzialmente originale e creativo. Tuttavia, nonostante noi parlanti possiamo esercitare la sua potenzialità creativa, non lo facciamo per forza. Al contrario, tendiamo a usare espressioni prefabbricate acquisite nel nostro lessico mentale anziché espressioni create *ex novo*. Nell'ottica di Sinclair, le collocazioni rientrano fra queste espressioni prefabbricate e sono un'istanza dell'*idiom principle*.

Riassumendo, i lavori di Halliday (1961, 1966) e Sinclair (1991), basati sugli studi di Firth (1957), hanno contribuito a sviluppare una nozione di collocazione come fenomeno testuale e probabilistico basata sulla frequenza. Da ciò è derivata la necessità di quantificare la probabilità di co-occorrenza di due parole per individuare vere e proprie collocazioni. Questo è stato reso possibile grazie allo sviluppo della linguistica dei corpora, che permette l'uso di strumenti quantitativi e qualitativi per esplorare aspetti linguistici che emergono da testi, per poi descrivere la struttura del linguaggio (Lenci et al., 2012).

1.2.2. *Le misure di associazione*

Si inserisce in questo contesto il lavoro di Evert (2008), che ha proposto un chiaro metodo per identificare e definire le collocazioni. Prima di tutto Evert (2008) distingue fra un concetto empirico e un concetto teorico di collocazione. Il primo definisce le collocazioni come combinazioni di parole ricorrenti e predicibili, mentre il secondo le identifica come espressioni multiparola lessicalizzate e idiosincratice (Sag et al., 2002). Evert (2008) sceglie di adottare il concetto empirico e definisce 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. Identifica tre diversi tipi di co-occorrenze:

1. *surface co-occurrence*: parole che ricorrono a una certa distanza, definita come *collocational span* (in genere da tre a cinque parole a destra e/o a sinistra del nodo³);

3. Evert propone una dimensione dello span diversa da quella di Sinclair (4 parole a destra e a sinistra del nodo) che meglio si adatta a lingue diverse dall'inglese.

2. *textual co-occurrence*: parole che ricorrono insieme in una stessa unità lessicale, come una frase, un paragrafo o un documento;
3. *syntactic co-occurrence*: le parole co-occorrono se sono legate da una certa relazione sintattica (ad esempio, il nome e il suo aggettivo modificatore, o il verbo e il suo oggetto diretto).

Basandosi sul concetto di *mutual expectancy* proposto da Firth (1957), Evert (2008) sostiene che le parole tendono a co-occorrere perché sono attratte l'una dall'altra, ovvero legate da un'associazione reciproca. Questa attrazione, secondo Evert, non può essere interpretata solo in termini di frequenza di ricorrenza, ma deve essere definita in termini statistici, usando quelle che sono chiamate misure di associazione. Le misure di associazione statistica utilizzano due diversi approcci per quantificare il legame associativo fra gli elementi di una collocazione. Il primo è l'approccio *threshold*, che mira ad individuare «vere collocazioni» impostando una soglia minima. Se una coppia di parole supera questa soglia, allora sarà definita come vera collocazione dal punto di vista statistico. Il secondo approccio, invece, è definito approccio *ranking*: in esso le collocazioni sono poste lungo una scala di *collocational strength*⁴ e a ognuna è assegnato un punteggio di associazione. A differenza dell'approccio *threshold*, le collocazioni nell'approccio *ranking* non sono distinte dicotomicamente in vere collocazioni e non-collocazioni, ma sono poste lungo un *continuum* di forza associativa, che va da coppie di parole meno associate a coppie di parole più associate.

Le misure di associazione, quindi, quantificano l'attrazione fra gli elementi di una collocazione e offrono un indice numerico di quanto due parole siano reciprocamente associate. Le misure di associazione confrontano la frequenza osservata (*O*) della coppia di parole, che è la frequenza di co-occorrenza in un testo, con la frequenza attesa (*E*) della coppia, che è la frequenza che dovremmo aspettarci se le due parole occorressero insieme per caso. Il rapporto fra frequenza osservata e frequenza attesa è indice del legame associativo fra le due parole; se la frequenza osservata è più grande della frequenza attesa, allora una coppia di parole può essere identificata come collocazione. Più l'indice ottenuto è grande, più il legame associativo fra le due parole è forte.

Come riporta Evert (2008), una delle misure di associazione maggiormente usate è la Mutual Information (MI; Church & Harris, 1989), la cui formula è la seguente:

4. Il concetto di *collocational strength* può essere interpretato come la forza associativa (in termini statistici) del legame fra il nodo e il collocato di una collocazione.

$$MI(u, v) = \log_2 \frac{p(u, v)}{p(u)p(v)}$$

La MI confronta la probabilità di osservare la coppia di parole $\langle u, v \rangle$ con la probabilità di osservare u indipendentemente da v . Se le due parole sono statisticamente indipendenti allora la probabilità di osservare le due parole $\langle u, v \rangle$ insieme sarà uguale alla probabilità di osservare u indipendentemente da v . I due termini sono, infatti, posti rispettivamente a numeratore e a denominatore; maggiore è il valore della frazione, più alto è il grado di dipendenza fra u e v , quindi più forte è la loro associazione lessicale.

Un'altra misura di associazione usata frequentemente per individuare vere collocazioni nel testo è il t -score, calcolata nel modo seguente:

$$t - score = \frac{O - E}{\sqrt{O}}$$

Sia la MI sia il t -score sono misure di associazione che fanno parte dell'approccio *threshold*; in letteratura si è soliti porre la soglia della MI a 3 e la soglia del t -score a 2. Ci sono però delle differenze fra queste due misure di associazione. Il t -score non è calcolato su una scala standardizzata e non può, quindi, essere usato per confrontare collocazioni in diversi corpora. Inoltre, tende a sovrastimare le alte frequenze; questo significa che combinazioni di parole che ricorrono in un corpus o in un testo con alta frequenza otterranno un alto valore di t -score. Frequenza e grado di associazione non sono, però, concetti interscambiabili; infatti, mentre le collocazioni con alto t -score sono frequenti, non tutte le combinazioni frequenti hanno un alto valore di t -score (Gablasova et al., 2017).

Al contrario, il valore di MI è normalizzato e questo rende le collocazioni comparabili in termini di associazione attraverso corpora diversi (anche se non è calcolato su una scala con un valore minimo e massimo). Il limite della MI è che tende a dare maggior peso alle basse frequenze. La MI è, quindi, negativamente correlata con la frequenza.

Una misura di associazione che supera i limiti delle due precedenti è il LogDice, una misura standardizzata che opera su una scala con il valore massimo fisso di 14. Questo rende il LogDice direttamente comparabile tra corpora diversi. Inoltre, permette di osservare più chiaramente quanto una combinazione sia lontana dal valore e , quindi, quanto le parole di una collocazione siano effettivamente associate fra di loro (Gablasova et al., 2017).

Lo sviluppo della linguistica dei corpora ha permesso di identificare «vere collocazioni» in un testo, attraverso l'analisi di frequenza e quanti-

ficando il legame associativo fra il nodo e il collocato della collocazione. Inserendosi in questa prospettiva, Evert (2008) definisce in modo più concreto le collocazioni, fornendo un metodo per individuarle all'interno del testo non basato unicamente sulla frequenza di co-occorrenza, ma anche sul legame associativo fra gli elementi di una collocazione.

Per riassumere, l'approccio frequentista, iniziato dai lavori di Firth (1957) e portato avanti e perfezionato dai lavori di Halliday (1961, 1966), Sinclair (1991) ed Evert (2008), definisce le collocazioni come co-occorrenze frequenti di due o più parole, entro una certa distanza in un testo, che ricorrono insieme più di quanto ci si aspetterebbe, caratterizzate da un'elevata associazione reciproca (quantificabile con le misure di associazione). L'adozione dell'approccio frequentista nella definizione e nella identificazione delle collocazioni non può essere separata, quindi, dall'analisi di corpora realizzata applicando strumenti di indagine quantitativi e qualitativi.

1.3. L'approccio fraseologico

La fraseologia nasce in Russia negli anni '40 del Novecento e possiamo definirla, molto sinteticamente, come lo studio delle combinazioni lessicali di una lingua (Vinogradov, 1947; Howarth, 1998). Nell'ottica fraseologica, le combinazioni lessicali sono distinte e classificate in base a due criteri principali: il criterio della fissità lessicale e il criterio della composizionalità. Il primo criterio si riferisce alla possibilità di sostituire un elemento della combinazione lessicale con un sinonimo o un termine affine; il secondo, invece, ha a che fare con quanto i significati dei singoli elementi di una combinazione contribuiscano al significato dell'intera combinazione lessicale.

Prendiamo ad esempio la combinazione *cucire una maglia*. In questo caso sappiamo dedurre il significato dell'intera combinazione dai significati dei suoi componenti, e possiamo sostituire *maglia* con *maglione* o *maglietta*, ottenendo una combinazione sensata senza cambiarne il significato. Invece, se prendiamo in considerazione *guerra fredda*, sappiamo che non stiamo parlando di una guerra scoppiata durante l'inverno, ma che *fredda* è usato in senso figurato. Inoltre, percepiamo che *fredda* non può essere sostituito da un sinonimo come *gelida*, perché ciò comporterebbe la perdita del significato della combinazione stessa.

Ciò suggerisce che le combinazioni lessicali variano nel grado di composizionalità e fissità lessicale. La fraseologia, infatti, distingue le combinazioni lessicali in tre tipologie principali: le combinazioni libere, le collocazioni e le espressioni idiomatiche. Nei modelli fraseologici le com-

binazioni lessicali sono disposte lungo un *continuum* di fissità lessicale e di composizionalità, che va dalle combinazioni libere alle espressioni idiomatiche, con le collocazioni a metà del *continuum*. Le combinazioni libere sono pienamente composizionali e i loro elementi possono essere liberamente sostituiti con sinonimi; al contrario, le espressioni idiomatiche sono le combinazioni meno composizionali e più fisse dal punto di vista lessicale. Le collocazioni, infine, si comportano in modo diverso sia dalle combinazioni lessicali sia dalle espressioni idiomatiche. Da una parte, sono caratterizzate da fissità lessicale, in quanto non permettono in tutti i casi una libera sostituzione dei propri elementi. Dall'altra parte, variano nel grado di composizionalità; molto spesso, infatti, uno dei due elementi della collocazione è usato in senso figurato.

1.3.1. *Il continuum fraseologico di Cowie*

Questa idea innovativa di non classificare le combinazioni in categorie distinte, ma di porle lungo un *continuum*, è stata sviluppata da A.P. Cowie (1981, 1988, 1994). Cowie (1988, 1994) fa una prima classificazione delle combinazioni lessicali in *formulae* e *composites*. Le *formulae* sono definite come quelle combinazioni che hanno una funzione pragmatica (p.e., *Grazie!*, *Come stai?*, *Buongiorno*, ecc.). Invece, i *composites* sono combinazioni i cui elementi sono legati da una relazione semantica. Cowie (1988, 1994) distingue i *composites* applicando due criteri principali: il criterio della trasparenza e il criterio della commutabilità. Individua in questo modo quattro diversi tipi di combinazioni lessicali, che pone lungo un *continuum* (Cowie, 1981):

1. *Free combinations*: gli elementi delle combinazioni libere sono liberamente sostituibili e il loro significato è composizionale (*to drink water*, 'bere l'acqua');
2. *Collocations*: sono definite come unità composite che permettono la sostituzione di uno solo dei propri elementi, mentre l'altro elemento e il suo significato rimangono costanti (*to explode an idea*, 'abbattere un'idea');
3. *Figurative idioms*: la sostituzione degli elementi è raramente permessa e queste combinazioni presentano un'interpretazione figurata dominante, anche se conservano un'interpretazione letterale (*to cloze ranks*, 'serbare le fila');
4. *Pure idioms*: gli elementi non possono essere sostituiti e la combinazione ha solo un'interpretazione figurata (*to spill the beans*, 'vuotare il sacco').

Approfondendo la definizione che dà Cowie di collocazione, sono distinti due principali tipi di collocazioni (1981): le *collocazioni aperte* e le *collocazioni ristrette* o *semifisse*. Cowie fa questa distinzione per distinguere le collocazioni dai *pure idioms*, che presentano una minor flessibilità rispetto alle collocazioni per quanto riguarda la sostituzione lessicale. Infatti, le collocazioni si dimostrano più flessibili delle espressioni idiomatiche e permettono modificazioni senza che venga alterato il significato dell'intera combinazione o dei suoi elementi.

Andando nello specifico, alcuni tipi di collocazioni, ovvero le collocazioni aperte, ammettono una sostituzione più libera. Cowie fa l'esempio di *run a business* ('gestire un affare'), dove il verbo *run* può essere combinato con altri sostantivi come *theatre* ('teatro') o *company* ('azienda'), mantenendo lo stesso significato di 'gestire'. Al contrario, nelle collocazioni ristrette o semifisse il collocato ha quasi perso il suo significato letterale o familiare, assumendone uno tecnico o figurato all'interno della collocazione. L'esempio che Cowie propone è quello di *explode an idea*, 'abbattere un'idea', dove il collocato *explode* è usato nel senso figurato di 'abbattere', e che difficilmente può essere combinato con altri nomi mantenendo lo stesso significato figurato (**explode a bomb*, 'abbattere una bomba'). Il grado di fissità delle collocazioni semifisse appare, dunque, più ristretto di quello che caratterizza le collocazioni aperte, in quanto l'insieme dei collocati con cui può ricorrere il nodo è più limitato nel caso delle collocazioni ristrette rispetto al caso delle collocazioni aperte.

1.3.2. *Il modello fraseologico di Howarth*

Il lavoro di Cowie (1981, 1988, 1994) ha ispirato il modello fraseologico di Howarth (1998), il quale classifica le combinazioni di parole in *lexical* e *grammatical composites* in base alla classe di parola dei costituenti, riprendendo il lavoro di Benson (1985). Benson (1985), infatti, definisce le collocazioni come gruppi di parole che occorrono ripetitivamente e si distinguono in *collocazioni grammaticali* e *lessicali*. Le collocazioni grammaticali sono costituite da una parola dominante (p.e., un verbo, un nome o un aggettivo), seguita da una preposizione (*adjacent to*, 'vicino a'). Invece, le collocazioni lessicali sono caratterizzate da due componenti lessicali (p.e., Aggettivo+Nome, Nome+Verbo, Verbo+Nome, ecc.).

Allo stesso modo, Howarth (1998) definisce i *lexical composites* come combinazioni costituite da due parole contenuto, ad esempio le combinazioni Verbo+Nome (*to make a claim*, 'presentare un reclamo') e le combinazioni Aggettivo+Nome (*ulterior motive*, 'ulteriore motivo'), e i *gram-*