

La competenza fraseologica in italiano lingua straniera tra i livelli elementare e intermedio

Un approccio per bigrammi

Simone Torsani

Università di Genova

Phraseology constitutes an indisputably important area of language proficiency. Among the methods devised to explore this domain, semiautomated analysis of learner corpora has yielded important findings, especially in the case of intermediate and advanced levels. This paper investigates phraseological competence in Italian as a Foreign Language between the A2 and B1 levels of the CEFR. We collected 160 texts based on different combinations of task difficulty and learner level and we compared, through two-way Anova tests, the different groups. Results show that A2 learners produce more non-collocational bigrams with respect to other groups, but when they attempt a B1 task they experience difficulty in producing high frequency bigrams. Implications for instruction, assessment and further research are discussed.

Keywords: learner corpora; bigram; phraseology; t-score; mutual information

1. Introduzione

La ricerca illustrata nel presente contributo indaga, tramite un approccio pseudo-longitudinale, la competenza fraseologica in apprendenti di italiano lingua straniera tra i livelli A2 (elementare) e B1 (intermedio) del Quadro Comune Europeo di Riferimento (QCER). Tale competenza è analizzata tramite misure di associazione applicate ai bigrammi di un testo. La ricerca confronta i valori delle misure in diversi gruppi di apprendenti per verificare alcune ipotesi formulate nella ricerca del settore, in particolare se gli apprendenti di livello elementare facciano un uso maggiore di bigrammi ad alta frequenza rispetto agli intermedi. Inoltre, poiché i gruppi sono rappresentativi di apprendenti di livello diverso che svolgono compiti di livello diverso, la ricerca indagherà come queste due variabili, livello del compito e livello valutato, influiscono sulla produzione di bigrammi.

1.1 Competenza fraseologica, bigrammi e misure di associazione

La ricerca e la pratica glottodidattica hanno ormai da tempo rivalutato il ruolo del lessico e della fraseologia nella competenza in lingua straniera e non sorprende che, oltre a lavori di taglio teorico (Pawley e Syder 1983; Sinclair 199; Howarth 1998) e metodologico/didattico (Lewis 1993; Wray 2000; in Italia, Porcelli 2004), siano comparsi negli anni studi volti a indagare la competenza fraseologica negli apprendenti (Ellis et al. 2008; Ohlrogge 2009; Spina 2018). Il motivo di tale interesse non è solo scientifico, ma anche pratico, dal momento che molti lavori hanno individuato le difficoltà di apprendenti, anche avanzati, in questo ambito (Laufer e Waldman 2011; Paquot 2019).

Tra i vari approcci proposti, un numero consistente di lavori ha adottato come oggetto di studio corpora di apprendenti e come unità di analisi gli *n-grammi*, sequenze di due o più parole contigue, cioè separate tra loro solo da spazio. Durrant e Schmitt (2009), in particolare, adottano per la loro ricerca due misure di associazione già usate nella lessicografia per quantificare, e quindi confrontare, il grado di collocatività e naturalezza di alcuni tipi di bigrammi, tra i quali le combinazioni aggettivo+sostantivo. Le due misure adoperate dagli autori sono il *t-score* e la *Mutual Information*, MI. I due autori, come anche lavori successivi (es. Granger e Bestgen 2014) e il presente, considerano il *t-score* come una misura che valorizza la frequenza di un bigramma, in particolare quanto il bigramma è più frequente rispetto alle attese. La MI, invece, è interpretata come una misura della forza, o meglio dell'esclusività tra i suoi componenti (cioè, quanto le due parole tendono a trovarsi insieme). Rimandando per approfondimenti all'ampia letteratura sulla materia, in particolare a Evert (2008), Gablasova et al. (2017) e Hunston (2002), l'Appendice 1 riporta alcuni dettagli sui limiti e sull'interpretazione delle due misure all'interno della presente ricerca.

L'approccio per bigrammi è stato adottato da ricerche successive, (Granger e Bestgen 2014; Bestgen e Granger 2014, 2018; Qian 2019). Bestgen e Granger (2014) si concentrano su apprendenti medio/avanzati e giungono, come già Durrant e Schmitt, alla conclusione che, rispetto ai parlanti nativi, non solo gli apprendenti di livello intermedio, ma anche gli avanzati usano meno combinazioni rare e forti (quindi con *t-score* basso e MI alta), come *immortal soul*, mentre tendono a sovra utilizzare combinazioni frequenti, come *hard work*. I due autori definiscono una tecnica, detta CollGram, nella quale ogni bigramma di un testo è associato alle due misure; attraverso tale tecnica gli autori intendono tracciare, tramite approcci longitudinali e pseudo-longitudinali, lo sviluppo della competenza fraseologica in apprendenti di lingua inglese. Il fondamento teorico di questa tecnica è quello di considerare grammatica e lessico come un *continuum* (v.

tra gli altri Granger e Bestgen 2014; Römer 2009; Ruegg et al. 2011). Poco, tuttavia, è detto circa i livelli inferiori e su altre lingue, come l'italiano (sul quale hanno lavorato, da prospettive diverse, Siyanova-Chanturia 2015; Siyanova-Chanturia e Spina 2015; Spina 2018). La presente ricerca intende applicare la tecnica CollGram a testi estratti da un corpus di apprendenti di lingua italiana tra i livelli elementare e intermedio, quindi A2 e B1.

1.2 Misure di associazione e livelli medio-bassi

Molti lavori sui bigrammi confrontano le due dimensioni di frequenza ed esclusività secondo il principio per cui, con l'evolversi della competenza linguistica (e fraseologica), queste dovrebbero comportarsi in maniera diversa. Nel caso di apprendenti avanzati, infatti, ci si attenderebbe di trovare, rispetto agli intermedi, un numero maggiore di bigrammi meno frequenti e più esclusivi (Bestgen e Granger 2014; Durrant e Schmitt 2009).

Nei livelli medio-bassi del QCER, tuttavia, è la dimensione della frequenza ad avere un ruolo primario, almeno secondo i descrittori del Quadro. A tali livelli, infatti, si lavora intorno a strutture e lessico ad alta frequenza; frequenza che diminuisce ai livelli superiori (v. Spinelli e Parizzi 2010, in particolare il cap. 5), quando i compiti diventano meno immediati e concreti. Dopo un'indagine su soggetti intermedi e avanzati, nella quale hanno constatato che gli apprendenti intermedi sovra utilizzano bigrammi ad alta frequenza, più salienti nell'input e facili da apprendere (Ellis et al. 2008), Granger e Bestgen (2014) si domandano se gli apprendenti di livello inferiore non facciano quindi un ricorso ancora maggiore a bigrammi con alta frequenza (quindi con *t-score* alto) rispetto agli intermedi. Questa ipotesi costituisce uno dei due argomenti della presente ricerca. Ai livelli medio-bassi, la dimensione dell'esclusività è, pur presente, secondaria rispetto alla frequenza e per tale motivo ipotizziamo che la MI abbia una minore capacità di discriminare tra i livelli.

Sulla base di tali argomenti, quindi, ci attendiamo che lo sviluppo nella competenza fraseologica tra i livelli A2 e B1 si manifesti soprattutto nella dimensione della frequenza.

1.3 Misure di associazione e difficoltà del compito

In generale, si possono individuare due obiettivi nelle ricerche su *n-grammi* e misure di associazione: il primo è quello di indagare l'interlingua degli apprendenti; il secondo è individuare metodi efficaci per discriminare tra livelli di competenza linguistica (Crossley et al. 2011; Kyle e Crossley 2015). Quest'ultimo obiettivo, che per ragioni di spazio non sarà considerato nella presente ricerca, spiega,

almeno in parte, il fatto che molti lavori prendano in esame testi di diverso livello, ma realizzati a partire da compiti di uno stesso livello.

Bestgen e Granger (2014) e Garner, Crossley e Kyle (2020) lavorano su testi prodotti a partire da compiti di difficoltà analoga che sono stati in seguito valutati da esperti. Un compito, tuttavia, è predisposto per elicitarne funzioni e forme di un dato livello. Per questo motivo pare lecito attendersi che testi valutati come di livello superiore o inferiore a quello del compito presentino proporzioni di bigrammi più o meno frequenti o esclusivi - ciò che qui definiremo profilo di bigrammi - diverse rispetto a elaborati che invece corrispondono al livello del compito.

Per esempio, un testo valutato come A2 (che definiremo qui “livello valutato” del soggetto), ma realizzato a partire da un compito B1 (che definiremo qui “livello del compito”), potrebbe contenere errori dovuti all’incapacità di produrre le forme necessarie allo svolgimento della consegna. Questi errori si potrebbero tradurre in bigrammi con scarsa attrazione (quando non del tutto assenti dal corpus di riferimento), che di conseguenza sarebbero più numerosi rispetto a un testo con meno errori.

Per le ragioni esposte, quindi, ci attendiamo che la relazione tra livello del compito e livello valutato possa influenzare il profilo di bigrammi di un testo. Nella presente ricerca definiamo come normoperformanti i soggetti che producono testi di un livello valutato *analogo* a quello del compito assegnato, sovraperformanti soggetti che producono un testo di livello *superiore* a quello della consegna e sottoperformanti quei soggetti che producono un testo di livello *inferiore* a quello del compito.

1.4 Domande di ricerca

Sulla base degli argomenti esposti la ricerca intende indagare il passaggio in apprendenti di italiano dal livello A2 al B1 del QCER confrontando tra loro diversi gruppi e rispondendo alle seguenti domande:

1. I soggetti sopra e sottoperformanti presentano un profilo di bigrammi analogo a quello del corrispondente livello del compito, a quello del livello valutato, oppure presentano invece un profilo originale?
2. Gli apprendenti a livello elementare (A2) producono più bigrammi ad alta frequenza (caratterizzati da *t-score* medio/alto) rispetto agli intermedi (B1)?

2. Materiali e metodo

2.1 Corpus di apprendenti, di riferimento e banca dati

Come corpus di apprendenti è stato scelto MERLIN (Boyd et al. 2014), un corpus di apprendenti di tre lingue: ceco, italiano e tedesco. Il corpus, come si può leggere nella documentazione disponibile sul sito del progetto¹ e alla quale si rimanda per approfondimenti, è realizzato a partire da prove di certificazione e, in particolare, da prove di produzione scritta: tra queste, contattare un amico dopo molto tempo (livello A2) oppure rispondere a un invito a un matrimonio (livello B1). Durante la fase di compilazione del corpus i testi sono stati rivalutati da esperti secondo scale incluse nella documentazione: il livello generale assegnato al singolo testo è quello che definiremo qui “livello valutato”. La scelta di lavorare su MERLIN è stata dettata, appunto, dalla possibilità di selezionare testi sulla base del livello del compito e del livello valutato.

Dal corpus è stato quindi estratto un campione di 160 testi suddivisi per livello del compito e livello valutato. Le dimensioni dei quattro gruppi risultanti sono piuttosto limitate, come si può osservare nella tabella sottostante, ma vi è comunque un numero sufficiente di documenti per realizzare test statistici. Non è stato inoltre possibile bilanciare le L1 dei soggetti, che risultano per la maggior parte di (63%) lingua tedesca, seguiti da francese (18%), ungherese (7%), polacco 7% e inglese e spagnolo, entrambi 1% (v. le conclusioni sui limiti della ricerca).

Tabella 1. Composizione del corpus

Gruppo	Liv. del compito	Liv. valutato	N. testi	Parole unità
Normoperformanti A2	A2	A2	40	3213
Sovraperformanti A2	A2	B1	40	3700
Normoperformanti B1	B1	B1	40	5363
Sottoperformanti B1	B1	A2	40	4040

Come corpus di riferimento, cioè il corpus sul quale sono calcolate le misure di associazione (v. *infra*), è stato scelto Paisà, un *web corpus* di circa 250.000.000 parole liberamente scaricabile (Lyding et al. 2014). Proprio il fatto di poter scaricare il corpus ha determinato la scelta di questo strumento, che è stato utilizzato anche in altre indagini su corpora di apprendenti di italiano (Spina 2018; Syianova-Chanturia 2015).

¹ <https://merlin-platform.eu/> (consultato il 10/06/2020).

Dal corpus Paisà sono stati estratti tutti i bigrammi e le singole parole insieme alle relative frequenze in modo da costituire una banca dati. La frequenza di bigrammi e parole nella banca dati utilizzata nella presente ricerca è analoga ai risultati ottenuti tramite l'applicativo Wordsmith Tools (Scott 2001) e può variare rispetto a quella restituita dall'interfaccia sul sito del corpus.

2.2 Metodo

2.2.1 Estrazione dei dati e applicazione delle misure

Sono stati quindi estratti i bigrammi dai vari documenti del campione e, per ognuno di essi, sono state calcolate le due misure di associazione (v. appendice 1 per dettagli sulle formule)². Sebbene in letteratura siano presentate diverse misure, come *z-score*, *Log Dice* o chi quadrato (v. Evert 2008), la presente ricerca si limita a *t-score* e MI perché utilizzate nei lavori ai quali essa fa riferimento.

La natura e l'uso delle due misure è materia di dibattito nel settore (v. *supra* e Appendice 1), tuttavia, come anticipato, il *t-score* valorizza la dimensione della frequenza di un bigramma e la MI quella della forza, o meglio l'esclusività tra i suoi componenti. Le due dimensioni sono calcolate confrontando frequenza attesa e osservata e pertanto vanno intese come più frequente/forte rispetto alle attese. Quando si parla di bigrammi frequenti con *t-score* alto, quindi, non si intende la frequenza assoluta, ma, appunto, una frequenza superiore alle attese. In particolare, è stato osservato (Gablasova et al. 2017) che non tutti i bigrammi con alta frequenza hanno *t-score* alto e pertanto l'uso del *t-score* come misura della frequenza, per quanto diffuso in letteratura, è problematico: la questione è affrontata in Appendice 1. I valori delle due misure sono stati calcolati secondo le formule presentate in Evert (2008), al quale si rimanda per approfondimenti.

2.2.2 Fasce di forza

Ogni valore (sia di *t-score*, sia di *MI*) è stato quindi inserito in una fascia di grandezza, secondo i criteri definiti da Bestgen e Granger (2014) e riportati nelle Tabelle 2 e 3. Le fasce raccolgono i bigrammi a seconda del loro valore in una determinata misura; per esempio, "mi piace" ha $MI=9,46$ e si trova nella fascia alta (A); mentre "questo inverno" ha $t-score=3,13$ e si trova nella fascia bassa (B) di questa misura³.

² L'Appendice 2 riporta, a titolo di esempio, l'analisi di un estratto da uno dei testi del corpus.

³ Esempi di bigrammi per ogni fascia nelle due misure sono riportati nell'Appendice 3.

Tabella 2. Valori soglia per le fasce del *t-score*

Fascia	T-score
Alta (A)	> 10
Media (M)	$6 \leq e \leq 10$
Bassa (B)	$2 \leq e \leq 6$
Non collocazionale (NC)	< 2
Sotto la soglia (SS)	Frequenza osservata < 5

Tabella 3. Valori soglia per le fasce della *Mutual Information*

Fascia	MI
Alta (A)	> 7
Media (M)	$5 \leq e \leq 7$
Bassa (B)	$3 \leq e \leq 5$
Non collocazionale (NC)	< 3
Sotto la soglia (SS)	Frequenza osservata < 5

Poiché, tuttavia, il *t-score* (ma non la *Mutual Information*) è influenzato dalla dimensione del corpus di riferimento, i valori di questa misura sono stati normalizzati, cioè calcolati come se le dimensioni del corpus di riferimento fossero 100.000.000 parole. La normalizzazione del *t-score* è un'operazione importante (Gablasova et al. 2017), perché influisce sulla possibilità di comparare risultati calcolati su corpora di riferimento diversi. Nel caso presente, le dimensioni del corpus di riferimento influiscono sulla distribuzione dei bigrammi nelle diverse fasce: infatti, più grande è il corpus di riferimento, più alto è il valore del *t-score* e, di conseguenza, maggiore il numero di bigrammi nelle fasce alte⁴. Per esempio, “chiama Maria” ha un *t-score* di 3,04 (e sarebbe perciò in fascia bassa) se calcolato sulle dimensioni del corpus Paisà, ma tale misura diventa 1,93 (e va quindi nella fascia non collocazionale) se normalizzata. Poiché i risultati della presente ricerca saranno confrontati, per quanto possibile, con lavori precedenti sullo stesso argomento (Durrant e Schmitt 2009; Granger e Bestgen 2014) che hanno utilizzato il *British National Corpus*, il *t-score* è stato normalizzato alle dimensioni di tale corpus.

⁴ Infatti, quanto più cresce la dimensione del corpus di riferimento, tanto più crescono (in teoria) frequenza attesa e osservata. Poiché il *t-score* prevede la sottrazione della frequenza attesa dalla frequenza osservata (diviso la radice quadrata della frequenza osservata, v. Appendice I), ne consegue che esso è correlato alle dimensioni del corpus.

La natura dei bigrammi contenuti nelle diverse fasce, per esempio le parti del discorso coinvolte (fattore che potrebbe essere determinante nel calcolo delle misure⁵), non sono stati oggetto di indagine e, al momento, manca un resoconto sistematico di cosa caratterizzi ognuna di esse. Forniamo tuttavia di seguito una sintetica panoramica delle diverse fasce.

La fascia alta (A) del *t-score* contiene bigrammi ad alta frequenza composti, per esempio, da parole grammaticali (preposizione+articolo, “con la”, *t-score*=269,34), o da combinazioni di parole grammaticali e lessicali (“della città”, 127,76 o “numero di”, 113,65), ausiliare+verbo “ha detto” (62,94) o aggettivo+sostantivo, “prossime settimane” (10,05). Si noti che, nonostante i limiti del *t-score*, tutti i bigrammi nella fascia alta di questa misura hanno frequenza alta. Scendendo nelle fasce, i bigrammi sono composti da una o due parole meno frequenti, come “di vederti” (3, 44) o “bella passeggiata” (3,39). Nella fascia non collocazionale (NC, *t-score*<2) i bigrammi sono, in genere, grammaticali, ma poco frequenti; il loro valore di *t-score* è infatti considerato per convenzione (Hunston, 2002) troppo basso per essere collocazionale, cioè non vi è attrazione tra i due membri; per esempio, “e amore” (-17,54) o “stati in” (-15,03). Poiché non vi è attrazione, questi bigrammi sono in genere considerati il prodotto di combinazioni libere che, riprendendo le note definizioni di Sinclair (1991), rispettano il principio della scelta libera, piuttosto che quello idiomatico.

Nel caso della MI si trovano nella fascia alta combinazioni esclusive, cioè coppie di parole che tendono a trovarsi insieme, spesso combinazioni di aggettivo e sostantivo “settimana scorsa” (11,10). Scendendo verso le fasce inferiori si incontrano via via bigrammi sempre meno esclusivi, fino alla fascia NC, che però segue regole diverse rispetto al *t-score*. Molti bigrammi nella fascia A del *t-score* (28% nei testi in esame nella presente ricerca), per esempio, sono combinazioni di parole ad altissima frequenza, come preposizioni o articoli (per es. “con la” MI=2,87), e quindi piuttosto deboli secondo la formula della MI.

I bigrammi categorizzati come sotto la soglia (SS), infine, sono quelli che hanno, nel corpus di riferimento, una frequenza osservata pari a 0 o talmente bassa (qui usiamo la soglia di 5) da rendere inaffidabili i valori della frequenza. All’interno di questa fascia possono ricadere errori di ortografia, morfologia e sintassi. L’analisi per bigrammi non è, va chiarito, un metodo di correzione automatica. Tuttavia, a livelli medio-bassi, dove errori di morfologia e ortografia sono frequenti, questa fascia può assumere una funzione importante nel discriminare la competenza di un soggetto nella produzione scritta. L’ortografia è, infatti, uno dei

⁵ V. Stubbs (2001) sulla critica all’idea di definire la frequenza attesa sulla base di una combinazione casuale delle parole.

criteri usati dagli esperti nella valutazione delle prove del corpus⁶. Questo è uno dei motivi per cui abbiamo scelto, a differenza di Bestgen e Granger, di non emendare i testi del corpus in esame. L'altro motivo è che il metodo adottato dai due autori, cioè correggere l'ortografia quando questa deviava di uno o due errori minori dalla forma corretta, è difficilmente applicabile ai testi del nostro campione, dal momento che molti casi non sembrano imputabili a sviste: per esempio, un singolo testo riporta parole come *congratulazioni*, *ringrazio* e *invitazione*, che lasciano intuire una carenza specifica nel soggetto. Al fine di verificare in via preliminare l'ipotesi secondo la quale in questa soglia si troverebbero molti bigrammi con errori abbiamo estratto da essa 100 bigrammi casuali, equamente divisi tra i quattro gruppi, per fare una statistica di quali sono grammaticali e quali no: i primi ammontano al 47%, i secondi al 53%. Quindi un testo con molti errori avrà in genere molti bigrammi in questa fascia.

Per ogni singolo testo abbiamo quindi provveduto a realizzare un profilo di bigrammi contenente una percentuale delle diverse fasce nelle due misure, per esempio:

nome documento: 1395_0000340.txt, livello del compito: A2, livello valutato: A2 (soggetto normo-performante A2)

Tabella 4. Profilo di bigrammi del documento 1395_0000340.txt (normoperformante A2)

Fascia	t-score	MI
A	40.98%	4.92%
M	9.84%	18.03%
B	13.11%	16.39%
NC	22.95%	47.54%
SS	13.11%	

2.2.3 Procedura

È stato, infine, realizzato un documento nel quale per ogni testo del campione sono riportati livello del compito, livello valutato e le percentuali delle singole fasce (quello che abbiamo definito “profilo di bigrammi”).

L'analisi consiste in un test Anova a due vie (e test alternativi quando le assunzioni non sono rispettate, v. Levshina 2015 sui parametri dei test Anova) nella quale la variabile dipendente, cioè la percentuale di una data fascia in un gruppo, è interpretata alla luce di due variabili indipendenti: livello del compito e livello valutato. Il test permette di calcolare sia l'impatto di ciascuna delle variabili

⁶ V. la documentazione sul sito.

indipendenti, sia la loro interazione, cioè se una amplifica o inibisce l'impatto dell'altra. Il modello statistico (v. Gries 2009 sui modelli) è:

percentuale di bigrammi nella fascia X ~ livello del compito * livello valutato

La dimensione dell'effetto per i test Anova è calcolata tramite la formula del η^2 che, secondo la definizione di Cohen (1988), stimiamo come debole (0,01), medio (0,06) e forte (0,14).

Abbiamo infine eseguito i necessari test *post-hoc*, per individuare i gruppi che presentano differenze significative rispetto agli altri. Per ogni test sono riportati, il grado di confidenza (* $p \leq 0,05$, ** $p \leq 0,01$ e *** $p \leq 0,001$), così come la dimensione dell'effetto, la d di Cohen ($d=0,2$ effetto debole; $d=0,5$, medio e $d=0,8$ forte).

3 Risultati

3.1 Bigrammi sotto la soglia (SS)

Le percentuali di bigrammi nella fascia SS sono state analizzate con un test Anova a 2 vie (Livello valutato: A2 vs. B1 e Livello del compito: A2 vs. B1), con correzione di White per campioni con varianza eterogenea.

L'effetto della difficoltà del compito (livello del compito) è estremamente significativo, $F(1, 156) = 21,44$, $p < 0,001$ (***), $\eta^2 = 0,08$, mentre l'effetto del livello valutato non lo è, $F(1, 156) = 0,07$, $p = 0,7$. Tuttavia, il livello valutato degli apprendenti intensifica l'effetto della difficoltà con $F(1, 156) = 13,44$, $p < 0,001$ (***), $\eta^2 = 0,06$.

La frequenza in questa fascia non cambia per i soggetti con competenza B1 (A2B1 e B1B1), mentre invece cambia, a seconda del compito, per i soggetti con competenza A2 (A2A2, B1A2) che producono molti più bigrammi in questa fascia quando devono svolgere un compito B1, questo spiega l'effetto intensificatore del livello del compito.

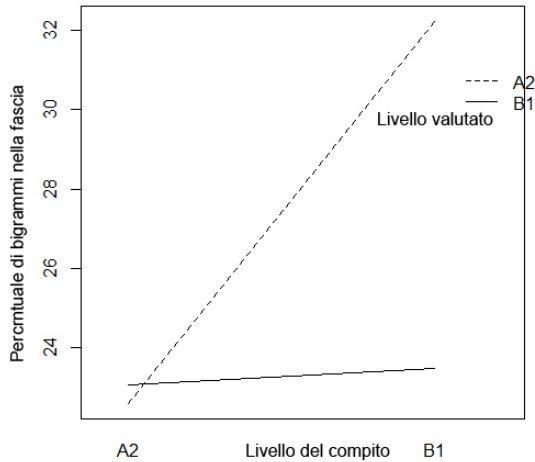


Immagine 1. Interazione tra le variabili indipendenti (livello del compito e livello valutato) nella fascia SS

Nel test post-hoc si osserva infatti che le differenze sono significative, e con una dimensione dell'effetto molto alta, tra B1A2 e tutti gli altri gruppi: B1A2 ($M=32,22$, $DS=10,54$) – A2A2 ($M=22,60$, $DS=7,56$), differenza=9,62, $p<0,001$ (***) , $d=1,04$; B1A2 – A2B1 ($M=23,05$, $DS=7,03$), diff. =9,17. $p<0,001$ (***) , $d=1,02$; B1A2 – B1B1 ($M=23,47$, $DS=5,23$), diff. =8,74, $p<0,001$ (***) , $d=1,05$.

3.2 T-score

Le percentuali di bigrammi nella fascia NC del t-score sono state analizzate tramite un test Anova a due vie (senza correzione, come i successivi test, perché la varianza è omogenea). L'effetto del livello del compito è estremamente significativo, $F(1, 156) = 18,9$, $p<0,001$ (***) , $\eta^2=0,10$, mentre l'effetto del livello valutato è molto significativo $F(1, 156) = 9,99$, $p=0,01$ (**), $\eta^2=0,06$; tuttavia, non vi è un'interazione significativa tra le due variabili indipendenti, $F(1, 156) = 2,98$, $p=0,08$.

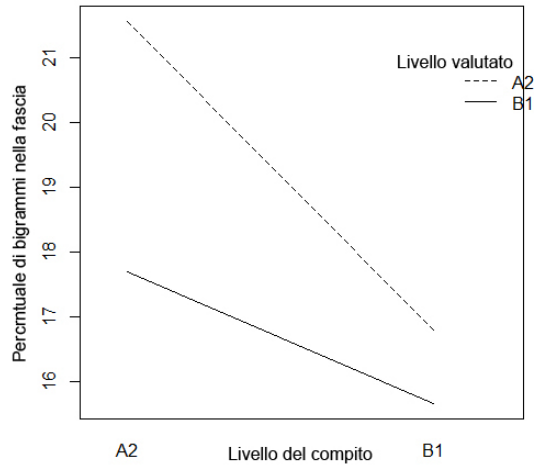


Immagine 2. Interazione tra le variabili indipendenti (livello del compito e livello valutato) nella fascia NC del *t-score*

La differenza è, in questo caso, tra normoperformanti A2 e tutti gli altri gruppi, che invece non mostrano differenze tra loro: A2A2 ($M= 21,56$, $DS= 5,21$) vs. A2B1 ($M= 17,69$, $DS= 5,37$), diff. 3,87, $p=0,003$ (**), $d= 0,73$; A2A2 vs. B1A2 ($M= 16,79$, $DS= 5,31$), diff. 4,76 $p<0,001$ (***), $d= 0,90$; A2A2 vs. B1B1, ($M= 15,66$, $DS= 3,99$), diff. 5,89 $p<0,001$ (***), $d= 1,27$.

La fascia bassa (B) presenta, se analizzata tramite lo stesso test, una sola differenza significativa, che è relativa alla variabile livello valutato $F(1, 156)=7,16$, $p=0,008$ (**), $\eta^2=0,06$ senza interazione con livello del compito e nessuna differenza significativa tra i gruppi nel test *post-hoc*. In altre parole, i soggetti con lo stesso livello valutato producono una percentuale di bigrammi simile in questa fascia sia in un compito A2, sia B1. La fascia M non presenta differenze significative.

Nel caso della fascia A la variabile più forte è costituita dal livello valutato $F(1, 156) = 13,55, p < 0,001 (***)$, $\eta^2 = 0,07$; il livello del compito non ha invece un impatto significativo di per sé, ma interagisce con l'altra variabile, intensificandone l'effetto $F(1, 156) = 9,44, p = 0,002 (**)$, $\eta^2 = 0,05$. La differenza significativa è anche in questo caso quella tra i sottoperformanti B1 e gli altri gruppi: B1A2 ($M = 33,75, DS = 7,36$) vs. A2A2 ($M = 39,01, DS = 7,38$), diff. 5,25 $p = 0,004 (**)$, $d = 0,71$; B1A2 vs. A2B1 ($M = 39,66, DS = 6,41$), diff. 5,90 $p < 0,001, d = 0,85$; B1A2 vs. B1B1 ($M = 41,05, DS = 6,07$), diff. 7,29 $p < 0,001, d = 1,08$;

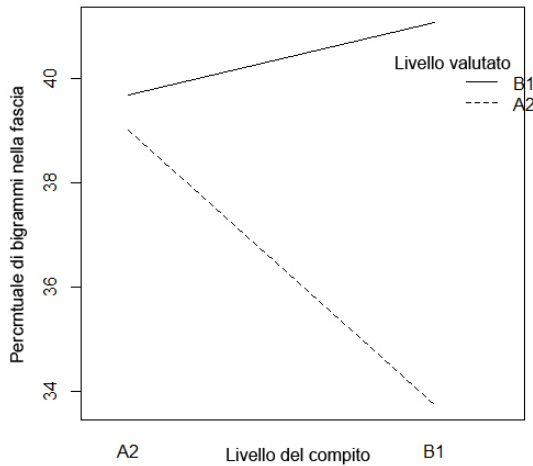


Immagine 3. Interazione tra le variabili indipendenti (livello del compito e livello valutato) nella fascia A del *t-score*

Nell'esecuzione di un compito B1 i soggetti A2, in altre parole, sono caratterizzati, rispetto ai soggetti B1, da una difficoltà verso i bigrammi di fascia alta. Vi è inoltre una correlazione negativa tra la fascia A del *t-score* e la fascia SS, $r = -0,71$, la correlazione è forte e significativa: $t = -13,03, df = 158, p < 0,001 (***)$: quanto più aumentano i bigrammi in una fascia, tanto più diminuiscono nell'altra.

3.3 Mutual Information

Relativamente alle percentuali di bigrammi nella fascia NC si osserva che l'effetto del livello del compito è molto significativo, $F(1, 156) = 8,08, p=0,005 (**)$, mentre l'effetto della livello valutato no; tuttavia, vi è interazione tra le due variabili indipendenti, $F(1, 156) = 8,18, p=0,04 (*)$, $\eta^2=0,04$, con la variabile livello valutato che intensifica l'effetto dell'altra. In maniera analoga a quanto avviene nella fascia SS, la percentuale della fascia NC è costante per i soggetti con competenza B1, mentre diminuisce sensibilmente per i soggetti A2 nel passaggio verso un compito più complesso.

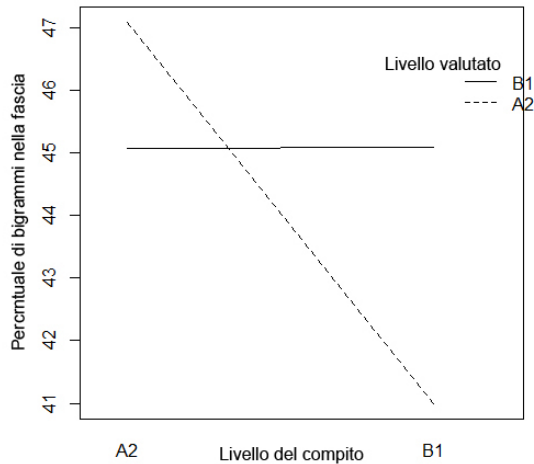


Immagine 4. Interazione tra le variabili indipendenti (livello del compito e livello valutato) nella fascia NC della *Mutual Information*

Anche in questo caso, la differenza è tra il gruppo B1A2 e gli altri, con B1A2 ($M=40,99, DS=7,84$) vs. A2A2 ($M=47,08, DS=5,65$), diff. 6,08 $p<0,001 (***)$, $d=0,88$; B1A2 vs. A2B1 ($M=45,07, DS=6,87$), diff. 4,07, $p=0,03 (*)$, $d=0,55$; B1A2 vs. B1B1 ($M=45,09, DS=6,42$), diff. 4,09, $p=0,03 (*)$, $d=0,57$.

Vi è una correlazione negativa tra la fascia NC della MI e la fascia SS, $r=-0,65$, la correlazione è moderata e significativa: $t=-10,92, df=158, p<0,001 (***)$: analogamente alla fascia A del *t-score*, quanto più aumentano i bigrammi in una fascia, tanto più diminuiscono nell'altra.

Nella fascia M della MI l'unica variabile che ha un effetto è il livello valutato con $F(1, 156) = 5,316, p = 0,02$ (*), mentre quella relativa al livello del compito non è significativa né interagisce con l'altra. Anche in questo caso, inoltre, il test *post-hoc* non mostra alcuna differenza significativa tra i gruppi. Non vi è, infine, alcuna differenza significativa nelle fasce B e A della MI.

4. Discussione

4.1 Risposte alle domande di ricerca

La risposta alla prima domanda di ricerca dipende dal gruppo in esame. I soggetti sovraperformanti A2 producono un profilo di bigrammi analogo a quello dei normoperformanti B1, cioè le percentuali delle diverse fasce delle due misure non presentano differenze significative. Ciò implica, rispetto al gruppo di riferimento per il livello del compito (A2A2, cioè apprendenti di livello A2 che svolgono un compito A2), una diminuzione della fascia non collocazionale del *t-score*. I soggetti sottoperformanti B1 presentano, invece, un profilo di bigrammi originale, che ha, rispetto a tutti gli altri gruppi, meno bigrammi nella fascia alta (A) del *t-score*, più bigrammi nella fascia sotto la soglia (SS) e meno bigrammi nella fascia NC della MI.

I sottoperformanti costituiscono un gruppo di particolare interesse, perché mostrano le difficoltà che caratterizzano i soggetti di livello A2 rispetto a un compito B1 e per questo motivo meritano particolare attenzione. Nelle ipotesi di partenza la fascia alta del *t-score* costituirebbe una sorta di zona sicura per gli apprendenti, perché raccoglie bigrammi molto frequenti e quindi facili da memorizzare. La fascia SS dovrebbe, inoltre, essere più ampia nei sottoperformanti a causa degli errori derivanti dalla difficoltà a svolgere il compito. Ci si poteva quindi attendere in questo gruppo una polarizzazione verso gli estremi, a scapito delle fasce intermedie (NC, B e M), il che però non avviene completamente.

I risultati della fascia A del *t-score* mostrano, per tutti i gruppi, una influenza determinante del livello valutato. I soggetti di livello B1 sono in genere più forti in questa fascia; tuttavia, mentre in un compito di livello A2 la differenza tra i due livelli valutati è, pur presente, non significativa, essa viene intensificata nel passaggio a un compito B1 e la difficoltà dei soggetti A2, qui sottoperformanti, si traduce in una diminuzione significativa in questa fascia. La capacità di memorizzare e riprodurre bigrammi ad alta frequenza costituisce quindi, per questo gruppo non una risorsa su cui fare affidamento nell'affrontare un compito più complesso, ma un obiettivo da perseguire.

Il secondo carattere del gruppo B1A2 è costituito, come invece previsto, dall'alta percentuale di bigrammi nella fascia SS. Questa situazione è simile al caso precedente. I soggetti B1 non mostrano alcuna differenza tra un compito A2 e uno B1 e anche i soggetti A2, finché lavorano su un compito A2, mostrano percentuali analoghe. Si noti, infatti, che il livello valutato dei soggetti non ha un effetto significativo a questo livello, ma, di nuovo, il passaggio a un compito B1 fa emergere differenze tra i livelli valutati, come suggerisce l'alto valore dell'interazione tra le variabili indipendenti. Come dimostrato dalla correlazione negativa tra fascia A del *t-score* e fascia SS, inoltre, sembra esserci in questi soggetti una connessione tra le due, anche se non è chiara la sua natura. Va infatti tenuto presente, come già anticipato, che i motivi per cui un bigramma si trova, a questi livelli, nella fascia SS sono molteplici (e non sono qui indagati) e sarebbe quindi azzardato concludere, senza ulteriori indagini, che i soggetti sottoperformanti devono il loro *status* a una minore competenza fraseologica. Questo dato conferma, inoltre, l'importanza di non correggere l'ortografia; non importa, infatti, il motivo per cui i sottoperformanti hanno difficoltà nella fascia alta del *t-score*, dal momento che, secondo le ipotesi iniziali, questa dovrebbe invece costituire il loro punto di forza.

Da un punto di vista metodologico, infine, va osservato che il risultato è analogo a quello di Bestgen e Granger (2014), i quali trovano una correlazione negativa tra valutazione globale di un elaborato e numero di bigrammi sotto la soglia. Tuttavia, come anticipato, il corpus su cui hanno lavorato i due studiosi è composto da documenti prodotti a partire da compiti dello stesso livello (Connor-Linton e Polio 2014) e solo in seguito valutati. Il confronto tra i loro risultati e quelli della presente ricerca sembra suggerire che la dimensione della fascia SS non dipenda tanto dal livello valutato dei soggetti, ma dal rapporto tra livello valutato e del compito. In altre parole, come ipotizzato nell'introduzione, un soggetto sottoperformante fa più errori e, di conseguenza, produce più bigrammi in questa fascia. Questo elemento va dunque tenuto in considerazione quando si conduce un'analisi per bigrammi e misure di associazione.

Il minor numero di bigrammi nella fascia NC della *Mutual Information* nel gruppo B1A2, infine, è di fatto l'unico risultato significativo prodotto da questa misura e conferma, *a latere*, lo scarso ruolo dell'esclusività a questi livelli, almeno relativamente all'analisi di tutti i bigrammi⁷. La diminuzione di bigrammi in questa fascia nei sottoperformanti costituisce un dato importante, ma non sorprendente perché, come già anticipato, circa un terzo dei bigrammi della fascia NC

⁷ Siyanova-Chanturia (2015) e Spina (2018), per esempio, utilizzano la MI con apprendenti elementari, ma lavorano su combinazioni specifiche, come sostantivo+aggettivo.

della MI si trovano nella fascia A del *t-score*: quindi una debolezza nella seconda si riflette nella prima (v. anche la correlazione negativa tra fascia NC della MI e fascia SS).

I soggetti sottoperformanti, in conclusione, presentano un profilo di bigrammi originale, diverso da quello dei normoperformanti (sia di livello valutato sia di livello del compito) che permette di determinare le loro debolezze (v. *infra* per una conclusione su questo gruppo). Si conferma, quindi, ma solo in parte, l'ipotesi iniziale secondo la quale due testi valutati come di analogo livello, ma realizzati a partire da compiti di livelli diversi, presentano un profilo di bigrammi diverso.

La risposta alla seconda domanda è no: i soggetti A2 non producono più bigrammi nelle fasce alte del *t-score* rispetto agli intermedi, ma la stessa percentuale in compiti corrispondenti al loro livello valutato e addirittura una percentuale inferiore in compiti più complessi. I risultati sembrerebbero escludere che i soggetti di livello basso tendano a sovra utilizzare bigrammi ad alta frequenza rispetto agli intermedi.

I dati sembrerebbero quindi suggerire che il passaggio dal livello A2 al livello B1 è caratterizzato, in primo luogo, da un percorso di (parziale) perdita e successiva riappropriazione della fascia alta del *t-score*, quindi di bigrammi ad alta frequenza.

4.2 Il caso della fascia NC del *t-score*

Un secondo elemento emerso durante l'analisi dei dati è costituito dalla fascia non collocazionale del *t-score*, con i normoperformanti A2 che producono più bigrammi in questa fascia rispetto tutti agli altri gruppi. Poiché un approccio pseudo-longitudinale permette di fare inferenze sullo sviluppo linguistico (Gass e Selinker 2008), la conclusione è che il passaggio tra i livelli A2 e B1 avviene con una diminuzione non nella fascia alta del *t-score*⁸, ma nella fascia non collocazionale. Questa diminuzione sembrerebbe costituire una prova, per quanto debole (dal momento che essa non coincide con un aumento significativo delle fasce collocazionali), a sostegno del ruolo della competenza fraseologica nella determinazione del livello linguistico.

Tuttavia, la diminuzione in questa fascia non si può, come visto, spiegare né con la difficoltà del compito, perché i sovraperformanti A2 (A2B1) usano, a parità di compito, meno bigrammi in questa fascia, né con il livello valutato, perché i

⁸ Non bisogna qui confondere le difficoltà dei sottoperformanti con il quadro generale. I sottoperformanti rappresentano soggetti che non hanno ancora raggiunto il livello B1, mentre i soggetti B1 (normo e sovraperformanti) presentano una percentuale di bigrammi in questa fascia analoga ai normoperformanti A2.

sottoperformanti B1 (B1A2) producono anch'essi meno bigrammi in questa fascia. Per comprendere quest'ultimo dato occorre osservare che nel test entrambe le variabili indipendenti influiscono su quella dipendente, ma, come visto, il risultato non significativo dell'interazione non permette di escludere l'ipotesi nulla secondo la quale esse non si influenzano a vicenda. L'alto valore della fascia NC del gruppo A2A2, in altre parole, è determinato sia dal livello valutato, sia dal livello del compito e non è perciò possibile stabilire se ve ne sia uno dominante (come nel caso, per esempio, della fascia SS). Infatti, già un compito più complesso (che implica, per esempio, una maggiore coesione) è sufficiente a far produrre a soggetti A2 meno bigrammi non collocazionali, mentre apprendenti di livello B1 producono meno bigrammi in questa fascia già in un compito più semplice.

Detto in termini più generali (e in ottica didattica), non sembra questa la fascia potenzialmente problematica per i soggetti A2, ma, come visto, la fascia A. I sottoperformanti non sembrano vincolati al principio della scelta libera, cioè sembrano cogliere la dimensione fraseologica, ma faticano a memorizzare e riprodurre bigrammi molto frequenti necessari a portare a termine un compito B1. Stando a questi risultati (ma v. le conclusioni sui limiti della ricerca) eventuali misure di supporto dovrebbero quindi essere orientate a una maggiore esposizione all'input in lingua straniera e ad attività volte alla memorizzazione piuttosto che a sviluppare la sensibilità verso la fraseologia.

4.3 Sintesi dei risultati

Le risposte alle due domande di ricerca possono essere, infine, confrontate con i risultati di lavori precedenti, in particolare con Granger e Bestgen (2014), che condividono con la presente l'approccio pseudo-longitudinale. Come visto, gli autori si domandano se non vi possa essere una sorta di correlazione negativa tra frequenza ed esclusività (attrazione), con i soggetti dei livelli più bassi che fanno maggior uso di strutture ad alta frequenza. I risultati della presente ricerca forniscono, pur nei limiti di un campione di dimensioni ridotte, alcuni dettagli in merito: la dimensione della frequenza (bigrammi nella fascia A del *t-score*), infatti, tende a essere satura già prima del livello intermedio⁹ e il passaggio verso il livello

⁹ Come detto nell'introduzione, i livelli medio/bassi sono caratterizzati da lessico e strutture ad alta frequenza, che si traducono anche in bigrammi con *t-score* molto alto, come "miei genitori", (*t-score*=11,63) "piace molto", (10,44). La fascia alta del *t-score* non cresce al livello valutato B1 perché la dimensione della frequenza dovrebbe semmai diminuire salendo nei livelli (v. *supra*), ma al livello B1, che tratta di compiti familiari e sempre abbastanza concreti, la virata verso un lessico e strutture meno frequenti non è ancora così forte da implicare differenze marcate rispetto al livello A2.

B1 è caratterizzato invece da uno sviluppo nella competenza fraseologica più diffuso, che vede la fascia NC assottigliarsi. Un risultato compatibile con le teorie *usage-based* spesso citate nelle ricerche in questo campo, perché potrebbe corrispondere a una fase nello sviluppo interlinguistico nella quale i soggetti analizzano formule memorizzate e ne estrapolano schemi produttivi (ing. *pattern*, Ellis e Wulff 2015 fanno l'esempio di schemi come aggettivo+sostantivo). Soprattutto, i risultati sembrerebbero confermare l'importanza della frequenza (Ellis et al. 2008), come si vede nel caso del gruppo B1A2, che è caratterizzato proprio da una carenza in questo senso. Anche questo dato è compatibile con una prospettiva *usage-based*, secondo la quale la difficoltà nel memorizzare e produrre bigrammi ad alta frequenza implica una più generale difficoltà linguistica.

In conclusione, non sembra ipotizzabile, secondo i dati presentati, una correlazione negativa tra frequenza ed esclusività nella progressione; il percorso sembra, piuttosto, non lineare e il livello dei soggetti è solo uno tra i fattori che lo determinano.

5. Conclusioni

La ricerca ha confrontato quattro gruppi di apprendenti, corrispondenti a quattro diverse combinazioni di competenza linguistica e difficoltà del compito, attraverso le percentuali di bigrammi prodotti nelle diverse fasce in due misure di associazione, il *t-score* e la *Mutual Information*.

Secondo l'analisi condotta la competenza fraseologica giocherebbe un ruolo importante nella definizione del livello linguistico di un soggetto, come si osserva nel decadimento della fascia non collocazionale del *t-score* da A2 a B1 o le difficoltà nella fascia alta nei sottoperformanti B1. Soprattutto, i risultati sembrerebbero indicare che tale competenza non costituisce un blocco monolitico, né presenta uno sviluppo lineare, dal momento che alcune fasce sono più o meno sviluppate a seconda della diversa interazione di livello valutato e del compito.

In prospettiva didattica, oltre a confermare la posizione, diffusa tra i lavori in quest'area (v. *supra*), secondo la quale la fraseologia costituisce un aspetto da valorizzare nell'insegnamento linguistico, i risultati sembrerebbero suggerire l'importanza, nel passaggio da A2 a B1, di attività di riconoscimento e memorizzazione di combinazioni ad alta frequenza. Tali combinazioni (v. Appendice 3) non sono, tuttavia, costituite solo da collocazioni nella forma sostantivo-aggettivo, ma anche bigrammi composti da altre classi di parole. Ricerche mirate sui tipi di bigrammi contenuti nelle diverse fasce (e nei diversi gruppi) potrebbero

fornire indicazioni più precise circa i bisogni degli apprendenti e, quindi, sui tipi di attività a loro più utili.

Bestgen e Granger (2014), inoltre, ipotizzano per la tecnica CollGram un ruolo come strumento per la valutazione di una competenza unitaria lessico-grammaticale. Rispetto al loro approccio (v. *supra*), i risultati della ricerca illustrata lascerebbero ben sperare in questo senso, perché la tecnica è stata qui in grado di distinguere non solo il livello valutato, ma anche di metterlo in relazione con il livello del compito. Come si è visto, per esempio, la percentuale di bigrammi nella fascia sotto la soglia è caratteristica di uno specifico gruppo, A2 che tentano di svolgere un compito B1. Si tratta, allo stato attuale, di ipotesi sulle quali occorrerà lavorare con attenzione e sulla base di risultati più ampi.

Da un punto di vista metodologico, infine, le conclusioni gettano una luce sul ruolo delle misure di associazione a livelli medio bassi, confermando l'ipotesi secondo la quale a tali livelli la frequenza costituisce la dimensione principale, almeno se si prendono in esame tutti i bigrammi e non categorie specifiche. Inoltre, come già osservato, la difficoltà del compito influisce sul profilo di bigrammi di un testo.

Tali conclusioni non possono, tuttavia, dirsi definitive visto il campione di indagine piuttosto limitato e squilibrato relativamente alla L1 dei soggetti, che rende necessario confermare i risultati su campioni più ampi e meglio bilanciati. Al netto dei limiti della presente ricerca, in ogni caso, molti sono gli ambiti sui quali lavorare. In primo luogo, manca per la lingua italiana un riferimento preciso sui testi ai quali gli apprendenti sono esposti; tali testi possono costituire un termine di paragone relativamente alla frequenza delle diverse fasce. Inoltre, al fine di approfondire la relazione tra competenza fraseologica e linguistica sarà necessario esaminare i bigrammi all'interno delle diverse fasce ed individuare eventuali differenze qualitative tra i diversi gruppi/livelli. Un aspetto importante, per esempio, è comprendere se i compiti del livello B1 implicino bigrammi di fascia alta diversi rispetto a quelli della stessa fascia, ma del livello A2. Un ultimo campo di indagine è costituito dalle sinergie tra bigrammi/misure di associazione e altri metodi: come dimostra il lavoro di Garner, Crossley e Kyle (2020), che prende in esame anche trigrammi, tali sinergie costituiscono un orizzonte di ricerca importante nel rendere sempre più efficaci le analisi semiautomatiche di testi di apprendenti.

L'analisi semiautomatica è, in conclusione, un campo innovativo e ricco di potenzialità che già dai suoi primi esperimenti offre risultati promettenti e utili non solo alla ricerca sull'apprendimento della lingua straniera, ma anche alla pratica didattica.

Riferimenti

- Bestgen, Y. & Granger, S. 2014. Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing* 26: 28-41.
- Bestgen, Y. & Granger, S. 2018. Tracking L2 writers' phraseological development using collgrams: Evidence from a longitudinal EFL corpus. In S. Hoffmann, A. Sand, S. Arndt-Lappe & L. M. Dillmann (eds), *Corpora and lexis*. Leiden: Brill Rodopi, 277-301.
- Boyd, A., Hana, J., Nicolas, L., Meurers, D., Wisniewski, K., Abel, A., Schöne K., Štindlová B. & Vettori, C. 2014. The MERLIN corpus: Learner language and the CEFR. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (eds.), *LREC 2014 Proceedings*. European Language Resources Association (ELRA), 1281-1288.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. New York: Academic press.
- Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. 2011. Predicting lexical proficiency in language learner texts using computational indices. *Language Testing* 28(4): 561-580.
- Connor-Linton, J. & Polio, C. 2014. Comparing perspectives on L2 writing: multiple analyses of a common corpus. *Journal of Second Language Writing* 26: 1-9.
- Durrant, P. & Schmitt, N. 2009. To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching* 47(2): 157-177.
- Ellis, N. C., Simpson-Vlach, R. & Maynard, C. 2008. Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *Tesol Quarterly* 42(3): 375-396.
- Ellis, N. C. & Wulff, S. 2015. Usage-based approaches to SLA. In B. VanPatten & J. Williams (eds.), *Theories in Second Language Acquisition*. New-York: Routledge, 75-93
- Evert, S. 2008. Corpora and collocations. In A. Lüdeling & M. Kytö (eds.), *Corpus linguistics*. Berlin: de Gruyter, 1212-1248.
- Gablasova, D., Brezina V. & McEnery, T. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language learning* 67(1): 155-179.
- Garner, J., Crossley, S. & Kyle, K. 2020. Beginning and intermediate L2 writer's use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching* 58(1): 51-74.
- Gass, S., Selinker, L. 2008. *Second language acquisition: An introductory course*. Mahwah: Lawrence Erlbaum Associates.
- Granger, S. & Bestgen, Y. 2014. The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching* 52(3): 229-252.
- Gries, S. .2009. *Statistics for linguistics with R: A practical introduction*. Berlin: de Gruyter.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Howarth, P. 1998. The phraseology of learners' academic writing. In A. P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press, 161-186.
- Kyle, K. & Crossley, S. A. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly* 49(4): 757-786.

- Laufer, B. & Waldman, T. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language learning* 61(2): 647-672.
- Levshina, N. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company.
- Lewis, M. 1993. *The lexical approach*. Hove: Language teaching publications.
- Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Dell'Orletta, F. Lenci, A. & Pirrelli, V. 2014. The Paisà corpus of Italian web texts., In F. Bildhauer & R. Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9) @ EACL 2014*, 36-43, URL <https://www.aclweb.org/anthology/W14-0400/> (visitato il 24 maggio 2020).
- Ohlrogge, A. 2009. Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (eds.), *Formulaic language* Vol. 2. Amsterdam: John Benjamins Publishing, 375-86.
- Paquot, M. 2019. The phraseological dimension in interlanguage complexity research. *Second language research* 35(1): 121-145.
- Pawley, A. & Syder, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. Schmidt (eds.), *Language and communication*. London and New York: Routledge, 191-227.
- Porcelli, G. 2004. *Comunicare in lingua straniera: il lessico*. Torino: UTET-Libreria.
- Qian, Y. 2019. Dynamism of collocation in L2 English writing: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*. Ahead of print. doi: <https://doi.org/10.1515/iral-2019-0012>.
- Ruegg, R., Fritz, E. & Holland, J. 2011. Rater sensitivity to qualities of lexis in writing. *Tesol Quarterly* 45(1): 63-80.
- Römer, U. 2009. The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7(1): 140-162.
- Scott, M. 2001. Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessy, A. Henry & R. L. Roseberry (eds.), *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins Publishing, 47-67.
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siyanova-Chanturia, A. 2015. Collocation in beginner learner writing: A longitudinal study. *System* 53: 148-160.
- Siyanova-Chanturia, A. & Spina, S. 2015. Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning* 65(3): 533-562.
- Spina, S. 2018. Lo sviluppo longitudinale della fraseologia in apprendenti cinesi di italiano L2. Uno studio preliminare su alcune categorie di errori. *Ricognizioni* 5(10): 97-119.
- Spinelli, B. & Parizzi, F. 2010. *Profilo della lingua italiana Livelli di riferimento del QCER A1, A2, B1, B2*. Firenze: La Nuova Italia.
- Stubbs, M. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell publishers.
- Wray, A. 2000. Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* 21(4): 463-489.
- Wulff, S. Ellis & N. C. 2018. Usage-based approaches to second language acquisition, In D. Miller, F. Bayram, J. Rothman & L. Serratrice (eds), *Bilingual Cognition and Language: The state of the science across its subfields*. Amsterdam: John Benjamins, 37-56.

Appendice 1. Dettagli su *t-score* e *Mutual Information*

Nella definizione di Durrant e Schmitt (2009) il *t-score* mette in evidenza collocazioni frequenti e la MI combinazioni meno comuni, ma le cui parole tendono a trovarsi raramente separate. Questa appendice riporta alcuni dettagli delle due formule utili ai fini di una giustificazione della loro interpretazione.

Entrambe le misure confrontano la frequenza osservata, o grezza, (O) di un bigramma, cioè il numero di volte in cui esso compare in un corpus di riferimento, con la frequenza attesa (E). E è calcolata come la probabilità di trovare insieme le due parole (w_1 e w_2) se tutte le parole del corpus di riferimento fossero mescolate. La formula è quella della probabilità congiunta data la frequenza delle due parole rispetto al numero totale di parole unità nel corpus di riferimento, cioè:

$$E = \frac{f_1 * f_2}{N}$$

Il *t-score* è calcolato da Evert (2008) con la formula:

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

Il *t-score* è da molti considerato una misura della frequenza¹⁰, cioè tende a valorizzare bigrammi molto frequenti e, in effetti, esso correla con O (ma v. Gablasova et al. 2017 per precisazioni in merito).

Affermare che il *t-score* correla con O, tuttavia, non significa che esso sia sinonimo di frequenza assoluta. A prescindere dai dubbi sollevati da diversi autori circa la sua validità (v. Evert 2008 che ne critica i fondamenti matematici e statistici), l'interpretazione del *t-score* come misura della frequenza è contestata, per esempio da Gablasova et al. (2017), i quali osservano che bigrammi molto frequenti possono avere un *t-score* basso. Per esempio, nel corpus Paisà il bigramma “da non” ha O=5414 (quindi una frequenza assoluta piuttosto alta), ma *t-score* molto basso (-45,18). Il motivo è che, data l'alta frequenza delle parole che compongono il bigramma (da, $f=1605559$; non, $f=1360704$), E (8738,76) è comunque più alta di O. Questo esempio spiega l'affermazione degli autori per cui, mentre tutti i bigrammi con *t-score* alto sono frequenti, non tutti i bigrammi frequenti hanno *t-score* alto.

¹⁰ Non, per esempio da Hunston, 2002 che lo considera una misura della certezza di una collocazione e usa una formula diversa rispetto a quella di Evert.

Il bigramma “da non” è molto frequente, ma meno di quanto atteso data la frequenza delle sue due parole le quali, quindi, non si attraggono (secondo questa formula) e il bigramma ricade perciò nella fascia non collocazionale, come molti altri bigrammi composti da parole ad altissima frequenza (per es. “e un”).

Il *t-score* è allora, forse più precisamente, una misura di quanto un bigramma è più frequente rispetto a quanto ci si attenderebbe (cioè, quanto maggiore è la differenza tra O ed E, tanto maggiore sarà il *t-score*), ma è comunque una misura della frequenza, sebbene non di quella assoluta. Quindi, è vero che non tutti i bigrammi frequenti hanno un *t-score* alto, ma questo potrebbe significare che essi non sono collocazionali o fraseologici (nella dimensione della frequenza) e quindi ricadono nella fascia NC. Si tratta di un’ipotesi che necessita di ulteriori e approfondite analisi.

La Mutual Information è, invece, calcolata secondo la formula:

$$MI = \log_2 \frac{O}{E}$$

Poiché è basata su una divisione, la MI non dipende tanto dalla grandezza di O ed E (ma v. Evert 2008 per il caso di $E < 1$), ma dal rapporto tra i due. Per esempio “sistema operativo” ha $O=4616$, $E=3,73$ ($f_1=100970$, $f_2=9235$), $MI=10,27$, in altre parole, circa la metà delle istanze di operativo compare insieme a sistema (che invece è molto più frequente). Se si prende il bigramma “protezione civile”, il risultato cambia di poco, anche in presenza di O ed E che corrispondono a circa un terzo di quella del bigramma precedente: $O=1490$, $E=1,65$ ($f_1=15529$, $f_2=26570$), $MI=9,82$ (fascia alta). Per questo MI valorizza l’esclusività, indipendentemente dalla frequenza del bigramma.

Appendice 2. Esempio di analisi con bigrammi.

Testo originale: “Anche fammi sapere se tutto sta in ordine con la tua famiglia.”.
Dall’elaborato 1395_0000549 (L1: tedesco; liv. del compito: A2; liv. valutato: A2).

Tabella 4. Esempio di analisi per bigrammi e misure di associazione

bigramma	Freq. osservata	t-score (normalizzato)	Fascia (t- score)	MI	Fascia (Mi)
anche fammi	0	0	SS	0	SS
fammi sapere	56	4,73	B	10,94	A
sapere se	1188	21,28	A	5,4	M
se tutto	609	9,09	M	1,26	NC
tutto sta	129	4,98	B	1,7	NC
sta in	757	-1,78	NC	-0,14	NC
in ordine	5171	37,85	A	2,58	NC
ordine con	210	-9,72	NC	-1,04	NC
con la	243225	269,34	A	2,87	NC
la tua	5382	44,66	A	4,74	B
tua famiglia	91	5,74	B	4,37	B
fino ad	16044	77,9	A	5,18	M
ad allora	4239	39,32	A	4,47	B
tanti saluti	94	6,12	M	9,6	A

Appendice 3. Esempi di bigrammi nelle diverse fasce.

La tabella riporta esempi di diversi bigrammi per ogni fascia nelle due misure.

Tabella 5. Esempi di bigrammi per le diverse fasce (dei bigrammi nella fascia SS non è riportata la misura di associazione, ma la frequenza osservata.)

Fascia	t-score	Mutual Information
Alta (A)	anche se: 164,19 mi piace: 36,19 al mese: 30,3 era difficile: 12,31 ho lavorato: 11,14	cari saluti: 10,86 mi piace: 9,46 sito internet: 8,72 ufficio turistico: 8,06 bella sorpresa: 7,21
Media (M)	ho studiato: 8,23 sto bene: 7,82 piace tanto: 7,69 bravo e: 6,59 allo sportello: 6,5	tempo libero: 6,93 sto bene: 6,72 siamo partiti: 6,43 i miei: 6,16 ci vediamo: 5,87
Bassa (B)	adesso lo: 4,09 alla posta: 3,08 coppia molto: 2,56 faccio per: 2,51 lavoro perché: 2,4	ho già: 4,64 una lista: 4,52 so che: 3,44 parlato con: 3,41 sei veramente: 3,28
Non collocazionale (NC)	scriverti per: 1,69 era magnifico: 0,6 ma vera: -7,83 pensi e: -8,21 e mia: -13,01	dove sono: 2,53 bravo e: 1,31 penso di: 0,48 sempre alla: 0,41 lavoro al: 0,01
Sotto la soglia (SS) I valori si riferiscono alla frequenza assoluta del bigramma nel corpus (O)	avere lavorare: $f=0$ incontriamo di: 3 italia te: 2 corso ci: 4 i ristorante: 1	