



Università  
Ca' Foscari  
Venezia

**Scuola Dottorale di Ateneo  
Graduate School**

**Dottorato di ricerca  
in Scienze del linguaggio  
Ciclo 24  
Anno di discussione 2013**

***Lexical bundles come tratto linguistico saliente***

**SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: L-LIN/02  
Tesi di Dottorato di Mojca Del Fabbro, matricola 955639**

**Coordinatore del Dottorato**

**Prof.ssa Alessandra Giorgi**

**Tutore del Dottorando**

**Prof. Paolo E. Balboni**

**Co-tutore del Dottorando**

**Prof.ssa Nives Zudič Antonič**

# Indice

<b>INTRODUZIONE</b> .....	<b>I</b>
<b>COORDINATE TEORICHE</b> .....	<b>3</b>
<b>I LO STUDIO SULLE COMBINAZIONI DI PAROLE E LA LORO</b>	
<b>IMPORTANZA</b> .....	<b>4</b>
I.1 I PRECURSORI DELLO STUDIO SULLE COMBINAZIONI DI PAROLE: GLI ANNI TRENTA E GLI ANNI CINQUANTA	5
I.2 GLI SVILUPPI DELLA TEORIA FIRTHIANA.....	10
I.3 LA TEORIA DEL <i>LEXICAL PRIMING</i> .....	18
I.4 L'IMPORTANZA DEL LINGUAGGIO FORMULAICO .....	19
I.5 GLI STUDI SUL LINGUAGGIO FORMULAICO PER LA LINGUA ITALIANA .....	25
I.6 <i>I LEXICAL BUNDLES</i> .....	32
I.6.1 <i>La definizione dei lexical bundles</i> .....	33
I.6.2 <i>La struttura grammaticale e la funzione discorsiva dei lexical bundles</i> ..	37
I.6.2.1 La struttura grammaticale dei lexical bundles .....	37
I.6.2.2 Le funzioni discorsive dei lexical bundles .....	39
I.6.3 <i>L'importanza dei lexical bundles nella produzione linguistica fluente</i> .....	44
I.6.4 <i>Le ricerche sui lexical bundles</i> .....	47
I.6.5 <i>Le ricerche in lingua inglese</i> .....	48
I.6.5.1 Le ricerche in altre lingue.....	57
I.7 CONCLUSIONI .....	60
<b>II STRUMENTI DI ANALISI DEL LINGUAGGIO FORMULAICO – LA</b>	
<b>LINGUISTICA DEI CORPORA</b> .....	<b>61</b>
II.1 DEFINIZIONE DI <i>CORPUS</i> .....	65
II.1.1 <i>La rappresentatività del corpus</i> .....	66
II.1.2 <i>La dimensione o l'estensione del corpus</i> .....	68
II.1.3 <i>Il riferimento standard</i> .....	70
II.2 IL <i>CORPUS</i> SPECIALISTICO .....	71
II.3 GLI APPROCCI <i>CORPUS BASED</i> E <i>CORPUS DRIVEN</i> .....	75
II.4 METODO COMPUTAZIONALE DI ESTRAZIONE DEI <i>LEXICAL BUNDLES</i> .....	79
II.5 CONCLUSIONI .....	85
<b>III ANALISI DEI CASI</b> .....	<b>87</b>
III.1 LA DIVERSITÀ TRA LINGUAGGIO PARLATO E SCRITTO NELLA DISTRIBUZIONE DEI NOMI E DEI VERBI	89
III.2 LA DISTRIBUZIONE STRUTTURALE DEI <i>LEXICAL BUNDLES</i> NEL PARLATO E NELLO SCRITTO..	96
III.2.1 <i>La distribuzione delle polirematiche nel parlato e nello scritto</i> .....	96
III.2.2 <i>I lexical bundles nel corpus parlato</i> .....	98
III.2.2.1 La struttura grammaticale dei lexical bundles .....	98
III.2.2.1.1 <i>Lexical bundles</i> che rappresentano frasi verbali complete o incorporano frammenti di frasi verbali .....	106
III.2.2.1.2 <i>Lexical bundles</i> che incorporano frammenti di sintagmi preposizionali .....	110
III.2.2.1.3 <i>Lexical bundles</i> che incorporano sintagmi nominali.....	111
III.2.2.1.4 <i>Lexical bundles</i> che sono congiunzioni .....	112
III.2.2.1.5 Altre espressioni.....	112
III.2.2.2 I tratti del parlato identificati nella struttura dei lexical bundles .....	113
III.3 <i>LEXICAL BUNDLES</i> NEL REGISTRO SCRITTO – NELLE MICROLINGUE SCIENTIFICO-PROFESSIONALI	119
III.3.1 <i>La struttura dei lexical bundles nel discorso accademico</i> .....	120
III.3.1.1 Il corpus di storia .....	121

III.3.1.1.1	La struttura dei <i>lexical bundles</i> nel <i>corpus</i> di storia .....	124
preposizionali	III.3.1.1.1.1 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi	125
sintagmi nominali	III.3.1.1.1.2 Lexical bundles che sono sintagmi nominali o incorporano frammenti di	128
	III.3.1.1.1.3 Lexical bundles che incorporano frammenti di frasi verbali .....	130
	III.3.1.1.1.4 Lexical bundles che sono congiunzioni.....	131
	III.3.1.1.2 I risultati emersi dall'analisi strutturale.....	131
III.3.1.2	Il <i>corpus</i> di filosofia.....	135
preposizionali	III.3.1.2.1 La struttura dei <i>lexical bundles</i> nel <i>corpus</i> di filosofia.....	138
sintagmi nominali	III.3.1.2.1.1 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi	139
	III.3.1.2.1.2 Lexical bundles che sono sintagmi nominali o incorporano frammenti di	141
	III.3.1.2.1.3 Lexical bundles che incorporano frammenti di sintagmi verbali.....	143
	III.3.1.2.1.4 Lexical bundles che sono congiunzioni o locuzioni congiuntive .....	144
	III.3.1.2.1.5 Lexical bundles che sono avverbi .....	144
	III.3.1.2.2 I risultati emersi dall'analisi strutturale.....	144
III.3.1.3	I <i>lexical bundles</i> nel <i>corpus</i> di medicina del lavoro.....	146
sintagmi nominali	III.3.1.3.1 La struttura dei <i>lexical bundles</i> nel <i>corpus</i> di medicina del lavoro.....	148
preposizionali	III.3.1.3.1.1 Lexical bundles che sono sintagmi nominali o incorporano frammenti di	150
	III.3.1.3.1.2 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi	154
	III.3.1.3.1.3 Lexical bundles che incorporano frammenti di sintagmi verbali.....	158
	III.3.1.3.2 I risultati emersi dall'analisi strutturale.....	160
<b>III.3.2</b>	<b>La struttura dei <i>lexical bundles</i> nella prosa accademica.....</b>	<b>161</b>
nominali	III.3.2.1 Lexical bundles che sono sintagmi nominali o incorporano frammenti di sintagmi	166
preposizionali	III.3.2.2 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi	168
	III.3.2.3 Lexical bundles verbali.....	169
	III.3.2.4 Lexical bundles che sono congiunzioni .....	170
<b>III.3.3</b>	<b>Comparazione del parlato e dello scritto .....</b>	<b>172</b>
<b>III.3.4</b>	<b>Variazione funzionale dei <i>lexical bundles</i> nel discorso accademico –</b>	
<b>microlingua della storia, della filosofia e della medicina del lavoro .....</b>		<b>176</b>
III.3.4.1	La funzione dei <i>lexical bundles</i> nel <i>corpus</i> di storia .....	178
III.3.4.1.1	La classificazione dei <i>bundles</i> in categorie funzionali.....	178
III.3.4.1.1.1	<b>I <i>lexical bundles</i> orientati alla ricerca</b> – aiutano lo scrivente a strutturare la	179
III.3.4.1.1.2	<b>I <i>lexical bundles</i> orientati al testo</b> – riguardano l'organizzazione del testo e	182
III.3.4.1.1.3	<b>I <i>lexical bundles</i> orientati al partecipante</b> – si concentrano sullo scrivente e	183
III.3.4.1.2	Risultati emersi dalla classificazione funzionale.....	183
III.3.4.2	La funzione dei <i>lexical bundles</i> nel <i>corpus</i> di filosofia .....	188
III.3.4.2.1	La classificazione dei <i>bundles</i> in categorie funzionali.....	188
III.3.4.2.1.1	<b>I <i>lexical bundles</i> orientati al testo</b> – riguardano l'organizzazione del testo e	189
III.3.4.2.1.2	<b>I <i>lexical bundles</i> orientati al partecipante</b> – si concentrano sullo scrivente e	190
III.3.4.2.1.3	<b>I <i>lexical bundles</i> orientati alla ricerca</b> – aiutano lo scrivente a strutturare la	191
III.3.4.2.1.4	Altri <i>bundles</i> .....	192
III.3.4.2.2	Risultati emersi dalla classificazione funzionale.....	192
III.3.4.3	La funzione dei <i>lexical bundles</i> nel <i>corpus</i> di medicina del lavoro .....	196
III.3.4.3.1	La classificazione dei <i>bundles</i> in categorie funzionali.....	196
III.3.4.3.1.1	<b>I <i>lexical bundles</i> orientati alla ricerca</b> – aiutano lo scrivente a strutturare la	197

III.3.4.3.1.2	<b>I lexical bundles orientati al testo</b> – riguardanti l'organizzazione del testo e il suo significato	202
III.3.4.3.1.3	<b>I lexical bundles orientati al partecipante</b> – si concentrano sullo scrivente e sul lettore del testo	204
III.3.4.3.1.4	Altri bundles .....	204
III.3.4.3.2	I risultati emersi dall'analisi funzionale .....	204
III.3.5	<i>Comparazione dei lexical bundles di tre ambiti scientifici – della storia, filosofia e medicina del lavoro</i> .....	209
III.4	CONCLUSIONI .....	213
	<b>APPENDICI</b> .....	<b>215</b>
	APPENDICE 1: <i>LEXICAL BUNDLES</i> DI 4 PAROLE IN SEQUENZA NEL <i>CORPUS</i> LIP .....	216
	APPENDICE 2: <i>LEXICAL BUNDLES</i> DI 4 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI STORIA.....	229
	APPENDICE 3: <i>LEXICAL BUNDLES</i> DI 4 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI FILOSOFIA.....	242
	APPENDICE 4: <i>LEXICAL BUNDLES</i> DI 4 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI MEDICINA DEL LAVORO .....	253
	APPENDICE 5: <i>LEXICAL BUNDLES</i> DI 4 PAROLE IN SEQUENZA NEL <i>CORPUS</i> SPECIALISTICO (STORIA, FILOSOFIA E MEDICINA DEL LAVORO).....	282
	APPENDICE 6: LE PRIME 40 POSIZIONI DEI <i>LEXICAL BUNDLES</i> DI 3, 4, 5 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI STORIA .....	290
	APPENDICE 7: LE PRIME 40 POSIZIONI DEI <i>LEXICAL BUNDLES</i> DI 3, 4, 5 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI FILOSOFIA .....	292
	APPENDICE 8: LE PRIME 40 POSIZIONI DEI <i>LEXICAL BUNDLES</i> DI 3, 4, 5 PAROLE IN SEQUENZA NEL <i>CORPUS</i> DI MEDICINA DEL LAVORO.....	294
	<b>BIBLIOGRAFIA</b> .....	<b>296</b>

## Introduzione

Questa tesi si pone come obiettivo principale quello di presentare la pervasività dei *lexical bundles* nel parlato italiano, basato sul *corpus* dell'*Italiano parlato* (LIP), e nel discorso accademico. La ricerca si propone di mostrare la variazione dei *lexical bundles* attraverso le modalità comunicative, parlato e scritto, e la loro variazione attraverso le diverse discipline scientifiche, allo scopo di mostrare la loro importanza nella formulazione del discorso.

Seguendo le ricerche svolte per la lingua inglese, il presente studio dei *lexical bundles* si basa sull'ipotesi di Biber (2006) che sostiene che i modelli ad alta frequenza non sono accidentali, ma non sono neanche di per sé esplicativi. Piuttosto, i modelli di frequenza *corpus-driven* comprendono fatti descrittivi che richiedono una spiegazione. I fatti che devono essere spiegati sono esistenze di sequenze di più parole che non rappresentano unità strutturalmente complete, svolgono, però, delle funzioni discorsive ben identificabili (Biber, 2006).

La prospettiva comparativa è stata messa in atto attraverso metodi e strumenti della linguistica dei *corpora*, raccogliendo da un lato il *Corpus dell'Italiano Parlato* (LIP), e dall'altro *corpora* di tre ambiti scientifici diversi, quali la storia, la filosofia e la medicina del lavoro.

Prendendo atto della pervasività dei *lexical bundles* nelle microlingue scientifico-professionali e della loro variazione attraverso le aree scientifiche, la presente ricerca sostiene che i *lexical bundles* debbano essere integrati tra le caratteristiche sintattico-lessicali delle microlingue in oggetto.

Dopo aver presentato l'idea portante di questo studio, si desidera ora illustrare la struttura in cui si articola la tesi. Nella prima parte, intitolata *Coordinate teoriche*, si riprendono le prospettive teoriche dello studio sulle combinazioni di parole, individuando in tale contesto la loro importanza. Nella seconda parte, intitolata *Analisi dei casi*, applicando il filone di ricerca sui *lexical bundles*, delineato nella prima parte, si è cercato di analizzare la struttura delle sequenze di 4 parole contigue emerse dall'estrazione automatica dei due *corpora*, il *corpus* LIP (*Lessico dell'Italiano Parlato*) e del *corpus* specialistico – il *corpus* di storia, filosofia e medicina del lavoro. I *lexical bundles* emersi dal primo *corpus*, il *corpus* dell'*Italiano Parlato* (LIP), sono stati

analizzati secondo la loro struttura, mentre dall'analisi del *corpus* specialistico si è inoltre proceduto all'esame funzionale di queste sequenze di parole.

Si anticipa di seguito il contenuto dei tre capitoli che compongono la tesi. Il primo capitolo, intitolato *Lo studio sulle combinazioni di parole e la loro importanza*, ripercorre la storia della teoria collocazionale, dai lavori pionieristici degli anni Trenta e Cinquanta fino agli orientamenti teorici attuali e alla teoria del *lexical priming*. Si procede, inoltre, con una panoramica storica della ricerca nella lingua italiana, per poter concentrarsi, nell'ultima parte, propriamente sui *lexical bundles*, analizzando le ricerche esistenti per poi proporre una classificazione strutturale e funzionale.

Il secondo capitolo, intitolato *Strumenti di analisi del linguaggio formulaico – la linguistica dei corpora*, cerca di delineare il contributo apportato dai computer e il processo di penetrazione dei *corpora* informatici nel lavoro linguistico. In questa prospettiva, si dà una definizione di *corpus*, per poi prendere in considerazione i due principali approcci all'analisi linguistica basata sui *corpora*: il *corpus-based*, ovvero la linguistica basata sui *corpora*; e l'approccio *corpus-driven*, vale a dire la linguistica diretta dai *corpora*. Nelle conclusioni al capitolo viene illustrato il metodo computazionale di estrazione dei *lexical bundles*, con una specificazione delle diverse fasi di analisi degli stessi.

Sulla base di quanto esaminato nella prima parte, il terzo capitolo intitolato *Analisi dei casi* si pone come obiettivo principale quello di presentare la pervasività dei *lexical bundles* sia nella conversazione quotidiana che nel discorso scientifico offrendo una panoramica delle strutture grammaticali e delle funzioni discorsive dei *lexical bundles*, allo scopo di mostrare la loro importanza nella formulazione del discorso. Analizzando due diversi registri – la conversazione e la prosa accademica – e concentrandosi propriamente sull'analisi di tre microlingue, relative alle discipline di storia, filosofia e di medicina del lavoro, si intende mostrare come la prospettiva di registro e di disciplina scientifica risultino importanti nello studio dei *pattern* formulaici.

**Parte prima**

**Coordinate teoriche**

## I Lo studio sulle combinazioni di parole e la loro importanza

La sintassi è per definizione una branca della linguistica che studia i diversi modi in cui le parole si uniscono tra loro per formare una frase e i vari modi in cui le frasi formano un periodo. Lo studio della struttura delle parole è invece oggetto di studio di un'altra branca della linguistica: la morfologia. Tale distinzione, tra parola da un lato e frase dall'altro, risale già all'antichità classica (*léxis* e *lógos* in Dionisio Trace), mentre il gruppo di parole come entità sintattica viene pienamente riconosciuta solo verso la fine del XIX secolo (Masini, 2009).

Farago Leonardi (1982) fa risalire già agli stoici, che per primi capirono la struttura semantica della lingua, l'idea che il lessico di una lingua non deve essere concepito come un inventario di parole. Il tema, oggi di grande attualità, è stato ripreso da molte e disparate scuole di pensiero: dai grammatici medioevali agli studiosi da Wilkins nel seicento fino all'antropologia funzionalista di Malinowski nel novecento (Farago Leonardi, 1982).

Le combinazioni di parole hanno cominciato ad essere oggetto di ricerca nei primi del novecento e sono oggetto quindi di una letteratura molto vasta. Le combinazioni di più parole sono state studiate sotto varie denominazioni. Nella tradizione linguistica inglese tali espressioni vengono denominate *collocations*, *lexical phrases*, *formulas*, *fixed expressions*, *prefabricated patterns*, *lexical co-occurrence*, *lexical bundles* (Biber ed altri, 2004), mentre nella tradizione linguistica italiana "collocazioni", "combinazioni polirematiche", "lessemi complessi", "unità lessicali complesse" o "espressioni multiparola" (Masini, 2009). Al di là dell'incertezza fra i ricercatori sul come definire queste espressioni multiparola, che possono essere riscontrate in un ampio raggio di termini ed etichette usati per descrivere e definire i diversi aspetti e le diverse categorie del linguaggio formulaico, si riscontra un certo disaccordo circa le loro caratteristiche specifiche e circa le metodologie atte a identificarle (Biber ed altri, 2004). Ne deduciamo pertanto che i confini di questo fenomeno linguistico non sono definibili in maniera assoluta e definitiva.

Il capitolo seguente si divide in sei parti. Le prime quattro sezioni ripercorrono la storia della teoria collocazionale, dai lavori pionieristici degli anni Trenta e Cinquanta fino agli orientamenti teorici attuali e alla teoria del *lexical priming*. Nella quinta parte viene illustrata la ricerca nella lingua italiana, mentre l'ultima parte si concentra



propriamente sui *lexical bundles*, analizzando le ricerche esistenti per poi proporre una classificazione strutturale e funzionale.

## **I.1 I precursori dello studio sulle combinazioni di parole: gli anni Trenta e gli anni Cinquanta**

Sebbene il primo ad utilizzare il termine scientifico *collocation* sia stato J.R. Firth, nel suo celebre *Papers in Linguistics 1939-1951* pubblicato nel 1957, negli anni cinquanta il concetto di collocazione non era qualcosa di nuovo. Secondo quanto esposto da Cortes, le origini del termine *collocation* in riferimento alla co-occorrenza lessicale risalgono al diciannovesimo secolo, quando l'*Oxford English Dictionary* (Simpson, Weiner, 1873; in Cortes, 2004) riporta che tutte le lingue usano in poesia una varietà più ampia di *collocazioni* che in prosa (Cortes, 2004).

Già nel 1924 Jespersen (in Cortes, 2004) aveva studiato le collocazioni e le espressioni fisse, distinguendo fra formule o unità formularie (*formular units*) ed espressioni libere. Jespersen (1924) porta ad esempio la frase formulaica *How do you do?* sottolineando che l'espressione viene sentita, considerata e anche imparata come un'intera unità. Da contro, le *free expressions* devono essere create dal nulla e caso per caso dal parlante, che seleziona le parole che meglio si inseriscono in una data situazione, rispettando le regole grammaticali e in tal modo coinvolgendo un'altro tipo di attività mentale (Jespersen, 1924). Il linguista sottolinea che il linguaggio consiste in una serie di azioni abituali, la maggior parte delle quali viene determinata da ciò che il parlante ha fatto precedentemente in condizioni simili, che è a sua volta determinato da ciò che egli aveva sentito dire d'abitudine da altri parlanti. Jespersen offre con il suo lavoro un primo spunto di riflessione circa le formule fisse da un lato e le espressioni libere dall'altro.

Uno dei pionieri dello studio sulle collocazioni fu Harold Edward Palmer. Palmer notò che, nel campo della linguistica, il termine *collocation* era già in uso come termine tecnico ma anche come espressione piuttosto vaga. Il linguista definì la collocazione

come "a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts"<sup>1</sup> (Palmer, 1933; in Cowie, 1999: 54).

Palmer distinse da un lato le collocazioni e dall'altro le combinazioni libere che sono manifestazioni del sistema e possono essere unite "by dint of the application of the commonest and best-known rules of grammar"<sup>2</sup> (Palmer, 1933: 5; in Cowie, 1999: 54). Nel 1933, Palmer fondò l'*Institute for Research in English Teaching* a Tokyo, dove si trovava con un incarico di insegnamento. L'obiettivo dell'istituto era di monitorare e migliorare i metodi di insegnamento dell'inglese come L2. Proprio ad inizio secolo, la diffusione della lingua inglese come L2 iniziò a spingere molti parlanti nativi a lasciare il paese per trovare lavoro come insegnanti di inglese all'estero (Cowie, 1999). Durante le sue ricerche, Palmer (1933: 13; in Cowie, 1999: 52-53) aveva rilevato che

"[...] that it is not so much the words of English nor the grammar of English that makes English difficult, but that vague and undefined obstacle to progress in the learning of English consists for the most part in the existence of so many odd comings-together-of-words"<sup>3</sup>.

Il suo lavoro *Second Interim report on English collocations* (Palmer, 1966) consiste in un'accurata classificazione delle combinazioni di parole presenti nella lingua inglese. Si tratta di un'opera fondamentale, che mette in evidenza sia l'importanza di particolari co-occorrenze di elementi lessicali, che pongono particolari difficoltà nell'apprendimento dell'inglese, sia la loro diffusione d'uso. Quest'opera tratta anche delle implicazioni delle co-occorrenze nel processo di insegnamento e apprendimento di una lingua straniera. Come notato da Cowie (1999), il lavoro di Palmer avrebbe avuto un impatto profondo sulla composizione dei dizionari per l'apprendimento dell'inglese come lingua straniera. Sempre Cowie nota che i concetti chiave delle opere di Palmer sono ancora oggi validi e ampiamente usati nell'approccio metodologico per l'insegnamento di lingue straniere. Si riconosce infatti che le combinazioni di parole giocano una parte cruciale nell'apprendimento di una lingua straniera (Pawley, Syder,

---

<sup>1</sup>Traduzione: "una successione di due o più parole che devono essere apprese come un'unità integrale e non connesse l'una all'altra a partire dai propri componenti".

<sup>2</sup>Traduzione: "per mezzo dell'applicazione delle regole grammaticali più comuni e meglio conosciute".

<sup>3</sup>Traduzione: "che non sono tanto le parole dell'inglese o la grammatica dell'inglese a rendere l'inglese difficile, ma che l'ostacolo vago e indefinito al progresso nell'apprendimento dell'inglese consiste perlopiù nell'esistenza di così tanti e così strani raggruppamenti di parole".

1983; in Cowie, 1999) e che sono parte integrante di tutti i programmi di apprendimento (Cowie, 1999).

L'opera di Palmer risulta quindi importante soprattutto da un punto di vista glottodidattico. Le collocazioni – *comings-together of words* o *collocations*, come il linguista le definisce formalmente – venivano analizzate nei termini della difficoltà di apprendimento linguistico e dell'approccio necessario da adottare per padroneggiarle.

In *A Grammar of English Words*, pubblicato nel 1938, l'autore ha per la prima volta potuto utilizzare il materiale classificato nel suo precedente lavoro, nell'*Interim Report* (1933), e dimostrare in modo pratico l'importanza del lavoro svolto, l'importanza delle collocazioni sia nell'apprendimento che nell'insegnamento dell'inglese (Cowie, 1999). Il lavoro include una serie di informazioni sulla sintassi, sul significato e sulla fraseologia di circa 1000 combinazioni di parole.

All'esperienza pionieristica di Palmer si aggiunge quella di Charles Bally, discepolo di de Saussure. Già de Saussure (1978), nella sua opera *Cours de linguistique générale*, accennava ad una distinzione tra *groupements de mots*, raggruppamenti che formano il lessico, e *relations sintagmaticques et associatives*, i legami responsabili della combinazione delle unità lessicali, che corrispondono a due forme della nostra attività mentale. Il rapporto sintagmatico è il rapporto fondato sul carattere lineare della lingua, che si compone sempre di due o più unità consecutive e che unisce i termini *in praesentia*. A differenza di questo, il rapporto associativo fa sorgere inconsciamente una serie di altre parole che fanno parte del patrimonio linguistico di ciascun individuo, costruendo in questo modo il rapporto *in absentia* (De Saussure, 1978).

Bally, nella sua opera *Traité de stylistique française* del 1909, mette in evidenza le combinazioni di parole, avvicinandosi al fenomeno lessicale definito da De Saussure. Bally propone due categorie di combinazioni di parole: da un lato *groupements usuels* e dall'altro *les séries phraséologique*. Il linguista distingue quindi due tipologie di combinazioni di parole, le unità fraseologiche vere e proprie o le espressioni fisse da un canto, e le serie fraseologiche o le combinazioni usuali dall'altro, offrendo in questo modo una descrizione sistematica di combinazioni libere in francese o *locutions phraséologiques*.

Secondo questa teoria, queste tipologie vengono definite sulla base della loro frequenza o diffusione nell'uso evidenziando l'apparente autonomia dei componenti. Bally descrive le combinazioni di questo tipo come segue:

[i]l y a une série ou groupement usuel lorsque les éléments du group conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du "déjà vu"<sup>4</sup> (Bally, 1951: 70).

Come delineato nella citazione sopra riportata, certe combinazioni di parole che hanno la tendenza di co-occorrere danno l'impressione di *déjà vu* perché vengono riconosciute come blocchi di parole abituali. A tal proposito Bally analizza la co-occorrenza *gravement malade* in cui le due componenti possiedono un loro significato in isolamento, formando però una *série usuelle*. Nell'analisi dell'uso degli avverbi che indicano il grado di intensità di una malattia la ricerca ha mostrato una predilezione per l'avverbio *gravement*, ma non per l'avverbio *grièvement*, che, pur avendo la stessa base etimologica, compare in diverse *séries usuelle*. L'avverbio *grièvement* viene usato con l'aggettivo *blessé*, con cui forma una *série usuelle*. Sulla scorta dell'esempio citato ne deduciamo che le combinazioni usuali sono espressioni che incorporano sempre lo stesso lessema principale, il quale seleziona il suo modificatore più usuale presentandosi non come una scelta obbligatoria, ma come una scelta preferenziale e non unica. La sostituzione del modificatore non produce un cambiamento semantico radicale, bensì determina quanto una combinazione di parole sia familiare o meno (Masini, 2009).

Il contributo di Bally (1951), oltre ad essere importante perché introduce l'idea di *déjà vu*, ovvero l'esistenza della preferenza per certe combinazioni di parole, risulta importante anche dal punto di vista glottodidattico. Secondo Bally l'apprendimento di una lingua e di conseguenza la produzione linguistica avvengono attraverso il meccanismo di associazione e raggruppamento di parole, nel processo della quale gioca un ruolo assai importante la componente affettiva del linguaggio, ovvero l'intervento del sentimento che agevolerebbe l'apprendimento linguistico. Ogni parola richiama molteplici associazioni, tra le quali alcune svaniscono subito, mentre altre vengono fissate nella memoria per diverse ragioni, una delle quali è anche l'alto livello di

---

<sup>4</sup>Traduzione: "c'è una serie o un raggruppamento normale in cui gli elementi del gruppo mantengono la propria autonomia pur lasciando vedere un'affinità evidente che li avvicina, di modo che l'insieme presenta dei contorni precisi e dà l'impressione del 'déjà vu'."

coesione interna. Quando ciò accade vengono create delle co-occorrenze lessicali. Le co-occorrenze lessicali possono essere solo temporanee o *passengers* oppure possono, attraverso la ripetizione e l'uso costante, venire percepite come abituali o *usuels* a tal punto da formare delle unità indissolubili, nelle quali i singoli componenti perdono il loro significato in isolamento, sia sintattico che semantico, avendo un senso compiuto solo nella combinazione.

Sul carattere abituale dei cosiddetti *groupements usuels* proposto da Bally viene costituita la teoria firthiana sulle collocazioni. Fu proprio J.R. Firth (1951) il primo a rendere popolare il termine *collocation* e lo slogan "You shall know a word by the company it keeps"<sup>5</sup>, con cui sottolineava l'importanza di conoscere le modalità attraverso le quali le parole si combinano tra loro. Il fenomeno delle collocazioni acquista vera rilevanza solo nel 1957 con la pubblicazione di *Papers in Linguistics, 1939-1951*, il lavoro principale di Firth, la figura di spicco della scuola britannica contestualista. Vari sono i contributi scientifici legati alla scuola firthiana, quali la fonologia prosodica, l'importanza del contesto situazionale, la distinzione tra struttura e sistema, e il contributo più importante per questa ricerca, la nozione di collocazione al livello lessicale. Con il termine collocazione, termine ripreso da Palmer, Firth si riferisce all'associazione abituale di una parola con altre determinate parole, entro gli enunciati di una lingua (Farago Leonardi, 1982).

Secondo Firth, infatti, il lessico di una lingua non deve essere concepito soltanto come un inventario di parole, ma anche come un insieme di possibilità di collocazioni, determinate non aprioristicamente dal sistema, ma dall'uso. Nel 1968 Firth scrive che la collocazione dovrebbe essere considerata come "an order of *mutual expectancy*. The words are mutually expectant and mutually prehended"<sup>6</sup> (Firth, 1968: 181). Secondo Hoey (1991), le collocazioni sono unità prefabbricate in cui una parola-chiave (*node*) determina la scelta del collocato – e viceversa – per effetti di *priming* lessicale, che sarà meglio analizzato nelle prossime sezioni.

A questo punto bisogna menzionare lo studio di Zgusta (1967), il quale si presenta particolarmente importante perché pone al centro dell'interesse linguistico non le singole parole, ma le unità di più parole. Il linguista fa, in base a nove criteri da lui

---

<sup>5</sup>Traduzione: "Si dovrebbe capire una parola dalla sua compagnia (dalle sue preferenze combinatorie)".

<sup>6</sup>Traduzione: "un ordine di reciproche aspettative. Le parole sono mutualmente in attesa e si mettono reciprocamente in relazione."

identificati, una prima distinzione per la lingua inglese di espressioni multiparola (*multiword lexical units*) da un lato e combinazioni libere di parole (*free combinations of words*) dall'altro. Lo studioso conclude che, nella linguistica, le unità di più parole dovrebbero andare a pari passo con le singole parole e che dovrebbero essere prese in considerazione nella creazione dei dizionari.

## I.2 Gli sviluppi della teoria firthiana

Alle intuizioni di Firth fecero seguito i lavori dei suoi allievi, i cosiddetti neo-firthiani della scuola sistemica britannica, rappresentata principalmente dai linguisti inglesi Halliday e Sinclair. Questi studiosi hanno sottolineato quell'aspetto della collocazione legato più all'analisi statistica e qualitativa, concentrandosi sul carattere di frequenza d'uso delle combinazioni.

All'interno della teoria sulla coesione lessicale, Halliday definisce la *collocation* come la co-occorrenza di unità lessicali che si manifestano regolarmente nella lingua, considerando la collocazione con un approccio puramente statistico (Halliday, Hassan, 1976: 285-286).

"There is always the possibility of cohesion between any pairs of lexical items which are in some way associated with each other in the language. So we will find a very marked and cohesive effect deriving from the occurrence in proximity with each other of pairs such as the following, whose meaning relation is not easy to classify in systematic semantic terms: *laugh ... joke, blade ... sharp* [...] The cohesive effect of such pairs depends not so much on systematic semantic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one another. In general, any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will generate a cohesive force if they occur in adjacent sentences<sup>7</sup>".

---

<sup>7</sup>Traduzione: "C'è sempre una possibilità di coesione fra qualsiasi paio di elementi lessicali che siano in qualche modo associati l'uno all'altro nel linguaggio. Così riscontreremo un effetto molto marcato e coesivo prodotto dall'occorrenza in prossimità di alcune coppie che, come le seguenti, hanno una relazione non facilmente classificabile in termini semantici: *laugh...joke, blade...sharp*[...]. L'effetto coesivo di queste coppie non dipende tanto da una relazione semantica sistematica, quanto dalla loro tendenza a condividere lo stesso ambiente lessicale, a ricorrere in COLLOCAZIONE l'una con l'altra. In

Dalla citazione di Halliday e Hassan possiamo dedurre che il fenomeno della collocazione delinea una co-apparizione frequente e lineare di lessemi nel discorso, raccogliendo sotto la stessa denominazione un gruppo abbastanza eterogeneo, dai sinonimi o quasi sinonimi (*climb ... ascent*), antonimi (*like ... hate*), iponimi dello stesso iperonimo (*elm ... tree*). Secondo i linguisti l'effetto di coesione non è limitato solo a due parole, bensì anche alle catene di parole più lunghe. I *pattern* che i linguisti chiamano *long cohesive chains* o *chains of collocational cohesion* sono composti da termini semanticamente legati tra loro. I linguisti indicano come esempi di *cohesive chains* le espressioni *candle ... flame ... flicker, hair ... comb ... curl ... wave, poetry ... reader ... writer ... style*, specificando che i *pattern* appaiono in una o in più frasi contestualmente legate, anche non presentando nessun legame sintattico. La coesione lessicale degli elementi rappresenta la forma singola più importante di legame coesivo, siccome ogni elemento lessicale è dotato di una storia contestuale. Il testo quindi fornisce il contesto per la creazione e l'interpretazione delle relazioni lessicali, proprio come le relazioni lessicali provvedono a creare la struttura del testo. "This environment determines the 'instantial meaning', or text meaning, of the item, a meaning which is unique to each specific instance"<sup>8</sup> (Halliday, Hasan, 1976: 289). I linguisti vedono la collocazione come un dispositivo coesivo e la descrivono come termine di copertura a partire dal tipo di coesione che risulta dalla co-occorrenza delle unità lessicali, che tendono a occorrere in ambienti simili (Hoey, 2005).

Secondo Hoey (2005), i linguisti non si riferiscono alla co-occorrenza regolare di parole: l'associazione alla quale si riferiscono deve essere di carattere psicologico, laddove le parole sono associate regolarmente nella mente del parlante, a causa del modo in cui vengono regolarmente rilevate in contesti testuali simili. Si tratta di un concetto fondamentale, già menzionato da altri studiosi (Jespersen, 1924; Bally, 1951).

Gli studi sulla coesione lessicale in lingua inglese saranno successivamente approfonditi da Hoey (1991: 10), secondo il quale:

---

generale, qualsiasi coppia di elementi lessicali che presentino delle tendenze di collocazione simili – cioè, che tendano ad apparire in contesti simili – genereranno una forza coesiva qualora occorran in frasi adiacenti".

<sup>8</sup>Traduzione: "Quest'ambiente determina 'il significato istantaneo', ovvero il significato testuale dell'elemento, un significato unico in ogni singola situazione".

"Lexical cohesion is the only type of cohesion that regularly forms multiple relationships [...] If this is taken into account, lexical cohesion becomes the dominant mode of creating texture. In other words, the study of cohesion in text is to a considerable degree the study of patterns of lexis in text<sup>9</sup>".

Hoey (2005) ha approfondito lo studio sui *pattern* lessicali nel suo lavoro, *Lexical priming: A new theory of words and language*. In questo lavoro, le frasi vengono concepite come *bundles* interconnessi; in maniera simile, le parole vengono "selezionate" mentalmente per essere usate insieme ad altre parole, grazie al fatto che le abbiamo esperite tramite associazioni frequenti. Si tratta quindi di una visione della grammatica radicalmente nuova (Hyland, 2008a), che sarà oggetto di discussione nelle sezioni seguenti.

Similmente a Halliday, la frequenza è considerata un carattere essenziale della collocazione anche da Sinclair, una delle figure più innovative e influenti della linguistica dei *corpora* informatici. Negli anni Settanta J.M. Sinclair studiò la co-occorrenza lineare di lessemi attraverso l'indagine statistica di vasti *corpora* testuali per mezzo di calcolatori elettronici, dando in questo modo un forte impulso allo sviluppo della linguistica computazionale e ponendo in questo modo lo studio del significato (*meaning*) al centro dell'interesse. A Sinclair e Jones (1996), che hanno prodotto la prima e fondamentale analisi computazionale della collocazione, è riconosciuto il merito di aver introdotto la nozione di frequenza – un carattere essenziale della collocazione – e quella di *node*, il nucleo, per specificare l'unità lessicale di cui si studia il comportamento nelle collocazioni. Il *corpus* che hanno analizzato era di sole 147.000 parole, ma era comunque abbastanza grande da poter dare una definizione circa l'ampiezza necessaria a identificare una collocazione. Secondo questi autori, quest'ampiezza arriva fino a quattro parole da entrambi i lati del nodo. I due studiosi considerano le collocazioni semantiche, quelle costituite da due unità lessicali, la cui frequenza di co-occorrenza è maggiore rispetto alla frequenza dei costituenti isolati e dalla lunghezza del testo in cui sono inseriti.

---

<sup>9</sup>Traduzione: "La coesione lessicale è l'unico tipo di coesione che forma regolarmente delle relazioni multiple. Se si prende in considerazione questo fatto, la coesione lessicale diventa il modo principale per creare consistenza. In altre parole, lo studio della coesione nel testo è in larga parte lo studio dei modelli del lessico nel testo".



Nel suo lavoro del 1991 Sinclair perfeziona la teoria collocazionale cercando di spiegare il funzionamento del linguaggio come un insieme di possibilità basate su due principi conflittuali e di come i significati vengono creati nel testo: *open-choice principle*, il principio di scelta libera, e *idiom principle*, il principio dell'idioma.

Il principio di scelta libera, chiamato anche modello *slot-and filler*, fa riferimento alla grammatica, le cui limitazioni riguardano la grammaticalità dell'enunciato. Il principio dell'idioma, da contro, agisce al livello lessicale. Attraverso questi due principi viene ancora una volta sottolineata l'importanza del criterio di frequenza come caratteristica fondamentale nella definizione delle co-occorrenze di parole. Il principio *open-choice*

"[...] is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, phrase, or clause), a large range of choice opens up and the only restraint is grammaticalness. This is probably the normal way of seeing and describing language. It is often called a 'slot-and-filler' model, envisaging texts as a series of slots which have to be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur<sup>10</sup>" (Sinclair, 1991: 109).

È chiaro che la scelta delle parole in un atto linguistico non viene operata secondo un ordine causale e in maniera completamente libera. Si basano su questo principio le grammatiche che presuppongono che gli *slots* di una frase vengano riempiti da determinate parole in maniera più o meno casuale (Sinclair, 1991). Se ci basassimo soltanto sul principio *open-choice*, infatti, non saremmo in grado di produrre un testo corretto. A questo riguardo, Sinclair specifica che l'*idiom principle* è una manifestazione del lessico, sul quale la lingua rende disponibili un'ampia serie di *semi-preconstructed phrases* o frasi semi-precostruite, di *chunks* o stringhe prefabbricate, a cui appartiene anche la collocazione o la locuzione idiomatica.

---

<sup>10</sup>Traduzione: "[...] è un modo di vedere il testo di linguaggio come risultato di un numero molto ampio di scelte complesse. A ogni punto in cui si completa un'unità (una parola, una frase o una clausola), si apre una vasta gamma di scelte e l'unico vincolo è la grammaticità. Probabilmente questo è il modo normale di vedere e descrivere il linguaggio. Spesso viene chiamato modello "slot-and-filler", perché rappresenta i testi come una serie di spazi da riempire con un lessico che soddisfa i vincoli locali. In ogni spazio, virtualmente, può occorrere qualsiasi parola".

"The principle of idiom is that 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. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. However it arises, it has been relegated to an inferior position in most recurrent linguistics, because it does not fit the open-choice model<sup>11</sup>" (Sinclair, 1991: 110).

Sinclair propone come esempio dell'*idiom principle* l'espressione *of course*, che, anche se costituita da due elementi indipendenti, opera come una singola parola, i cui due componenti perdono le loro principali caratteristiche semantiche e grammaticali. Come si evince dal passaggio citato, secondo Sinclair l'*idiom* può riflettere il ripetersi di situazioni simili nell'esperienza umana, perché la natura del mondo attorno a noi viene riflessa nell'organizzazione del linguaggio. In altre parole, le cose che avvengono assieme nel mondo fisico hanno una forte probabilità di occorrere insieme anche nella lingua e in tal modo contribuiscono alla non casualità.

Inoltre, l'*idiom principle* può illustrare la tendenza naturale della lingua al principio di economia, poiché tende a minimizzare lo sforzo mnemonico, e può essere in parte motivato dai condizionamenti temporali nell'ambito di una conversazione, vale a dire dall'uso che i parlanti fanno della lingua.

Secondo Sinclair, la pervasività dell'*idiom* è evidente, nonostante il fatto che la linguistica tradizionale ne abbia a lungo sottovalutato l'importanza, trascurandone lo studio e relegandolo ad un livello inferiore perché non rientra nei meccanismi del modello *open choice* non può essere quindi analizzato seguendo le regole grammaticali. Il principio dell'idioma si riferisce alla "tendenza fraseologica" del linguaggio, poiché le parole non si presentano in maniera isolata, ma "go together and make meanings by

---

<sup>11</sup> Traduzione: "Il principio dell'idioma è che l'utilizzatore di un linguaggio ha a sua disposizione un ampio numero di frasi semi-precostruite che costituiscono scelte singole, anche se possono sembrare analizzabili come segmenti. In certa misura, questo può riflettere la ricorrenza di situazioni simili nelle questioni umane; può illustrare una tendenza naturale all'economia dello sforzo; o può essere in parte motivato dalle esigenze della conversazione in tempo reale. Comunque questo accada, questo principio è stato relegato a una posizione inferiore nella maggior parte della linguistica ricorrente, perché non è adatto al modello a scelta aperta".

their combinations<sup>12</sup>" (Sinclair ed altri, 2004: 29). Sulla scorta dell'esempio *of course*, Sinclair (1991) indica che questa frase opera effettivamente come una parola unica e che lo spazio fra le due parole, che è un artificio strutturale, potrebbe scomparire nel tempo, così come è accaduto con *maybe, anyway and another* (Sinclair, 1991). Secondo questa interpretazione, avrebbe senso aggiungere frasi come *of course* alla lista dei composti, come *cupboard*, i cui elementi hanno perso la propria identità semantica. Secondo Sinclair, centinaia di frasi simili possono essere analizzate con questo metodo: idiomi, clichés, proverbi, termini tecnici, espressioni gergali, verbi frasali – tutte circostanze in cui una decisione porta all'uso di più di una parola nel testo e dove tutto può essere ben espresso tramite un enunciato relativamente semplice. Sinclair (1991) sostiene che il modello *default* della costruzione di un testo "normale", includendo i registri speciali come quelli microlinguistici, è *idiom principle*. Secondo Sinclair (ibidem) i due principi si escludono a vicenda e sono incompatibili.

Questa linea teorica supporta l'uso di programmi informatici per l'estrazione di collocazioni, intese come "the occurrence of two or more words within a short space of each other in a text<sup>13</sup>" (Sinclair, 1991: 170), a significare una sequenza frequente di parole.

Lo studio delle co-occorrenze linguistiche basato sullo strumento statistico rappresenta una visione nuova della lingua e della grammatica. I *corpora* di grandi dimensioni permettono di rivalutare le convenzioni grammaticali tradizionali e aprono nuove possibilità di studiare certe unità lessicali che altrimenti andrebbero perse. Prima dello sviluppo della linguistica dei *corpora*, questo tipo di dati non erano a disposizione e i linguisti potevano confidare solo nella loro capacità di un'analisi testuale dettagliata e nella loro capacità intuitiva di cogliere le forme tipiche del linguaggio sia nella vita quotidiana che professionale (Cortes, 2008).

Sinclair (2004) reitera questa tesi anche nei lavori successivi, dove si sottolinea che se la co-occorrenza di due parole è statisticamente significativa, questo parametro indica che la probabilità di accidentalità è minima e che certamente esiste una ragione che spiega questa co-occorrenza.

---

<sup>12</sup>Traduzione: "occorrono insieme e traggono significato dalle proprie combinazioni".

<sup>13</sup>Traduzione: "l'occorrenza di due o più parole entro un breve spazio di distanza l'una dall'altra in un testo".

Sinclair (1991) asserisce inoltre che i *corpora* di piccole dimensioni hanno solo confermato le tesi più affermate circa la grammatica inglese, mentre i nuovi dati risultanti da grandi *corpora* testuali suggeriscono che tali generalizzazioni grammaticali non si basano su principi rigidi, ma sono piuttosto il risultato di un'accumulazione di modelli costituiti da centinaia di parole e frasi individuali. A proposito della dimensione del *corpus* Sinclair (2004: xxii) precisa che

"[...] The shorter the *corpus*, the poorer, the less reliable the results of statistical analyses will be. In order to get all the 'good' collocations you have to take a lot what seems to be rubbish into the bargain. Setting the significance level too high means you get only a few collocations. With a much larger *corpus*, the patterning becomes clearer<sup>14</sup>".

Emerge con chiarezza che la lingua appare assai diversa quando viene esaminata nel suo insieme (Sinclair, 1982). Le nuove opportunità offerte dall'informatizzazione dei *corpora* consistono nel poter valutare delle istanze effettive e nel poter selezionare le più tipiche fra esse. Una serie completa di istanze tipiche dovrebbe esemplificare i modelli strutturali dominanti del linguaggio senza far ricorso all'astrazione né alla generalizzazione (Sinclair ed altri, 2004). Ogni massa di istanze contiene infatti solo un piccolo elemento di tipicità, ma un piccolo numero di masse deve contenere molte caratteristiche tipiche (Sinclair ed altri, 2004). Sinclair (2004) asserisce che, paradossalmente, gli esempi a più alta tipicità spesso non sono molto comuni e devono essere identificati tramite metodi statistici in un *corpus* vasto, anche perché l'uso linguistico è assai specifico ed è difficile trovare molti esempi effettivamente rappresentativi. Questo fatto vale sia per la grammatica che per il lessico, perché la grammatica non è costituita semplicemente da modelli di parole grammaticalmente comuni, ma si basa sull'intero vocabolario di una lingua. Sinclair (1991: 108) osserva che

---

<sup>14</sup>Traduzione: "[...] Quanto più breve è il *corpus*, tanto più poveri e meno affidabili saranno i risultati dell'analisi statistica. Al fine di ottenere tutte le collocazioni "buone" si dovrà includere nell'affare anche molte cose che sembrano scarti. Stabilire il livello di significatività ad un valore troppo alto significa ottenere solo poche collocazioni. Con un *corpus* molto più ampio, le tendenze – il *patterning* – diventano più chiare".

"By far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns. Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text. This is totally obscured by the procedures of conventional grammar<sup>15</sup>".

Come si evince dal passaggio sopra citato, la grammatica è il risultato di raggruppamenti collocazionali ripetuti. Sinclair guarda al lessico come sistematicamente strutturato tramite modelli di utilizzo ripetuti, invece di vedere le scelte lessicali come costrette dagli *slots* che la grammatica mette a disposizione (Hyland, 2008a). Una serie di studi sui *corpora* per la lingua inglese (Kjellmer, 1984; Baayen e Lieber, 1991; Altenberg, 1993; Barkema, 1993; in Wray, Perkins, 2000), ha mostrato che il *patterning* delle parole e delle frasi nel linguaggio comune manifesta meno variabilità di quanto potrebbe essere prevista solo sulla base della grammatica e del lessico (Wray, Perkins, 2000). Quest'osservazione conferma la riflessione di Sinclair circa la grammatica come prodotto di unità multiparola ripetute. Kjellmer (1994: IX-XI) sostiene che "there is no doubt that natural language has a certain block-like character. Words tend to occur in the same clusters again and again. [...] A living language is largely made up of recurring clusters of words<sup>16</sup>".

È stato anche notato che la pervasività delle combinazioni di parole nel linguaggio comune ha condotto linguisti come Sinclair e Hoey a proporre radicali nuove teorie sul linguaggio (Hyland, 2008a).

---

<sup>15</sup>Traduzione: "La grande maggioranza del testo è costituita da occorrenze di parole comuni in modelli comuni, o da leggere varianti su quei modelli comuni. Gran parte delle parole quotidiane non hanno un significato indipendente, ma sono componenti di un vasto repertorio di modelli multiparola che compongono un testo. Questo fatto viene posto completamente in ombra dalle procedure della grammatica convenzionale".

<sup>16</sup>Traduzione: "non esiste alcun dubbio sul fatto che il linguaggio naturale ha un certo carattere "a mattoni". Le parole tendono a ricorrere negli stessi gruppi ripetutamente. [...] una lingua viva è fatta in gran parte di blocchi ricorrenti di parole".

### 1.3 La teoria del *lexical priming*

Michael Hoey è un altro studioso di tradizione britannica contestualista che è stato influenzato dalle idee di John Sinclair, sulla base delle quali ha sviluppato le sue teorie. Nel suo lavoro *Lexical Priming*, pubblicato nel 2005, ha esposto "a new theory of the lexicon, which amounts to a new theory of language"<sup>17</sup> (Hoey, 2005: 1), mettendo la componente lessicale al centro dell'interesse. Il linguista afferma: "The theory reverses the roles of lexis and grammar, arguing that lexis is complexly and systematically structured and that grammar is an outcome of this lexical structure"<sup>18</sup> (ibidem). Le collocazioni vengono infatti interpretate come un fattore chiave nella naturalezza del linguaggio, come una proprietà del linguaggio per mezzo della quale due o più parole sembrano apparire frequentemente in compagnia l'una dell'altra. Hoey si spinge oltre con asserzioni psicolinguistiche, quando definisce la collocazione come una associazione psicologica fra parole, composta da un massimo di quattro parole, che si evidenziano dalla loro occorrenza contigua nei *corpora*, in una frequenza tale da essere inspiegabile in termini di distribuzione casuale (Römer, 2009). Un'osservazione fondamentale nella teoria di Hoey è che "every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word"<sup>19</sup> (Hoey, 2005: 13). Secondo Hoey, le frasi sono abitualmente composte da *bundles* interconnessi mentre le parole sono selezionate mentalmente per l'utilizzo assieme ad altre parole esperite assieme in associazioni frequenti. Tutto ciò che sappiamo circa una parola è il risultato dei nostri contatti con essa; le enunciazioni che scegliamo sono informate dal modo in cui le riscontriamo regolarmente in testi simili e dipendono completamente dal nostro esperirle. In questa visione, il *lexical priming* è, secondo Hoey (2005), un ponte fra quella che secondo Chomsky è la "lingua esternalizzata" (*E – language*) e la lingua interiorizzata (*I-language*), ovvero la lingua che risiede nella mente del parlante. Sempre Hoey nota che le affermazioni circa i modi in cui la lingua viene acquisita e utilizzata in situazioni specifiche sono costrette quindi dal dominio e/o dal genere, poiché si prende in considerazione l'identità del parlante o dello scrivente, ciò di cui si

---

<sup>17</sup> Traduzione: "una nuova teoria del lessico, che ammonta ad una nuova teoria del linguaggio".

<sup>18</sup> Traduzione: "La teoria inverte i ruoli del lessico e della grammatica, affermando che il lessico è completamente e sistematicamente strutturato e la grammatica è un esito di questa struttura lessicale".

<sup>19</sup> Traduzione: "ogni parola è, per effetti del *priming* lessicale, pronta per essere usata in discorso – tale uso è il risultato degli effetti cumulativi degli incontri individuali con la parola".

sta parlando o scrivendo e a che genere si sta partecipando. In maniera automatica, quindi, il parlante seleziona e adotta i modelli di utilizzo e apprende le strutture linguistiche nei quali queste strutture si presentano. Secondo Römer (2009), questo implica che gli effetti di *priming* sono specifici per ogni registro e che bisogna studiare separatamente le ampie collezioni di testi di diverse tipologie, perché le combinazioni di parole sono usate in maniera differente a seconda del registro. Come abbiamo già accennato in precedenza, Hoey sostiene un'altra interpretazione della lingua, che si basa sugli elementi del vocabolario e poi analizza le associazioni preferenziali, ponendo l'accento su ciò che in una lingua è naturale.

#### **I.4 L'importanza del linguaggio formulaico**

Cortes (2008) sostiene che i primi studi sulle combinazioni di parole ricorrenti si basavano più su criteri intuitivi che su prospettive empiriche o quantitative. È dagli anni Settanta che gli studi quantitativi e basati sui *corpora* hanno acquisito una certa importanza. Si è già accennato al fatto che i ricercatori non concordano sulla definizione di cosa deve essere definito come linguaggio formulaico e cosa non può esserlo. Secondo alcuni, la formulaicità è un termine ampio che include un'ampia varietà di stringhe di parole che hanno la funzione di scorciatoie funzionali (Wray, Perkins, 2000), poiché piuttosto che essere generate nuovamente a ogni occasione, esse vengono archiviate e recuperate per intero o come catene automatiche dalla memoria a lungo termine al momento dell'utilizzo (Pawley, Syder, 2000; in Wray, Perkins, 2000).

Wray e Perkins (2000: 1) definiscono la sequenza formulaica come:

"a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar<sup>20</sup>".

---

<sup>20</sup> Traduzione: "una sequenza, continua o discontinua, di parole o di altri elementi portatori di significato, che è o appare prefabbricata: cioè, archiviata e recuperata tutta intera dalla memoria al momento di utilizzo, piuttosto che soggetta a generazione o analisi dalla grammatica della lingua".

Per gli autori la sequenza formulaica include diverse sottocategorie, quali proverbi, collocazioni, espressioni idiomatiche, formule del discorso, che possono essere classificate in completamente fisse – come, ad esempio, gli idiomi, le collocazioni e le espressioni fisse – e in composizionali, come ad esempio le frasi semi-precostituite, i modelli, i costruttori di frase (Nekrasova, 2009). I linguisti Pawley e Syder, ripresi da Wrey e Perkins (2000), parlano del linguaggio formulaico includendo sia le stringhe strettamente idiomatiche e immutabili, che quelle trasparenti e flessibili. D'altro canto, Sinclair (1991) include nel linguaggio formulaico anche l'ampia serie di collocazioni lessicali "semplici", i cui modelli sono, da un punto di vista grammaticale, sia notevoli sia possono essere costruiti come un puzzle. Ciò rende difficile se non impossibile discriminare circa la misura in cui un linguaggio è formulaico o meno. Secondo Altenberg (1990), attenendosi al criterio di Sinclair, circa il 70% del linguaggio adulto è formulaico. D'altro canto, lo studioso Howarth (1998) suggerisce che il linguaggio è prefabbricato in una percentuale che varia da 37% a circa la metà del linguaggio, come sostenuto anche dagli studi svolti da Cowie (1991; in Howarth 1998) circa il linguaggio giornalistico. Nell'analisi di più di 5.000 combinazioni di parole di un *corpus* scritto di 240.000 parole, Howarth (1998) ha scoperto che più di un terzo delle espressioni composte da nomi o verbi ha la struttura formulaica. Nel suo lavoro precedente Howarth (1996) aveva analizzato la prosa accademica e scoperto la stessa percentuale della formulaicità del linguaggio, un dato che si presenta importante per la presente analisi. I risultati menzionati sono supportati anche dall'analisi svolta da Erman e Warren (2000) sul *London Lund Corpus of Spoken English*, la quale ha mostrato che in media più del 50% di un testo consiste delle costruzioni prefabbricate.

Questi dati sono ulteriormente sostenuti dagli studi sui *corpora* condotti dai linguisti Renouf (1992) e Renouf e Sinclair (1991), i quali hanno mostrato che la maggior parte del linguaggio naturale può essere strutturato in *set* collocazionali o *framework* (in Wray e Perkins, 2000).

Le unità multiparola giocano un ruolo fondamentale in vari campi della ricerca sul linguaggio. Queste analisi possono essere organizzate in diverse categorie: gli studi sul dominio dell'insegnamento della lingua e dell'acquisizione della lingua; gli studi descrittivi che si occupano dell'uso delle combinazioni di parole nel linguaggio; gli studi che presentano classificazioni funzionali e strutturali delle combinazioni di parole; gli



studi nel campo della lessicografia, in cui le unità multiparola sempre più frequentemente hanno un ruolo centrale; e gli studi basati sui *corpora*, che considerano i *lexical bundles* come categorie particolari di combinazioni di parole (Cortes 2002; in Jalali, 2008; Spina, 2010).

Molti studi sulle espressioni multiparola sono stati essenzialmente teorici e si differenziano anche nelle metodologie utilizzate per identificarle. Secondo Biber ed altri (2004), questi studi empirici non rendono bene l'intera situazione, poiché portano avanti solo un tipo di ricerca empirica. Infatti, si pongono diversi scopi di ricerca: alcuni descrivono l'intera gamma di combinazioni multiparola, mentre altri selezionano piccole serie di sequenze "importanti". Si danno anche criteri discordanti nell'identificare le combinazioni multiparola (ad esempio, il criterio percettivo, quello di frequenza). Inoltre, tengono conto di caratteristiche formali distinte (come modelli lessico-grammaticali); differiscono nei campioni testuali utilizzati (da piccoli a grandi *corpora*). Molti di questi studi ignorano completamente il registro, altri ancora analizzano solo testi parlati o scritti, e solo un piccolo numero di studi paragona le unità multiparola attraverso registri differenti (per approfondimenti cfr. Biber ed altri, 2004).

D'accordo con Biber (2004), sosteniamo che solo conducendo studi empirici da diverse prospettive si può comprendere a pieno l'importanza e le funzioni delle unità multiparola.

Dato che da lungo tempo non c'è accordo sulla definizione di combinazioni di parole ricorrenti, si riscontra anche una certa incertezza sul come chiamarle, cosa che si riflette anche nell'ampio raggio di termini e di etichette usate per descrivere e coprire i diversi aspetti e categorie del linguaggio formulaico, come illustrato da Wrey e Perkins (2000). Le ricerche indicano l'esistenza di più di quaranta termini usati per riferirsi a uno o più tipi e sottotipi di linguaggio formulaico (Wrey, Perkins, 2000). Nonostante non ci sia accordo fra i ricercatori circa le caratteristiche fondamentali delle sequenze multiparola, sul come analizzarle e perfino sul come chiamarle, un certo consenso generale esiste circa la loro importanza (Biber ed altri, 2004). Ad esempio, Nattinger e De Carrico (1992) hanno sottolineato l'importanza delle combinazioni multiparola come tecnica di sostegno alla comunicazione, che rende il linguaggio più predicibile per l'ascoltatore. I linguisti sostengono che una parte notevole del linguaggio a cui siamo esposti nella vita quotidiana rappresenta una routine così come la rappresentano le

situazioni giornaliere. Secondo gli autori la ritualizzazione non presenta altro che una caratteristica del comportamento umano e del suo ambiente, come già da tempo riconosciuto dai sociologi e dagli antropologi. Nel lavoro del 1992 i due linguisti cercano di mostrare come le sequenze prefabbricate, prodotti di questa ritualizzazione, giocano un ruolo importante sia nella produzione linguistica che nell'apprendimento linguistico.

Pawley e Syder (1983) hanno dibattuto circa i vantaggi del linguaggio precostituito nella riduzione dello sforzo elaborativo. La natura prefabbricata fa sì che le sequenze necessitino poca elaborazione e il parlante può dirigere i suoi pensieri verso altre attività. Sostengono che questa sia la ragione per cui una persona o perfino un'intera comunità preferisce certe combinazioni di parole al posto di altre altrettanto accettabili per le stesse situazioni comunicative. Come asserisce Nesselhauf (2005) le sequenze formulaiche facilitano la comprensione, da contro quelle non formulaiche irritano il parlante nativo, orientando la sua attenzione su qualcos'altro. Confermando l'ipotesi di Pawley e Syder (1983) circa la preferenza di una comunità di sequenze di parole, la linguista (2005) afferma, inoltre, che le sequenze formulaiche indicano l'appartenenza ad un certo gruppo.

L'ampio uso di queste sequenze prefabbricate, come ad esempio *as you can see* nel genere accademico scritto aiuta a segnalare il registro del testo al lettore, riducendo così i tempi di interpretazione tramite l'uso di modelli familiari che collegano fra loro gli elementi di nuova informazione (Hyland, 2008). Queste combinazioni ricorrenti di parole "appear to represent a psychological association between words and reflect a very real part of users' communicative experience"<sup>21</sup> (Hyland, 2008a: 5). Come sostiene Hyland i dati di frequenza basati su uno studio di *corpus* hanno ancora più rilevanza rispetto alle sequenze multiparola perché sono un aspetto della misura in cui una sequenza di parole viene archiviata e utilizzata come un pezzo prefabbricato. L'estensione di frasi formulaiche alle collocazioni regolari come *Have a nice day* o *I want to make three points* indica la misura della formulaicità nell'uso del linguaggio (Hyland, 2008a).

A questo punto si presenta particolarmente importante lo studio di Tremblay, Derwing e Libben (2009), che ha confermato le ipotesi che i *lexical bundles* vengono

---

<sup>21</sup> Traduzione: "sembrano rappresentare un'associazione psicologica tra parole e riflettere una parte reale dell'esperienza comunicativa dell'individuo".

immagazzinati ed elaborati olisticamente ovvero come un'unica unità. Gli esperimenti di lettura di *lexical bundles*, i quali sono stati a loro volta presi in considerazione dal lavoro di Biber ed altri (1999), hanno mostrato che le frasi contenenti *lexical bundles* sono state lette più velocemente delle frasi che non includevano *lexical bundles*. Secondo gli autori questo fatto suggerisce che i *lexical bundles* facilitano l'elaborazione delle frasi e vengono immagazzinati olisticamente.

Secondo Wray e Perkins (2000), le sequenze formulaiche sono più che semplici unità linguistiche. Esse sono infatti strumenti che operano sulla base di cambiamenti di sviluppo, che avvengono momento – per – momento, nel corso del programma comunicativo e del processo elaborativo dell'individuo. I ricercatori sostengono che il riconoscimento del ruolo centrale delle sequenze formulaiche e l'ammissione della loro presenza nel lessico non implica dover negare la nostra capacità creativa, né tantomeno implica il vedere la creatività, né la formulaicità, come elementi che si escludono a vicenda. I due linguisti vedono la formulaicità come uno strumento di interazione sociale e come una scorciatoia per l'elaborazione. Indicano infatti tre funzioni centrali dell'interazione sociale, ovvero la manipolazione degli altri, l'asserzione di una identità separata e l'asserzione di un'identità di gruppo, che può essere raggiunta tramite l'uso del linguaggio formulaico (Wray, Perkins, 2000).

Spiegano inoltre che non è necessario sostenere che queste funzioni non possano essere raggiunte anche usando delle strutture nuove, create dal niente al momento dell'enunciazione, ma che sono le sequenze formulaiche che sono generalmente associate a queste funzioni.

La seconda funzione principale del linguaggio formulaico indicata dagli studiosi è quella compensatoria: le sequenze formulaiche sono meccanismi che compensano i limiti della memoria. Secondo questi studiosi, sembra che l'uso delle frasi precostituite serva a minimizzare gli effetti di possibili errori di corrispondenza fra le nostre capacità linguistiche potenziali e quella che è di fatto la nostra capacità di memoria a breve termine.

Le ricerche di Raichle e McCrone (in Wray, Perkins, 2000) mostrano che una volta che il cervello riconosce un compito linguistico come familiare, esso è capace di andare oltre la via di elaborazione utilizzata per apprenderlo, operando in questo modo come una scorciatoia di elaborazione.

La forza motrice delle scorciatoie elaborative è la loro capacità di assicurare che la produzione del parlante sia fluente e che l'informazione sia disponibile nella maniera e nei tempi richiesti; d'altro canto, la forza motrice delle formule socio-interazionali viene dal far sì che il parlante ottenga ciò che vuole e sia percepito come individuo all'interno del gruppo.

Sulla scorta dell'esempio dei comandi militari, Wray e Perkins (2000) indicano che il linguaggio formulaico si adatta meglio alle due funzioni summenzionate, poiché è più probabile che un ascoltatore comprenda un messaggio se questo arriva in una forma che egli/ella ha già sentito in precedenza, e che egli/ella possa elaborare senza ricorrere a una piena decodificazione analitica – cosa che dipende anche dalla capacità del parlante di anticipare correttamente la conoscenza del linguaggio dell'ascoltatore.

I ricercatori hanno suggerito un modello che indica i ruoli del linguaggio formulaico nell'assistere il parlante nella produzione e nella comprensione. Il punto di intersezione del modello sono le sequenze formulaiche, che sono utilizzate per strutturare il discorso, dando priorità al bisogno di decodificare in maniera soddisfacente quegli *input* che si presentano in un formato nuovo; al bisogno di elaborare velocemente e accuratamente, senza sovraccaricare il sistema; e al bisogno di far sì che il messaggio abbia la massima opportunità di essere compreso correttamente (Wray, Perkins, 2000: 18).

Gli autori affermano che in un'interazione normale la formulaicità è l'impostazione di base, sia nella produzione che nella comprensione. Come sostengono Wray e Perkins (2000) il modello sui ruoli del linguaggio formulaico identifica una dinamica complessa, secondo la quale la scelta dell'individuo circa una sequenza formulaica in un dato momento viene determinata da molti fattori, quali la comprensione generale del linguaggio e/o il livello di sviluppo cognitivo, lo scopo del discorso, la complessità e la novità dell'idea, il contesto dell'interazione e del discorso e la competizione con attività concomitanti (Wray, Perkins, 2000).

Risulta dunque evidente che le frasi si compongono di blocchi di parole perché queste vengono impresse nella mente non come singole unità ma insieme ad altri termini, in un contesto che è condizionato dalla nostra esperienza, dalle associazioni di parole che abbiamo già incontrato (Hoey, 2005). Tutto ciò che sappiamo di una parola è il risultato delle nostre esperienze dei contesti in cui abbiamo incontrato quella data

parola. Questo implica che quando vogliamo formulare qualcosa, le formulazioni e lo stile che scegliamo vengono condizionati dal contesto in cui abbiamo incontrato certe parole (Jespersen, 1924; Bally, 1951; Hoey, 2005). Sinclair (1991) e Hoey (2005) evidenziano che in base alle nostre esperienze testuali le nostre concordanze mentali si differenziano e i *pattern* vengono richiamati dalla memoria in un contesto a noi noto e più volte incontrato.

## **I.5 Gli studi sul linguaggio formulaico per la lingua italiana**

In termini generali, si può affermare che la ricerca linguistica italiana è meno produttiva di quella inglese. Simone (1997) è abbastanza critico nei riguardi della ricerca lessicografica italiana, osservando come questa consti delle classi di parole identificate dalla grammatica tradizionale (nomi, verbi, aggettivi, avverbi, preposizioni e pronomi), basandosi sul presupposto che queste categorie siano monorematiche. A questo proposito, è evidente che praticamente tutti i dizionari italiani sono stati compilati su questa base, così come anche molti dizionari di altre lingue. Questa scelta dovrebbe essere riesaminata in maniera critica perché il lessico è strutturato secondo principi complessi, i quali vanno ben oltre le tradizionali classi di parole (Simone, 1997). Il ragionamento solo in termini di parti del discorso e la loro concezione come unità composte da una sola sequenza di componenti isolati tra due spazi vuoti è insoddisfacente, perché non rispecchia "né la reale organizzazione della lingua né la competenza che il parlante ne ha" (Simone, 1997: 156). Come avviene nella lingua inglese, anche nella lingua italiana esistono combinazioni di parole, formate da più parole che non possono co-apparire l'una senza l'altra. Queste combinazioni di parole hanno caratteristiche specifiche e meritano di essere prese in considerazione perché, come già evidenziato prima, rispecchiano il linguaggio naturale.

Un primo riconoscimento delle parole sintagmatiche in italiano è stato condotto da Voghera (2004), che riconosce alcune espressioni polirematiche nella sua terminologia, distinguendole dalle collocazioni nella loro accezione tradizionale. Questa studiosa ha notato che in italiano non esiste un unico termine accettato dai linguisti per indicare queste combinazioni di parole. La linguista indica che nell'ambito lessicografico, come

ad esempio nel dizionario GRADIT o nel *Lessico di frequenza dell'italiano parlato* (LIP), si trova il termine *unità polirematica* o semplicemente *polirematica*, in morfologia il termine *composto sintagmatico* (Scalise, 1994; in Voghera, 2004), e infine i termini *lessema complesso* (Voghera, 1994; De Mauro, Voghera, 1996) o *unità lessicale superiore* (Dardano, 1978; in Voghera, 2004) con i quali vengono descritte le unità lessicali percepite come un unico lessema. Ciascuna di queste denominazioni sottolinea un aspetto importante di queste formazioni di parole.

In questa prospettiva, le polirematiche vengono collocate tra la formazione delle parole e la sintassi, specificando che la loro posizione è determinata dal fatto che le une possono condividere alcune delle caratteristiche dei composti, mentre le altre si avvicinano più ai sintagmi, rappresentando in questo modo un gruppo eterogeneo. Il fenomeno, dai confini sfumati, si presenta ampio e vario e questa varietà si riflette nell'assenza di un unico termine per descrivere queste combinazioni di parole, così come riscontrato per la lingua inglese (Voghera, 2004).

Occorre inoltre precisare che già Lo Cascio (1971) aveva avuto modo di menzionare le costruzioni o espressioni polirematiche, però inserendole tra le collocazioni. Secondo questo studioso, le espressioni polirematiche non devono essere intese nel loro significato tradizionale come "l'accoppiamento di due specifici lemmi o parole che si incontrano in modo fisso e sistematico creando un concetto unitario e preciso", ma come manifestazioni delle possibilità combinatorie standard (Lo Cascio, 1997: 71). Un aspetto particolarmente importante del contributo di Lo Cascio (1997: 72) è la sua affermazione che

"la lingua deve essere considerata in altri termini come interamente caratterizzata da collocazioni, essa è cioè a mio parere molto più formulaica di quanto non si voglia ammettere. Dietro la completa esigenza di libertà creativa del parlante c'è anche l'abitudine all'uso stereotipo della lingua, determinato dagli automatismi acquisiti, dalla costrizione e limitazione imposta dalla sintassi, dalla coerenza semantica ed enciclopedica e dalla limitazione nella competenza linguistica. Si può quasi affermare, anche se con il rischio di offendere la libertà creativa dei parlanti, che l'essere umano usa pacchetti linguistici prefabbricati e apprende una lingua anche per 'pacchetti'."

Quello che per il linguista rappresenta solo un'intuizione viene sostenuto dalle ricerche in lingua inglese, cui abbiamo accennato precedentemente. Secondo Lo Cascio, il carattere formulaico è evidente soprattutto nei linguaggi specialistici, un'osservazione che risulta fondamentale per questa ricerca. Inoltre, merita di essere presa in considerazione la sua ipotesi che sia la formulaicità a rendere possibile un'organizzazione sistematica del linguaggio e una sua sistematica memorizzazione, aderendo così anche a un principio di economia. Di fatti, la memorizzazione e l'elaborazione delle combinazioni di parole dipendono proprio dalla loro struttura semantica e sintattica, come hanno dimostrato molte ricerche in campo psicolinguistico (Cardona, 2008). A tale riguardo, Lo Cascio precisa che la comunicazione quotidiana viene organizzata per lo più per combinazioni di parole *default*, per veri e propri pacchetti prefabbricati, arrivando a sostenere che il lessico dovrebbe essere corredato dall'elenco delle collocazioni.

In alcuni recenti contributi, Simone si è avvicinato agli studi sulle combinazioni di parole nell'ambito della sua teoria del significato grammaticale. La sua *Grammatica di Costruzioni e Categorie*, a differenza della *Construction Grammar*, si concentra principalmente sulla natura della grammatica e sul suo modo di organizzarsi (Simone, 2007, 2008).

Simone ha proposto una nuova classificazione che si codifica in due canali: categorie e costruzioni. Questa classificazione di aggregazioni di parole distingue tra combinazioni e costruzioni. Nelle combinazioni, "le parole si collegano occasionalmente per disaggregarsi fuori dell'enunciato" (Simone, 2006: 393), mentre nelle costruzioni "le parole [...] occorrono in forma di combinazione già fatta, stabile o preferenziale" (Simone, 2006: 393). Secondo questa interpretazione, le parole si combinano in aggregazioni perché nelle lingue opera il principio sintagmatico, che contrae le parole in sintagmi "di varia natura, varia forza attrattiva e varia estensione" (Simone, 2008: 4). Lo studioso suddivide i sintagmi in combinazioni e in costruzioni. Le combinazioni vengono a loro volta ripartite in combinazioni volatili, combinazioni preferenziali e collocazioni (Simone, 2006).

Simone (2008: 5) raggruppa sotto ciascun gruppo di sintagmi le combinazioni di parole come segue:

- a. *Combinazioni volatili*: aggregazioni una volta combinate in sintagmi, le parole sciolgono i loro legami e tornano al loro posto nel lessico;
- b. *Combinazioni preferenziali*: sono aggregazioni di parole che "occorrono insieme frequentemente" (Simone, 2008: 5), costituite di unità lessicali indipendenti (per esempio: caro amico, bella donna);
- c. *Collocazioni*: sono combinazioni che rispettano le due seguenti condizioni:
  1. condizione sintattica: incorporano sempre la stessa testa;
  2. condizione lessicale: bloccano la sinonimia, ciò significa che l'alternanza del termine che entra nella combinazione non è libera (per esempio: disco rigido, caffè lungo);
- d. *Costruzioni in senso proprio* – rispondenti ad un insieme di requisiti specifici.

Secondo Simone (1997), le costruzioni sono aggregazioni di parole piuttosto stabili e sono trattate come unità, ovvero sono archiviate nella memoria, richiamate ed elaborate sia formalmente che semanticamente. Come nota Simone, le costruzioni sono state senza dubbio centrali nella storia della linguistica, poiché già le grammatiche latine e greche usavano le costruzioni sia come concetto classificatorio che come strumento pedagogico di grande utilità. Se ne presupponeva lo studio e la memorizzazione in quanto unità che andavano a formare un archivio atto a facilitare il processo ulteriore di apprendimento.

In generale, le costruzioni dispongono di una forza pragmatica e alcune di esse sono portatrici di una specifica informazione pragmatica, essendo in questo singolari. Inoltre, una parte delle costruzioni è riconoscibile da marche dedicate. Tutte le costruzioni sono specifiche di una data lingua (*language specific*) e infine appaiono nel lessico di una lingua come unità lessicali e non come parole isolate. Simone (2007) precisa che, prendendo in considerazione tutti questi criteri, non sorprende che il comportamento linguistico possa apparire in larga misura come costituito da costruzioni. A tale riguardo, si ricorda che Erman e Warren (2000; in Simone 2008) hanno calcolato che sia nel discorso scritto che parlato, fino a circa la metà del linguaggio è costituita da costruzioni di varia natura. I linguisti Pawley e Syder (1983) sostengono che le costruzioni memorizzate rappresentano dei blocchi linguistici della comunicazione



fluente e allo stesso tempo costituiscono dei modelli per la creazione delle nuove stringhe di parole, le quali, a loro volta entrano a far parte delle espressioni familiari.

Come notato da Simone (2007), non tutte le aggregazioni di parole si collocano sullo stesso piano: combinazioni volatili, preferenziali e collocazioni si attivano quando le parole co-appaiono nella catena, mentre le costruzioni, diversamente dalle combinazioni, si stabilizzano a livello paradigmatico, in altre parole hanno un significato costruzionale, un'espressione costruita sintagmaticamente viene trasferita sul piano paradigmatico. Secondo questa interpretazione, le aggregazioni di parole dispongono di una variabile forza coesiva, definita come l'effetto di forze attrattive che costituiscono un legame tra le parole. Questa forza è presente in grado crescente a partire dalle combinazioni volatili, che dispongono di una minima forza attrattiva, fino al suo livello massimo di coesione: le costruzioni. Simone (2007: 212) le distribuisce lungo una scala, che copre i diversi livelli compresi fra la libertà relativa di combinazioni e costruzioni fisse. Lo studioso pone ad un estremo le combinazioni volatili, seguite dalle combinazioni preferenziali e collocazioni, all'altro, invece, le costruzioni fisse.

Masini (2009) prosegue lungo questa linea di ricerca, fornendo una classificazione delle combinazioni di parole di natura lessicale. Riprendendo la distinzione proposta da Bally già nel 1909, Masini distingue tra le unità fraseologiche vere e proprie o le espressioni fisse da un lato e le serie fraseologiche o combinazioni usuali dall'altro.

Sulla scorta dell'analisi per la lingua inglese delle espressioni multiparola da un lato e delle combinazioni libere dall'altro, proposta da Zgusta (1967), Masini analizza diversi tipi di combinazioni di parole in italiano.

Come esempio di una combinazione usuale, Masini propone l'espressione *pioggia torrenziale*, in cui la sostituzione del modificatore, in questo caso *torrenziale*, "determina una differenza nel grado di familiarità che l'espressione assume" (Masini, 2009: 81).

In questo caso, si prendono in considerazione aggregazioni talmente convenzionalizzate da costituire delle frasi fatte e si distinguono quelle combinazioni di parole in cui la scelta del modificatore è obbligatoria – le espressioni multiparola – dalle combinazioni usuali, le quali presentano una forte coesione interna e una semantica non trasparente, come per esempio *anno accademico* e *guerra fredda*. Nella tradizione

linguistica italiana, queste combinazioni di parole vengono definite in molti modi diversi: espressioni polirematiche (Simone, 1997; Lo Cascio, 1997); lessemi complessi (Voghera, 1994; De Mauro, Voghera, 1996); unità lessicali superiori (Dardano, 1978). Masini (2009) specifica che la distinzione tra le due tipologie di parole non è sempre così netta, poiché le espressioni multiparola possono presentare anche una variabilità paradigmatica, se pur minima, come nel caso di *fare luce/chiarizza*. Nella classificazione di Masini (2009: 82) si può osservare che nella cornice delle combinazioni di parole che si distinguono dalle espressioni multiparola rientrano anche le combinazioni preferenziali e collocazioni vere e proprie.

Le espressioni multiparola, come per esempio *l'anima gemella*, vengono immagazzinate nel nostro lessico mentale come blocchi di parole, presentando una costruzione fissa, non variabile. Da contro, le collocazioni, come *aprire un conto*, e le combinazioni preferenziali, come per esempio *pioggia torrenziale*, si distinguono dalle espressioni multiparola proprio per la mancanza sintagmatica, le combinazioni preferenziali anche per la mancanza paradigmatica.

Sia le collocazioni che le combinazioni preferenziali sono soggette a convenzionalizzazione e pertanto possono diventare nel tempo espressioni multiparola (Masini, 2009). Esistono molte espressioni multiparola che possono essere classificate secondo la loro idiomaticità e secondo la loro invariabilità. Nel quadro generale delle espressioni multiparola rientrano diversi tipi di espressioni: le espressioni idiomatiche (*tirare le cuoia*); quelle metaforiche (*parlare dietro le spalle*); le formule conversazionali (*piacere di conoscerLa*); i proverbi (*si salvi chi può*); le citazioni (*dire qualcosa di sinistra*) e persino interi testi (canzoni, poesie, ecc.) (Masini, 2009).

Sulla base della classificazione proposta da Masini possiamo ascrivere i *lexical bundles* alla tipologia di espressioni multiparola perché rientrano nella cornice dei parametri che le delineano: essi sono infatti dotati di una fissità sintagmatica degli elementi interni alla combinazione di parole; presentano una fissità paradigmatica, ovvero l'obbligatorietà degli elementi che li costituiscono; infine, la loro frequenza d'uso ne sancisce il grado di familiarità.

Dobbiamo però puntualizzare che il significato dei *lexical bundles* è facilmente deducibile dal significato dei loro componenti. Questa caratteristica li distingue da alcuni sottogruppi di espressioni multiparola.

Nella letteratura inglese i *lexical bundles* vengono intesi come delle collocazioni estese. Biber ed altri (1999) parlano di *lexical bundles* nel caso di sequenze di parole che co-occorrono in maniera molto frequente in un dato registro. Generalmente si tratta di espressioni fisse e non è possibile sostituire la sequenza con una parola singola (ad esempio *on the basis of* e non *on a basis of*).

Gli studiosi distinguono due estremi delle espressioni multiparola: su un lato collocano gli idiomi, sull'altro le collocazioni. Gli idiomi sono espressioni relativamente non variabili, con significati che non possono essere desunti dai significati delle singole parti e devono essere appresi come un tutto unico, anche se conosciamo i significati di ogni singola parola che li compone. Da contro le collocazioni sono associazioni fra parole lessicali, parole che co-occorrono con una frequenza non spiegabile tramite criteri di casualità. A differenza degli idiomi, queste combinazioni di parole sono associazioni statistiche, piuttosto che espressioni relativamente fisse. Oltre a presentare questa caratteristica, in una collocazione le parole individuali mantengono il proprio significato e tendono a co-occorrere con serie specifiche di collocati. Ciononostante, parte del significato esteso di una parola proviene dal fatto che questa tende a co-occorrere con una serie specifica di collocati. Difatti le parole con simili significati si distinguono spesso tramite la loro collocazione preferita (Biber ed altri, 1999).

Tra le espressioni multiparola, piuttosto studiate negli ultimi anni si possono considerare le espressioni idiomatiche e metaforiche, seppure gli studi sulla loro frequenza in lingua inglese hanno mostrato che le espressioni idiomatiche occorrono circa cinque volte ogni milione di parole e la loro occorrenza è specialmente rilevata nella narrativa e raramente in altri registri, ivi compresa la conversazione, mentre i *bundles* lessicali sono molto più frequenti in tutti i registri.

Biber ed altri (1999) sostengono che i *lexical bundles* non sono combinazioni di parole che i parlanti nativi riconoscerebbero come espressioni idiomatiche o come altre espressioni lessicali fisse; la maggior parte di esse non sono affatto complete nella struttura e rappresentano piuttosto frasi frammentarie o clausole che comprendono nuovi frammenti.

Sulla base di queste premesse, possiamo desumere che i *lexical bundles* rappresentano un gruppo di combinazioni di parole a sé stante. I *lexical bundles* si distinguono da altre combinazioni di parole dal modo in cui vengono identificati: essi

vengono identificati empiricamente, solo sulla base della frequenza e dell'ampiezza d'uso, senza necessitare di conoscenze pregresse circa le co-occorrenze di altre unità di parole (Cortes, 2002).

Rispetto al panorama inglese, l'ambito di ricerca sulle espressioni multiparola nella lingua italiana è ancora agli inizi. Poiché i *lexical bundles* in lingua italiana non sono stati ancora studiati, il presente studio rappresenta un primo tentativo in materia.

## **I.6 I *lexical bundles***

Oltre alla tradizione contestualista britannica e al lavoro di Sinclair e Hoey, il lavoro sui *lexical bundles* di Douglas Biber rappresenta un'altra importante corrente di ricerca che si colloca nell'interfaccia fra lessico e grammatica mostrandone l'interdipendenza (Römer, 2009). Questo filone di ricerca emerge dal Nord America e analizza le combinazioni ripetute di parole contigue, o le sequenze di tre o quattro parole nei registri scritti e parlati.

Gli approcci descritti sopra – *idiom principle* di Sinclair e il *lexical priming* di Hoey che su di esso si basa – hanno alcune questioni in comune con l'approccio dei *lexical bundles* usato in questa ricerca. La prima osservazione comune è che lo studio del linguaggio è empirico e si basa su grandi quantità di testi prodotti in maniera naturale (Römer, 2009). La frequenza delle occorrenze è cruciale e i *corpora*, così come gli strumenti del *corpus*, sono utilizzati per identificare gli elementi comuni ai vari contesti e ai vari tipi di discorso (Römer, 2009). Tutti e tre i settori collocano al primo posto i dati e l'osservazione, e la teoria viene in secondo luogo, poiché le scoperte teoriche derivano induttivamente dai dati. I *lexical bundles* rilevati sono così frequenti che i dati prodotti dall'investigazione di un *corpus* richiedono una spiegazione (Biber ed altri, 2004).

Tutti e tre gli approcci vedono ciascuna parola come portatrice di una propria fraseologia, di proprie collocazioni preferite, collegamenti preferiti, prosodia semantica, ecc. (Granger, 2011). In tutti e tre gli approcci l'unità di significato in una lingua non è la parola presa in maniera isolata ma una costruzione o un'unità frasale.

Inoltre, la forma viene vista come inseparabile dal significato. Per concludere, la terza osservazione condivisa è che il lessico e la grammatica dovrebbero essere considerati come interdipendenti e inseparabili, e non come due unità isolate, e che le combinazioni ricorrenti di parole dovrebbero avere un ruolo maggiore nella teoria linguistica (Römer, 2009).

Biber (1999) e Cortes (2002) concordano sul fatto che la differenza principale fra i *lexical bundles* e le altre combinazioni ricorrenti di parole è il modo in cui i *bundles* vengono identificati. Questa identificazione non ha bisogno di conoscenze pregresse o di intuizioni circa le possibili frequenze in cui le unità di parole possono occorrere (Cortes, 2002).

Lo scopo principale di questa ricerca è comparare i modelli generali di utilizzo dei *lexical bundles* nella lingua scritta e parlata, in particolare in tre discipline di lingua italiana nell'ambito delle microlingue, esplorando la misura in cui i *bundles* vengono usati, insieme alle associazioni principali di *bundles* strutturali e funzionali.

In base alle ricerche svolte per la lingua inglese, sia nel parlato che scritto (Biber ed altri, 1999, 2004) e specialmente nell'ambito del discorso accademico (Biber, 2006; Cortes, 2002, 2004; Hyland, 2008a) ma anche in base agli studi condotti per lingua spagnola e per la lingua coreana, una lingua con delle caratteristiche tipologiche completamente diverse, si ipotizza la stessa frequenza dei *lexical bundles* per la lingua italiana. Vengono inoltre previste delle funzioni discorsive dei *lexical bundles* specifiche e peculiari di una data microlingua.

### **I.6.1 La definizione dei *lexical bundles***

L'oggetto principale della presente ricerca si basa sui modelli lessicali, su sequenze di quattro parole, *bundles* di parole, anche conosciuti come gruppi (*clusters*) o blocchi (*chunks*) (Hyland, 2008a) che mostrano una tendenza statistica alla co-occorrenza.

Queste combinazioni di parole possono essere guardate come espressioni multiparola e vengono chiamate *lexical bundles*. Il termine *lexical bundles* è stato

coniato da Biber insieme ad altri autori (1999: 90) in un'innovativa e approfondita trattazione della grammatica inglese, in cui i *lexical bundles* vengono definiti come:

"the most frequent recurring lexical sequences in a register, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse<sup>22</sup>".

Nel *Longman Grammar of Spoken and Written English*<sup>23</sup> (Biber ed altri, 1999) – un lavoro basato esclusivamente sul *British National Corpus* di 100.000 milioni di parole – i *lexical bundles* vengono identificati in base alla loro lunghezza. Nei primi studi, Biber insieme ad altri studiosi (1999) ha analizzato *bundles* composti da tre parole, mentre nei lavori successivi (Biber ed altri, 2004, Biber, 2006) i ricercatori si concentrano anche sulle sequenze di 4 parole, specificando che i *lexical bundles* più lunghi sono di solito estensioni o formazioni di uno o più *bundles* più brevi.

Una gamma di studi di *corpus* ha mostrato quanto i *lexical bundles* siano onnipresenti nella conversazione e nei generi accademici. Biber ed altri (1999) hanno suggerito che i *bundles* di tre parole ricorrono più di 60.000 volte nella prosa accademica e più di 80.000 volte nella conversazione, mentre i *bundles* di quattro parole ricorrono più di 5.000 volte su un milione di parole nella prosa accademica e 8.500 volte nella conversazione. Hanno riportato che i *bundles* individuali di tre parole sono leggermente più frequenti che i *bundles* di quattro parole: il *bundle* di tre parole medio ricorre 25 volte su un milione di parole, mentre il *bundle* di quattro parole medio ricorre 20 volte su un milione di parole.

Circa il 21% della componente accademica del *Longman Spoken and Written English corpus* rappresenta questi *bundles* comuni e circa il 30% delle parole nella conversazione si presenta in un *lexical bundle* ricorrente, mentre le combinazioni di tre

---

<sup>22</sup> Traduzione: "le sequenze lessicali che ricorrono più frequentemente in un registro, indipendentemente dalla loro idiomaticità, e indipendentemente dal loro status strutturale. Cioè, i *lexical bundles* sono semplicemente delle sequenze di forme di parole vanno comunemente insieme nel discorso naturale".

<sup>23</sup> Il *Longman Spoken and Written English Corpus* comprime quattro milioni di parole da conversazioni in inglese britannico; circa tre milioni di parole da conversazioni in inglese americano; circa 5.3 milioni di parole in prosa accademica. Il *corpus* di prosa accademica comprende sia articoli di ricerca accademica (1.7 milioni di parole) e libri accademici avanzati (circa 2.6 milioni di parole). Per approfondimenti vedi Biber ed altri (1999).

parole più frequenti appaiono più di 200 volte su un milione di parole e le combinazioni di quattro parole circa 100 volte. Le ricerche considerano la frequenza come la caratteristica più saliente e determinante dei *bundles*: i *lexical bundles* sono semplicemente espressioni fisse determinate dalla frequenza, che è specificata dal programma informatico. L'utilità dei dati di frequenza e dell'analisi del *corpus* risiede nell'identificazione di modelli di utilizzo che altrimenti non verrebbero notati, rendendo possibile lo studio di parole particolari e delle loro funzioni particolari all'interno di date classi grammaticali, un aspetto del linguaggio che è spesso ignorato (Biber ed altri, 1999).

In verità il punto limite usato per identificare i *lexical bundles* varia a seconda del ricercatore ed è in qualche modo arbitrario, poiché può variare da 10 a 40 su un milione di parole. Per esempio, in uno studio Biber ed altri (1999) considerano 10 occorrenze su un milione di parole come punto limite, mentre nello studio dei *lexical bundles* nei libri di testo e nell'insegnamento universitario il punto limite è di 40 occorrenze su un milione di parole (Biber ed altri, 2004).

Biber ed altri (1999) hanno riportato che alcune combinazioni di parole sono molto più comuni e occorrono più di 200 volte su un milione di parole. Altri studiosi che hanno analizzato le microlingue come Cortes (2002, 2004) e Hyland (2008a, 2008b) hanno usato un punto limite di frequenza di 20 volte per identificare una combinazione di parole come *lexical bundle*.

Sulla base degli studi precedenti sui *lexical bundles* per la lingua inglese, il presente lavoro adotta un approccio conservativo: si considera come punto limite la frequenza di 10 volte su un milione di parole.

Un'ulteriore restrizione è data dal fatto che una combinazione di parole deve essere usata in almeno cinque testi differenti per poter essere considerata un *lexical bundle*, questo per evitare usi idiosincratici da parte di parlanti individuali o autori o anche usi ripetitivi da parte di autori che possono mancare di buona maestria, almeno nell'uso delle sequenze di parole preferite dagli esperti (Jalali ed altri, 2008). Biber e altri (2004) hanno rilevato che questa restrizione ha poco effetto mentre anche i *lexical bundles* meno comuni sono usati in almeno 20 testi differenti. La stessa restrizione è stata usata anche da altri ricercatori e viene adoperata anche nel presente lavoro.

Anche se la serie identificata di *lexical bundles* è basata solo su criteri distribuzionali (tasso di occorrenza e distribuzione nei testi), sembra che i *bundles* presentino diverse caratteristiche che li distinguono da altri tipi di espressioni formulaiche: sono estremamente comuni, non sono idiomatici e non rappresentano un'unità strutturale completa (Biber ed altri, 2004). Come già menzionato, la struttura è un'altra caratteristica principale dei *bundles*; solitamente non è una struttura grammaticale completa.

Come riportato da Biber ed altri (1999), solo il 15% dei *lexical bundles* nella conversazione può essere guardato come composto da frasi o clausole complete, mentre meno del 5% dei *lexical bundles* nella prosa accademica è composto da unità strutturali complete, ma piuttosto da frasi frammentate o clausole con nuovi frammenti incorporati.

I *lexical bundles* hanno forti correlazioni grammaticali, cosa che facilita il loro raggruppamento in tipi strutturali (Biber ed altri, 1999). Secondo Biber ed altri (1999), la maggior parte dei *lexical bundles* in inglese collega due unità strutturali: per esempio i *bundles* nella prosa accademica collegano due frasi (ad esempio, *the nature of the, as a result of*), mentre i *bundles* nella conversazione collegano due clausole, ad esempio *I don't know why*.

Come riportato dai ricercatori, nelle conversazioni molti *bundles* contengono l'inizio di una clausola principale, seguita dall'inizio di una clausola complemento incorporata. Lo studio ha mostrato che la maggioranza dei *lexical bundles* nella prosa accademica è composta da unità strutturali incomplete, tendendo ad essere composta da blocchi nominali piuttosto che da blocchi clausali. Circa il 90% dei *lexical bundles* nella conversazione include parte di una frase verbale e circa il 60% di tutti i *bundles* nella prosa accademica sono parti di frasi nominali o di frasi preposizionali (Biber ed altri, 1999). Come vedremo in prosieguo, i risultati ottenuti dall'analisi dei *lexical bundles* confermano le ipotesi tradizionali sulla distinzione tra la lingua parlata e lingua scritta.

Sulla base di quanto premesso, proponiamo la seguente definizione di *lexical bundles*: I *lexical bundles* sono combinazioni di parole ricorrenti in un registro. Le loro caratteristiche salienti sono la fissità, la non idiomaticità, la trasparenza del significato, che è completamente deducibile dal significato dei singoli membri che compongono un *lexical bundle*, e la struttura grammaticale frammentata. A titolo d'esempio possiamo



indicare le seguenti combinazioni di parole: *allora vediamo un po', e che ne so, non ci sono più, io gli ho detto, se non ricordo male, ecc.*

## **I.6.2 La struttura grammaticale e la funzione discorsiva dei *lexical bundles***

I *lexical bundles* sono diventati oggetto di grande interesse dal 1999, specialmente nei recenti studi basati sui *corpora*, da un lato per il fatto che sono molto frequenti, dall'altro a causa del loro contributo funzionale alla coerenza e all'organizzazione di testi diversi, sia scritti che parlati (Cortes, 2004; Biber ed altri, 2004; Hyland, 2008a, 2008b). I *lexical bundles* hanno la funzione di "building blocks of coherent discourse"<sup>24</sup> (Hyland, 2008b: 1) e servono a una gamma di funzioni discorsive molto vasta, che riguarda l'organizzazione del discorso, l'espressione dell'opinione dello scrivente e il riferimento a entità testuali o esterne.

Come vedremo nelle sezioni seguenti, i *lexical bundles* sono stati classificati sia secondo la loro struttura (Biber ed altri, 1999, 2003, 2004; Biber, 2006) che secondo la loro funzione discorsiva (Biber, Conrad, Cortes, 2003; Biber ed altri, 2004; Biber, Barbieri, 2007; Hyland, 2008a, 2008b).

### **I.6.2.1 La struttura grammaticale dei *lexical bundles***

Queste sequenze prefabbricate si distinguono l'una dall'altra nella forma. La struttura dei *lexical bundles* è stata analizzata per la prima volta da Biber ed altri (1999) in uno studio che compara le combinazioni di parole ricorrenti nella prosa accademica e nella conversazione del *Longman Corpus of Written and Spoken English*, che suggerisce una categorizzazione strutturale e grammaticale di tipi di *lexical bundles*. Questo quadro analitico è la tipologia di *bundles* più ampiamente usata ed è stata applicata in numerosi studi successivi (Biber, Conrad, 1999; Biber, Conrad, Cortes, 2003, 2004; Cortes, 2004; Partington, Morley, 2004) e anche in altre lingue, in quella spagnola e coreana. I tipi

---

<sup>24</sup>Traduzione italiana: "blocchi costruttivi del discorso coerente".

strutturali di *lexical bundles* sono stati organizzati usando le prime parole del *bundle* come *marker* del loro raggruppamento grammaticale correlato. I *lexical bundles* sono identificati empiricamente come combinazioni di parole che ricorrono più comunemente in un registro. Biber e Conrad (1999) hanno spiegato che sebbene i *lexical bundles* non costituiscano delle unità grammaticali complete, ma piuttosto delle frasi frammentate o delle clausole con nuovi frammenti incorporati, hanno delle forti correlazioni grammaticali, cosa che rende facile raggrupparli in diversi tipi strutturali di base. Biber ed altri (1999) hanno sviluppato una tassonomia iniziale raggruppando i *lexical bundles* in tre categorie principali, secondo la loro struttura: i *lexical bundles* che incorporano frammenti di frasi verbali; i *lexical bundles* che incorporano i frammenti di clausole dipendenti; e i *lexical bundles* che incorporano le frasi nominali e i frammenti di frasi preposizionali, e le diverse sottocategorie.

Lo stesso *framework* è stato adoperato anche per la lingua spagnola e coreana. Come segnalato da Kim (2009) alcune modificazioni della tassonomia sono state necessarie per la lingua coreana, in quanto si tratta di una lingua tipologicamente completamente diversa.

Sulle orme degli studi precedenti per la classificazione strutturale dei *lexical bundles* (Biber ed altri, 1999, 2003, 2004; Cortes, 2004, 2008; Tracy-Ventura ed altri, 2007) il raggruppamento dei *lexical bundles* per la lingua italiana è stato svolto usando le prime parole del *bundle* come indicatori del loro raggruppamento. I *bundles* sono stati suddivisi rispettando la tipologia specifica della lingua italiana, mentre la classificazione elaborata per la lingua inglese è servita soltanto come strumento di base.

I *lexical bundles* nella lingua italiana sono stati raggruppati in cinque categorie principali e suddivisi in sottocategorie. Dobbiamo precisare che non tutte le categorie sono state identificate dall'analisi dei singoli *corpora*. Elenchiamo di seguito i cinque principali gruppi: a) *lexical bundles* che incorporano frasi verbali; b) *bundles* che contengono segmenti di frasi preposizionali; c) *bundles* che incorporano elementi di sintagmi nominali; d) *clusters* che sono congiunzioni; e) *bundles* che vengono classificati come "altre espressioni".

### *1.6.2.2 Le funzioni discorsive dei lexical bundles*

Anche se i *lexical bundles* non sono né espressioni idiomatiche né espressioni strutturalmente complete, hanno funzioni del discorso sistematiche e sono blocchi costituenti importanti nel discorso (Biber, 2006). Infatti forniscono "a kind of pragmatic 'head' for larger phrases and clauses, where they function as discourse frames for the expression of new information"<sup>25</sup> (Biber, 2009: 284-285).

Molti studi hanno sviluppato una classificazione funzionale di questi gruppi di parole, che hanno mostrato come queste possano servire a una vasta gamma di funzioni discorsive come organizzazione di discorso, espressione di posizione, e riferimento a entità testuali o esterne (Culpeper, Kytö, 2002; Cortes, 2002, 2004, 2008; Biber, Conrad, Cortes 2004; Biber, 2006; Biber, Barbieri, 2007; Hyland, 2008a, 2008b).

Come indicato da Cortes (2004), nel loro *corpus* di *Early Modern English*, Culpeper e Kytö hanno prodotto una classificazione funzionale dei *lexical bundles* di tre parole, mentre Cortes ha mosso i primi passi nel creare una tassonomia per la classificazione funzionale dei *lexical bundles* di quattro parole con un gruppo di *bundles* identificati in un *corpus* di scrittura compositiva di studenti al primo anno di università (Cortes, 2002, 2004). La sua tassonomia è stata migliorata in seguito da Biber e altri autori (2006) che hanno esaminato le liste di concordanze per analizzare le funzioni dei *bundles* comuni nei contesti di discorso allargato, usando principalmente un approccio induttivo. Questi studiosi hanno raggruppato i *bundles* che servono a funzioni simili a seconda del significato tipico e dell'uso di ogni *bundle*. Le funzioni del discorso associate ad ogni gruppo si basano su studi teoretici precedenti circa le funzioni del discorso delle caratteristiche linguistiche (Hymens, 1974; Halliday, 1978; Brown e Fraser, 1979; Biber, 1988, 1995; in Biber ed altri, 2004). Questa tassonomia è stata pensata per migliorare quella precedente riferita ai tipi di *lexical bundles* trovati nella prosa e nella conversazione accademica (Biber ed altri, 1999) che era stata sviluppata per includere le funzioni potenzialmente realizzabili in qualsiasi registro.

Il quadro analitico qui proposto è stato applicato anche, da un lato, a domini più specializzati: Cortes per esempio l'ha applicato allo studio dei *lexical bundles* nella

---

<sup>25</sup>Traduzione italiana: "Infatti forniscono "una specie di 'cappello' pragmatico a frasi e clausole più lunghe, nelle quali fungono da cornici discorsive per l'espressione di informazioni nuove.".

scrittura accademica di professionisti e studenti nel campo della storia e della biologia (Cortes, 2004), mentre Partington e Morley (2004) hanno esplorato l'uso dei *lexical bundles* nel discorso politico parlato, mostrando come i *lexical bundles* possono essere usati per analizzare la distintività di particolari registri; d'altra parte questo quadro analitico è stato usato anche in lingue diverse dall'inglese, come lo spagnolo e il coreano. Per la lingua coreana l'esistente tassonomia è stata modificata in base alla struttura dei *lexical bundles* identificati, tenendo conto delle caratteristiche tipologiche della lingua coreana.

Biber ed altri (2004) hanno identificato tre categorie principali: i *bundles* referenziali, gli organizzatori di testo e i *bundles* interazionali, che potrebbero essere legati alle funzioni linguistiche descritte da Halliday (Halliday, 1971; in Cortes, 2004). La tassonomia funzionale proposta può essere così riassunta:

- a. *stance bundles* o *bundles* posizionali. Svolgono funzioni interpersonali e sono costituiti da due gruppi principali: i *bundles* epistemici, che possono esprimere certezza o incertezza, probabilità e possibilità (per esempio *I don't know if, I don't think so*) e *bundles* attitudinali/di modalità, che esprimono gli atteggiamenti del parlante verso le azioni e gli eventi descritti nella preposizione seguente. I *bundles* attitudinali/di modalità potrebbero essere raggruppati in sottocategorie: desiderio, obbligo/direttiva, intenzione/previsione, e capacità;
- b. *discourse organizing bundles* o *bundles* che organizzano il discorso, riflettono la relazione fra il discorso precedente e quello successivo, svolgendo così due funzioni principali: introduzione all'argomento/elaborazione e restringimento dell'attenzione sull'argomento/spiegazione (per esempio: *I want to talk about* per introduzione all'argomento e *has to do with* per elaborazione dell'argomento);
- c. *referential bundles* o *bundles* referenziali aiutano lo scrivente a strutturare la propria esperienza e a determinare il modo in cui questi guarda alle cose. Questa categoria include quattro sottocategorie principali: identificazione/focus (per esempio *those of you who*), *bundles* di imprecisione (ad esempio *or something like that*), i *bundles* che specificano gli attributi (per esempio *have a lot of*) e i *bundles* di tempo/ luogo/deissi del testo (ad esempio *in the United States*).

Nel suo studio sulla scrittura accademica di storia e biologia, Cortes (2004) ha riscontrato dei *lexical bundles* che non sono stati identificati come combinazioni frequenti di parole nella classificazione menzionata in precedenza. La studiosa ha quindi creato nuove categorie per classificare tutti i *bundles* trovati nei *corpora*. Questi *bundles* erano: a) nella scrittura accademica in storia: *bundles* referenziali quantitativi, *bundles* referenziali legati al soggetto e altri *bundles*; b) nella scrittura accademica in biologia: *bundles* referenziali quantitativi, *bundles* referenziali statistici, *bundles* referenziali descrittivi, altri *bundles* posizionali, e altri *bundles*.

Queste scoperte confermano quelle di Hyland (2008a) circa il fatto che i *corpora* specifici richiedono un quadro analitico più specifico che comprenda tutte le caratteristiche della prosa accademica. A partire dalla classificazione di Biber ed altri (2004) e Biber (2006), Hyland (2008a) ha sviluppato un quadro analitico che raggruppa i *bundles* in tre ampi ambiti: a) *bundles* orientati alla ricerca; b) quelli orientati al testo; e c) quelli orientati ai partecipanti. Secondo Hyland (2008a), il quadro analitico proposto da Biber non potrebbe essere applicato in toto a *corpora* specifici composti da articoli di ricerca, tesi dottorali e tesi di master nelle scienze sociali e naturali, quali l'ingegneria elettrica, la biologia, le scienze aziendali e la linguistica applicata.

Come riportato dall'autore, il *corpus* di registri scritti e parlati analizzati da Biber ed altri (2004) che includevano conversazioni casuali, libri di testo, compendi universitari, rapporti di servizio, testi istituzionali e altro, sembra aver prodotto più *bundles* personali, referenziali e direttivi rispetto al *corpus* di prosa accademica (Hyland, 2008a).

Nel *corpus* di registri scritti e parlati, una delle categorie più importanti è la posizione (*stance*), al di là dei *bundles* referenziali e degli organizzatori di testo, mentre Hyland l'ha incorporata in un raggruppamento in cui i *bundles* si riferiscono o allo scrivente o al lettore. Inoltre ha introdotto delle sottocategorie che riflettono in maniera specifica le preoccupazioni della scrittura orientata verso la ricerca.

La tassonomia proposta da Hyland (2008a), la quale ha raggruppato i *lexical bundles* del linguaggio scientifico, è stata adoperata anche per la presente ricerca. La classificazione sotto riportata indica le funzioni proposte da Hyland e gli esempi per la lingua inglese, tratti dalla ricerca di Hyland (2008a: 12).

- a. ***Lexical bundles orientati alla ricerca*** – aiutano lo scrivente a strutturare la propria attività e la propria esperienza del mondo reale:

- collocazione: indica il tempo e il luogo (*at the same time, in the present study*);
- procedura: *the use of the, the operation of the*;
- quantificazione: *a wide range of, one of the most*;
- descrizione: *the structure of the, the size of the*;
- argomento della ricerca: *in the Hong Kong*;

**b. *Lexical bundles orientati al testo*** – riguardanti l'organizzazione del testo e il suo significato in quanto messaggio o argomento:

- segnali di cornice (specificano argomenti delimitandone le condizioni): *in the case of, with respect to the*;
- segnali di transizione (stabiliscono collegamenti addizionali o contrastivi fra gli elementi): *on the other hand, in addition to the*;
- segnali di risultato (segnalano relazioni inferenziali o causative fra elementi): *as a result of, it was found that*;
- segnali strutturali (segnali di riflessione sul testo che organizzano parti del discorso o lettura diretta in altri punti del testo): *in the next section, as shown in the figure*;

**c. *Lexical bundles orientati al partecipante*** – si concentrano sullo scrivente e sul lettore del testo e includono:

- caratteristiche posizionali (trasmettono le valutazioni e gli atteggiamenti dell'autore): *are likely to be, it is possible that*;
- caratteristiche di coinvolgimento (si riferiscono in maniera diretta al lettore): *as can be seen, it should be noted that*.

Come già accennato in precedenza, la classificazione proposta da Hyland (2008a) è stata adottata anche per il presente studio. I *lexical bundles* sono stati raggruppati in base alle loro funzioni discorsive, considerando le caratteristiche specifiche di ciascun ambito microlinguistico in questione – non tutte le categorie sono state identificate in tutti e tre gli ambiti scientifici.

Nel *corpus* di storia, per esempio, la maggior parte dei *lexical bundles* identificati rispecchia una connessione intrinseca all'essenza della disciplina stessa. Questi *bundles*

si riferiscono soprattutto al tema centrale della ricerca o agli argomenti inerenti alla ricerca, come ad esempio *la storia dell'immigrazione, i figli di immigrati*, oppure indicano il periodo – *alla fine del secolo, nel corso dell'Ottocento* – e il luogo dell'evento storico – *in tutto il mondo, nel paese d'origine*. Tenendo conto della numerosità dei *lexical bundles* temporali e locali si è dovuto creare due classi separate, mentre le categorie che denotano la procedura e descrivono gli oggetti di ricerca sono state omesse.

Inoltre, dai *lexical bundles* orientati al testo, ovvero riguardanti l'organizzazione testuale e il suo significato, è stato creato un sottogruppo per tutti e tre gli ambiti scientifici. Questa classe riguarda l'organizzazione testuale al microlivello, più in particolare indica delle connessioni interne ai paragrafi o tra due frasi contigue, stabilisce collegamenti addizionali di tipo causativo, contrastivo o consecutivo o enfatizza elementi adiacenti. A titolo d'esempio possiamo indicare i *bundles la necessità di un, la possibilità di un, l'idea di un, l'esistenza di un/una*. Gli stessi *bundles* – *la idea de la, la existencia de un/una* – sono stati riscontrati anche in lingua spagnola. La ricercatrice (Cortes, 2008) li ha raggruppati sotto *Referential Intangible Framing Attributes*, ovvero sotto *clusters* che si riferiscono alla ricerca e la delimitano. Nella presente ricerca si è optato per una classe separata per poter meglio definire le funzioni discorsive dei *lexical bundles* settoriali.

Oltre a ciò, sono stati identificati anche *bundles* – seppure in un numero abbastanza esiguo – senza una funzione discorsiva individuabile, come ad esempio *in cui ci si, per il quale il*. Questo gruppo è stato riscontrato nel *corpus* di filosofia e medicina del lavoro. *Lexical bundles* di questo tipo sono stati rilevati solo in lingua spagnola, indicando una somiglianza tra le due lingue romanze.

Sulla base delle ricerche condotte, Biber e altri (2004) hanno riscontrato l'esistenza di un'associazione diretta fra forma e funzione nei *lexical bundles*. In qualità di esempio, hanno elencato le frasi nominali e proposizionali, i meccanismi grammaticali principali usati in inglese per le funzioni referenziali, che sono diventati fissi come *lexical bundles* referenziali. L'analisi ha dimostrato che la maggior parte dei *bundles* posizionali sono composti da frammenti di clausole dipendenti, poiché in inglese in generale i *lexical bundles* composti da clausole dipendenti sono fra i meccanismi grammaticali più produttivi quando si vuole esprimere la posizione

(*stance*). Nella ricerca precedente, datata 1999, i *bundles* più frequentemente rilevati nella prosa accademica e nella conversazione erano stati classificati secondo la loro struttura, giungendo alla conclusione che la conversazione, in quanto registro orale, per la funzione di posizione usa prevalentemente *lexical bundles* verbali e di clausole dipendenti; d'altro canto, la prosa accademica, in quanto registro scritto, per le funzioni referenziali usa prevalentemente frasi nominali e frasi preposizionali basate sui *bundles*. I ricercatori vedono una complessa interazione fra la forma strutturale, la funzione del discorso e gli scopi tipici e le caratteristiche situazionali dei registri.

### **I.6.3 L'importanza dei *lexical bundles* nella produzione linguistica fluente**

Secondo Biber ed altri (2004) i *lexical bundles* dovrebbero essere considerati come costrutti linguistici di base, che svolgono funzioni importanti per la costruzione del discorso. Anche se definiti strettamente sulla base della frequenza, e non corrispondendo al tipo di strutture grammaticali riconosciute dalla teoria linguistica tradizionale, tutti i *lexical bundles* sono chiaramente interpretabili in termini sia strutturali che funzionali e la maggior parte di questi hanno strutture correlate ben definite. I ricercatori (Biber ed altri, 2004) vedono i *lexical bundles* come "cornici" strutturali, che funzionano come ancore del discorso per le informazioni che seguono in forma di "blocchi" che dicono al lettore o all'ascoltatore come interpretare una data informazione circa i criteri di posizione (*stance*), di organizzazione del discorso e di status referenziale. La prova più generale circa l'importanza dei *lexical bundles* proviene dalla loro frequenza d'uso e da ovvie funzioni del discorso. Un esame di questi modelli di alta frequenza nel contesto testuale mostra che essi sono blocchi costitutivi del discorso estremamente importanti e che servono come strumenti di inquadramento del discorso: forniscono un tipo di cornice che esprime la posizione, l'organizzazione del discorso o lo status referenziale associato con un dato blocco per l'espressione di nuove informazioni relative a quella data cornice (Biber ed altri, 2004).



Ciò che la ricerca di Biber e altri (2004: 399) indica è che "lexical bundles are a fundamentally different kind of linguistic construct from productive grammatical constructions<sup>26</sup>".

Anche se solitamente i *lexical bundles* non sono unità strutturali complete e si può risalire al loro significato tramite le parole che compongono il *bundle*, questi aiutano a dare forma ai significati del testo e contribuiscono al nostro senso di distintività in un dato registro, funzionando quindi come unità del discorso. La presenza di *lexical bundles* come *it should be noted that, as can be seen* aiuta a identificare un testo come appartenente al registro accademico, mentre i *lexical bundles* come *with regard to, in accordance with* distinguono più tipicamente un testo legale (Hyland, 2008b). Secondo Nattinger e DeCarrico (1992), si ritiene che le combinazioni frequenti di parole che svolgono determinate funzioni del discorso abbiano un ruolo importante nella produzione linguistica fluente. I linguisti sostengono che questi siano segnalatori di primaria importanza che indicano la direzione del linguaggio scritto e parlato. Si è considerato l'uso delle collocazioni e delle espressioni fisse come un segnale di uso avanzato del linguaggio di un dato registro, che segnala la partecipazione competente a una data comunità e a un dato registro (Howarth, 1998a). Come sostengono Simpson-Vlach e Ellis (2010) la conoscenza delle espressioni formulaiche viene vista come parte cruciale nella produzione linguistica fluente. I linguisti affermano che la conoscenza e la produzione automatica di tali espressioni è la caratteristica del linguaggio dei parlanti nativi.

Secondo Cortes (2004: 398), l'uso frequente dei *lexical bundles* sembra segnalare un uso competente del linguaggio in un dato registro al punto che "that learning conventions of register use may in part consist of learning how to use certain fixed phrases<sup>27</sup>". Come sostenuto da Pawley e Syder (1983), i benefici di un linguaggio precostituito nel ridurre lo sforzo produttivo possono spiegare il perché un individuo o invero un'intera comunità di parlanti arriva a preferire certe collocazioni ed espressioni di un'idea ad altre espressioni ugualmente permissibili. Alla competenza collocazionale

---

<sup>26</sup>Traduzione: "i *lexical bundles* sono un costrutto linguistico completamente diverso dai costrutti grammaticali produttivi".

<sup>27</sup>Traduzione: "l'apprendimento di un registro possa in parte consistere nello studio di certe combinazioni di parole".

viene attribuito un ruolo fondamentale nell'apprendimento delle lingue (Nesselhauf, 2005). Haswell (1991: 236; in Hyland, 2008b: 42) suggerisce che:

"there can be little doubt that as writers mature they rely more and more on collocations and that the lesser use of them accounts for some characteristic behaviour of apprentice writers. Gaining control of a new register therefore requires a sensitivity to expert users' preferences for certain sequences of words over others that might seem equally possible<sup>28</sup>".

Così, se imparare a utilizzare le espressioni fisse più frequenti di una disciplina può contribuire ad acquisire competenza comunicativa nel campo di studio, ci possono essere dei vantaggi nell'identificare questi *clusters* così da poter aiutare gli studenti ad acquisire delle capacità retoriche specifiche. Secondo Hyland (2008b), gli scriventi hanno bisogno di una certa familiarità sia con i *clusters* specifici della disciplina e con quelli valutati nei generi particolari.

In altre parole, acquisire il controllo di un registro richiede una sensibilità verso le preferenze degli utilizzatori esperti circa alcune combinazioni di parole e collega l'importanza di un loro uso appropriato nell'acquisizione di competenze comunicative e pragmatiche. L'identificazione di queste combinazioni di parole ricorrenti può contribuire ad aiutare gli studenti ad acquisire le tecniche retoriche specifiche di ciascuna comunità, come sottolinea Cortes 2004.

Inoltre, secondo Granger (2011), la natura formulaica di molte regole pragmlinguistiche ha contribuito allo studio di *pattern* prefabbricati, poiché a partire dagli anni '80 la competenza pragmatica viene vista come parte essenziale dell'apprendimento linguistico.

Gli studi svolti sui *lexical bundles* hanno mostrato che i *bundles* sono molto comuni sia nel linguaggio comune che specialistico. Essendo delle espressioni molto ricorrenti si potrebbe supporre che siano delle espressioni semplici e facili da imparare (Biber, Conrad, 1999: 188).

---

<sup>28</sup>Traduzione: "Non c'è ombra di dubbio sul fatto che, mano a mano che gli scrittori maturano, tendono ad affidarsi sempre più alle collocazioni e che un loro più sporadico uso è spiegabile dal comportamento caratteristico degli scrittori apprendisti. Raggiungere il controllo di un nuovo registro perciò richiede una sensibilità per le preferenze degli utilizzatori esperti verso certe sequenze di parole rispetto ad altre che possono sembrare ugualmente possibili".

Cortes (2004) ha provato a verificare questa ipotesi nel caso della produzione scritta di autori pubblicati e di studenti universitari, nel campo della scrittura accademica in storia e biologia. I risultati hanno mostrato come l'uso dei *bundles* da parte degli studenti di biologia e di storia nei corsi universitari di tre livelli differenti (primo anno del corso di laurea, secondo anno del corso di laurea e livello di master) è molto lontano dall'uso di quegli stessi *bundles* fatto dagli autori che pubblicano nelle suddette discipline. Le scoperte di queste ricerche hanno mostrato che molti *bundles* che risultano frequentemente usati nella scrittura accademica pubblicata in biologia non erano mai o quasi mai usati dagli studenti di quella disciplina a tutti i livelli, e che anche la maggioranza dei *lexical bundles* usati dagli autori che pubblicano nel campo della storia erano usati raramente o affatto dagli studenti della disciplina a tutti i livelli universitari.

Questi risultati dimostrano che sebbene gli studenti abbiano incontrato frequentemente queste espressioni nelle loro letture accademiche, la semplice esposizione all'uso frequente dei *lexical bundles* nella scrittura accademica pubblicata non causa e non si riflette nell'acquisizione di queste espressioni da parte degli studenti universitari. Cortes (2004) ha riportato che non solo la frequenza era bassa, ma anche l'uso dei *lexical bundles* da parte degli studenti non sempre trasmetteva quelle funzioni tipicamente associate con la scrittura accademica pubblicata in storia e biologia. L'ipotesi proposta da Biber ed altri (1999) è stata confutata dai risultati di questa ricerca, rilevando, in questo modo, un punto interessante che dovrebbe essere integrato nell'ambito dell'insegnamento delle microlingue per sviluppare le abilità della lettura e scrittura da parte degli studenti, come mette in rilievo Cortes (2004).

#### **I.6.4 Le ricerche sui *lexical bundles***

Come già accennato precedentemente, la maggior parte delle ricerche sui *lexical bundles* si è concentrata sulla lingua inglese. Per quanto riguarda le altre lingue, esistono alcuni studi, anche se in numero esiguo. Tremblay, Derwey e Libben (2007) riferiscono che il concetto di *lexical bundles* risale a Salem (1987) e alle ricerche da lui

condotte su un *corpus* di testi governativi francesi – questo è l'unico studio svolto per la lingua francese di cui siamo a conoscenza.

Butler ha in seguito utilizzato la nozione di sequenze ripetute per analizzare la lingua spagnola, mentre più recentemente i ricercatori hanno iniziato a comparare i *lexical bundles* in lingua inglese e spagnola (Butler, 1997; Tracy-Ventura, 2007; Cortes, 2008). Kim (2009) che ha studiato i *lexical bundles* in lingua coreana ha scelto di integrare la ricerca con una lingua con diverse caratteristiche tipologiche.

Nelle sezioni seguenti proponiamo una sintetica rassegna delle ricerche svolte, concentrandoci soprattutto sui risultati emersi dagli studi più rilevanti per la presente analisi.

### **I.6.5 Le ricerche in lingua inglese**

Per quanto riguarda la lingua inglese, lo studio sui *lexical bundles* comincia già negli anni '90 con Altenberg e Eeg-Olofsson (1990) e con lo studio di Altenberg (1998). Lo studio preliminare di Altenberg e Eeg-Olofsson (1990) circa le sequenze di più parole nel *London Lund corpus*, rappresenta una piattaforma per lo studio di Altenberg (1998), svolto sulle sequenze di tre parole che occorrono per almeno dieci volte nel *London Lund Corpus*. Seguendo la metodologia di questo studio, si è sviluppata la metodologia di ricerca sui *lexical bundles*.

Il termine *lexical bundles* entra nella letteratura linguistica con l'opera Longman Spoken and Written English nel 1999. Dal 1999, è stata lanciata una serie di studi basati sui *corpora* e per lo più comparativi volti a esplorare le possibili differenze e similitudini nell'uso dei *bundles* attraverso i registri parlati e scritti, fra diversi campi disciplinari, e fra diversi gradi di maestria di scrittura, da un lato fra autori pubblicati e studenti universitari, parlanti non nativi, dall'altro fra parlanti nativi, esperti e novizi (Jalali ed altri, 2010).

Alcuni studi sono di interesse particolare poiché si occupano dei *lexical bundles* nella scrittura accademica generale (Biber, Conrad, 1999; Biber ed altri, 1999) o in discipline accademiche specifiche (Cortes, 2002; Cortes, 2004; Partington e Morley, 2004; Hyland, 2008a, 2008b).

Una caratteristica importante dei *bundles* è la loro variabilità attraverso generi diversi. Impiegando metodi d'inchiesta empirici sviluppati da loro stessi, Biber e i colleghi (1999) hanno identificato dei *lexical bundles* di diversa lunghezza, da 3 a 6 parole, identificando quelli di 4 parole come il tipo predominante e hanno prodotto una serie di liste dei *lexical bundles* più frequenti in diversi registri, compresa la scrittura accademica. Nello studio delle variazioni fra registri diversi, Biber e colleghi (1999) hanno paragonato la conversazione e la prosa accademica, usando dei campioni testuali estratti dal *Longman Corpus of Written and Spoken English*. La ricerca ha dimostrato che i correlati grammaticali dei *lexical bundles* nella conversazione sono decisamente diversi da quelli della prosa accademica.

Nella conversazione, almeno il 90% dei *lexical bundles* incorporano frasi verbali, la cui maggioranza inizia con un pronome personale + frase verbale (per esempio *I was going to*); a seguire vengono i *lexical bundles* che consistono di un frammento di frase verbale estesa (come *have a look at*) e i *lexical bundles* che sono frammenti di domanda (per esempio *I have a*). I *lexical bundles* che appaiono nella prosa accademica sono frasali piuttosto che clausali. Circa il 70% dei *bundles* nella prosa accademica è costituita da espressioni a frase nominale (esempio *the nature of*) o da una sequenza che fa da ponte fra due frasi preposizionali (esempio *as a result of*).

Il lavoro pionieristico di Biber e colleghi ha posto le fondamenta della ricerca sui *lexical bundles* e ha portato a numerosi e fruttuosi studi sull'argomento, da parte di altri studiosi. Per esempio, la ricerca di Partington e Morley (2004) che hanno studiato l'uso dei *lexical bundles* nel discorso politico parlato ha mostrato come i *lexical bundles* possano essere usati per analizzare la distintività di certi registri particolari.

Nella presente ricerca, si dà particolare importanza agli studi di Cortes (2002, 2004, 2008) e di Hyland (2008a, 2008b). Cortes ha analizzato i *lexical bundles* nella prosa accademica, su dei *corpora* costituiti da articoli di ricerca e ha riscontrato che gli articoli di ricerca di biologia utilizzano *lexical bundles* molto di più che quelli di storia, attribuendo la differente salienza al diverso campo scientifico, alla scienza naturale da un lato e scienza umanistica dall'altro.

Cortes ha anche rilevato che molti *lexical bundles* sono circoscritti a una data disciplina e svolgono funzioni discorsive e contestuali differenti in discipline accademiche differenti. I risultati di questa ricerca mostrano che i *lexical bundles* trovati

in storia e biologia si differenziano strutturalmente – in storia, i *lexical bundles* appartengono a due gruppi strutturali (frasi nominali e frasi preposizionali), mentre in biologia i *lexical bundles* coprono un'ampia gamma di gruppi strutturali e funzionali (Cortes, 2004: 410). I diversi *lexical bundles* identificati in entrambi i *corpora* erano utilizzati per svolgere funzioni simili.

Un'altra differenza fra le discipline menzionata da Cortes è la varietà dei *markers* di posizione epistemica-impersonale/probabile-possibile usati in biologia (esempio *are likely to be, is likely to be, it is possible that, the probability that the*), che venivano usati in molti casi per proteggere l'effetto di un'affermazione o per rendere più temporanea un'affermazione o un argomento mentre, dall'altra parte, questi *bundles* non sono usati frequentemente da scrittori che pubblicano articoli di storia, né lo sono altri *lexical bundles* con queste stesse funzioni. Secondo Cortes, questi risultati suggeriscono che gli autori di storia fanno uso di singoli elementi di vocabolario o di altre espressioni per esprimere la stessa funzione.

Nel suo studio successivo e complementare, Cortes (2008) ha paragonato l'uso dei *lexical bundles* nella scrittura accademica di storia in inglese e in spagnolo, utilizzando *corpora* paragonabili, ovvero collezioni di testi autentici per ciascuna lingua. Le analisi hanno mostrato che i *bundles* identificati in ogni lingua presentano molte caratteristiche comuni. Mentre un primo gruppo può essere visto come il risultato di traduzioni dirette, un secondo gruppo dimostra che c'è un alto livello di accordo fra i tipi strutturali identificati in entrambi i *corpora*. La maggior parte dei *lexical bundles* in entrambi i *corpora* erano frasi proposizionali e la seconda categoria più frequente era la frase nominale. Questi risultati coincidono con degli studi di ricerca precedenti nel campo dei *lexical bundles* che sostengono che la maggior parte dei *bundles* nella scrittura accademica in generale, e negli articoli di ricerca in particolare, sono essenzialmente frasali. In conclusione, un'analisi funzionale ha mostrato che i *lexical bundles* in entrambi i *corpora* hanno svolto funzioni simili; in entrambi i *corpora* la maggior parte dei *bundles* erano referenziali. La ricerca ha anche mostrato che alcuni *bundles* sono storia-dipendenti, vale a dire che hanno delle funzioni connesse alla prosa accademica e all'essenza della disciplina e che presentano parole di contenuto che rivelano gli argomenti più frequentemente discussi nei testi del *corpus*.

Il lavoro di Hyland (2008a) è particolarmente importante poiché mostra come e fino a che punto i *lexical bundles* variano da disciplina a disciplina. Hyland analizza un *corpus* di 3.5 milioni di parole composto da articoli di ricerca, tesi di dottorato e tesi master di quattro discipline, quali la biologia, la linguistica applicata, gli studi aziendali e l'ingegneria elettrica.

I risultati della sua ricerca mostrano non solo che i *lexical bundles* costituiscono un punto centrale nella produzione del discorso accademico ma anche che variano di disciplina: diversi tipi di *lexical bundles* vengono adoperati nelle rispettive discipline. L'analisi ha mostrato che i testi di ingegneria elettrica contenevano la gamma più ampia di *bundles*, con 213 *bundles* diversi composti da 4 parole che soddisfano la soglia di 20 per milione di parole e anche con la più alta proporzione di parole nei testi che si presentano in *bundles* di 4 parole. Molti *lexical bundles* trovati nel *corpus* di ingegneria elettrica non sono stati trovati in altre discipline, il che potrebbe essere, secondo Hyland, una conseguenza della natura grafica e astratta della comunicazione tecnica. Secondo Hyland la densità dei *lexical bundles* nel *corpus* di ingegneria elettrica potrebbe riflettere la dipendenza della retorica ingegneristica dalla rappresentazione visuale e grafica, così, mentre gli argomenti si basano su interpretazioni plausibili dei dati, si rifanno in ultima analisi ai risultati che sono spesso presentati in forma visiva. Dall'altro lato la biologia adopera un numero minimo di *lexical bundles* – la proporzione è simile a quella nelle due discipline delle scienze sociali. Secondo Hyland le ragioni non sono del tutto chiare. Difatti la biologia come l'ingegneria elettrica adopera il visuale per argomentare, ma è una disciplina più discorsiva e descrittiva con un programma meno attivo e meno applicato, una disciplina che nomina e codifica più di altre scienze. Tra i top 50 della linguistica applicata ci sono 28 *lexical bundles* che sono legati alla disciplina e che non servono in nessun'altra disciplina; tra i top 50 dell'ingegneria elettrica ce ne sono 28, inoltre più della metà dei *lexical bundles* non serve in nessun'altra disciplina e solo il 30% dei *lexical bundles* di ogni singola disciplina si può trovare in due altre discipline. I *lexical bundles* riscontrati in tutte e quattro le discipline sono solo quattro, in tre invece ce ne sono solo 14. Gli studi aziendali e la linguistica applicata hanno 18 *lexical bundles* in comune, mentre la biologia e l'ingegneria elettrica ne hanno 16. Tra i *lexical bundles* più frequenti nel campo della biologia e

dell'ingegneria elettrica ci sono soprattutto sequenze di parole che si riferiscono ai grafici, alle tabelle o ai risultati della ricerca.

Hyland specifica che questi risultati indicano l'esistenza delle somiglianze nei *lexical bundles* fra campi affini – ingegneria elettrica e microbiologia da un lato e studi commerciali e linguistica applicata dall'altro. Oltre alle frequenze dissimili, i *corpora* mostrano che le strutture principali dei *bundles* si differenziano anche attraverso i campi, ma d'altro canto i dati confermano le precedenti ricerche della letteratura sui *lexical bundles* che indicano che la maggior parte dei *lexical bundles* nella scrittura accademica sono essenzialmente frasali. I dieci *bundles* più ricorrenti in tutta la scrittura accademica sono parti di nomi o frasi preposizionali. In tutti i *corpora* i *bundles* nominali sono generalmente più comuni, comprendendo circa un quarto di tutte le forme del *corpus*. I *bundles* coprono un'ampia gamma di significati nel discorso accademico e sono utilizzati in particolare per identificare la quantità, una collocazione locale o la taglia di specifiche qualità. Più interessante è la differenza fra le discipline, con i *corpora* di scienze sociali che fanno ben più ampio uso di *bundles* che iniziano con una frase preposizionale, con un'enfasi sull'esplorazione discorsiva di condizioni di possibilità e di limitazione, identificando ed elaborando le relazioni nell'argomento.

D'altro canto, i testi di ingegneria e biologia impiegano molti più *bundles* passivi, tramite i quali gli scriventi cercano o di guidare i lettori attraverso il testo, o di identificare la base di un'asserzione in un argomento. Sono state riscontrate anche differenze funzionali fra le discipline. Una delle differenze è la più alta concentrazione di *bundles* orientati alla ricerca in biologia e ingegneria elettrica, una preferenza che ammonta a quasi la metà di tutti i *bundles* in questi *corpora*. Hyland indica che nelle scienze naturali la scala di quest'uso serve a impartire alla scrittura un senso maggiore di aderenza al reale e di maggiore importanza ai dati di laboratorio. Al contrario, i *corpora* di linguistica applicata e di economia e commercio erano dominati da *bundles* orientati al testo. Come indicato da Hyland, questo risultato riflette le tendenze più discorsive e valutative dell'argomentazione nelle scienze umanistiche, laddove la persuasione è più esplicitamente più interpretativa e meno basata sull'empirismo, poiché produce discorsi promuovendo la tolleranza fra i lettori tramite una progressione etica piuttosto che cognitiva (Hyland, 2004: 145).



Lo studio di Hyland (2008b) indica che gli scriventi in diversi campi si rifanno a risorse diverse per sviluppare i propri argomenti, per consolidare la propria credibilità e per persuadere i propri lettori – una scoperta confermata anche da altri suoi studi complementari più tardivi.

Un'altra pubblicazione importante e degna di nota è il recente studio di Liu (2012) sui *bundles* più comuni nella scrittura accademica generale, basato su sottocorpora di scrittura accademica del *corpus* COCA, composto da 82.91 milioni di parole e sul *corpus* BNS, composto da 15.33 milioni di parole. Questo studio mira a identificare i costrutti multiparola più comuni nella scrittura accademica, ivi compresi i *lexical bundles*, gli idiomi e i verbi frasali e proposizionali. Poiché i due *corpora* possono essere interrogati solo tramite i propri motori di ricerca interni, la ricerca di espressioni multiparola è stata fatta usando le funzioni di ricerca messe a disposizione, che non includono l'estrazione automatica delle espressioni (Liu, 2012). Come riportato da Liu (2012), le fonti pubblicate in inglese accademico scritto trovate in altri autori sono state usate come banca dati per individuare quali costruzioni multiparola dovessero essere usate, usando distribuzioni di frequenza e di gamma nel selezionare le espressioni più importanti – sono stati usati i risultati di Biber ed altri, 1999; Carter, McCarthy, 2006; Simpson-Vlach, Ellis, 2010).

Per quanto riguarda la distribuzione per gamma, lo studio ha seguito il principio basilare di Simpson-Vlach e di Ellis (2010): un'espressione multiparola deve avere 20 entrate per milione di parole in sei su otto divisioni accademiche nel COCA (American English) oppure in cinque su sei divisioni accademiche nel BNC (British English) (Liu, 2012: 29). Nella ricerca, vi sono due risultati particolarmente importanti: lo studio concorda con studi precedenti sul fatto che le costruzioni nominali e preposizionali costituiscono le due tipologie più ampie di espressioni multiparola nell'inglese accademico scritto, come nel caso degli studi di Biber e colleghi (1999). Il test di gamma ha evitato che alcune costruzioni multiparola frequenti costituissero la lista delle espressioni multiparola, poiché non avevano una frequenza sufficientemente alta o nelle divisioni di scienze umane/sociali o in quelle delle scienze naturali/ingegneria, confermando così le precedenti scoperte circa la specificità disciplinare dei *lexical bundles*, come indicato da Cortes, 2002, 2004, 2008; Hyland, 2008a, 2008b). Si tratta di una scoperta fondamentale ai fini della nostra ricerca.

Come riportato dalla ricercatrice, un altro nuovo risultato è che non solo la lista di Liu non comprendeva un notevole numero di elementi negli elenchi esistenti delle espressioni multiparola più comuni nella scrittura accademica generale, ma questi mostravano anche una frequenza molto bassa nella scrittura accademica generale. Secondo Liu (2012: 31), la discrepanza fra i risultati sulla frequenza delle espressioni multiparola specializzate sembra suggerire l'importanza e il bisogno di usare dei *mega corpora* e di controllare la distribuzione di gamma per ogni divisione nello studio delle espressioni multiparola nella scrittura accademica generale. D'altro canto, alcuni risultati precedenti concordano con questi nuovi risultati sul fatto che le espressioni multiparola sono legate alla disciplina specifica – come gli studi di Cortes 2002, 2004, 2008; Hyland, 2008a, 2008b.

Importanti sono anche le pubblicazioni che si concentrano sulla comparazione dell'uso dei *lexical bundles* adoperati da esperti in materia e da futuri esperti, quali gli studenti universitari (Cortes, 2004; Hyland, 2008b). Lo studio di Cortes (2004) si presenta particolarmente importante mettendo in evidenza la differenza nell'uso dei *lexical bundles* nella scrittura accademica pubblicata e studentesca in storia e in biologia, con riferimento all'uso delle combinazioni di parole frequenti specifiche della disciplina, cercando di contribuire alla descrizione della scrittura in queste discipline, esaminando l'uso degli studenti a tre livelli: laurea di primo livello e di secondo livello e al livello dottorale. Il suo studio mostra che molti *lexical bundles* usati dagli esperti in questi due campi, così come esaminati dai rispettivi articoli di ricerca, non erano mai o affatto usati da nessuno dei due gruppi di studenti. Si riscontrava, infatti, solo nel caso degli studenti di biologia una tendenza graduale verso lo sviluppo dell'uso di questi *bundles*. Anche la ricerca di Hyland (2008a), basata su *corpora* di articoli di ricerca e sui *corpora* di studenti universitari di lingua madre cantonese presso le università di Hong Kong, ha dimostrato che i tre generi analizzati – articoli di ricerca, tesi di master e tesi di dottorato – si affidavano a diversi tipi e numeri di *bundles*, laddove le tesi di master impiegano i *bundles* più che le tesi di dottorato e molto più degli articoli di ricerca (Hyland, 2008a).

Questo studio ha dimostrato che, strutturalmente, e a differenza che negli articoli di ricerca, i *bundles* nei generi studenteschi erano più frasali che clausali. Funzionalmente i *bundles* nelle tesi di master erano fortemente orientati alla ricerca,

mentre i *bundles* negli articoli di ricerca erano in gran parte orientati al testo. I *bundles* nelle tesi di dottorato erano più simili a quelli degli articoli di ricerca, ma allo stesso tempo i *clusters* temporali negli articoli di ricerca erano più orientati al partecipante rispetto a quelli delle tesi di dottorato. Lo studio conclude dicendo che gli autori meno competenti e sicuri di sé tendono ad affidarsi di più alle espressioni formulaiche. Secondo Jalali e colleghi (2008), i generi accademici di ogni disciplina dovrebbero essere analizzati separatamente al fine di identificare le somiglianze e le differenze nell'uso dei *clusters* lessicali da parte dei novizi e degli esperti. Questi studiosi hanno deciso di analizzare, comparare e mettere a confronto da un lato la frequenza, la struttura e la funzione dei *clusters* di 4 parole negli articoli di ricerca e dall'altro le tesi di master e di dottorato come campione della scrittura a livello di istruzione specialistica nell'ambito di una sola disciplina: la linguistica applicata.

Questo studio rappresenta il primo tentativo di studiare esclusivamente l'uso dei *bundles* all'interno di una singola area disciplinare, indicando le somiglianze e le differenze fra esperti e novizi – gli studenti di dottorato iraniani di madrelingua persiana presso l'università di Isfahan, in Iran.

Lo studio menzionato ha mostrato alcune differenze e alcune somiglianze fra gli articoli di ricerca e la scrittura a livello dottorale. Entrambe si assomigliano molto poiché fanno forte affidamento sui *bundles* frasali piuttosto che su quelli clausali. La differenza principale sta nel fatto che il numero di *lexical bundles* usati nella scrittura dottorale è due volte più grande che quello usato negli articoli di ricerca. L'altra discrepanza risiede nella varietà di alcuni gruppi strutturali di *bundles* utilizzati.

Come riportato dai ricercatori, lo studio ha mostrato che nella scrittura dottorale le gamme di *bundles* di frasi nominali o di frasi preposizionali sono molte di più di quelle degli articoli di ricerca. Funzionalmente, gli articoli di ricerca e la scrittura dottorale sono simili poiché entrambi attingono a *bundles* orientati alla ricerca più che a *bundles* orientati al testo o al partecipante, ma in termini di *bundles* orientati alla ricerca la differenza principale fra i due generi è che la scrittura dottorale si affida a una gamma di *bundles* che è due volte maggiore di quella degli articoli di ricerca. I risultati della ricerca confermano che i novizi adoperano un numero maggiore dei *lexical bundles*. Va inoltre ricordato che la sola esposizione ai *lexical bundles* non è sufficiente per la loro

interiorizzazione – come evidenziato dagli studi di Cortes (2004) –, un punto che dovrebbe essere approfondito nell'insegnamento delle microlingue.

Secondo Jalali e colleghi (2008) questo studio ha mostrato che i *lexical bundles* potrebbero essere discriminatori di genere relativamente forti poiché sia gli articoli di ricerca che le tesi dottorali comprendevano delle serie di *bundles* specifici e quasi diversi nel proprio discorso. I risultati di questo studio, confermando lo studio di Biber e altri (1999, 2003, 2004), hanno mostrato che gli accademici pubblicati usano più *bundles* frasali – una caratteristica distintiva della scrittura accademica. Anche la classificazione strutturale e la comparazione dei *bundles* ha corroborato i risultati di studi precedenti (Biber ed altri, 1999, 2004), cioè che a differenza della conversazione, la prosa accademica mostra una forte tendenza all'uso di *bundles* frasali piuttosto che di *bundles* clausali (Biber ed altri, 1999, 2004). Lo studio ha evidenziato che almeno l'80% dei *bundles* totali utilizzati in entrambi i *corpora* comprendevano diverse frasi nominali e preposizionali.

Gli studi svolti nell'ambito del discorso scientifico scritto da un lato da esperti in materia e dall'altro da futuri esperti hanno confermato i risultati delle ricerche condotte per il discorso accademico, le quali sostengono che una delle peculiarità del linguaggio scientifico è il carattere nominale dei *lexical bundles*.

Sembra che la padronanza di una microlingua consista anche nella scelta di espressioni multiparola caratteristiche di un dato ambito scientifico. Come già precedentemente accennato, il contesto in cui incontriamo le parole gioca un ruolo assai importante nella memorizzazione e nel richiamo delle combinazioni di parole. Come sostenuto da Hyland (2008b: 46), il contesto collocazionale "normale" o le combinazioni di parole che riteniamo "normali" dipendono dalle comunità di cui facciamo parte e dai generi testuali a cui siamo esposti. Secondo Hyland (2008b), i linguisti applicati e gli insegnanti di lingue straniere sono sempre più inclini a vedere i *clusters* come costituenti fondamentali di ogni discorso coerente e come caratteristici dell'uso della lingua in circostanze particolari.

### *I.6.5.1 Le ricerche in altre lingue*

Come già più volte evidenziato, lo studio sui *lexical bundles* si è limitato soprattutto all'analisi della lingua inglese, anche se alcune analisi dei registri in altre lingue sono state svolte per la lingua spagnola, francese e coreana.

Nel 1997, Butler ha analizzato le sequenze ripetute in un *corpus* per la lingua spagnola, composto da una varietà di testi appartenenti a diversi registri, da interviste a giornali quotidiani. L'analisi funzionale delle combinazioni di parole ricorrenti identificate nei registri menzionati ha mostrato un più alto numero di sequenze ripetute di quattro parole nei registri parlati che nel singolo registro scritto dei quotidiani.

La ricerca complementare di Tracy-Ventura, Cortes e Biber (2007) ha esaminato l'uso dei *lexical bundles* in due registri diversi: il registro accademico e le interviste sociolinguistiche. La ricerca ha mostrato che il discorso accademico presenta più tipi di *lexical bundles* e che, inoltre, i *bundles* nel discorso accademico sono più frequenti. La ricerca strutturale dei *lexical bundles* ha seguito la tassonomia proposta da Biber ed altri (1999), mostrando che il discorso accademico predilige i *lexical bundles* nominali e preposizionali ai verbali. Da contro, le interviste sociolinguistiche hanno una preferenza per i *bundles* verbali. Questo tratto linguistico conferma gli studi precedenti sulla differenza tra la lingua scritta e quella parlata nella distribuzione dei nomi e verbi. A differenza dello studio di Butler (1997), la ricerca ha mostrato che il registro scritto usa un numero maggiore di *bundles* rispetto al registro parlato.

Tracy-Ventura e altri (2007) hanno paragonato i modelli di utilizzo attraverso i registri con la lingua inglese, riscontrando che la somiglianza più ovvia fra lo spagnolo e l'inglese era una forte associazione fra caratteristiche funzionali e caratteristiche strutturali dei *bundles*. A mo' di esempio, i ricercatori indicano l'uso basato sui *bundles* verbali per le funzioni di posizione e di organizzazione del discorso e quello basato sui *bundles* nominali per le funzioni referenziali. Inoltre le due lingue sono simili negli usi comunicativi dei *bundles*: la prosa accademica in inglese e in spagnolo usa i *lexical bundles* prevalentemente per funzioni referenziali, mentre i registri parlati usano i *bundles* in tutti i tre gruppi principali: posizione, organizzazione del discorso e funzione referenziale.

Lo studio ha mostrato che la prevalenza dei *lexical bundles* nominali non dipende solo dal registro ma anche dalla struttura della lingua stessa: a titolo d'esempio i

ricercatori indicano in lingua inglese il bundle nominale *the majority of* e in lingua spagnola due corrispondenti nominali *la mayoría de los, la mayoría de las*.

Circa l'uso dei *lexical bundles* nel campo dell'analisi contrastiva, un altro studio è quello condotto da Cortes (2008), che ha analizzato le strutture e le funzioni dei *lexical bundles* riscontrati nella scrittura accademica di storia in inglese e in spagnolo, usando *corpora* paragonabili costituiti da articoli di ricerca. Una scoperta interessante risiede nel fatto che il numero dei *lexical bundles* ritrovati nel *corpus* spagnolo di scrittura accademica di storia è più di due volte più grande di quello individuato nel *corpus* inglese. Secondo Cortes, questa differenza è causata da diversi fenomeni linguistici: i nomi spagnoli vengono premodificati solo raramente, poiché la modificazione del nome in spagnolo viene solitamente fatta usando gli aggettivi e le frasi preposizionali che post-modificano il nome; la modificazione nome-nome, estremamente frequente nella prosa accademica inglese, non è possibile in spagnolo e viene espressa tramite frasi preposizionali post-modificanti. La ricerca ha inoltre mostrato che molte combinazioni di quattro parole in lingua spagnola hanno un'equivalente espressione inglese in un'espressione più corta, poiché il plurale dei nomi in spagnolo viene formato con l'articolo determinativo (per esempio *de los derechos humanos*), la formazione del quale però non avviene in inglese (per esempio *of human rights*). Come stabilito da Tracy-Ventura ed altri (2007), è evidente che una volta identificata la diversa struttura linguistica, la distribuzione dei *lexical bundles* in due registri dell'inglese e dello spagnolo risulta più analoga.

L'analisi di Cortes (2008) ha mostrato che sebbene il numero dei *lexical bundles* nelle due lingue (l'inglese, lo spagnolo) fosse tanto disparato, c'era un alto livello di accordo fra i tipi strutturali identificati in entrambi i *corpora*. Le combinazioni di parole più ricorrenti in entrambi i *corpora* erano frasi preposizionali, per esempio *as a result of* e per lo spagnolo *en el contexto de*.

La seconda categoria più frequente era la frase nominale, con frase nominale + frase preposizionale come sottogruppo più frequente (per esempio *the beginning of the* e per lo spagnolo *el seno de la*). Come riportato da Cortes, in entrambi i *corpora* c'erano molti pochi *bundles* clausali – un'importante scoperta che conferma le precedenti ricerche sui *lexical bundles*, le quali hanno sostenuto che la maggior parte dei *bundles* nella scrittura accademica in generale, e negli articoli di ricerca in particolare, sono

essenzialmente *bundles* frasali. L'analisi funzionale ha mostrato che i *lexical bundles* identificati in entrambi i *corpora* svolgevano funzioni simili: gran parte di essi erano referenziali.

In vista di quanto detto precedentemente e in base agli studi svolti in lingua inglese e spagnola, possiamo concludere che i diversi tipi di *lexical bundles* e la loro distribuzione non dipendono solo dal registro adoperato, bensì dalle caratteristiche strutturali della lingua stessa.

Questi due studi hanno comparato i principali modelli strutturali e funzionali per i *lexical bundles* trovati nelle precedenti ricerche per la lingua inglese con quelli trovati per lo spagnolo, suggerendo che lingue diverse possono mostrare strutture o funzioni aggiuntive di *lexical bundles*.

Lo studio di Kim (2009) ha scelto di integrare la ricerca sui *lexical bundles* con l'esame dell'uso dei *lexical bundles* in coreano, una lingua con caratteristiche tipologiche differenti. Come in studi precedenti, è stato adottato un approccio basato sulla frequenza per identificare i *lexical bundles* nella prosa accademica e nella conversazione naturale in coreano, sulla base del *Corpus Sejong* sia scritto che orale. Il *corpus* di conversazione naturale usato per la ricerca comprendeva 2.604.054 parole e consisteva di conversazioni quotidiane prodotte spontaneamente e raccolte in contesti diversi, come le conversazioni studentesche all'università, gli incontri religiosi fra parrocchiani, le conversazioni fra studenti delle scuole superiori, i dialoghi nei ristoranti – mentre il *corpus* accademico conteneva 3.406.020 parole. La collezione di testi accademici consisteva di libri accademici pubblicati e di libri di testo che trattano diversi argomenti (Kim, 2009: 143). È stato scelto un punto di discriminazione della frequenza di 20 volte su un milione di parole. Una volta identificati, i *lexical bundles* sono stati esaminati a seconda della loro classificazione strutturale e funzionale, sulla base degli studi di Biber e altri (2004). È stato necessario modificare alcune classificazioni strutturali per poter descrivere alcune caratteristiche tipiche del coreano. In questo studio, l'analisi ha indicato che la prosa accademica coreana presenta un numero più alto di *lexical bundles* rispetto alla conversazione, conformemente agli studi di Tracy-Ventura (2007) per la lingua spagnola. Un'altra somiglianza riscontrata riguarda l'occorrenza dei *bundles* nominali, che sono risultati essere più comuni dei *bundles* verbali. La distribuzione dei tipi funzionali di *lexical bundles* in coreano si

differenzia in qualche modo dalla ricerca sull'inglese e sullo spagnolo, mostrando più somiglianze con le funzioni dei *bundles* inglesi (Kim, 2009: 146-156).

I risultati di questo studio sulla lingua coreana hanno fornito ulteriori indizi sull'occorrenza dei *lexical bundles* in lingue diverse, dimostrando che i *lexical bundles* sono componenti importanti del discorso, anche se non necessariamente idiomatiche né strutturalmente complete (Biber ed altri, 2006; Biber, Barbieri, 2007). Lo studio ha mostrato che l'occorrenza dei *lexical bundles* in una lingua è influenzata sia dalle risorse strutturali disponibili nella lingua che dalle caratteristiche dei registri. L'analisi della lingua coreana ha mostrato che, rispetto all'inglese e allo spagnolo, i *lexical bundles* sono generalmente rari nel complesso, data l'ampia gamma e varietà delle terminazioni delle parole (Kim, 2009). Come riportato da Kim (2009: 157), nella prosa accademica, che ha meno inflessioni rispetto alla conversazione, è stato riscontrato un ampio numero di diversi tipi di *bundles*.

## **I.7 Conclusioni**

Gli studi svolti in lingua inglese e spagnola, e anche quelli condotti su una lingua tipologicamente completamente diversa come il coreano, hanno mostrato che i *lexical bundles* non sono solo indicatori di registro diverso, bensì anche di disciplina scientifica diversa (Biber ed altri, 1999, 2004; Cortes, 2004, 2008; Hyland, 2008a, 2008b; Jalali ed altri, 2008; Tracy-Ventura ed altri, 2007). Questo significa che ogni registro e ogni disciplina attinge a una serie distinta di *bundles* al fine di organizzare il discorso, esprimere atteggiamenti e punti di vista e riferirsi a diverse parti del testo nel suo evolvere o a diversi elementi fuori dal testo. Inoltre, gli studi di Biber e altri (1999, 2003, 2004), di Cortes (2004, 2008), di Hyland (2008a, 2008b) e di Jalali ed altri (2008) hanno mostrato che, contrariamente alla lingua parlata, il registro accademico predilige i *lexical bundles* nominali, una caratteristica propria del discorso scientifico e della lingua scritta.

Sulla base di studi precedenti, condotti su altre lingue, si ipotizza che nella lingua italiana si riscontrerà da un lato la stessa frequenza dei *lexical bundles* e dall'altro la



stessa variazione attraverso i registri, nonché un'ulteriore variazione strutturale e funzionale attraverso gli ambiti scientifici diversi.

## **II Strumenti di analisi del linguaggio formulaico – la linguistica dei *corpora***

Il contributo apportato dai computer è enorme. Grazie agli strumenti informatici, i ricercatori possono estrarre dati che sarebbero altrimenti difficilmente prevedibili. Inoltre le informazioni possono essere ricavate in tempo reale, senza dover ricorrere alle analisi manuali.

Sono state messe in evidenza tre fasi nel processo di penetrazione dei *corpora* informatici nel lavoro linguistico. Dopo una prima fase in cui il computer è visto come un semplice strumento, sopravviene la fase in cui il computer ridefinisce il quadro metodologico dell'inchiesta poiché velocizza le operazioni, le sistematizza e le rende applicabili a grandi quantitativi di dati in tempo reale (Tognini-Bonelli, 2001: 47-48).

Gli strumenti informatici, grazie alla loro velocità e potenza, consentono di analizzare in tempo reale grandi quantità di dati. Questo non era nemmeno immaginabile fino agli anni Novanta, che hanno visto l'inizio della terza fase di penetrazione dell'informatica negli studi linguistici. Questa terza fase ha preso via come miglioramento metodologico ma ha implicato anche una manifestazione quantitativa dei dati emersi dall'analisi dei *corpora*. La grande quantità dei dati ha offerto delle intuizioni circa il linguaggio che hanno scosso gli assunti fondamentali alla base di molte posizioni teoriche dominanti nel settore (Tognini-Bonelli, 2001: 47-48). A tale riguardo, il computer ha influito sull'intero apparato metodologico, rendendo possibile l'osservazione su ampia scala dell'uso linguistico e modificando la natura stessa dell'osservazione (Tognini-Bonelli, 2000: 99).

Uno dei punti maggiormente dibattuti riguarda la definizione stessa di linguistica dei *corpora*, poiché alcuni linguisti la considerano una metodologia fra le tante, mentre altri la vedono come una vera e propria teoria.

Secondo alcuni linguisti la linguistica dei *corpora* non è una branca della linguistica come possono essere, ad esempio, la lessicologia, la semantica o la psicolinguistica – tutte branche che si concentrano sulla spiegazione o sulla descrizione di un aspetto dell'uso linguistico. Al contrario, la linguistica dei *corpora* consiste in una metodologia (Leech, 1992; in Tognini-Bonelli, 2000). Leech vede la linguistica dei *corpora* come una piattaforma metodologica per lo studio della lingua, come per esempio per lo studio di fonetica, sintassi, sociolinguistica; una base metodologica e non un dominio di studio. Anche i linguisti McEnery e Wilson (2001) sostengono che essa non delimiti un'area della linguistica ma piuttosto definisca il metodo di studio in una data ricerca, come per esempio lo studio della semantica attraverso l'approccio *corpus-based* o *non corpus-based*.

Dall'altro lato, secondo i linguisti che operano con i *corpora*, come per esempio Halliday e Sinclair, la linguistica dei *corpora* è una vera e propria disciplina che va ben oltre la pura metodologia di ricerca ed offre enormi possibilità alla ricerca quantitativa ed all'integrazione teorica (Tognini-Bonelli, 2000).

Poiché modifica la concezione tradizionale della lingua, la linguistica dei *corpora* offre una nuova visione della lingua e determina un cambiamento qualitativo, sia nell'ambito teorico che lessicografico, che viene largamente discusso nel lavoro di Sinclair (Tognini-Bonelli, 2000). Sinclair (1991: 4; in Tognini-Bonelli, 2000: 93) sottolinea che "the ability to examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before"<sup>29</sup>.

La linguistica dei *corpora*, possedendo i mezzi per stabilire le proprie regole e affermazioni teoriche, è stata definita come un nuovo sistema di misura per la descrizione della lingua atto a spiegare e descrivere i fatti linguistici successivamente utilizzati (Tognini-Bonelli, 2000: 94).

A tale proposito, molte aree di ricerca hanno beneficiato della linguistica dei *corpora* (Tognini-Bonelli, 2000; McEnery, Wilson, 2001) come per esempio la lessicografia (Sinclair, 1987), la traduzione (Baker, 1993; Tognini-Bonelli, 1996), la grammatica (Francis, 1993; Francis, Hunston, 2000), la semantica, la psicolinguistica, la pedagogia e la didattica delle lingue.

---

<sup>29</sup>Traduzione: "l'abilità di esaminare *corpora* di grandi dimensioni in una maniera sistematica consente l'accesso ad una qualità di prove finora non disponibili".

Ciononostante, il lessico e l'interazione con la grammatica sono stati un campo di studio privilegiato, fin dai primi tempi della diffusione della linguistica dei *corpora* (Spina, 2001). Fra i principi fondamentali che formano la base teorica della linguistica dei *corpora* figura il metodo dello studio del linguaggio, il quale viene esaminato nel suo uso reale, autentico, e non come un insieme di frasi isolate e basate sull'intuito (Spina, 2001: 57). Il testo nella sua unità è invece l'oggetto di studio della disciplina.

La linguistica dei *corpora*, sviluppatasi soprattutto a partire dagli anni '60 in riferimento alle lingue più studiate, come l'inglese, appare come un'impresa nuova e "foriera di una nuova visione filosofica del suo oggetto di studio" (Tognini-Bonelli, 2000: 94).

La linguistica dei *corpora* sembra essere una vera rivoluzione destinata a modificare le basi della linguistica stessa. A tale proposito Tognini-Bonelli (2000) fa menzione della *lexico-grammar* di Halliday, che conferma l'interdipendenza tra il fenomeno grammaticale e quello lessicale; dell'*idiom principle* di Sinclair, che rappresenta un cambiamento qualitativo della nostra concezione della lingua; e della testimonianza delle differenze qualitative e quantitative tra registri parlati e scritti, evidenziate da Biber ed altri.

La combinazione di quantitativo e di qualitativo, di dati empirici e sistematici che riflettono l'uso autentico della lingua e di generalizzazioni teoriche, ovvero l'interpretazione qualitativa dei dati emersi, sono elementi fondamentali per la consapevolezza della struttura e del funzionamento del linguaggio (Prat Zagrebelsky, 2002).

Il metodo di raccolta di testimonianze linguistiche è stato largamente usato in diverse forme fino agli anni Cinquanta. Anche se i *corpora* usati erano solamente grandi raccolte di testi e alcune nozioni fondamentali, come la rappresentatività, non erano state usate per costruire questi *corpora*, i linguisti hanno impiegato i *corpora* in studi sull'acquisizione del linguaggio (Preyer, 1889; Stern, 1924); in studi sulle convenzioni di sillabazione, ad esempio Kading ha usato un vasto *corpus* per il tedesco, composto da circa 11 milioni di parole, per collocare le distribuzioni di frequenza delle lettere in tedesco (Kading, 1987) ; in studi di pedagogia delle lingue straniere (Fries, Traver, 1940; Bongers, 1947); in studi di linguistica comparativa, come ad esempio quello di Eaton (1940) che paragona la frequenza dei significati delle parole in olandese,

francese, tedesco e italiano; e in studi che si concentrano sull'analisi della sintassi e della semantica – come il lavoro di Fries (1952), che precede le grammatiche dei *corpora* per l'inglese, scritte negli anni '80, mentre Gougenheim ed altri (1956) hanno usato un *corpus* trascritto per il francese parlato, allo scopo di descrivere le scelte lessicali ad alta frequenza e le scelte grammaticali (in McEnery, Wilson, 2007: 2-5).

È importante notare che, fino agli anni Cinquanta, l'approccio metodologico dominante si basava sull'osservazione del linguaggio e sulla metodologia dell'analisi del *corpus*, in uso dai primi del Novecento.

Negli anni Cinquanta, l'approccio empirico basato sulla raccolta di testi autentici è stato messo in discussione e screditato dal linguista americano Noam Chomsky (McEnery, Wilson, 2007). Chomsky ha criticato la finitezza e l'incompletezza dei *corpora* sulla base del fatto che essi rappresentano solo una piccola parte di una popolazione più grande, sminuendo quindi la natura aperta e creativa del linguaggio. Data l'influenza di queste critiche, lo studio dei *corpora* ha attraversato un periodo di crisi.

Con l'avvento dell'era informatica, le nuove tecniche di trattamento dei dati hanno aumentato sia la quantità che la qualità dell'elaborazione dei dati testuali. Negli anni Sessanta e Settanta, in risposta alla critica chomskyana, si è verificato un proliferare di progetti della linguistica dei *corpora*.

La linguistica dei *corpora* si oppone all'intuizione e al giudizio di approvazione di una forma linguistica da parte di un parlante nativo ideale, rivolgendo il suo interesse verso indagini linguistiche basate su larghe raccolte di dati linguistici reali (Chiari, 2007).

Nelle sezioni seguenti si darà una definizione di *corpus*, per poi prendere in considerazione i due principali approcci all'analisi linguistica basata sui *corpora*: il *corpus-based*, ovvero la linguistica basata sui *corpora*; e l'approccio *corpus-driven*, vale a dire la linguistica diretta dai *corpora*. In seguito si giustificherà l'approccio adottato in questo studio. Nella conclusione si illustrerà il metodo computazionale di estrazione dei *lexical bundles*, illustrando le diverse fasi di analisi degli stessi.

## II.1 Definizione di *corpus*

Come ricordato da Marengo (1996), i primi esperimenti nella pratica delle ricerche informatiche in linguistica sono stati fatti proprio in Italia su un insieme di opere in latino di San Tommaso d'Aquino. Il termine latino *corpus* (*pl. corpora*), designa infatti anche una raccolta di dati e di opere e coincide con il termine usato nella *corpus linguistics*.

"In principle, any collection of more than one text can be called a *corpus*: the term 'corpus' is simply the Latin for 'body', hence a *corpus* may be defined as any body of text. It need imply nothing more. But the term 'corpus' when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for"<sup>30</sup> (McEnery, Wilson, 2001: 29).

Come si evince dalla citazione un *corpus* è qualcosa di più di una semplice campionatura di testi da analizzare. Un *corpus* deve possedere quattro requisiti fondamentali: rappresentatività, dimensione, forma elettronica (*machine-readable form*), formato standard (McEnery, Wilson, 2001); e deve essere inoltre composto da testi autentici (McEnery, Xiao, Tono, 2006).

Partiamo innanzitutto dallo specificare cosa si intende per forma elettronica. Il formato elettronico indica l'archiviazione digitale di grandi quantità di materiali e la loro consultazione attraverso appositi *software* di interrogazione di linguistica computazionale. I *corpora* elettronici permettono di gestire i dati in maniera complessa e di operare sofisticate ricerche su di esse, le quali risultano impossibili utilizzando altri formati, come quello cartaceo o con schede perforate, che consentirebbero di effettuare solo lente e laboriose ricerche manuali.

In generale, un *corpus* può essere circoscritto come una collezione di testi considerata rappresentativa di una data lingua, dialetto o altra varietà linguistica. I presupposti della raccolta sono la spontaneità linguistica, definita secondo espliciti criteri e la rappresentatività di diverse tipologie testuali (Tognini-Bonelli, 2000).

---

<sup>30</sup> Traduzione: "In linea di principio, si può chiamare *corpus* qualsiasi raccolta composta da più di un testo singolo: il termine "corpus" è semplicemente la parola latina per "corpo", da ciò un corpus può essere definito come qualsiasi corpo di un testo. Non implica necessariamente niente di più che questo. Ma più frequentemente il termine "corpus" quando viene usato nel contesto della linguistica moderna tende ad avere connotazioni più specifiche di quelle fornite da questa semplice definizione".

### II.1.1 La rappresentatività del *corpus*

L'elemento chiave dalla costruzione di un *corpus* è dato dal concetto di rappresentatività. Sinclair (1991) sottolinea l'importanza di includere in un *corpus* l'utilizzo di un numero grande di scrittori ordinari e non il genio isolato o il giornalista scaltro, se vogliamo avvicinarci a una visione realistica del modo in cui viene utilizzato il linguaggio.

In altre parole, si deve tener conto della varietà linguistica e comunicativa che si intende delimitare. Secondo McEnery e Wilson (2001: 30) l'obiettivo della realizzazione di un *corpus* di una varietà linguistica sarebbe:

"a sample which is maximally representative of the variety under examination, that is, which provides us with as accurate a picture as possible of the tendencies of that variety, including their proportions<sup>31</sup>".

Gli autori sopra citati affermano che la rappresentatività di un *corpus* è legata inescandibilmente sia alla selezione dei testi che all'estensione del *corpus* stesso. Anche se le analisi quantitative possono essere eseguite su qualsiasi tipo di campione testuale, queste possono essere fuorvianti qualora si voglia generalizzare i risultati di un dato campione rispetto a popolazioni più ampie, ad esempio, rispetto a un intero genere (McEnery, Wilson, 2001). Secondo gli autori (McEnery, Wilson, 2001), questo è il punto di partenza della critica di Chomsky rispetto alle prime analisi dei *corpora*, in particolare rispetto alla possibile parzialità e non rappresentatività di un dato *corpus*, poiché un *corpus* costituisce solo un piccolo campione di una popolazione grande e potenzialmente infinita. Per Chomsky una linguistica dei *corpora* è inadeguata perché un dato *corpus* è sempre un insieme finito e incompleto. I suoi enunciati sono sempre condizionati da fattori casuali e, data la sua incompletezza, ci dà solo indicazioni sulla frequenza con le quali determinate peculiarità linguistiche si manifestano (Chiari, 2007: 30).

---

<sup>31</sup> Traduzione: "un campione che sia massimamente rappresentativo della varietà sotto esame, cioè, che ci fornisca un'immagine, la più accurata possibile delle tendenze di quella varietà, comprese le loro proporzioni".

Chiari (2007) osserva che questa critica perde in parte validità, visto che al giorno d'oggi esistono tecnologie informatiche capaci di analizzare centinaia di migliaia di parole in tempo reale. Secondo McEnery e Wilson (2001: 30) la critica di Chomsky non dovrebbe essere ignorata, piuttosto, le considerazioni circa la critica di Chomsky dovrebbero orientarsi verso l'elaborazione di tecniche di costruzione dei *corpora* che permettano di aumentare la rappresentatività.

Ciononostante, quando si analizza un *corpus* si deve sempre tenere a mente che si tratta sempre di un'analisi basata su un campione estratto da una popolazione molto più grande.

A tale proposito, Chiari (2007: 43) nota che un *corpus* dovrebbe essere costruito in modo che ogni fenomeno linguistico abbia la stessa probabilità di occorrere nel campione e nella popolazione, presentando caratteristiche simili alla popolazione esaminata. Analogamente Biber (1993; in McEnery, Wilson, 2001) nella sua particolareggiata inchiesta circa la rappresentatività di un *corpus*, sottolinea che il primo passo nella campionatura di un *corpus* è quello di definire il più chiaramente possibile i limiti della popolazione che si intende studiare, prima di poter procedere a definire le procedure di campionatura.

La rappresentatività di un *corpus* viene raggiunta tramite una ricca varietà testuale, campioni proporzionalmente selezionati dalla popolazione che si desidera esaminare (McEnery, Xiao, Yukio, 2006). Secondo gli autori un *corpus* di questo tipo può essere definito come bilanciato.

Prima di definire scrupolosamente i confini della varietà linguistica da analizzare, *il sampling frame*, la campionatura dei testi dovrebbe venire determinata in base alla struttura gerarchica della popolazione. Dopo aver determinato la popolazione da analizzare, si devono precisare i campioni che la definiscono al meglio, sia nel numero ottimale di testi inclusi nel *corpus*, sia nella lunghezza ottimale di ciascun testo (McEnery, Wilson, 2001). A tale proposito Chiari (2007: 44) nota che la rappresentatività è una caratteristica che varia secondo l'aspetto linguistico che si vuole studiare, assumendo aspetti relativi e dinamici.

Questo studio prende in esame le microlingue scientifico-professionali, in particolare la microlingua della storia, della filosofia e della medicina. Verranno infatti esaminati degli articoli di ricerca che rappresentano il genere accademico più

prestigioso e sono "the principle site of disciplinary knowledge-making" (Hyland, 2008b: 5), ovvero l'ambito d'uso più significativo di questa varietà linguistica. Come si vedrà in seguito, si è scelto di adottare l'approccio *corpus-driven*, in cui "the issue of the representativeness of the *corpus* can be seen in its true importance, since the information provided by the *corpus* is placed centrally and accounted for exhaustively<sup>32</sup>" (Tognini-Bonelli, 2001: 88).

### II.1.2 La dimensione o l'estensione del *corpus*

La dimensione, che riguarda il numero di parole di cui è costituito un *corpus*, dovrebbe essere definita all'inizio della costruzione di un *corpus*, rispettando la varietà linguistica e testuale oggetto di studio. In realtà, la questione di come definire una dimensione ideale di un *corpus* rimane tuttora aperta.

Secondo Sinclair (1991) un *corpus* dovrebbe essere il più grande possibile e in continua espansione. La dimensione "aperta" si riferisce ai cosiddetti monitor *corpora*, *corpora* in continua espansione -- una caratteristica propria della lingua stessa. Un esempio di monitor *corpus* o *corpus* di monitoraggio è la *Bank of English* diretta da John Sinclair.

Tranne per i monitor *corpora*, la dimensione di un *corpus* dovrebbe essere finita perché anche il termine *corpus* tende ad implicare una quantità finita di testi (McEnery, Wilson, 2001). Circa le dimensioni del *corpus*, secondo Chiari (2004: 45) un *corpus* di meno di 15.000 occorrenze non può essere considerato rappresentativo.

Anche Biber (2006) sostiene che un *corpus* deve essere abbastanza grande da rappresentare in maniera adeguata le occorrenze delle caratteristiche studiate. Come spiega il linguista, negli studi grammaticali questo non rappresenta un problema per le caratteristiche comuni, mentre le caratteristiche meno comuni richiedono *corpora* ben più grandi, poiché un *corpus* limitato può contenere un campione accidentale delle caratteristiche rare, portando a conclusioni incoerenti sulla frequenza. Lo studioso

---

<sup>32</sup> Traduzione: "il problema della rappresentatività del corpus può essere visto in tutta la sua importanza dato che l'informazione fornita dal corpus è messa al centro e resa in maniera esauriente".



considera la dimensione del *corpus* come una variabile ancora più importante negli studi del vocabolario, poiché gran parte delle parole singole tendono a occorrere molto meno frequentemente rispetto alle costruzioni grammaticali.

Come notato da Chiari (2007: 44), sia la rappresentatività che l'estensione del *corpus* sono caratteristiche relative, sottoposte alla varietà linguistica in esame. A tale riguardo, si nota che la grandezza del *corpus* per analisi lessicali, morfologiche o sintattiche dovrebbe essere più ampia di quella necessaria alle analisi fonologiche e fonetiche.

L'ampiezza del *corpus*, misurata con numero di parole o *tokens* diversi, ha subito dagli anni Novanta in poi un forte incremento. Al tempo, lo standard per i *corpora* di riferimento si aggirava intorno a 500.000 parole, mentre successivamente la rivoluzione informatica ha innalzato il limite superiore. Per le lingue maggiormente studiate, i *corpora* contemporanei più diffusi possiedono un'estensione di circa 100 milioni di parole, come per esempio per l'inglese il *British National Corpus*, mentre altri *corpora*, tra cui il *corpus COBUILD (Bank of English)* raggiungono circa 500.000 milioni di parole. Per la lingua italiana, il *Corpus di italiano scritto contemporaneo* comprende 120.000 milioni di parole.

Le valutazioni sulla dimensione dei *corpora* si basano sul numero di occorrenze (*tokens*) di parole grafiche presenti nel *corpus*. Cortes (2004), discutendo la dimensione ideale di un *corpus* per l'identificazione dei *lexical bundles* in un *corpus* specialistico, ha sostenuto che nella prosa accademica è assai consigliabile lavorare con *corpora* di almeno 1.000.000 parole. Se viene usato un *corpus* di meno di 1.000.000 di parole, il numero di volte in cui occorre una certa combinazione di parole dovrebbe venir normalizzato a un milione di parole, utilizzando la formula di normalizzazione, che secondo Cortes (2004) non sempre dà risultati esatti. Dato che la frequenza con cui occorrono i *lexical bundles* dovrebbe essere riportata in un *corpus* di almeno un milione di parole (Cortes, 2004; Biber, 2006), l'uso di statistiche normalizzate può pregiudicare la validità e il valore generale della ricerca in maniera sostanziale (Cortes, 2004; Biber, 2006).

Altri studi hanno dimostrato l'innegabile influenza della dimensione del *corpus* e della varietà dei testi usati sulla varietà e sui tipi di *bundles*, riconosciuti come appartenenti al campo della linguistica applicata (Jalali ed altri, 2008). Prima di questi

studi, solo Hyland (2008a) aveva esplorato l'uso dei *lexical bundles* nella linguistica applicata per la lingua inglese, su un *corpus* che non soddisfaceva la dimensione ottimale minima di un milione di parole. Oltre ad alcune somiglianze, i ricercatori (Jalali ed altri, 2008) hanno rilevato alcune differenze importanti: ad esempio, alcuni *bundles* nello studio di Hyland non risultavano essere così frequenti come in quello di Jalali ed altri, e viceversa. Inoltre, si rilevava anche una certa discrepanza circa la gamma dei *bundles*. Secondo Jalali ed altri (2008) queste differenze potrebbero essere attribuite perlopiù alla dimensione del *corpus* e anche, ma solo in parte, all'effetto della prima lingua e della cultura nel caso delle tesi esaminate: le tesi, negli studi di Hyland, erano state scritte perlopiù da studenti di madrelingua cinese, mentre le tesi nello studio di Jalali ed altri erano state scritte da studenti iraniani.

### II.1.3 Il riferimento standard

Secondo Chiari (2007: 47) il riferimento standard segnala:

"la possibilità per i grandi *corpora* di diventare dei punti di riferimento per lo studio della varietà che rappresentano, attraverso l'esplicazione delle metodologie di analisi, facilitando la comparazione tra *corpora* diversi ed evitando la necessità di ripetere i computi per ogni analisi successiva".

Anche se il formato standard non rappresenta una caratteristica essenziale del *corpus*, segnala una misura standard concernente l'ambito studiato o la varietà del linguaggio che rappresenta (McEnery, Wilson, 2001). I nuovi risultati di una data ricerca possono essere facilmente paragonati a quelli già pubblicati, per questa ragione è necessario specificare esplicitamente la metodologia di raccolta dati. Il riferimento standard implica anche l'utilizzo di una base continua di dati; per questo la variazione fra studi è imputabile all'adeguatezza degli assunti e delle metodologie contenute nello studio (McEnery, Wilson, 2007).

## II.2 Il *corpus* specialistico

Le diverse tipologie di *corpus* possono essere identificate a seconda della modalità comunicativa (per esempio *corpora* di lingua scritta, parlata); del numero di lingue analizzate, quali per esempio *corpora* bilingue, monolingue, paralleli; o ancora a seconda che i testi in una determinata lingua siano prodotti da parlanti nativi o non nativi (per esempio *learner corpora*); in base al tipo di varietà linguistica – *corpus* sincronico e diacronico; e in base al tipo di comunità linguistica raffigurata, quali *corpus* generale di riferimento e *corpus* specialistico (McEnery, Xiao, Yukio, 2006; Bowker, Pearson, 2002). Nell'ambito delle microlingue e della terminologia specialistica, si utilizzano i *corpora* specialistici o *special-purpose corpora*. I *corpora* specialistici sono collezioni di testi che esaminano solo un particolare aspetto della lingua, come per esempio una microlingua, un particolare genere testuale – articoli scientifici, commenti, riassunti, articoli di giornale, ecc.

Prima di entrare nel merito dei *corpora* specialistici analizzati in questa ricerca, prendiamo in esame la definizione di *corpus* specialistico data da Bowker e Pearson (2002: 12):

"[...] a special purpose corpus is one that focuses on a particular aspect of a language. It could be restricted to the LSP of a particular subject field, to a specific text type, to a particular language variety or to the language used by members of a certain demographic group [...]"<sup>33</sup>.

Un *corpus* specialistico è un *corpus* che analizza un aspetto particolare della lingua. La sua natura specialistica non permette, quindi, di formulare delle osservazioni sulla lingua generale.

Secondo Hoey (2005: 14), c'è bisogno di un *corpus* specialistico per prendere coscienza di caratteristiche più specifiche dell'utilizzo del linguaggio. Hoey sostiene che un computer non può dirci quali siano le preselezioni che operano per un qualsiasi utilizzatore del linguaggio, ma può però suggerire i modi in cui la preselezione si

---

<sup>33</sup>Traduzione: "[...] un corpus a scopi speciali è un corpus che si concentra su un particolare aspetto di una lingua. Può essere ristretto alla LSP di un campo di competenza particolare, o a una specifica tipologia testuale, a una varietà particolare di una lingua o al linguaggio usato dai membri di un certo gruppo demografico [...]"

verifica, i tipi di caratteristiche per le quali le parole o le sequenze di parole possono essere preselezionate e può indicare che alcune preselezioni sono potenzialmente condivise da un ampio numero di parlanti.

Secondo Sinclair (2005: 14) in linea di principio un *corpus* dovrebbe essere ideato e costruito da un esperto linguistico, che usa il linguaggio rispecchiato nel *corpus*. Nel nostro caso i *corpora* di storia, filosofia e medicina dovrebbero essere costruiti da esperti in materia, rispettivamente da storici, filosofi e medici.

Sulla base di una comparazione fra le frequenze terminologiche di un *corpus* generale e quelle di uno specializzato, l'autore discute le proporzioni di un *corpus* in una data area specialistica. Tale comparazione ha dimostrato che il numero delle diverse forme terminologiche è di gran lunga minore nel testo specialistico che in quello generale (per una differenza di circa il 40%). Un'altra indicazione della diffusione del vocabolario è la proporzione delle occorrenze singole, con il corpo specialistico che presenta solo il 31% del totale dell'altro *corpus*. Anche le forme terminologiche che ricorrono due volte sono molto meno comuni nel *corpus* specialistico, ma guardando ai dati per venti o più occorrenze le proporzioni si invertono. Alla frequenza di 200 e oltre, il *corpus* generale presenta solo il 69% del numero di queste parole nel gruppo specialistico. Sinclair (ibidem) conclude che la comparazione suggerisce che il *corpus* specialistico sottolinea un vocabolario tecnico e limitato, rappresentato dalle liste di frequenza, e perciò necessita di un *corpus* ben meno esteso rispetto a quello necessario allo studio generale di una lingua.

Oltre a specificare il numero totale di parole o *tokens* da includere nel *corpus* e la lunghezza di ciascun testo, si deve inoltre stabilire quanti articoli di autori diversi saranno inclusi nel *corpus*. Secondo Bowker e Pearson (2002), un *corpus* multiautoriale dà una buona idea di quali termini e concetti vengono usati più comunemente in una microlingua, mentre nel caso di *corpus* costituiti da singoli autori siamo di fronte ad aspetto idiosincratico.

Uno dei problemi da affrontare nella costruzione di un *corpus* specialistico è costituito dall'acquisizione di testi autentici in formato elettronico. La selezione degli articoli scientifici per i *corpora* analizzati in questa ricerca è stata condizionata da un lato dall'opinione da parte degli esperti in materia e dall'altro dalla disponibilità in formato elettronico dei testi.

Molti testi, infatti, non sono facilmente accessibili e nella maggior parte dei casi la disponibilità è condizionale al pagamento. Tutti gli articoli scientifici analizzati sono stati scritti da esperti di madrelingua italiana, perché si suppone che contengano degli esempi rappresentativi di una microlingua.

Per il *corpus* di storia sono state pertanto considerate quattro riviste scientifiche liberamente consultabili in formato elettronico, ovvero *Altreitalie*, *Reti medievali*, *Storicamente* e *Storia delle donne*. Per il *corpus* di filosofia, si sono presi in considerazione tre periodici quali *Noema*, *Philosophical news*, *Segni e comprensione*. Per il *corpus* di medicina del lavoro, si è inclusa la rivista scientifica *Giornale italiano di medicina del lavoro ed ergonomia* (per la descrizione dei *corpora* vedi Tabella 2, 3, 4).

Le discipline scientifiche, la storia e la filosofia da un lato e la medicina del lavoro dall'altro sono state scelte perché rappresentano diverse tradizioni di ricerca scientifica, appartenendo rispettivamente alle discipline umanistiche e alle scienze applicate.

In ambito medico, a causa della scarsa reperibilità del materiale scientifico online, si è scelto il settore di medicina del lavoro – "la branca della medicina che si occupa dello studio, della prevenzione e della terapia di malattie derivanti dalle attività lavorative. La medicina del lavoro è strettamente collegata, per gli aspetti normativi e prescrittivi, al diritto del lavoro<sup>34</sup>".

Il periodico specialistico *Giornale di medicina del lavoro ed ergonomia* è stato scelto anche perché disponibile gratuitamente in formato elettronico. Si tratta di una rivista multidisciplinare, che include sia la scienza della medicina del lavoro che l'ergonomia, rappresentando così la scienza naturale. Si è cercato, inoltre, di includere in tutti e tre i *corpora* articoli scientifici recenti, per poter esaminare lo stato dell'arte attuale, sia a livello linguistico che concettuale.

Le tabelle seguenti indicano i dati dei *corpora* presi in esame in questa ricerca.

Tabella 1: *Corpus* di storia

<b>Riviste scientifiche</b>	<b>Annate</b>	<b>Numero articoli/autori</b>	<b>Numero totale di parole (tokens)</b>
Altreitalie	1993	5 / 5	
	2000	2 / 2	
	2001	6 / 7	

<sup>34</sup>Definizione tratta dall'Enciclopedia italiana Treccani reperibile online ([www.treccani.it](http://www.treccani.it)).

	2003	5 / 5	
	2004	11 / 12	
	2005	14 / 19	
	2006	32 / 35	
	2008	60 / 64	
	2009	7 / 8	
	2010	3 / 4	
Reti medievali			
	2004	1 / 1	
	2005	3 / 3	
	2006	5 / 5	
	2007	9 / 9	
	2010	7 / 7	
	2011	4 / 4	
Storicamente			
	2005	3/3	
	2006	6/6	
	2007	3/3	
	2008	5/5	
	2009	4/4	
	2010	15/15	
	2011	5 / 5	
Storia delle donne			
	2005	12 / 12	
	2006	12 / 12	
	2007	10 / 10	
	2008	6 / 6	
	2009	7 / 7	
	2011	7 / 7	
<b>Totale</b>		160/291	1.605.974

Tabella 2: *Corpus* di filosofia

Riviste scientifiche	Annate	Numero articoli/autori	Numero totale di parole ( <i>tokens</i> )
Noema	2012	8 / 8	
Philosophical news	2012	4 / 4	
Segni e comprensione	2007	37/37	
	2008	10 / 10	
	2009	11 / 11	
	2010	22 / 22	
	2011	6/6	
<b>Totale</b>		100/100	780.989

Tabella 3: *Corpus* di medicina del lavoro

Riviste scientifiche	Annate	Numero articoli/autori	Numero totale di parole ( <i>tokens</i> )
----------------------	--------	------------------------	---

Giornale di medicina del lavoro ed ergonomia	2008 2009 2010	82 / 340 38 / 180 376 / 1895	
<b>Totale</b>		496/2415	1.047.442

La reperibilità del materiale online ha condizionato la dimensione del *corpus* di filosofia, in cui, per il calcolo dei *lexical bundles*, si è dovuta adoperare la formula di normalizzazione. Da contro, la grandezza degli altri due *corpora* è stata definita in base al numero di parole suggerito nelle ricerche svolte per la lingua inglese nell'ambito delle microlingue (Cortes, 2002, 2004; Hyland, 2008a, 2008b).

### II.3 Gli approcci *corpus based* e *corpus driven*

Le più recenti ricerche di linguistica dei *corpora* si sono concentrate su come studiare il linguaggio formulaico.

Si evidenziano due approcci allo studio sul linguaggio formulaico: l'approccio *corpus-based*, ovvero lo studio basato su *corpora*; e quello *corpus-driven* ovvero lo studio indotto da *corpora*.

Tognini-Bonelli (2001: 65) definisce così l'approccio *corpus-based* o approccio basato sui *corpora*:

"The approach *corpus based* is used to refer to a methodology that avails itself of the *corpus* mainly to expound, test or exemplify theories and descriptions that were formulated before large *corpora* became available to inform language study. Traditionally, linguistic theories are the result of reflection by a scholar after absorbing a great deal of experience of language and languages, and testing the implications and consequences with reference to the intuition of competent or native speaker<sup>35</sup>".

<sup>35</sup> Traduzione: "L'approccio basato sul corpus viene usato per riferirsi a una metodologia che si avvale del corpus principalmente per esporre, mettere alla prova o esemplificare delle teorie e delle descrizioni che erano state formulate prima che i grandi corpora divenissero disponibili per dare informazioni allo studio della lingua. Per tradizione, le teorie linguistiche sono il risultato di una riflessione da parte di uno studioso che ha assorbito una lunga esperienza in materia di linguaggio e di lingue, e che ha messo alla prova le implicazioni e le conseguenze con riferimento all'intuito del parlante competente o nativo".

Secondo la linguista, la ricerca *corpus-based* presuppone la validità delle strutture linguistiche derivate dalla teoria linguistica; lo scopo principale della ricerca è analizzare i modelli di utilizzo sistematici per quelle caratteristiche linguistiche predefinite in cui la relazione fra un elemento e il suo contesto non è intesa come sistematica e determinante per la definizione delle categorie linguistiche.

Questa metodologia rappresenta la norma negli studi della variazione grammaticale; questi di solito adottano le categorie grammaticali definite da precedenti ricerche linguistiche e poi utilizzano la ricerca sui *corpus* per descrivere i modelli di variazione e di utilizzo associati a queste caratteristiche grammaticali. Per questo, negli studi *corpus-based* del linguaggio formulaico, il ricercatore preseleziona le espressioni formulaiche e poi analizza il *corpus* per scoprire in che modo quelle espressioni vengano utilizzate (Tognini-Bonelli, 2001).

L'approccio *corpus-based* è quindi un approccio metodologico che tende a integrare l'approccio quantitativo con quello qualitativo, includendo sia le metodologie *top-down*, che approfondiscono le teorie preesistenti senza mettere in discussione le unità di analisi fondamentali e utilizzano tassonomie già date; sia metodologie *bottom-up*, basate su modelli casuali, che procedono dalla verifica puntuale dei presupposti formulati.

Questo approccio fa affidamento su un gruppo di espressioni selezionate prima dello studio e funziona sempre all'interno di cornici già stabilite. Le espressioni analizzate vengono riconosciute come familiari dai parlanti nativi della lingua oppure sono state considerate come espressioni frequenti nella letteratura correlata e possono essere considerate significative da quel punto di vista (Nattinger e DeCarrico, 1992).

Per l'italiano, un progetto analogo a quello di Sinclair è il CODIS (*Corpus dinamico dell'italiano scritto*), elaborato e coordinato dal 1998 da R. Rossini Favretti al Centro di linguistica applicata e teorica dell'Università di Bologna. Il *corpus* viene aggiornato ogni due anni con nuovo materiale e contiene 100.000 milioni di parole.

L'altro possibile approccio alla linguistica dei *corpora* è l'approccio *corpus-driven*, ovvero indotto da *corpora*, che è di più recente elaborazione. Si tratta di un approccio da cui:



"the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to *corpus* evidence. The *corpus*, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system. [...] recurrent patterns and frequency distributions are expected to form the basic evidence for linguistic categories; the absence of a pattern is considered potentially meaningful<sup>36</sup>" (Tognini-Bonelli, 2001: 84).

Questo tipo di approccio è più induttivo, poiché i costrutti linguistici emergono dall'analisi di un *corpus* e i dati vengono considerati nella loro interezza per trarre conclusioni teoriche. Di fatto, le argomentazioni teoriche derivano direttamente dai dati ricavati dal *corpus*, dai quali vengono elaborate teorie e ipotesi linguistiche (Biber, 2009).

L'approccio *corpus-driven* si differenzia dalle pratiche linguistiche standardizzate poiché minimizza gli assunti aprioristici circa i costrutti linguistici che devono essere impiegati nell'analisi. Nella sua forma più elementare, l'analisi *corpus-driven* ha come assunto solo l'esistenza delle parole; i modelli di co-occorrenza fra le parole, scoperti dall'analisi del *corpus*, forniscono la base per le successive descrizioni linguistiche (Biber, 2009).

Secondo Chiari (2007: 42), questo tipo di approccio rappresenta "in modo più fedele le dinamiche, le variazioni e le tendenze di una lingua parlata e scritta".

Tognini-Bonelli (2001) ha sottolineato che è importante comprendere che l'approccio *corpus-driven* non è un approccio meccanico, ma una metodologia costantemente mediata dal linguista, che si comporta ancora come un linguista, trasferendo le sue conoscenze e le sue esperienze ad ogni stadio di questo processo. La linguista identifica come primo studio di lessicografia *corpus-driven* il progetto *Cobuild Project*, che riflette le posizioni di Sinclair sulla teoria del linguaggio e sulla metodologia descrittiva (Tognini-Bonelli, 2001). Questo progetto ha influenzato altre

---

<sup>36</sup> Traduzione: "L'impegno del linguista è verso l'integrità dei dati nel loro complesso e le descrizioni mirano ad essere esaurienti rispetto alle prove fornite dal corpus. Il corpus, perciò, è visto più come un deposito di esempi che suffragano le teorie preesistenti o come una estensione probabilistica di un sistema già ben definito. [...] ci si aspetta che i modelli ricorrenti e le distribuzioni di frequenza costituiscano la prova fondamentale per la formazione delle categorie linguistiche; l'assenza di un modello è considerata potenzialmente significativa".

aree di studio, come la grammatica, la collocazione, l'insegnamento della lingua e ha modificato il modo in cui il significato viene identificato e definito. L'inaspettatezza dei risultati derivati dai dati del *corpus* porta a concludere che l'intuizione non è una fonte di informazioni sul linguaggio completamente affidabile (Tognini-Bonelli, 2001).

Come evidenzia Tognini-Bonelli, la metodologia essenziale della linguistica *corpus-driven* consiste nell'esercizio dell'intuito del ricercatore circa la presenza di dati nuovi e di fare affidamento sul punto di vista esposto.

Lo studio di grammatica inglese *Longman Grammar of Spoken and Written English* (Biber ed altri, 1999), basato su un *corpus* di inglese britannico e americano di circa 40 milioni di parole, nasce da questo approccio che introduce nella teoria linguistica un nuovo filone di ricerca: lo studio dei *lexical bundles*.

L'approccio *corpus-driven* ha avuto il merito di individuare il linguaggio come strutturato e composto da *pattern* lessicali.

Gli studi sulle collocazioni lessicali sono di solito considerati *corpus-driven* perché i *pattern* lessicali di una lingua vengono identificati attraverso l'analisi della loro frequenza d'uso in un dato *corpus*.

Per quanto riguarda la lingua inglese, i pionieri di questi studi sulle sequenze di più parole sono stati Altenberg e Eeg-Olofsson, nel loro studio del *London-Lund Corpus*, comprendente 500.000 parole di inglese parlato (Altenberg, 1998; Altenberg, Eeg-Olofsson, 1990; in Biber, 2006).

Per quanto riguarda le altre lingue, i primi studi sulla lingua francese sono stati condotti da Salem (1987), che ha analizzato un *corpus* dei documenti governativi in lingua francese. Per la lingua spagnola, è stata fondamentale la ricerca di Butler (1998), di cui approfondiremo lo studio nelle sezioni successive.

Tuttavia alcuni studi spesso hanno teso a combinare gli approcci *corpus-driven* con quelli *corpus-based*, ovvero l'approccio ibrido (Biber, 2009). Come esempio, l'autore indica lo studio di Renouf e Sinclair (1991), che è allo stesso tempo *corpus-driven* – poiché le sequenze di riempitori lessicali erano state scoperte tramite l'analisi del *corpus* – e *corpus-based*, perché le cornici collocazionali erano state preselezionate da Renouf e Sinclair (1991).

Secondo Biber (2009: 281) un approccio radicalmente *corpus-driven* allo studio del linguaggio formulaico deve avere tre caratteristiche generali: deve essere basato

sull'analisi delle vere forme lessicali che si presentano nel *corpus*; deve analizzare le sequenze delle forme lessicali, senza considerazione alcuna dello status sintattico-grammaticale delle parole; e deve concentrarsi sulle combinazioni di forme lessicali frequenti e ricorrenti.

Il presente studio adotta la metodologia *corpus-driven* così come proposta da Biber (2009) che presuppone solo l'esistenza delle forme lessicali senza referenza alcuna alle categorie grammaticali e che incorpora tutte e tre le caratteristiche che definiscono la ricerca *corpus-driven*, al fine di identificare i *lexical bundles* più comuni, in due registri – la conversazione e la prosa accademica – e in tre discipline scientifiche diverse, al fine di investigare i diversi modi in cui quei modelli sono variabili nei due registri e nelle tre discipline scientifiche, usando le distribuzioni a frequenza semplice per identificare le sequenze di parole ricorrenti.

Si è optato per l'approccio *corpus-driven* seguendo gli studi precedenti svolti per la lingua inglese e spagnola circa le combinazioni di più parole (Altenberg, Eeg-Olofsson, 1990; Altenberg, 1998; Butler, 1997; Biber ed altri, 1999; Biber, Conrad, 1999; Biber, Conrad, Cortes, 2003, 2004; Conrad, Biber, 2004; Cortes, 2004, 2006; Hyland, 2008a, 2008b).

Seguendo le orme di Tognini-Bonelli (2000: 97) possiamo concludere che:

"Il punto di vista che il corpus offre all'analista è anzi privilegiato, in quanto garantisce l'accesso sia al singolo elemento, che può essere letto ed espanso sull'asse orizzontale della concordanza, sia all'uso generalizzato riscontrabile dall'osservazione di ripetuti fenomeni di co-selezione sull'asse verticale della concordanza. Essendo la frequenza con cui una parola è presente nel *corpus* indicativa della frequenza d'uso, questo tipo di analisi offre un utile punto di partenza nell'analisi del profilo di una parola, struttura o espressione in relazione ad una norma socialmente stabilita".

#### **II.4 Metodo computazionale di estrazione dei *lexical bundles***

Come già accennato in precedenza, il criterio della considerazione quantitativa del linguaggio non è qualcosa di nuovo, bensì risale già agli antichi Greci e Romani che

già consideravano il fattore quantitativo soprattutto nella distinzione tra le parole di minore o maggiore frequenza, tra quelle di uso raro e tra *hapax legomena*, ovvero le parole molto frequenti (Bortolini, Taglavini, Zampolli, 1972: 5-10). Gli eruditi medievali compilarono le prime liste di concordanze per i testi biblici, mentre le prime liste di frequenza della lingua comune vennero alla luce nel XIX secolo. La prima lista di frequenza per la lingua tedesca, *Häufigkeitwörterbuch der deutschen Sprache*, fu pubblicata già nel 1898, mentre le liste di frequenza per la lingua italiana sono apparse negli anni '80 del Novecento. Di seguito riportiamo gli esempi più importanti, come il LIF – *Il lessico italiano di frequenza*, composto da 500.000 occorrenze tratte da cinque gruppi (Teatro, Romanzi, Cinema, Periodici e Sussidiari); il VELI – il *Vocabolario elettronico della lingua italiana* – costituito da circa 10.000 lessemi ordinati per frequenza e tratti da un *corpus* di più di 25 milioni di parole; il LIP – il *Lessico di frequenza dell'italiano parlato*; i due *database* elaborati dall'Università di Venezia negli anni '90; la SIWL – la *Spoken Italian Word List* – con un totale di 43.534 forme lemmatizzate; lo IWL – *Italian Word List* – composta di 30.000 forme molto ricca di *hapax legomena* (Delmonte, 1997).

L'obiettivo primario di questo studio è fornire un'indagine empirica di *pattern* lessicali o *lexical bundles* che più frequentemente compaiono in un registro. Lo studio dei *lexical bundles* sui *corpora* permette di indagare la co-occorrenza di combinazioni di parole determinate dalla loro frequenza d'uso.

A livello metodologico, gli articoli scientifici selezionati hanno richiesto un trattamento informatico di conversione dal formato originale *.pdf* al formato *word*. Gli articoli sono stati inseriti nel *database* testuale e analizzati con uno dei più diffusi software per l'analisi dei *corpora*: il WordsmithTools Version 5.0, creato da Mike Scott. A questo scopo, i *file* sono stati ulteriormente convertiti in *file* di testo (*.txt*), la conversione diretta da *file .pdf* in *file .txt* ha creato notevoli modifiche testuali. Un'altra operazione preliminare all'analisi è stata la normalizzazione grafica dei *corpora*, che consiste nella correzione ortografica dei testi e nell'eliminazione delle righe vuote e degli spazi bianchi. Nella versione definitiva del *corpus* analizzato non sono stati inclusi titoli, sottotitoli, note, citazioni, formule scientifiche, grafici, tabelle, referenze bibliografiche e riassunti. Sono state anche escluse le parti in lingua straniera.

L'operazione iniziale all'analisi del *corpus* del parlato (LIP) è stata la conversione dalla forma di dialogo alla forma narrativa. La lettura dei *lexical bundles* è stata svolta considerando la frase come unità massima e il punto come la fine di ogni sezione.

Grazie al sofisticato *software* Wordsmith 5.0, è possibile individuare le concordanze e creare liste di frequenza, ovvero elenco di tutte le parole presenti nel *corpus* con il numero totale delle relative occorrenze. Inoltre, cosa particolarmente importante per questa ricerca, con questo *software* è possibile creare *Wordlist-index* calcolando i *lexical bundles*, cioè la distribuzione delle co-occorrenze delle unità lessicali; è anche possibile estrarre combinazioni di più parole consultando il testo originale in cui queste compaiono, un'operazione importante per l'analisi delle funzioni discorsive dei *lexical bundles*. Le potenzialità di questo *software* sono notevoli e hanno avuto un ruolo fondamentale nel produrre i risultati indicati in questa ricerca.

L'analisi quantitativa dei *lexical bundles* è stata svolta con il comando *WordList* che consente di identificare e di memorizzare tutte le sequenze di 4 parole presenti nel *corpus* analizzato. L'analisi per i *lexical bundles* non garantisce una totale esaustività e completezza: i *lexical bundles* che prevedono al loro interno forme flesse – per esempio *è in corso di*, *sono in corso di* – sono stati considerati nella presente ricerca come due tipi di *bundle*.

Per l'analisi linguistica del *corpus* è stata scelta la funzione *WordList* di Wordsmith Tools, che consente di generare due tipi di *wordlist*: la *wordlist* composta da singole parole e la *Wordlist-index* che offre la possibilità di computare dei *clusters* o *lexical bundles*, composti da almeno due o più parole. Entrambe le liste producono sempre la rispettiva incidenza, in ordine di frequenza discendente. La funzione *WordList* genera i testi fornendo inoltre i dati riguardanti la statistica del *corpus* (Tabelle 4, 5, 6). Questi dati riguardano molteplici variabili: la grandezza del *corpus* (*file size*); il numero di *tokens* (*running words*) o di diverse parole che compaiono nel *corpus*, indipendentemente dalla loro ripetitività o dal loro significato contestuale; il numero di *types* (*distinct words*) o dei vocaboli tipo che comprendono tutte le parole con la medesima sequenza grafica, trascurando, però, la polisemia; la proporzione tra *type* e *token* (*type/token ratio*), misura che indica il rapporto tra il numero totale di parole presenti nel *corpus* (*tokens*) e il numero di parole diverse effettivamente

utilizzate (*type*), indicando la varietà nella scelta dei termini; il numero delle frasi (*sentences*) e il numero medio di parole in una frase (*mean in words*). Il programma offre anche la possibilità di vedere il contesto originale dei *lexical bundles*. Riportiamo di seguito solo i dati rilevanti per la presente ricerca.

Tabella 4: Estratto dalla *wordlist* del *corpus* di storia

<b>Text file</b>	<b>Overall</b>
<b>File size</b>	10.598.023
<b>Tokens (running words in text)</b>	1.605.974
<b>Types (distinct words)</b>	64.019
<b>Type/token ratio</b>	4,07
<b>Sentences</b>	44.218
<b>Mean (in words)</b>	35,59

Tabella 5: Estratto dalla *wordlist* del *corpus* di filosofia

<b>Text file</b>	<b>Overall</b>
<b>File size</b>	5.139.996
<b>Tokens (running words in text)</b>	780.989
<b>Types (distinct words)</b>	41.727
<b>Type/token ratio</b>	5,37
<b>Sentences</b>	42.184
<b>Mean (in words)</b>	18,41

Tabella 6: Estratto dalla *wordlist* del *corpus* di medicina del lavoro

<b>Text file</b>	<b>Overall</b>
<b>File size</b>	7.390.184
<b>Tokens (running words in text)</b>	1.047.442
<b>Types (distinct words)</b>	38.365
<b>Type/token ratio</b>	3,79
<b>Sentences</b>	31.184
<b>Mean (in words)</b>	32,43

La funzione *Wordlist-index* offre la possibilità di computare i *clusters* o *lexical bundles*. Prima di effettuarla è stato necessario impostare alcuni parametri che determinano il calcolo dei *lexical bundles*:

- a. la lunghezza dei *lexical bundles* (*cluster size*), cioè la dimensione degli n-grammi: nella presente ricerca si è optato per l'analisi dei *lexical bundles* di quattro parole perché molti *lexical bundles* di quattro parole contengono nella loro struttura *lexical bundles* di tre parole (per esempio il *bundle* di 4 parole *dal punto di vista* contiene nella sua forma il *bundle* *punto di vista*; inoltre si presenta nella versione più estesa *da un punto di vista, da questo punto di vista*) e sono molto più frequenti dei *lexical bundles* di cinque parole (Cortes, 2004); inoltre, presentano una più vasta varietà di strutture e funzioni da analizzare (Hyland, 2008a, 2008b);
- b. il numero minimo di frequenza dei *lexical bundles* nel *corpus* analizzato (*min. frequency*): in questa ricerca la ricorrenza minima dei *lexical bundles* di quattro parole è di almeno 10 volte su un milione di parole o *tokens* e la loro presenza deve essere individuata in almeno 5 articoli scientifici di autori diversi, al fine di non essere interpretati come la scelta personale di un autore (Biber ed altri, 1999; Conrad, Biber, 2004; Cortes, 2004, 2006; Hyland, 2008a, 2008b).

Biber e il suo gruppo di ricerca (1999) considerano come *lexical bundles* tutte le combinazioni di tre o quattro parole che ricorrono più di dieci volte su un milione di parole e che sono ripetute in almeno cinque o più testi nel *Longman corpus*. La presente ricerca segue il *cut-off point* proposto da Biber ed altri (1999), che individua come *lexical bundles* ogni combinazione di quattro parole che ricorre per almeno dieci volte su un milione di parole.

Il programma WordSmith legge attraverso tutti i testi presenti nel *corpus*, immagazzinando tutte le sequenze a partire dalla prima parola di ognuna e avanzando una parola alla volta. Ogni volta che una sequenza viene identificata, il programma la verifica automaticamente contrastandola con le sequenze identificate in precedenza. Se una data sequenza è stata già utilizzata nel *corpus*, il programma esegue un conteggio computazionale delle frequenze, per mostrare quanto spesso la data sequenza si ripete

(Scott, Tribble, 2006; Biber ed altri, 2004). Solo le sequenze ininterrotte di quattro parole sono state trattate come *lexical bundles*.

Nel *corpus* di filosofia il numero di ricorrenze di una parola o *token* è stato normalizzato, usando la formula di normalizzazione. La grandezza del *corpus* di filosofia è condizionata dalla difficoltà di accesso alle fonti elettroniche degli articoli scientifici. Dato che, per necessità operative, i *lexical bundles* dovrebbero essere analizzati in un *corpus* di almeno 1.000.000 di *tokens* si è ricorso a due tipi di procedure per affrontare questo problema (Cortes, 2002, 2004; Hyland, 2008a).

Nella prima fase, nel *corpus* di filosofia (780.989 parole) ogni combinazione di quattro parole che ricorre per almeno 7 volte è stata considerata come un *lexical bundle*.

Nella seconda fase, per poter comparare l'uso dei *lexical bundles* in tre *corpora* analizzati seguendo la procedura della normalizzazione, tutti i risultati ottenuti sono stati normalizzati su un milione di parole. Si è deciso di adottare questa procedura nonostante i suoi limiti e i suoi possibili effetti sui risultati della ricerca siano noti (Cortes, 2002; Biber, 2006).

Il calcolo dei *lexical bundles* è stato, quindi, effettuato seguendo questi parametri:

- lunghezza dei *lexical bundles*: combinazioni di 4 parole;
- ricorrenza minima: 10 volte su un milione di parole;
- presenza in articoli scientifici diversi: 5 articoli scientifici di autori diversi

su un milione di parole.

L'analisi si è sviluppata in tre fasi. Nella prima fase, tutti i *lexical bundles* con le loro rispettive occorrenze nel testo sono stati individuati, normalizzati dove necessario, e strutturati seguendo la tassonomia strutturale. Nella seconda fase, usando il *software* Word Smith e la tassonomia funzionale per i linguaggi scientifici proposta da Hyland (2008a), si è impostata l'analisi del contesto d'uso dei *lexical bundles*, individuando le loro funzioni predominanti e classificandoli seguendo le loro funzioni discorsive. Infine, i dati di tutti e quattro i *corpora* sono stati comparati allo scopo di stabilire le similitudini e le differenze, sia tra la conversazione e il discorso accademico che tra diverse discipline accademiche.



## II.5 Conclusioni

In questo capitolo si è cercato di descrivere la linguistica dei *corpora*, la disciplina che si serve del computer e dei programmi informatici per l'elaborazione di grandi quantità di dati linguistici, al fine di analizzare il più grande numero possibile di dati linguistici.

L'avvento dell'era informatica e della composizione a stampa elettronica ha permesso l'elaborazione di grandi quantità di dati, offrendo grandi opportunità alle utilizzazioni linguistiche.

Questo cambiamento ha permesso di vedere la lingua da un'ottica assai diversa. Secondo Sinclair (1982: 6), i *corpora* hanno consentito "new kinds of access to the patterns of the language [...] which are inaccessible to direct observation<sup>37</sup>".

Tramite specifici programmi informatici, è possibile costruire dei *corpora* e portare alla luce tratti linguistici che sono costanti nella produzione linguistica. I programmi permettono di classificare questi tratti e di descriverli in modo approfondito.

L'approccio utilizzato in questa ricerca è l'approccio *corpus-driven*, in cui i dati ricavati dal *corpus* vengono considerati nella loro interezza per trarne descrizioni esaurienti.

Le affermazioni ricavate dal *corpus* riflettono l'uso reale e autentico della lingua in esame. Analizzando i *corpora* microlinguistici di storia, filosofia e medicina del lavoro per *lexical bundles* si è cercato di individuare quelle peculiarità linguistiche, la cui pervasività indica un tratto saliente microlinguistico. Queste peculiarità possono essere evidenziate solo da un'analisi statistica dei dati.

---

<sup>37</sup>Traduzione: "nuovi modi di accesso ai *pattern* linguistici [...], i quali sono inaccessibili all'osservazione diretta".

## **Parte seconda**

### **Analisi dei casi**

### III Analisi dei casi

Partendo da questi presupposti, la presente ricerca si pone come obiettivo principale quello di presentare la pervasività dei *lexical bundles* sia nella conversazione quotidiana che nel discorso scientifico, offrendo una panoramica delle strutture grammaticali e delle funzioni discorsive dei *lexical bundles*, allo scopo di mostrare la loro importanza nella formulazione del discorso.

Analizzando due diversi registri – la conversazione e la prosa accademica – si intende mostrare come la prospettiva di registro risulti importante nello studio dei *pattern formulaici*. Nell'analisi dei *lexical bundles* è, inoltre, necessario tenere in massima considerazione un'altra caratteristica, la loro variazione a seconda delle discipline scientifiche. Prendendo atto della pervasività dei *lexical bundles* nelle microlingue scientifico-professionali e della loro variazione attraverso le aree scientifiche, la presente ricerca sostiene che i *lexical bundles* debbano essere integrati tra le caratteristiche sintattico-lessicali delle microlingue in oggetto. Il presente lavoro si concentra propriamente sull'analisi di tre microlingue, relative alle discipline di storia, filosofia e di medicina del lavoro.

Il presente studio è modellato seguendo le ricerche svolte per la lingua inglese e segue l'approccio dell'analisi dei *lexical bundles*. I *lexical bundles* vengono identificati utilizzando esclusivamente l'approccio basato sulla frequenza d'uso. Nei *corpora* microlinguistici e nel *corpus Lessico di frequenza dell'italiano parlato* (LIP) vengono analizzate tutte le combinazioni di quattro parole in sequenza. I *clusters* prevedono che i componenti ripetuti siano adiacenti e che la loro sequenza non sia interrotta da nessun altro segmento. D'accordo con quanto sostenuto da Spina (2001), le espressioni di questo tipo costituiscono una relazione più stretta di quella delle collocazioni, dato che le ultime comprendono sia elementi adiacenti che co-occorrenze di una parola in prossimità di altre.

Poiché fino ad oggi l'analisi dei *lexical bundles* non è mai stata applicata alla lingua italiana, il presente studio comincia con l'analisi dei *lexical bundles* ad un livello più globale, esaminando da un lato la conversazione e dall'altro il discorso accademico.

In un secondo momento, la ricerca si pone la seguente domanda: nella conversazione e nel linguaggio accademico della lingua italiana, esistono combinazioni

di 4 parole particolarmente frequenti? La risposta a questa domanda risulta essere positiva: espressioni di più parole o *lexical bundles* sono molto frequenti sia nella conversazione che nel linguaggio scientifico.

Inoltre, i risultati della ricerca mostrano che i *lexical bundles* variano a seconda della disciplina. La pervasività e la distribuzione diversa di queste combinazioni di parole attraverso registri e discipline richiama l'attenzione su una seconda domanda di ricerca: che struttura hanno questi *lexical bundles* e che funzioni discorsive svolgono?

Il frequente uso dei *lexical bundles* suggerisce inoltre una terza domanda che merita di essere presa in considerazione: se i *lexical bundles* sono così ricorrenti, dovrebbero essere considerati come una peculiarità microlinguistica?

L'ipotesi principale della presente ricerca è la seguente: i *lexical bundles* svolgono delle funzioni discorsive ben individuabili anche nelle microlingue analizzate. Pertanto essi possono essere considerati come un tratto microlinguistico saliente e come tale dovrebbero rappresentare una metodologia di analisi linguistica.

Seguendo le ricerche svolte per la lingua inglese, il presente studio dei *lexical bundles* si basa sull'ipotesi generale di Biber (2006) che sostiene che i modelli ad alta frequenza non sono accidentali, ma non sono neanche di per sé esplicativi. Piuttosto, i modelli di frequenza *corpus-driven* comprendono fatti descrittivi che richiedono una spiegazione (Biber, 2006). I fatti che devono essere spiegati sono esistenze di sequenze di più parole che non rappresentano unità strutturalmente complete, svolgono, però, delle funzioni discorsive ben identificabili (Biber ed altri, 1999).

Questo studio si propone quindi di esaminare sia la variazione dei *lexical bundles* attraverso i registri sia la loro variazione attraverso le diverse discipline scientifiche, facendo luce sulle seguenti domande:

- a) Quali sono i più frequenti *lexical bundles* nella conversazione e nelle tre discipline del registro accademico?
- b) Ci sono similitudini o differenze nelle strutture dei *lexical bundles* nella conversazione e nel discorso accademico?
- c) In che modo le variazioni nella struttura e nelle funzioni discorsive possono essere spiegate nei termini della variazione attraverso le discipline scientifiche?

- d) Quali sono le funzioni discorsive dei *lexical bundles*; che ruolo ricoprono i *lexical bundles* nel processo comunicativo?
- e) I *lexical bundles* sono discriminatori di genere microlinguistico e debbono come tali rappresentare una metodologia di analisi linguistica?

### **III.1 La diversità tra linguaggio parlato e scritto nella distribuzione dei nomi e dei verbi**

La diversità lessicale ed organizzativa del lessico del parlato rispetto allo scritto è stata ampiamente discussa in linguistica, con studi fondamentali per la lingua inglese Halliday (1989), Biber (1995), Biber ed altri (1999, 2004) e per la lingua italiana Laudanna, Voghera, Gazzellini (2002), Giordano e Voghera (2002), Voghera e Laudanna (2002), Laudanna, Gazzellini, De Martino (2004).

Gli studi di Halliday (1989) hanno mostrato come la diversa distribuzione delle categorie grammaticali possa essere un indizio di strategie di registro diverso. La diversa distribuzione delle categorie grammaticali è stata confermata in maniera puntuale dagli studi di Biber (1995) e di Biber ed altri (1999, 2004).

Le serie di accurate analisi dei *pattern* grammaticali tipici per diversi registri della lingua inglese (Biber ed altri, 2004) – quali la conversazione, il linguaggio scientifico, le lezioni universitarie e i libri di testo – hanno mostrato che i registri scritti si differenziano dai registri parlati nell'uso dei sostantivi, dei verbi e dei pronomi personali.

Lo studio svolto sul *Corpus Longman Spoken and Written English* del 1999 ha rilevato che i registri scritti mostrano una predilezione per i nomi in misura molto maggiore che per qualsiasi altra categoria grammaticale, mentre nei registri parlati domina l'uso dei verbi e dei pronomi personali. Nella *Longman Grammar* (Biber, 1999: 65) si nota che sia il linguaggio scritto, sia il linguaggio scientifico che la narrativa (*fiction*) mostrano una preferenza per l'uso dei nomi rispetto ai verbi, con una distribuzione di quattro nomi ad un verbo. Come emerge dalla ricerca, gli aggettivi e gli avverbi sono distribuiti in una maniera molto simile: gli aggettivi vengono utilizzati più comunemente nei registri scritti, mentre gli avverbi vengono preferiti nei registri parlati.

La presente sezione della ricerca parte dai risultati emersi da un lato da due studi svolti per l'italiano parlato nell'ambito dei *corpora* orali, dall'altro dall'analisi svolta sul *corpus Lessico di frequenza della lingua italiana contemporanea* (LIF), ovvero sul *corpus* dell'italiano scritto (Bortolini ed altri, 1972).

Il LIF rappresenta il primo grande progetto di costruzione di un lessico di frequenza per la lingua italiana. Si tratta di un *corpus* costituito da 500.000 parole della lingua italiana contemporanea, selezionate dal 1945 al 1968, ed è organizzato in cinque gruppi: teatro, romanzi, cinema, periodici e sussidiari.

La presente sezione dello studio si basa su *corpora* orali, quali raccolte di trascrizioni di testi orali, concentrandosi sul testo trascritto, sia dialogico che monologico.

La ricerca parte dal progetto COR-AL-ROM, all'interno della quale è nato il *corpus* LABLITA. Questo *corpus* di italiano parlato si compone di tre parti: un *corpus* di parlato spontaneo adulto che raccoglie circa 120 testi che riguardano situazioni comunicative diafasiche diverse per un totale di 60 ore; un *corpus* della lingua dei media (cinema, radio e televisione); e un *corpus* di 100 ore di italiano registrato nella fase del primo apprendimento (in bambini di 18-36 mesi). Inoltre, si è consultato anche il *Lessico di frequenza della lingua italiana* (LIP) del 1993.

Il LABLITA è stato incluso perché questo *corpus* è corredato da una dettagliata pubblicazione dei risultati emersi durante il progetto (Cresti, Moneglia, 2005). Il LIP è stato scelto perché rappresenta la più importante raccolta dell'italiano parlato, il cui materiale è liberamente consultabile sul sito BADIP della Karl-Franzens-Universität Graz<sup>38</sup>, rappresentando in questo modo una risorsa preziosa per le ricerche linguistiche.

Oltre a questi due progetti meritano di essere menzionati altri importanti progetti di ricerca svolti per l'italiano parlato e scritto. Per il registro parlato, si prende nota del progetto AVIP (*Archivio delle Varietà di Italiano Parlato*), conclusosi nel 1999; del progetto API (*Archivio di Italiano Parlato*), terminato nel 2001; del CLIPS (*Corpora linguistici per l'Italiano parlato e scritto*) compiuto nel 2003; dell'IPAR (*Italiano parlato*) del 2001; dei Lessici di Frequenza dell'Italiano Radiofonico del 1995; e del CIT (*Corpus di Italiano televisivo*) cominciato nel 1998 e ancora in svolgimento.

---

<sup>38</sup>Il *corpus* è liberamente consultabile sul sito <http://badip.uni-graz.at/>.

Per l'italiano scritto, si fa menzione del *corpus* CODIS/CODIS, *corpus* di *Italiano Scritto Contemporaneo*, nato nel 1998 e aggiornato tramite un *corpus* di monitoraggio inglobato con cadenza triennale; del *Corpus Penelope* (*corpus* di italiano scritto e orale, antico e contemporaneo); del TLIO (*Tesoro della lingua italiana delle origini*, composto da testi scritti in lingua italiana prima del 1375, in poesia e prosa) nato nel 1995; del LIZ (*Letteratura Italiana Zanichelli*), un *corpus* di testi della letteratura italiana; del CLIPS (*corpora* linguistici per l'italiano parlato e scritto) concluso nel 2004; del COLFIS (*Lessico di Frequenza dell'Italiano Scritto*).

Il progetto C-ORAL-ROM, "*Integrated reference corpora for spoken romance languages*", coordinato da Cresti e Moneglia<sup>39</sup> e svolto dal 2001 al 2004, rappresenta una risorsa linguistica del registro orale in quattro lingue romanze: l'italiano, lo spagnolo, il portoghese e il francese. Il progetto offre una raccolta della variazione del parlato spontaneo, ovvero della lingua quotidiana in tutta la sua variabilità e con tutte le sue peculiarità. Questo progetto costituisce quattro *corpora* rappresentativi, multimediali, che contengono più di 30 ore di registrazioni spontanee per ciascuna delle lingue, per un totale di 1.300.000 parole di trascrizione (Cresti, Moneglia, 2005).

Secondo questi studiosi, il lessico del parlato diverge dal lessico dello scritto nella densità lessicale, che è più bassa nel parlato, e soprattutto nella proporzione dei verbi rispetto ai nomi, la quale si presenta più alta nel parlato. In particolare, per quanto riguarda l'italiano viene confermato uno scarto a favore dei verbi. La ricerca conferma le ipotesi linguistiche sulla composizione del lessico del parlato, quali la prevalenza dei verbi rispetto ai nomi e il carattere verbale degli enunciati (Cresti, Moneglia, 2005).

Secondo questa interpretazione, il verbo mostra di essere il nucleo primario di costruzione testuale della lingua parlata. I testi sottoposti a spoglio sono stati raggruppati in sei gruppi. Il massimo della concentrazione dei verbi è individuabile nel nodo Informale Dialogico (Informale Dialogico e Telefono), con una progressiva diminuzione verso l'Informale Monologico e Media, per raggiungere il livello minimo nel Formale Monologico, che, oltre al gruppo Media, rappresenta il numero massimo dei nomi (Cresti, Moneglia, 2005: 216).

---

<sup>39</sup> Il progetto è stato realizzato da un consorzio europeo di istituti di alta formazione: Università di Firenze, Université de Provence, Fundação da Universidade de Lisboa, Universidad Autónoma de Madrid, Pitch Instruments France, Istituto Trentino di Cultura, European Language Association, Agency e Instituto Cervantes.

Come già menzionato, il Formale Monologico e i Media mostrano una minore presenza dei verbi e una maggiore presenza dei nomi. Questo dato è coerente con quelli raccolti da Voghera (2005), che attribuisce la diversa distribuzione dei nomi e verbi alla pianificazione minore o maggiore del discorso e alla presenza quantitativa dei dialoghi nel discorso. Difatti, il Formale Monologico e i Media prediligono i nomi ai verbi, con una pianificazione maggiore e una presenza minore dei dialoghi. La predominanza degli enunciati verbali rispetto ai non verbali è stata confermata in maniera puntuale dalle ricerche condotte in tutte e quattro le lingue romanze.

Secondo Biber ed altri (1999: 1071) il 38% delle unità di riferimento per l'inglese americano parlato risulta privo di verbo o di una struttura di clausola. Questo dato coincide in maniera quantitativa con il *corpus* romanzo, pur adoperando un metodo di analisi differente. Infatti, per la lingua inglese è stata svolta un'analisi morfosintattica, mentre nelle quattro lingue romanze gli enunciati sono stati identificati attraverso il riconoscimento percettivo dei loro *pattern* prosodici (Cresti, Moneglia, 2005).

I risultati delle ricerche hanno mostrato che più di un terzo delle unità prese a riferimento risulta privo di un verbo, un tratto che è uno dei più caratteristici del parlato (Cresti, Moneglia, 2005: 224). Secondo Cresti (2005) la percentuale più bassa degli enunciati non verbali nella lingua francese sarebbe da attribuire al modo d'identificazione degli enunciati, su base di pausa e non di *pattern* prosodico, impiegato dai ricercatori francesi<sup>40</sup>.

I risultati ottenuti nella ricerca COR-AL-ROM confermano le ipotesi sulla diversità del linguaggio parlato verso il linguaggio scritto nella distribuzione dei nomi rispetto ai verbi, mostrando una preferenza per i verbi nella conversazione.

Per quanto riguarda i *corpora* orali per la lingua italiana, al gruppo di ricerca coordinato dal professor De Mauro va il merito di aver aperto la strada con il progetto LIP (*Lessico di frequenza dell'Italiano Parlato*). Il *corpus* LIP, risalente al 1993, non è solo la prima e la più importante raccolta dell'italiano parlato, ma rimane senza dubbio anche la più adoperata nelle ricerche linguistiche.

Il *Lessico di frequenza dell'italiano parlato* (LIP) (De Mauro ed altri, 1993) è stato realizzato sulla base di un *corpus* dell'ampiezza complessiva di 475.883 parole

---

<sup>40</sup> Secondo quanto esposto da Cresti (2005: 224), i dati che fanno riferimento all'identificazione dell'enunciato presentano nel *corpus* francese delle percentuali discordanti; d'altro canto, tutti gli altri dati sono pienamente comparabili con quelli delle altre lingue romanze.



grafiche. Questa raccolta comprende registrazioni di testi parlati che sono state effettuate in quattro grandi città italiane: Milano, Firenze, Roma e Napoli. Questo permette di ottenere una raccolta di dati equilibrata che vede confluire tutto il territorio demografico italiano – toscano, romano, meridionale e settentrionale – evitando di basarsi su una varietà specifica dell'italiano.

Il materiale è stato raccolto secondo un criterio di gradazione dialogica, procedendo da una maggiore a una minore naturalezza, ovvero dal più dialogico al meno dialogico e dal meno monologico al più monologico.

I metodi di raccolta sono i seguenti: a) scambio comunicativo bidirezionale con presa di parola libera faccia a faccia (conversazioni in casa, al lavoro, in ambito scolastico e universitario, sui mezzi di trasporto); b) scambio comunicativo bidirezionale con presa di parola a distanza (conversazioni telefoniche in casa e al lavoro); c) scambio comunicativo faccia a faccia in cui i parlanti devono intervenire seguendo un ordine (assemblee, dibattiti, incontri di lavoratori, interrogazioni a scuola ed esami universitari, interrogatori processuali, interviste); d) scambio comunicativo unidirezionale in presenza del destinatario (lezioni scolastiche e universitarie, relazioni a congressi, comizi, omelie, conferenze non specialistiche, arringhe giudiziarie); e) scambio comunicativo unidirezionale o bidirezionale a distanza o differito su testo non scritto (trasmissioni televisive e radiofoniche).

L'analisi lessico-grammaticale del *corpus* consente di dare uno sguardo alla differenza tra la lingua parlata e scritta: la prima ricorre soprattutto a verbi, pronomi e avverbi, con una bassa percentuale di nomi e aggettivi, mentre nella seconda prevalgono sostantivi e aggettivi, seguiti da preposizioni e verbi (D'Agostino, 1998: 18). L'analisi di D'Agostino illustra la diversa distribuzione delle categorie morfosintattiche nei testi analizzati compresi nel *corpus* LIP (*Lessico di frequenza dell'italiano parlato*) e LIF (*Lessico di frequenza della lingua italiana contemporanea*), basandosi sulle analisi di frequenza del parlato e dello scritto. Come ha sottolineato De Mauro (1993) nelle conclusioni al LIP, nel parlato dominano il verbo e il pronome, seguiti da congiunzioni e avverbi, ovvero gli operatori di apertura, chiusura e variazione del discorso, mentre i nomi sono molto più rari che nell'uso scritto.

La differenza tra parlato e scritto considerata dalla frequenza d'uso di alcune categorie morfosintattiche è stata anche confermata dallo studio di Voghera (2005), che

ha sottoposto a spoglio il *corpus* LIP e LIF. La linguista ha analizzato dati provenienti da 10 diversi tipi di testi scritti e parlati, estraendo la frequenza dei nomi e dei verbi dei primi 2.000 lemmi di entrambi i *corpora*. In questo studio, la limitazione ai primi 2.000 lemmi deriva dal fatto che questi coprono il 91,1% del *corpus* LIP e il 90,7 % del *corpus* LIF (Voghera, 2005). Comparando la frequenza di occorrenza dei nomi e dei verbi nei testi, è emerso che la frequenza d'uso è l'unico criterio che ci può dare un'indicazione sul ruolo dei nomi e dei verbi nel linguaggio scritto e parlato, poiché la numerosità e frequenza di occorrenza presentano due tipi di misure diverse.

Solo il conteggio delle occorrenze ci permette di individuare l'uso reale delle categorie morfosintattiche in questione e di individuare somiglianze e differenze tra i diversi tipi testuali (Voghera, 2005: 135).

Dalle ricerche di Voghera (2005: 130) emerge che in tutti i tipi di parlato i verbi vengono usati più spesso dei nomi. Secondo Voghera, la diversa salienza di nomi e verbi è da attribuire al grado di pianificazione del discorso e alla quantità di dialogo. Nella lingua parlata il parlante non pianifica la conversazione in anticipo, prestando poca attenzione all'organizzazione del discorso. La non pianificazione del discorso rappresenta secondo Berruto (1985) uno dei fattori principali che caratterizzano il parlato.

Lo studio di Voghera (2005) ha mostrato che la conversazione, la quale rappresenta la produzione spontanea non pianificata e in cui i verbi usati sono circa il doppio, si colloca ad un estremo. All'estremo opposto si trovano i monologhi, frutto di una pianificazione maggiore e portatori di un numero minore di occorrenze verbali. Secondo Voghera (2005) la quantità di dialogo influisce sull'uso dei nomi e verbi nel parlato: i monologhi in cui il dialogo è assente presenta il numero minimo di verbi usati. Altresì si assiste ad una variazione nell'uso delle categorie morfosintattiche attraverso una modalità più o meno formalizzante del parlato (De Mauro, 1993).

Come mostrato dall'analisi del *corpus* LIF, la stessa diversità nella distribuzione di nomi e verbi viene riscontrata anche nello scritto (Voghera, 2005: 131).

Secondo Voghera (ibidem), la differenza nell'occorrenza dei nomi e verbi tra giornali, periodici e sussidiari da una parte e opere teatrali e copioni cinematografici dall'altra dipende dal fatto che le opere teatrali e copioni cinematografici imitano la conversazione spontanea, in cui l'uso dei verbi è più alto. La linguista attribuisce l'alta

frequenza dei verbi nei romanzi alle parti dialogiche nei romanzi e alle parti narrative in cui il testo si evolve attraverso clausole verbali.

Un'attenta valutazione del peso che i verbi e i nomi hanno nella lingua parlata e scritta è stata discussa da Voghera, in collaborazione con Laudanna (2002, 2003). Le osservazioni dei due linguisti consentono di delineare con più precisione la diversità nell'uso delle categorie morfosintattiche nei due registri.

I due studiosi attribuiscono un maggior numero di verbi usati nel parlato alla struttura sintattico-discorsiva del parlato, visto come una successione di clausole brevi, per lo più costituite da sintagmi verbali o da clausole con copula. Secondo questa interpretazione, il parlato rappresenta un processo linguistico interrotto e per questo mostra una preferenza per segmenti linguistici brevi.

Il dialogo tra il parlante e l'ascoltatore è costituito da clausole brevi con pochi dati nuovi, in modo da poter contribuire ad una conversazione fluente. Oltre a questo elemento, Voghera e Laudanna assegnano la diversa ripartizione alla diversa modalità linguistica – parlato verso scritto –, alle diverse abilità linguistiche – produzione scritta, comprensione orale, lettura – e alla specifica intersezione fra modalità e abilità linguistiche.

A questo punto, è opportuno accennare che nel 1993 De Mauro aveva sostenuto che il parlato può e deve percorrere delle scorciatoie: esso può avvalersi del contesto extralinguistico e, della prosodia, oltre a essere contrassegnato dall'immediata correggibilità nel caso l'ascoltatore non percepisca bene l'informazione. Secondo De Mauro, le scorciatoie possono causare un diverso uso delle forme morfosintattiche – uso evidente soprattutto sulle categorie grammaticali che operano come operatori di apertura, chiusura e variazione dell'enunciato, ovvero verbi, pronomi e avverbi.

Nella nostra analisi, è risultato utile anche uno studio comparativo di testi parlati e scritti appartenenti a registri diversi, di quattro lingue diverse, quali l'inglese, il tuvulan, il coreano e il somalo (Biber, 1995).

La ricerca conferma la variazione di registro in tutte e quattro le lingue che, dal punto di vista tipologico, sono completamente diverse. Oltre alla similarità indicata, le quattro lingue analizzate presentano somiglianze a diversi livelli: dalle similitudini nell'uso e nelle funzioni dei *pattern* tra le lingue, ai rapporti sincronici tra registri

all'interno di una lingua e ai diversi modelli diacronici all'interno di un registro e attraverso i registri (Biber, 1995).

In base a quanto esposto, possiamo concludere che anche se la frequenza d'uso dei nomi e dei verbi varia in rapporto a ciascuna tipologia testuale, essa è sempre profondamente connessa alla differenza tra registro parlato e registro scritto.

## **III.2 La distribuzione strutturale dei *lexical bundles* nel parlato e nello scritto**

I dati appena esposti ci permettono di sostenere che anche la struttura delle espressioni di più parole, o *lexical bundles*, sarà legata al registro diverso. Si suppone, pertanto, che la maggioranza dei *lexical bundles* legati al registro parlato avrà una struttura verbale, mentre la struttura delle espressioni di più parole appartenenti al registro scritto sarà nominale. Nelle due sezioni successive cercheremo di verificare questa ipotesi.

### **III.2.1 La distribuzione delle polirematiche nel parlato e nello scritto**

Il LIP dà spazio a diversi tipi di parole sintagmatiche e riporta 1254 parole polirematiche per 5.579 occorrenze su un totale di circa 16.000 lemmi, esclusi i nomi propri, i cognomi, i nomi geografici e gli esotismi.

Il *corpus* fornisce una lista di frequenza di lessemi complessi o polirematiche che include: a) espressioni il cui uso si è cristallizzato, per esempio *donna di servizio*; b) espressioni il cui significato non è deducibile dalla somma di quello delle loro componenti, per esempio *faccia di bronzo*; c) espressioni la cui funzione grammaticale è data dall'insieme dei loro membri, per esempio *nella misura in cui* (Voghera, 1994).

Come fa notare Voghera (1994: 186) l'occorrenza delle polirematiche non è da trascurare né dal punto di vista qualitativo né quantitativo, ma sembra indicare uno dei meccanismi generali del lessico di una lingua. Questo fatto viene dimostrato sia dall'uso

che i parlanti fanno delle polirematiche, sia dal fatto che queste intrattengono rapporti semantici con parole semplici o con altri lessemi complessi in modo del tutto uguale ai lessemi semplici (Voghera, 1994).

Inoltre, dal punto di vista della categoria grammaticale non esiste nessuna restrizione alla creazione delle polirematiche. Questo dato conferma che le polirematiche costituiscono uno dei meccanismi generali della creazione del lessico (Voghera, 1994).

Dallo studio di Voghera (1994) si può evincere che il gruppo più numeroso è quello dei lessemi complessi nominali, che costituiscono il 39,43% di tutti i lessemi complessi, mentre al secondo posto si trovano i verbi con il 14,35% e al terzo gli avverbi con una percentuale abbastanza simile. Oltre al numero di lemmi, per ciascuna categoria grammaticale viene indicata anche l'occorrenza d'uso. La frequenza d'occorrenza indica il peso che i lessemi complessi hanno nell'uso che ne fanno i parlanti.

Dai dati forniti la nostra intuizione che i *lexical bundles* verbali rappresentino una percentuale maggiore ai *lexical bundles* nominali sembrava essere messa in dubbio. L'intuizione viene però in parte sostenuta dalle osservazioni di Lo Cascio (1994: 63) – le polirematiche verbali, quali per esempio *mi sembra che sia stato giusto, ci hanno fatto aspettare, senti una cosa, vi raccontano un sacco di balle, sono profondamente convinto, ma va'*, sono escluse dall'elenco delle polirematiche nel LIP. Secondo Lo Cascio, un'analisi più approfondita ci porterebbe a scoprire una notevole quantità di strutture polirematiche non ancora prese in considerazione nel *corpus* LIP.

Il LIP, mettendoci a disposizione un vasto *corpus* di parlato, permette di analizzare il parlato dal punto di vista dei *lexical bundles* e di verificare se esso sia caratterizzato da espressioni formulaiche, da *clusters* lessicali, come già intuito da Lo Cascio (1994).

Come potremo osservare più avanti, l'analisi del *corpus* LIP circa i *lexical bundles* ci porta a scoprire una notevole quantità di combinazioni di parole che formano i modi in cui i parlanti italiani si esprimono e che dovrebbero essere prese in considerazione anche dal punto di vista lessicologico.

### III.2.2 I *lexical bundles* nel *corpus* parlato

Prima di dare avvio all'esposizione dei dati, è opportuno precisare la terminologia adoperata. Nel corso di questa ricerca verranno usati i termini seguenti: a) *lexical bundles* per riferirsi alla sequenza di 4 parole; b) *frequenza d'occorrenza* per specificare la frequenza d'uso delle varie forme nei testi.

I *bundles* – *ha detto che non, ha detto che sì, lui mi ha detto, mi ha detto che* – non sono stati raggruppati sotto la forma canonica – o "testa". Si è optato per questa scelta, perché si è voluto mostrare l'uso reale dei *lexical bundles* comparsi dall'estrazione computazionale.

Le forme scisse *c'è, c'era, c'erano, c'entra, c'entrano, l'altro, l'altra, d'altro, d'altra, dell'altro*, essendo singole forme grafiche, sono state considerate come un'unica parola, senza tentativi di separazione grafica. Per quanto riguarda le forme scisse *l'ho, l'hanno, l'ha, c'ha, m'ha, l'è*, queste sono state elencate come due parole, dato che nei corpora analizzati sono state riscontrate le loro controparti non scisse e si è preferito attenersi a un criterio unico.

#### III.2.2.1 La struttura grammaticale dei *lexical bundles*

Consideriamo a questo punto l'estrazione automatica dei *lexical bundles* dal *corpus* LIP. Dal *corpus* è stata estratta la sequenza di segmenti di 4 parole. La frequenza dei *lexical bundles* è stata normalizzata, seguendo la regola di normalizzazione come proposta già da altri ricercatori (Cortes, 2002, 2004; Biber, 2006).

Prima di procedere all'analisi strutturale, consideriamo alcuni dati statistici del *corpus* LIP e la ricchezza formulaica del *corpus* espressa nei termini di *lexical bundles* di 4 parole in sequenza.

Tabella 7: Dati statistici del *corpus* LIP – occorrenza 5 volte, *clusters* di 4 parole

<b>Dati statistici <i>corpus</i> LIP</b>	
Tokens	438.811
Types	24.741

Numero conversazioni	302
<i>Lexical bundles</i> diversi – occorrenza minima 5 volte	403
Frequenza dei <i>bundles</i>	3118
% delle parole in <i>bundles</i>	2,9%

I dati emersi dall'estrazione automatica del *corpus* sono dati attesi, nel senso che sia il numero alto di diversi *lexical bundles* che la loro effettiva frequenza d'uso nel parlato sono fenomeni già riscontrati nella lingua inglese e spagnola. I dati rappresentano una conferma dei dati ipotizzati.

L'analisi computazionale ha mostrato che nel *corpus* appaiono 403 tipi di *lexical bundles* diversi con una frequenza superiore o uguale a 5 volte su 24.741 parole, in almeno 3 diversi tipi del testo parlato. Il numero minimo di occorrenza è stato scelto seguendo la regola di normalizzazione. Sono stati inclusi nell'analisi solo i *bundles* riscontrati in almeno 3 diversi tipi di parlato trascritto per evitare l'idiosincrasia. Il numero delle conversazioni trascritte è stato scelto seguendo le analisi svolte sia per la lingua inglese che per quella spagnola.

Come già indicato in precedenza, la presente analisi si concentra sui *lexical bundles* di 4 parole in sequenza. L'estrazione automatica di *clusters* di 3 segmenti in sequenza rivela che questi sono estremamente numerosi, mentre il numero di quelli di 5 parole in sequenza cala notevolmente. Possiamo osservare questa tendenza nella tabella 8.

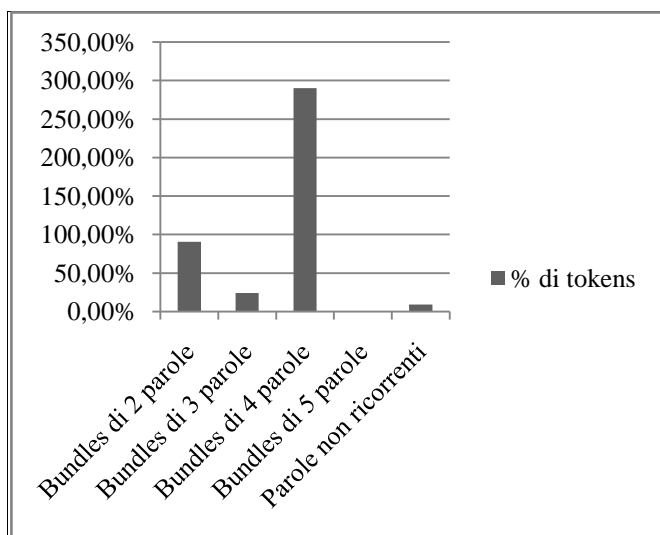
Tabella 8: *Clusters* diversi di lunghezza 3, 4 e 5 parole in sequenza con occorrenza 5 volte

<i>Lexical bundles</i>	Diversi	Occorrenze
<b>2 parole</b>	11.640	199.163
<b>3 parole</b>	3.512	35.291
<b>4 parole</b>	403	3.118
<b>5 parole</b>	50	348

Il dato ci suggerisce che, nella lingua parlata, le espressioni formulaiche ricorrenti sono abbastanza brevi, essendo costituite perlopiù da 3 o 2 parole. L'estrazione automatica delle sequenze da 2 e 5 parole consente dunque di delineare con più precisione la percentuale circa la formulaicità del linguaggio parlato. Il calcolo delle

espressioni di 2 e 3 parole in sequenza che figurano nel *corpus* LIP ha mostrato chiaramente la pervasività delle costruzioni prefabbricate. L'atteso risultato può vedersi bene nel grafico 1.

Grafico 1: Percentuale delle formule precostituite nel *corpus* LIP di 438.811 parole



Come si può evincere dal grafico 1, i dati coincidono con le percentuali indicate da ricerche sulla lingua inglese, secondo le quali la formulaicità del linguaggio varia da 38% fino a 70% (Howarth, 1996, 1998a; Cowie, 1991).

Si può osservare che la percentuale dei *lexical bundles* di 2 parole rappresenta il 90,7% di tutte le parole presenti nel *corpus*, mentre la percentuale di 3 parole costituisce il 24,1 % di tutti i *tokens* presenti nel *corpus*.

Si deve inoltre precisare che per il calcolo della percentuale formulaica nella lingua inglese sono state incluse le sequenze da 2 a 6 parole in sequenza (Howarth, 1998a). Il dato emerso dall'estrazione automatica supporta l'osservazione dei linguisti che la maggior parte del linguaggio sia strutturata in *set* prefabbricati (Howarth, 1996, 1998a; Cowie, 1991; Renouf, 1992; Renouf e Sinclair, 1991).

Per farsi un'idea dell'uso effettivo dei *clusters* è necessario distinguere tra la numerosità dei *bundles* e la loro reale occorrenza nel *corpus* (Voghera, 2005). Il conteggio delle occorrenze dei *bundles* mostra una percentuale di *set* prefabbricati usati dai parlanti. Le sequenze di 2 parole compaiono nel *corpus* 199.163 volte, pari a una frequenza del 90,7%; quelle composte da 3 parole occorrono 35.291 volte, pari al 24,1% di tutte le parole presenti nel *corpus*, mentre i *bundles* di 4 e 5 parole mostrano una



frequenza di 2,9% e 0,4% rispettivamente. Come si può vedere dal grafico 1, la distribuzione dei *clusters* evidenzia un andamento graduale, che cresce quando la sequenza è più breve. Lo stesso andamento di frequenze si riscontra nel *London-Lund Corpus*, il *corpus* di inglese parlato.

Gli studi di *corpora* di Biber ed altri (1999) hanno suggerito che i *bundles* di tre parole ricorrono più di 80.000 volte nella conversazione, mentre i *bundles* di quattro parole ricorrono più di 8.500 volte nella conversazione. I linguisti hanno riportato che il *bundle* di tre parole medio ricorre 25 volte su un milione di parole, mentre il *bundle* di quattro parole medio ricorre 20 volte su un milione di parole.

L'analisi del *corpus* LIP ha mostrato che i *bundles* di tre parole ricorrono più di 35.000 in un *corpus* di circa 500.000 *tokens*; quelli di quattro parole più di 3.000 volte. Lo spoglio ha ulteriormente rilevato che il *cluster* medio di tre parole occorre 20 volte su un milione di parole e il *lexical bundle* di quattro parole 15 volte su un milione di parole. Per il calcolo dell'occorrenza media delle espressioni formulaiche è stata adoperata la formula di normalizzazione.

Nel *corpus* di inglese parlato, circa l'80% delle parole fa parte di una combinazione di parole ricorrente (Altenberg, 1998). Nel *corpus* LIP, il 90,7% delle parole o *tokens* fa parte di una combinazione che viene ripetuta nel *corpus* – la lunghezza di queste sequenze varia da 2 a 5 segmenti.

L'alta frequenza delle sequenze ripetute rappresenta una prova della pervasività del linguaggio preconstituito, che dovrebbe essere preso in considerazione dall'analisi linguistica.

Dobbiamo inoltre precisare che la restrizione imposta all'inizio dell'analisi computazionale, circa l'uso dei *lexical bundles* di 4 segmenti in sequenza in almeno 3 testi di parlato trascritto, risulta di poco effetto. Come si può vedere dalla tabella (vedi Appendice 1), i *lexical bundles* che coprono i primi cinquanta posti compaiono in un numero che varia da 28 a 5 testi di conversazione trascritta.

Nella fascia successiva, dal 50° fino al 200° posto, i *bundles* compaiono in media in 6 conversazioni del parlato trascritto. Va inoltre menzionato che l'estrazione automatica mette in luce che solo a quota 271 *lexical bundle* si raggiunge il punto di occorrenza minima (almeno 5 volte a 438.811 parole) stabilita prima dell'analisi.

A questo punto possiamo esaminare i diversi tipi di parlato inclusi nel LIP. Data la proporzione quasi uguale del numero di parole ricorrenti (*tokens*) nei sottocorpora del LIP, si è ricorso alla normalizzazione uguale per tutti e cinque i sottocorpora.

Sono state considerate pertanto come *lexical bundles* tutte le espressioni di 4 parole che compaiono in ciascun sottocorpus per almeno 3 volte e in almeno 2 testi diversi.

Nell'analisi ci si è limitati ai *lexical bundles* presenti in un sottocorpus per almeno 3 volte anche se, seguendo la regola di normalizzazione, tutti i *bundles* con la frequenza di almeno 2 volte a 100.000 *tokens* avrebbero dovuto essere presi in considerazione.

Si è optato per questa scelta per facilitare la loro analisi strutturale, considerando il fatto che i *bundles* che ricorrono 2 volte sono estremamente numerosi. Le limitazioni imposte sono in qualche modo arbitrarie, tuttavia ci danno un'idea circa la preferenza di certe espressioni formulaiche da parte dei parlanti.

Tabella 9: Dati statistici dei sottocorpora del *corpus* LIP<sup>41</sup>

<b>LIP</b>	<b>Conversazioni (A)</b>	<b>Telefono (B)</b>	<b>Dialoghi (C)</b>	<b>Monologhi (D)</b>	<b>Media (E)</b>
<i>Tokens</i>	87.773	59.939	96.711	109.700	84.688
<i>Types</i>	7.842	5.715	9.468	11.807	9.603
<b><i>LB, occ. 2</i></b>	1584	824	2734	551	396
<b><i>LB, freq.</i></b>	3440	1913	8210	1435	1013
<b><i>LB, occ. 3</i></b>	127	124	2432	158	109
<b><i>LB, freq.</i></b>	525	513	7606	648	435
<b>% occ. 2</b>	15,7%	12,8%	34%	5,2%	4,8%
<b>% occ. 3</b>	2,4%	3,4%	31,5%	2,4%	2,1%
<b>Testi</b>	56	79	43	68	56

Il ricorso al linguaggio formulaico risulta dalla tabella 9 piuttosto evidente. La tabella mostra come alcune co-occorrenze di parole si ripetano con una frequenza che non è possibile sottovalutare. A tale proposito, Voghera e Laudanna (2002, 2003)

<sup>41</sup> Spiegazione delle abbreviazioni: LB = *lexical bundles*, occ. 2 = occorrenza di 2 volte, occ. 3 = occorrenza di 3 volte, freq. = frequenza d'occorrenza.

asseriscono che il parlato si compone di clausole brevi con pochi dati nuovi, in modo da poter contribuire ad una conversazione fluente.

Dall'altro canto, i linguisti Simpson-Vlach e Ellis (2010) suggeriscono che la conversazione prodotta in tempo reale richiede delle esigenze cerebrali maggiori rispetto allo scritto e per questo una maggiore necessità di fare affidamento alle formule precostituite. Come già in precedenza menzionato, la formulaicità viene vista dai linguisti Wray e Perkins (2000) come una scorciatoia per l'elaborazione dell'informazione.

Dalla lettura empirica dei dati del *corpus* LIP si ricava che la tipologia C, ovvero lo scambio comunicativo faccia a faccia in cui i parlanti devono intervenire seguendo un ordine (assemblee, dibattiti, incontri di lavoratori, interrogazioni a scuola ed esami universitari, interrogatori processuali, interviste), presenta un numero alto di occorrenza dei *bundles* di 4 parole che spicca su tutti gli altri. Dall'analisi statistica emerge che i *bundles* diversi con occorrenza 3 volte rappresentano il 31,5% di tutte le parole, mentre quelli con occorrenza 2 volte il 34% di tutte le parole presenti nel sottocorpus. Il dialogo non libero caratteristico per assemblee, dibattiti, incontri di lavoro, interrogazioni scolastiche, interrogatori processuali, interviste, è contraddistinto da una spontaneità della conversazione, anche se richiede una certa preparazione al discorso, la quale contribuisce ad una conversazione fluente. Da ciò si deduce che entrambi i fattori, sia la pianificazione della conversazione che l'uso massiccio di una serie di espressioni abituali o formule precostituite, le quali agevolano la comprensione e la produzione linguistica, contribuiscono a un dialogo fluente. L'alta occorrenza di *set* prefabbricati potrebbe, dunque, indicare che più i parlanti tendono ad un dialogo fluente, più confidano nelle formule precostituite, che fungono come scorciatoie di elaborazione dell'informazione. Inoltre, gli studi di Norrick (1987) hanno mostrato che l'uso di formule precostituite contribuisce alla coerenza del testo.

Il dialogo che, secondo Simpson-Vlach e Ellis (2010), richiede un'elaborazione cerebrale maggiore, confida, dunque, in una routine abituale, che opera come un'informazione linguistica familiare. Come tale costituisce una scorciatoia di elaborazione e soddisfa da un canto la necessità che l'informazione sia disponibile nella maniera e nei tempi richiesti e dall'altro il bisogno del parlante di ottenere ciò che vuole.

Tutti questi presupposti emergono nella tipologia del parlato che sta a metà strada tra il parlato libero da un lato e il parlato unidirezionale di alta formalità dall'altro.

Un altro fattore che spiega questa differenza nell'uso è l'integrazione diversa della prosodia, del contesto extralinguistico e dell'immediata modifica del discorso nel caso dell'incomprensione del messaggio. Questa è infatti al massimo grado nei dialoghi faccia a faccia e nelle conversazioni telefoniche e al minimo grado nelle trasmissioni radiofoniche e televisive senza dibattito o in altri contesti unidirezionali formali. D'altro canto, la tipologia del parlato unidirezionale formale si avvicina di più allo scritto e come tale si avvale di più delle risorse sintattico-grammaticali e lessicali diverse, in altre parole di un patrimonio linguistico più ricco e variegato e più accuratamente scelto.

Da queste premesse si deduce pertanto che se da un lato i dialoghi faccia a faccia confidano di più nel contesto extralinguistico e nella dimensione fonologia, dall'altro le conversazioni con una modalità più formalizzante fanno affidamento sulla ricchezza linguistica e morfosintattica. Viene inoltre confermata l'ipotesi di Simpson-Vlach e Ellis (2010) che il dialogo in tempo reale ricorre alle formule precostituite per facilitare l'elaborazione mentale.

I dialoghi formali che si trovano a metà strada tra formale e informale dialogico ricorrono di più alle formule precostituite, agevolando l'elaborazione cerebrale e garantendo in questo modo una comunicazione fluente e coerente, necessaria per il parlato di alta formalità.

In un secondo punto di analisi si è esaminata la struttura grammaticale dei *bundles* di 4 parole, comparsi dall'estrazione automatica dei *corpora* presenti. La tabella sotto riportata illustra la struttura dei *lexical bundles* identificati nel *corpus* LIP. Anche se la maggioranza dei *lexical bundles* non presenta una struttura grammaticalmente completa, i *clusters* possono essere raggruppati in cinque gruppi principali diversi.

Sulla scorta delle ricerche svolte per la lingua inglese, i *lexical bundles* sono stati classificati in base alle prime parole dell'espressione e raggruppati in cinque categorie principali seguendo le caratteristiche specifiche della lingua italiana.

Le principali categorie sono: a) *lexical bundles* che incorporano frasi verbali; b) *bundles* che contengono segmenti di frasi preposizionali; c) *bundles* che incorporano elementi di sintagmi nominali; d) *clusters* che sono congiunzioni; e) *bundles* che vengono classificati come "altre espressioni". Le categorie sono state scelte in base alle

forme dei *clusters* emersi dall'estrazione automatica. Si è cercato di raggruppare i *bundles* in tanti sottogruppi quante sono le forme diverse che presentano, anche rischiando di avere dei gruppi con una sola forma. Tra parentesi viene indicata la rispettiva frequenza di ciascun *lexical bundle* identificato.

Il gruppo I rappresenta i *bundles* verbali e viene suddiviso in 19 sottogruppi: i *clusters* che cominciano con un pronome personale (per esempio *io gli ho detto, io non l'ho*); i *bundles* che cominciano con un verbo attivo alla prima persona singolare (per esempio *farei nel primo giorno*) o in tutte le altre persone verbali (*non è riuscito a*); le espressioni che incorporano un verbo seguito da una clausola frammentata (*io non credo che*); quelle con il verbo *essere* che funge da copula ed è seguito da un nome (*è una cosa che*), da una clausola frammentata (*è anche vero che*), da un sintagma preposizionale frammentato (*è per quello che*), o il verbo *esserci* come copula e seguito da un sintagma nominale frammentato (*c'è anche un'altra*), da un pronome indefinito (*c'è niente da fare*) o da un avverbio (*c'è un po' di*), e infine vengono elencate le altre forme con il verbo *essere*, come per esempio *sì sì sì è*; i *bundles* che incorporano la forma impersonale del verbo (*mi sembra che non*); quelli con un verbo seguito da un sintagma preposizionale frammentato (*è in grado di*); un pronome (dimostrativo – *quello che ho detto*, o indefinito – *ognuno ha le sue*); un avverbio (ad esempio *ecco questo è un*); i *bundles* che cominciano con *che* seguito da un sintagma verbale frammentato (come per esempio *che non l'ho*); con un marcatore del discorso (*e, eh, ah, poi, quindi, ma, anche*); un nome (*il problema è che*); una congiunzione (*perché se non è*); una preposizione (*per far sì che*); un interrogativo (*che cosa vuol dire*); quelli che rappresentano un sintagma verbale frammentato alla forma passiva (*va fatto da un*); i *bundles* che sono frammenti di proposizioni relative (*in cui usciamo da*) e infine quelli che iniziano con un participio, come per esempio *capito sì sì sì*.

In base alla numerosità delle espressioni di quattro parole comparse dall'estrazione automatica del *corpus*, abbiamo deciso di elencare come secondo gruppo i frammenti preposizionali, anche se la differenza tra il numero dei sintagmi preposizionali e quelli nominali è minima. Abbiamo suddiviso il gruppo in quattro sottogruppi, considerando sempre il primo elemento dell'espressione. Sono stati pertanto identificati i *bundles* che rappresentano un frammento di una frase preposizionale (per esempio *dal punto di vista, all'interno di questo*) – distribuiti in sottocategorie a seconda della preposizione

adoperata – e inoltre quelli che iniziano o includono frammenti di complemento *conche* (per esempio *di una persona che, che a questo punto*); quelli che iniziano con un marcatore del discorso (come per esempio *anche dal punto di, e a un certo punto*) o incorporano frammenti di frasi relative (*nel momento in cui*).

Il terzo gruppo rappresenta i frammenti che incorporano sintagmi nominali, suddivisi in sei sottogruppi. Il primo sottogruppo rappresenta i *bundles* contenenti la preposizione *di* (*un gruppo di lavoro, la maggior parte di*). Nel secondo sono stati classificati i *clusters* che incorporano una preposizione diversa dalla preposizione *di* (*la parola al presidente, il cinquanta per cento*); un complemento frammentato (*l'unica cosa che*); sintagmi nominali completi o frammentati (per esempio *un milione e mezzo, il sabato e la*), e quelli che iniziano con un marcatore del discorso o un clitico (*eh una serie di*) o una congiunzione (*se un gruppo di*).

Le espressioni di quattro parole delle ultime due categorie rappresentano congiunzioni (*in modo tale che, a meno che non*) e "altre espressioni", come per esempio *uno due tre quattro, sì sì sì sì*.

Proponiamo di seguito la classificazione strutturale dei *lexical bundles* di 4 parole del corpus LIP.

### III.2.2.1.1 *Lexical bundles* che rappresentano frasi verbali complete o incorporano frammenti di frasi verbali

- a. *Lexical bundles* che incorporano il pronome personale *io* + sintagma verbale frammentato

io gli ho detto (12); io non l'ho (8); io non so se (6); io non mi sento (6); io mi rendo conto (7); io non lo so (35); va be' ma io (7); ce l'ho io (6); io ce l'ho (12); non lo so io (10);

- b. *Bundles* verbali con il verbo alla prima persona singolare

farei nel primo giorno (6); l'ho detto prima (6); te l'ho detto (6); gli ho detto ma (5); l'ho già detto (5); ve l'ho detto (5); direi che per ogni (6); ho capito va bene (9); ho capito ho capito (14); lo so lo so (8); ho capito va be' (7); sì, ce l'ho (6);

non ce l'ho (15); e che ne so (9); non l'ho sentito (5); no non l'ho (6); non l'ho mai (6); non me lo ricordo (6); più non lo so (5); non lo so eh (12); non lo so se (12); non lo so ma (11); non lo so non (8); no, non lo so (12); non lo so comunque (6); ma non lo so (6);

c. *Bundles* in tutte le altre persone oltre alla prima persona singolare

non è riuscito a (5); non me l'ha (5); te l'ha detto (5); veniamo a casa vostra (5); me l'ha detto (6); ha le sue idee (6); ha qualcosa da dire (6); va in classe normalmente (6); usciamo da questa scuola (9); ci troviamo di fronte (7); non ce la fa (6); allora vediamo un po' (8); chi l'ha detto (6); sì, guarda chi è (6); non ce l'ha (7); farà dei gruppi di (6); non ci sta no (5);

vuol dire vuol dire (5); va bene ciao ciao (9); bene va bene va (11); non è non è (16); va be' ma non (8); no ma non è (7); va bene va bene (46); chi è chi è (5); sì sì va bene (10); va be' va bene (21); no no va be (8); sì sì va be' (5); bene va bene ciao (5); no no va be' (5); va be' e va (5); va be' va bene (5); succede al momento in (6);

d. Verbo + *che* - clausola frammentata

io non credo che (6); io credo che il (5); io credo che questo (5); io penso che sia (5); ho l'impressione che (5); non è detto che (13); mi ha detto che (11); mi rendo conto che (9); ha detto che non (7); mi hanno detto che (5); non lo so che (5); mi sembra che non (7); mi pare che sia (5); s'era detto che (5); m'ha detto che (10);

e. *Bundles* con il verbo essere che funge da copula

- essere/non essere + sintagma nominale frammentato

è un'altra cosa (15); è la stessa cosa (12); è una cosa molto (9); non è una cosa (9); è un problema di (8); è una cosa importante (6); è una cosa un (6); è la cosa più (9); è un'altra un (5); è il caso di (5); è la prima volta (7); è il caso di (5); è un po' di (10); non è il caso (5); non è solo la (5); non è un caso (5); non è solo un (7); non è un problema (9);

- essere + sintagma preposizionale frammentato

è per quello che (6); è in grado di (6);

- essere + *che* - clausola frammentata

è anche vero che (5); non è che lo (5); non è che si (9); non è che ci (7); tant'è vero che (11); noi non è che (5); non è che non (11); è un caso che (6); è il fatto che (6); è un problema che (9); è un discorso che (9); è il fatto che (6); non è che io (6); io non è che (6); è vero che non (5); è una cosa che (27); questo è quello che (11);

- esserci + frase nominale frammentata

ci sono anche delle (5); c'è anche un'altra (6);

- esserci + pronome indefinito

c'è niente da fare (6);

- esserci + avverbio

c'è un po' di (9); è stato un po' (5); non ci sono più (5); è sempre stato così (6);

- Altre forme con essere

sì sì sì è (8); no no no è (6); è vero è vero (6);

- f. *Lexical bundles* con la forma impersonale del verbo

si tratta di un (6); si tratta di eh (5); si tratta di una (5); non si sa mai (6); si farà dei gruppi (6);

- g. Verbo + sintagma preposizionale frammentato

troviamo di fronte a (5); è in grado di (5); vediamo un po' di (5);

- h. Infinito + clausola dipendente

guardare chi è assente (6).

- i. *Lexical bundles* verbali che cominciano con un pronome

- pronome dimostrativo + *che* – complemento frammentato

quello che è il (9); quello che è stato (8); quello che è successo (6); quello che può essere (6); quello che è l' (5); quello che ha fatto (5); quello che ho detto (5); quello che sta succedendo (5); quello che ti pare (5); quelli che sono i (14); quelli che vogliono andare (6); quella che è la (7); quella che è stata (5); quelle che sono le (13);

- pronome dimostrativo + verbo essere + frammento

questo è un altro (11); questo è un po' (6); questo è un problema (9);  
questa è una cosa (11); questa è la mia (6); questa è un'altra (6); questa è stata la (6);

- pronome indefinito + sintagma verbale frammentato

ognuno ha le sue (6);

- l. Sintagmi verbali frammentati che cominciano con un avverbio

ecco questo è un (6); ora so nove argomenti (6);

- m. *Che* + sintagma verbale frammentato

che ci sia un (8); che è una cosa (8); che non è che (6); che non è solo (6); che fanno parte della (5); che non l'ho (5); che è quello che (5); che vedete in sovrimpressione (5); che ne so io (5); che non è un (5);



n. Marcatori del discorso + clausola verbale frammentata

ah e va be' (5); ah non lo so (11); ah va be' va (6); poi ci sono anche (5); ma non è vero (5); ma non è che (9); ma questo è un (7); e quindi questo è (5); e che si chiama (5); e questo è il (9); e non lo so (10); e va be' e (8); e va be' ma (6); e che ci sono (7); e questa è una (5); e che ne so (9); cioè non è che (6); quindi questo è un (7); quindi non è che (6); eh non è che (7); eh vediamo un po' (5); eh non lo so (7); sì sì sì va (8); be' e va be' (5); be' ma non c'è (5); e niente da fare (6); e questa è la (14); eh per quanto riguarda (8);

o. Nome + frase verbale frammentata

il problema è che (15); questa scuola può essere (6); scuola può essere un (6); un'altra cosa è (5); una cosa che è (5); un discorso che riguarda (5); numeri che vedete in (5); le persone che sono (5); l'unica cosa è (5); il discorso è questo (10); lavoro è formato da (6); il problema è questo (6);

p. Congiunzione + frase verbale frammentata

se non ricordo male (5); come si è detto (5); quando si parla di (6); perché è una cosa (7); perché se non è (6); no perché non è (5);

q. Interrogativo + frase verbale frammentata

che cos'è che (7); che cosa vuol dire (12); che cos'è e (6); come si può dire (11); com'è che si (6); cosa succede al momento (6);

r. Finale frammentata

per far sì che (5);

s. Frase passiva frammentata

va fatto da un (6); è formato da sei (6);

- *che* + frase passiva

che è stato fatto (5); che vogliono andare di (6);

t. Frase relativa frammentata

in cui usciamo da (9); momento in cui usciamo (9); cui usciamo da questa (9);

u. Preposizione + sintagma verbale frammentato

di quella che è (7); di quelli che sono (7); a quello che è (6); di quello che è (14); per quello che riguarda (8); per quanto riguarda la (17); per quanto riguarda il (12); per

quanto riguarda l' (10); per quanto riguarda le (9); per quanto riguarda i (7); per ora so nove (6); da questa scuola può (6); a quella che è (6); a che fare con (7);

v. Participio + frammento verbale

fatto da un alunno (6); formato da sei persone (6); capito sì sì sì (5).

### III.2.2.1.2 *Lexical bundles* che incorporano frammenti di sintagmi preposizionali

a. Frammenti di sintagmi preposizionali

- la preposizione A: all'ordine del giorno (8); a parte il fatto (5); a questo punto di (5); a questo punto eh (5); a questo punto non (5); alle prime dieci persone (5); a un certo punto (25); al cento per cento (16); all'una e mezza (10); alle due e mezza (10); ad un certo momento (6); a portata di mano (6);
- locuzioni preposizionali con A: al di là di (21); al di là del; alla fine dell'anno (10); all'interno di un (6); all'interno di questo (8);
- la preposizione DI: dei gruppi di lavoro (6); d'altra parte non (6); di là di questo (5); di una serie di (5); di che cosa si (6); di un milione e (5); di questo tipo di (12); di lavoro e formato (6); di nomi tipo eh (6);
- locuzioni preposizionali con DI: di fronte a un (5); di fronte a una (5);
- la preposizione IN: in un altro modo (6); in fin dei conti (5); in questo caso non (5); in tutti i modi (5); in un modo o (5); in una situazione di (5); e nello stesso tempo (8); in tutto il mondo (16); in un certo senso (15); nel modo più assoluto (10); nei luoghi di lavoro (7);
- la preposizione DA: dal punto di vista (43); da un punto di (20); da parte dei professori (6); da un'altra parte (6); da questo punto di (15); da tutte le parti (9);
- la preposizione PER: per un po' di (5); per una serie di (5); per la prima volta (9); per tutta la vita (7);

b. Frammenti di clausola preposizionale con *che*

- frase preposizionale + complemento frammentato

di una persona che (6); di quello che si (5); di tutto quello che (5); nel senso che eh (5); nel senso che non (5); di una cosa che (5);

- che all'inizio

che in questo momento (5); che a questo punto (7);

c. Clausola preposizionale con frase relativa frammentata

nel momento in cui (25); al momento in cui (9); un momento in cui (8);

d. Marcatore del discorso + frammento di clausola preposizionale

anche dal punto di (5); ma al di là (5); però a questo punto (5); e a questo punto (6); allora a questo punto (6).

### III.2.2.1.3 *Lexical bundles* che incorporano sintagmi nominali

a. Sintagma nominale con frammento di una frase preposizionale che comincia con la preposizione DI:

un punto di vista (22); un gruppo di lavoro (19); il numero di telefono (17); un quarto d'ora (15); una cosa del genere (13); una lista di nomi (6); questo punto di vista (15); un dato di fatto (5); un punto di riferimento (5); una regola di condotta (5); un po' di più (10); un po' di tempo (9); un po' di musica (5); cose di questo genere (7); il disegno di legge (15); i gruppi di lavoro (9); vari gruppi di lavoro (6); tutta una serie di (11); la maggior parte dei (5); un certo numero di (5); un certo tipo di (5); una serie di eh (5); punto di vista della (9); centinaia di migliaia di (6); punto di vista di (5); punto di vista eh (5); lista di nomi tipo (6); gruppo di lavoro e (6); i vari gruppi di (6); quarto d'ora ma (6);

b. Sintagma nominale con frammento di una frase preposizionale che comincia con una preposizione diversa da DI

situazione in medio oriente (9); anche situazione in medio (6); la violenza negli stadi (6); come in questo caso (5); il cinquanta per cento (10); la parola al presidente (5);

c. Sintagma nominale con complemento frammentato

l'unica cosa che (11); la prima volta che (11); un'altra cosa che (8); tutte le volte che (7); il fatto che eh (6); il fatto che non (6); l'ultima cosa che (6); una cosa che non (6); tutte le persone che (6); la prima cosa che (5); una cosa che io (5); il gruppo di lavoro che (6);

d. Sintagma nominale frammentato e non

un milione e mezzo (8); un anno e mezzo (7); la cosa più importante (9); una cosa un po' (7); il sabato e la (8); sabato e la domenica (7); le stesse cose e (5); la politica un altro (6);

e. Marcatore del discorso + clausola nominale frammentata  
eh una serie di (5); poi un'altra cosa (5); e anche un'altra (6);

f. Congiunzione + clausola nominale frammentata se un gruppo di (6).

#### III.2.2.1.4 *Lexical bundles* che sono congiunzioni

in modo tale che (7); a meno che non (6).

#### III.2.2.1.5 Altre espressioni

ma io non mi (5); no no no no (36); no no no io (8); no no ma non (6); no no no non (7); non ha non ha (5); ah sì sì sì sì (6); due o tre volte (7); un po' un po' (5); grazie ciao ciao ciao (5); eh sì sì sì (5); e poi e poi (5); due un milione e (5); ciao ciao ciao ciao (17); e allora e allora (12); sì sì sì sì (21); di di di di (13); sì sì no no (6); uno due tre quattro (6); di un di un (5); un'altra un'altra (5); e un e un (6); che cosa che cosa (7); zzz zzz zzz zzz (16).

Come si può osservare nella tabella 10, il gruppo maggiormente rappresentato delle espressioni di quattro parole nel *corpus* LIP è quello dei *bundles* verbali. Seguono i *clusters* preposizionali e nominali. È stato identificato un numero abbastanza esiguo di espressioni congiunzionali e di altre espressioni.

Tabella 10: Distribuzione strutturale dei *lexical bundles* di 4 parole in sequenza nel *corpus* LIP

<b><i>Lexical bundles</i> di 4 parole</b>	<b><i>Bundles diversi</i></b>
<b>Verbali</b>	249
<b>Preposizionali</b>	68
<b>Nominali</b>	60
<b>Congiunzionali</b>	2
<b>Altre espressioni</b>	24

L'analisi dei *lexical bundles* svolta nel *corpus* LIP ha rilevato che la stragrande maggioranza dei *bundles* nel parlato presenta la struttura verbale che costituisce più del 62% di tutti i *bundles* presenti nel *corpus*. Da tutti i *bundles* identificati il 29,5% ha la struttura di una clausola dipendente frammentata o completa, mentre tutti gli altri hanno la struttura frasale. Da contro, nella lingua inglese la maggioranza dei *lexical bundles* ha la struttura di una clausola dipendente, costituita da un frammento verbale.

Sia l'alto numero dei *bundles* verbali che la loro frequenza d'uso che copre il 61% di tutte le frequenze d'uso dei *clusters*, coincidono con le ricerche condotte per la lingua inglese. La percentuale riscontrata nella lingua inglese è ancora più alta – copre il 90% dei *lexical bundles* identificati nella conversazione (Biber ed altri, 1999, 2004).

Questo dato non solo conferma la nostra ipotesi posta all'inizio della ricerca, ma si accorda anche con i risultati delle analisi svolte per la lingua italiana, le quali dimostrano che una delle peculiarità del parlato è la sua struttura verbale.

Come riportato da Biber ed altri (1999), solo il 15% dei *lexical bundles* nella conversazione può essere guardato come composto da frasi o clausole complete, mentre il resto è composto da frasi frammentate o clausole con nuovi frammenti incorporati.

Dallo spoglio svolto sul *corpus* LIP è emerso che il 18,4% delle espressioni di quattro parole è composto da sintagmi completi. Si può osservare che la percentuale delle clausole complete risulta in lingua italiana un po' più alta – la preferenza del parlato per clausole brevi viene espressa anche in termini dei *lexical bundles* di 4 parole in sequenza che presentano una forma completa. A mo' di esempio riportiamo alcuni *lexical bundles* completi: *io non lo so, ho capito va bene, ho capito ho capito*.

### ***III.2.2.2 I tratti del parlato identificati nella struttura dei lexical bundles***

È importante constatare che i tratti tipici del dispositivo del parlato si possano osservare anche nella struttura dei *lexical bundles*.

La frammentarietà del discorso costituisce un tratto caratteristico dei testi parlati. A tale proposito, è interessante notare che la categoria dei *bundles* verbali include un gruppo abbastanza ampio dei *bundles* – il 26,6% – che non sono frammenti ma formano una frase completa, come per esempio *io non lo so, non ci sono più, non ce la fa*.

Come già messo in evidenza da Laudanna e Voghera (2002, 2003), la lingua parlata è contraddistinta da un procedere per piccoli blocchi semantico-sintattici, per lo più costituiti da sintagmi verbali o da clausole con copula.

Il parlato predilige le clausole brevi perché è contrassegnato da un processo linguistico interrotto. Quest'osservazione trova riscontro nei *lexical bundles* che rappresentano delle frasi complete, come per esempio *io non lo so, non è un problema,*

*allora vediamo un po'*. Sono stati identificati 73 *clusters* che possono essere considerati come frasi complete, la cui effettiva frequenza d'uso è del 20,2% di tutte le occorrenze dei *bundles*. Le espressioni individuate hanno la forma di un sintagma verbale o di una clausola con copula.

Un altro tratto caratteristico del parlato è la ripetizione, volta ad aggiustare l'enunciato o semplicemente a mettere in risalto la conversazione (Berrutto, 1993). Questo tratto si riflette soprattutto nel gruppo dei *bundles* che formano delle frasi complete, come per esempio *ho capito ho capito, va bene va bene, chi è chi è, che cosa che cosa*, o da quelli che sono stati classificati sotto "altre espressioni", come per esempio *ah no no no, no no ma non, e un e un*.

Riportiamo a mo' di esempi alcune sezioni dei testi parlati:

Dialogo 1 – Telefono (Tipologia B)

A: Come no voi c'andate?

B: Ci andiamo domani sera.

A: E quando ritornate?

B: Ma penso sabato sera o domenica.

A: *Ah ho capito ho capito*.

B: Probabilmente sabato.

A: Andate tu e l'ingegnere?

Dialogo 2 – Dialogo informale (Tipologia A)

A: Sì, certo quelli di quelli di quelli di dell'anno scorso.

B: Dell'anno scorso o quella.

A: Okay okay *va bene va bene*.

A: Quindi dobbiamo mandare anche quest'altro importo tanto non ha questo capitolo non ha non ha problemi né di capienza.

Il *lexical bundle va bene va bene* costituisce inoltre un segnale di conferma e contatto da parte del parlante, un tratto tipico del parlato.

Nel parlato, il ricorso alle clausole brevi viene bilanciato con il cumolo di vari tipi di particelle discorsive, quali segnali di apertura, di chiusura, di conferma o con le riformulazioni del discorso (Berrutto, 1993).

Questo dispositivo tipico del parlato appare anche nella struttura dei *lexical bundles*. In questo gruppo troviamo: a) esempi che fungono come segnali di apertura del discorso, come per esempio *allora vediamo un po'*, *ah vediamo un po'*, *ecco questo è un*;

Questo tratto lo possiamo osservare nella sezione del dialogo seguente, tratto da una conversazione formale (Tipologia C).

Dialogo 3 – Conversazioni formali (Tipologia C)

A: Cioè lei non ha fatto il biennale di inglese no senta io lo devo sapere perché devo farla leggere un po' in inglese e vedere com'è capito no per questo glielo chiedevo no. Di quale testo preferisce parlare.

B: Di Benito Cereno.

A: Allora eh *allora vediamo un po'*. Incominci a tradurre però vediamo un po' perché insomma anche se l'ha letto in italiano però qui adesso sulla traduzione vediamo.

b) quelli che introducono riformulazioni o spiegazioni, come per esempio *cioè non è che* – osserviamo l'uso nel dialogo tratto da una conversazione informale (Tipologia A):

Dialogo 4 – Conversazioni informali (Tipologia A)

A: Signora ma io l'aiuto questi ragazzi.

B: E lei lo so.

A: Però bisogna che loro si mettano in condizioni da sé. Di studiare. Di fare il triennio perché poi dopo non l'aiuta più nessuno *cioè non è che* non l'aiuti più nessuno.

B: Sì.

A: Però la cosa si fa più difficile e più complicata.

B: Che cosa devo fare con questa materia.

c) fatismi o intercalari che fungono da segnali di apertura o hanno valore di conferma come per esempio *ah non è che, ah e va be', si tratta di eh* – riportiamo a titolo d'esempio un dialogo tratto da una conversazione telefonica:

Dialogo 5 – Telefono (Tipologia B)

A: Ah e n'ho capito ah aho' ma so' terribili sa'.

B: Ah. *Ah e va be'* io dico la verità lì so' stata cretina pure io perché io non ci ho pensato Rino me l'ha detto dopo due tre giorni.

A: E lo so ma.

B: Ah hai capito.

d) la risalita del clitico *eh* appare evidente anche nella struttura dei *bundles* – *eh vediamo un po', eh non lo so, eh per quanto riguarda, eh non è che* – come possiamo osservare nella sezione seguente di un dialogo informale:

Dialogo 6 – Conversazioni informali (Tipologia A)

A: Allora vediamo un attimo *eh vediamo un po'* che numero è dodici dunque allora all'inizio del secondo quadrimestre ha avuto un sei più un sei negli scritti va bene è uno dei pochi che hanno dei veramente dei buoni scritti e ha avuto un sei e mezzo e un sette meno meno nei temi.

f) e anche quelli che enfatizzano o attenuano il valore dell'enunciazione vediamo un po' di, è stato un po', dove un po' ha il valore attenuativo, come in:

#### Dialogo 7 – Media (Tipologia E)

A: Katuscia eh aspetta cara perché ecco fino a un momento fa avevo una penna che adesso non mi trovo più che bello che bello che bello Katuscia abbi pazienza intanto *vediamo un po' di* scegliere solerte regia ecco grazie infinite eh mi porta una penna eh che numero vuoi cara che numero vorresti.

In generale, nel campo delle forme morfosintattiche vi è una maggiore frequenza dei pronomi personali e dimostrativi (Berrutto, 1997). Dall'analisi strutturale dei *lexical bundles* è stato riscontrato un numero notevole di *clusters* che cominciano con il pronome personale *io*, come per esempio *io gli ho detto, io mi rendo conto, io non so se, io non credo che*, o il pronome personale tonico, come in *mi ha detto che, mi sembra che non*.

È interessante notare che sono state identificate poche espressioni di più parole che contenevano un pronome personale diverso da *io*. Questo fatto è da collegare all'enfasi espressiva del dialogo, che accentua la posizione del parlante. Molte delle espressioni identificate sono di natura interpersonale e servono per esprimere un commento del parlante, la sua volontà di affermare o rifiutare qualcosa, come per esempio *io gli ho detto, ah sì l'ho, io non lo so, io non credo che*, come in:

#### Dialogo 8 – Media (Tipologia E)

A: [...] ma questo è un è un altro discorso secondo lui non è che può accusare così impunemente impunemente un giocatore o la o i dirigenti o la presidenza e non ricevere quello che gli che gli spetta *io gli ho detto* ma guardi lei intanto cominci a dare gli esempi coi fatti non mica a parlare venir qua cominci a dimostrarlo sul campo e non andare al cinema.

oppure sono combinazioni che indicano la posizione del parlante, come per esempio *io mi rendo conto, io non so se*, riscontrabili nel seguente esempio:

#### Dialogo 9 – Conversazioni formali (Tipologia C)

A: [...] capito puoi esprimere eh problemi di comprensione perché *io mi rendo conto* me ne rendo conto bene perché mentre in un libro come il loro non me ne rendo conto perché quello io



lo capisco bene quello che hai letto tu non lo capisco bene nemmeno io e quindi sono automaticamente più eh capisco [...]

La preferenza del parlato si rispecchia anche nell'uso dei pronomi dimostrativi – *questo/quello* rimpiazzano il pronome neutro *ciò* (Berrutto, 1993). Lo stesso tratto viene riscontrato anche nella struttura dei *clusters*. È stata rilevata una quantità vasta di espressioni che incorporano il dimostrativo *questo* o *quello*, come per esempio *quello che ho detto, quella che è stata, quelli che vogliono andare*.

Dal punto di vista sintattico c'è tra gli studiosi ampio accordo sulla predilezione per la coordinazione nel registro parlato. Secondo Berrutto (1993), le congiunzioni di cui si avvale il parlato rappresentano una gamma minore rispetto allo scritto. A differenza dello scritto, le congiunzioni che compaiono nel parlato hanno una frequenza d'occorrenza più alta. Tra le congiunzioni coordinanti predominano *e, dopo, poi, allora, ma, però* (Berrutto, 1993) e le stesse congiunzioni sono state identificate nella struttura dei *lexical bundles* emersi dall'analisi. A titolo d'esempio possiamo elencare *poi ci sono anche, ma questo è un, allora vediamo un po', però a questo punto*.

Il parlato predilige tra le congiunzioni causali *perché* e *siccome*, tra le temporali *quando* e *mentre*, mentre *se* e *come* sono connettivi decisamente più riscontrati (Berrutto, 1993). Nella struttura delle espressioni di 4 parole sono stati riscontrati *bundles* che contenevano le congiunzioni più adoperate nella testualità del parlato, come per esempio *perché se non è, quando si parla di, come si è detto, se non c'è un*.

È utile notare che le preposizioni che maggiormente ricorrono nei frammenti preposizionali sono *di* e *a*, mentre nei frammenti nominali la preposizione *di*, vale a dire le preposizioni semplici più frequenti in italiano (Voghera, 2004).

Il carattere della sintassi del parlato si riflette anche nella struttura dei *bundles*, mostrando che le combinazioni di più parole dovrebbero essere prese in considerazione dall'analisi linguistica.

Possiamo concludere che anche se i *clusters*, nella maggioranza dei casi, non sono delle strutture grammaticalmente complete, la loro forma consente di classificarli in categorie che coincidono con le principali categorie morfosintattiche che vengono predilette nella testualità del parlato. La maggioranza dei *bundles* emersi dall'analisi ha una struttura verbale; a questi seguono quelli con la struttura preposizionale, nominale e infine "altre espressioni" e in un numero abbastanza esiguo i *clusters* congiunzionali.

Essendo il verbo il nucleo primario del parlato, il nostro risultato coincide in maniera puntuale con le ricerche precedenti.

Un diverso sfruttamento delle categorie morfosintattiche si rispecchia anche nell'uso dei nomi e preposizioni. Come sostenuto da De Mauro (1993), le preposizioni e i nomi sono ben più rari che nell'uso scritto. I risultati della nostra analisi coincidono con questa osservazione, mostrano però una preferenza dei parlanti sia verso i *bundles* preposizionali che nominali. La maggioranza delle combinazioni di quattro parole preposizionali rappresenta un'espressione che funge da indicatore di una posizione personale rispetto all'informazione indicata in seguito – una caratteristica già riscontrata nella lingua inglese. Il fattore dell'egocentrismo (Berrutto, 1985) spiega al meglio l'alto uso delle espressioni preposizionali incentrate sul parlante.

Senza addentrarci qui nelle funzioni discorsive dei *lexical bundles*, questa sezione ha voluto mostrare che le sequenze di 4 parole, pur non possedendo una struttura grammaticale identificabile dalla linguistica tradizionale, possono essere classificate strutturalmente.

Una buona parte di queste, infatti, coincide con caratteri tipici del parlato, mostrando i modi in cui i parlanti si esprimono e la loro preferenza strutturale dei *lexical bundles*.

La stragrande maggioranza dei *lexical bundles* incorpora il pronome personale *io* o indica una posizione del parlante, centrando il discorso verso il parlante. La struttura dei *lexical bundles* è connotata dall'organizzazione discorsiva del parlato, a sua volta caratterizzata dall'egocentrismo linguistico, dalla non pianificazione del discorso, dalla percettività e dalla semplificazione, intesa anche come vera e propria semplificazione linguistica (Berrutto, 1985) e inoltre dalla dialogicità (Voghera, 1992).

La ricerca ha rilevato una quantità assai vasta di espressioni di 4 parole nel *corpus* LIP, le quali non erano state identificate dalla ricerca primaria sul *corpus* (De Mauro ed altri, 1993). Su un campione di 403 *lexical bundles* diversi, che erano emersi dall'analisi computazionale, solo 2 sono stati inclusi nella lista delle polirematiche nel *corpus* LIP. L'elenco al LIP include 1933 polirematiche presenti nel *corpus*, la cui frequenza d'occorrenza varia da 1004 a 1 volta sull'intero *corpus*. Le polirematiche presenti nel LIP sono espressioni dalla lunghezza da 2 a 6 segmenti in sequenza. Il dato supporta l'osservazione di Lo Cascio (1994), il quale ha sottolineato che solo un'attenta analisi ci

porterebbe a scoprire una notevole quantità di strutture polirematiche non ancora prese in considerazione nel *corpus* LIP. Questo dato è stato di fatto riscontrato nella nostra ricerca.

Come sottolineato da Biber e il suo gruppo di ricerca (2006), in larga parte i linguisti non hanno notato queste combinazioni ricorrenti di parole e, secondo gli autori, questo è stato causato dal fatto che la maggior parte della ricerca linguistica si è concentrata sulle clausole e le frasi grammaticali strutturalmente complete, trascurando la possibile esistenza di unità lessicali non complete che attraversano trasversalmente le strutture grammaticali.

Nella sezione seguente ci proponiamo in primo luogo di aprire la strada della verifica sperimentale sul piano dell'analisi strutturale dei *lexical bundles* nel discorso accademico, in tre ambiti scientifici diversi – storia, filosofia e medicina del lavoro.

Siamo partiti dall'analisi strutturale dei *lexical bundles* di ciascun ambito scientifico, procedendo in un secondo momento con l'analisi complessiva di tutti e tre gli ambiti, al fine di poter comparare la loro struttura con la struttura dei *bundles* identificati nel *corpus* LIP e di poter trarre delle conclusioni più ampie.

### **III.3 *Lexical bundles* nel registro scritto – nelle microlingue scientifico-professionali**

Il presente lavoro, basato su uno spoglio di 3.434.405 parole o *tokens* del linguaggio scientifico, fornisce una prima raccolta di *lexical bundles*, con la quale ci si propone di offrire un utile strumento di lavoro sia per lo studio del lessico microlinguistico che per l'insegnamento delle microlingue in questione. Per meglio chiarire la natura dell'uso dei *lexical bundles*, si è ritenuto necessario lavorare su un *corpus* che fosse il più possibile rappresentativo, considerando il numero di *tokens* imposto dai ricercatori inglesi (Cortes, 2004, 2006; Hyland, 2008a, 2008b) per lo studio dei *lexical bundles* nel linguaggio scientifico – il numero minimo di 1.000.000 di parole.

I dati per l'analisi dei *lexical bundles* sono costituiti da tre *corpora* elettronici di testi scritti, ovvero degli articoli scientifici di tre ambiti disciplinari diversi così composti: a) il *corpus* di storia si basa su 1.605.974 *tokens*; b) il *corpus* di filosofia

consiste di 780.989 *tokens*; c) quello di medicina del lavoro è composto da 1.047.442 *tokens*.

Abbiamo estratto le sequenze tipiche da un *corpus* di 3.434.405 *tokens*, assumendo che i dati siano rappresentativi della tendenza d'uso che gli esperti in materia fanno circa le combinazioni di parole precostituite.

Proponiamo di seguito la struttura grammaticale dei *bundles* emersi dall'estrazione automatica dei *corpora* in questione.

### III.3.1 La struttura dei *lexical bundles* nel discorso accademico

Una prima fondamentale distinzione da tenere presente nell'analisi dei *lexical bundles* è quella tra il numero della frequenza d'uso dei *lexical bundles* che compongono un *corpus*, e il loro contesto d'uso.

Il criterio legato alla quantità di *bundles* e alla loro effettiva frequenza d'uso ci permette di evidenziare all'interno del *corpus* alcuni elementi caratteristici di un ambito scientifico, mentre solo l'analisi relativa al contesto di occorrenza di una forma ci indica la sua funzione discorsiva.

Dopo l'estrazione automatica dei *bundles*, si è proceduto col conteggio dei *bundles* diversi e della loro effettiva frequenza d'uso. I *clusters* sono stati ordinati in senso decrescente in base alla loro frequenza d'uso (vedi Appendice 2, 3, 4, 5). In un secondo momento si è passato al raggruppamento strutturale dei segmenti ripetuti, sulla base dei primi elementi del *cluster*.

I *lexical bundles* sono stati infine comparati attraverso le discipline scientifiche. Il confronto è stato eseguito paragonando il diverso numero dei *clusters*, la loro frequenza d'uso e la loro struttura.

Nelle sezioni seguenti, i tre ambiti scientifici vengono esaminati in un primo luogo separatamente, per poi passare all'analisi sommativa delle microlingue in questione. Nelle tre sezioni seguenti proponiamo l'analisi strutturale delle tre discipline analizzate.

### III.3.1.1 Il corpus di storia

Il *corpus* di storia analizzato consiste di 1.605.974 *tokens* e di 64.019 *types* o parole distinte. Per il *corpus* sono state considerate quattro riviste scientifiche liberamente consultabili in formato elettronico, ovvero *Altreitalie*, *Reti medievali*, *Storicamente* e *Storia delle Donne*. Ci si è limitati alle annate più recenti – dal 1993 al 2011 – per consentire uno sguardo attuale dell'uso delle sequenze precostituite. Nell'analisi sono stati inclusi 291 articoli scientifici e 160 autori diversi, per evitare l'aspetto idiosincratico di singoli autori.

Tabella 11: Dati statistici del *corpus* di storia

<b>Dati statistici del <i>corpus</i> di storia</b>	
Tokens	1.605.974
Types	64.019
Numero articoli scientifici	291
<i>Lexical bundles</i> diversi – occorrenza minima 13 volte	399
Frequenza dei <i>bundles</i>	9.182
% delle parole in <i>bundles</i>	2,3%

Come si evince dalla tabella 11, nel *corpus* di storia appaiono 399 tipi di *lexical bundles* di quattro parole in sequenza con un'occorrenza superiore o uguale a 13 su 1.605.974 parole, in almeno 5 articoli scientifici diversi. Il numero di frequenza dei *bundles* e il numero di occorrenza in articoli scientifici diversi ricalcano le scelte dei ricercatori inglesi (Cortes, 2004, 2006; Hyland, 2008a, 2008b).

Come punto limite è stata scelta l'occorrenza in almeno 5 articoli scientifici diversi, pur se questa scelta produce pochi effetti sui risultati. Nell'appendice 2 possiamo osservare che i *lexical bundles* di quattro parole che coprono i primi 200 posti appaiono in media in 23 articoli scientifici differenti; il loro uso varia da 15 fino a 78 articoli scientifici.

Come già più volte indicato, il seguente studio si concentra sulle espressioni di 4 parole in sequenza. Seguendo gli studi precedenti si è voluto concentrare sui *clusters* di 4 parole, poiché questi *bundles* contengono nella loro struttura sequenze di 3 parole e,

nella forma più estesa, rappresentano combinazioni di 5 segmenti – come già riscontrato nella lingua inglese (Cortes, 2004, 2006).

L'appendice 6 illustra le prime 40 posizioni dei *bundles* di 3, 4 e 5 parole in sequenza. Come possiamo osservare, molti *bundles* di 3 parole sono incorporati nelle espressioni di 4 parole, mentre quelli di 5 parole in sequenza sono delle estensioni dei *lexical bundles* di 4 parole. A titolo d'esempio, riportiamo il *bundle* di 4 parole *dal punto di vista*, che appare nel *corpus* 203 volte e la cui struttura si può riscontrare nel *cluster* di 3 parole *punto di vista*, con la frequenza d'occorrenza 350, e *dal punto di*, con l'occorrenza di 204 volte, rappresentando dall'altro lato un'espressione più estesa nel *bundle da questo punto di vista* e *da un punto di vista*, con ricorrenza 60 e 39 rispettivamente.

Inoltre, va menzionato che i primi 10 *bundles* di 4 parole appaiono in media 42 volte su un milione di parole e in 35 articoli scientifici diversi. La tabella sotto riportata ci indica come le sequenze di più parole siano onnipresenti anche nell'ambito scientifico di storia.

Tabella 12: Il numero dei *lexical bundles* diversi e la loro frequenza d'occorrenza nel *corpus* di storia

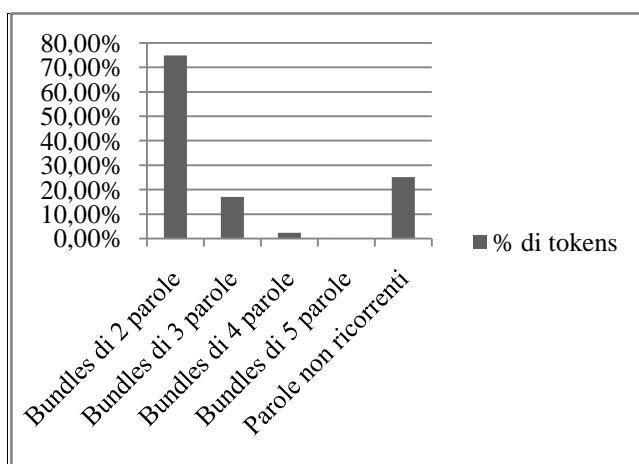
<i>Lexical bundles</i>	Diversi	Occorrenze
<b>2 parole</b>	13.994	601.539
<b>3 parole</b>	3.439	91.038
<b>4 parole</b>	399	9.182
<b>5 parole</b>	42	839

I dati appena esposti riguardano il fenomeno già riscontrato nella lingua parlata circa la pervasività del linguaggio precostituito. Sommando le espressioni di 3 e 4 segmenti in successione, ci sono quasi 4.000 differenti *bundles* su circa 1.600.000 *tokens*, che rappresentano una percentuale del 19,3% nei termini della proporzione di tutte le parole presenti nel *corpus*.

I *bundles* di 3 parole occorrono circa 60.000 su un milione di parole, mentre quelli di 4 segmenti circa 6.000 su un milione di parole. Il *bundle* di 4 parole più frequente – *dal punto di vista* – presenta una ricorrenza di 126 su un milione di parole o 203 occorrenze su 1.605.974.

La sequenza di 3 parole più frequente – *negli Stati Uniti* – mostra una frequenza di occorrenza di 362 su 1.605.974 *tokens*, ovvero 225 su un milione di parole. La pervasività e l’alta frequenza di occorrenza sono i dati indicati già da ricercatori sulla lingua inglese. A questo riguardo, i risultati del nostro spoglio presentano una conferma dell’ipotesi sull’estensione del linguaggio prefabbricato.

Grafico 2: Percentuale delle parole in *lexical bundles*



I dati nel grafico 2 indicano che le formule precostituite sono brevi, mostrando la possibilità di articolare il *corpus* per fasce della formulaicità: possiamo osservare, quindi, l’estensione del linguaggio formulaico, ossia con quante espressioni prefabbricate è composto il *corpus* di storia.

I dati indicati coincidono in maniera quantitativa con i dati svolti sulla lingua inglese. Le ricerche per la lingua inglese hanno mostrato che i *lexical bundles* – di 3 e 4 parole in sequenza – nella prosa accademica coprono una percentuale del 20% di tutte le parole presenti nel *corpus*, con i *bundles* di 3 parole che occorrono più di 60.000 volte su un milione di parole, e i *clusters* di 4 parole con un’occorrenza di 5.000 su un milione di parole.

Solo il conteggio delle effettive occorrenze dei *bundles*, ci consente di dare uno sguardo all’uso effettivo delle sequenze precostituite. I dati emersi risultano piuttosto congruenti con le ricerche svolte per la lingua inglese. Nella lingua italiana, la percentuale si aggira sul 75%, mentre nella lingua inglese sull’80%.

A questo punto possiamo proseguire con l’analisi strutturale dei *lexical bundles* identificati nel *corpus* di storia.

### III.3.1.1.1 La struttura dei *lexical bundles* nel *corpus* di storia

Anche se nella maggioranza dei casi non presentano delle strutture grammaticalmente complete, i *lexical bundles* possono essere raggruppati in base ai primi segmenti del *bundle* (Biber ed altri, 1999). La numerosità delle sequenze diverse ha suggerito una classificazione in quattro categorie principali: a) *lexical bundles* che sono clausole preposizionali o incorporano segmenti di sintagmi preposizionali; b) *lexical bundles* che sono sintagmi nominali completi o incorporano segmenti di clausole nominali; c) *clusters* verbali frammentati o completi; d) *clusters* congiunzionali.

Come per i *lexical bundles* identificati nel *corpus* LIP, anche nella classificazione dei *clusters* settoriali si è cercato di raggrupparli in tanti sottogruppi quante forme diverse presentavano, per poter al meglio definire le caratteristiche strutturali dell'ambito scientifico. In alcuni casi si hanno dei gruppi che contengono solo alcune espressioni, o una singola espressione.

Nel primo gruppo sono stati classificati i *bundles* che incorporano sintagmi preposizionali. Sono state identificate sei sottocategorie principali. Il primo sottogruppo rappresenta i *bundles* che cominciano con una preposizione – il gruppo maggiormente rappresentato è quello che comincia con la preposizione DI (come per esempio *della seconda guerra mondiale, degli italiani all'estero*); segue la preposizione IN (*nel mercato di lavoro, nel mondo del lavoro*), DA (per esempio *dal punto di vista, da parte degli immigrati*) e A (*all'interno della comunità, all'interno dell'unione*); e infine le preposizioni PER (*per la prima volta, per la maggior parte*), SU (*sul mercato del lavoro, su un totale di*), TRA (*tra i due paesi, tra le due guerre*), CON (*con l'intento di, con il compito di*), VERSO (*verso gli Stati Uniti, verso la fine degli*), DOPO (*dopo la caduta del, dopo la seconda guerra*) e DURANTE (*durante il regno di, durante la seconda guerra*).

Nel secondo sottogruppo sono stati identificati i frammenti preposizionali che contengono *che* sia all'inizio o alla fine, come per esempio *che a sua volta, di origine italiana che*, mentre nel terzo le espressioni che cominciano con un coordinatore, come nel caso di *e allo stesso tempo, ben al di là*. L'ultimo gruppo contiene frammenti di frasi relative (*in cui si sono, nel caso in cui*).



In base alla numerosità, i sintagmi nominali vengono classificati come seconda categoria. Sono state riscontrate sei classi principali. La classificazione inizia con i sintagmi nominali che incorporano la preposizione DI – *la storia dell'emigrazione, i figli di immigrati*; seguono i frammenti delle clausole nominali contenenti la preposizione A (*italiani residenti all'estero, un ritorno alle radici*), la preposizione IN (*immigrazione italiana in Germania, italiani negli Stati Uniti*), la preposizione CON (*i legami con la*) e la preposizione PER (*una volta per tutte*).

Dall'analisi sono stati inoltre riscontrati sintagmi nominali composti da un nome e un aggettivo oppure da un nome e da una parte di un secondo sintagma nominale (*la seconda guerra mondiale, prima guerra mondiale e*), i quali sono stati classificati come sintagmi nominali complessi, e infine quelli che includevano un complemento frammentato (*il fatto che le*), o un coordinatore (*e i loro discendenti*).

Nel terzo gruppo sono stati classificati i *bundles* verbali. Sono state identificate sette sottocategorie, ovvero: *clusters* che incorporano il verbo *essere* come copula (*non è un caso di, sono in grado di*); quelli con un verbo impersonale (*si tratta di un, si può dire che*); le espressioni che iniziano con un infinito e vengono classificate come *bundles* complessi (*fare i conti con*), quelli che cominciano con una preposizione (*a che fare con; a partire dagli anni*) o con una congiunzione (*se è vero che*) e infine le sequenze che contengono frammenti di frasi passive, come per esempio *come si è visto, come si è detto*, o relative – *di cui si è; in cui si sono*.

Come l'ultimo gruppo sono stati classificati i *clusters* che sono congiunzioni – *nella misura in cui, in modo tale che*.

Tra parentesi viene indicata la rispettiva occorrenza d'uso di ciascun *lexical bundle*.

#### III.3.1.1.1.1 *Lexical bundles che sono clausole preposizionali o frammenti di sintagmi preposizionali*

a. *Lexical bundles* preposizionali completi o frammentati

- la preposizione DI:

della seconda guerra mondiale (86); degli italiani all'estero (67); del mercato del lavoro (62); della prima guerra mondiale (27); degli italiani a Berlino (24); degli italiani in America (23); degli italiani in Germania (30); del muro di Berlino (20); del paese d'origine (20); degli italiani nel mondo (19); degli italiani in Francia (17); del diritto di voto (15); del paese di accoglienza (14); dell'immigrazione di massa (14); delle condizioni di vita (14); di Anna Maria Martellone (14); dei diritti dell'uomo (13); della seconda generazione di (13); di prigionieri di guerra (13); della prima metà del (13);

di fronte a un (45); di una serie di (44); di un gruppo di (27); degli anni settanta e (23); degli immigrati italiani in (23); del lavoro e della (27); dell'immigrazione italiana in (32); della comunità italiana in (33); della comunità italiana di (22); di lavoro e di (20); di origine e di (20); di vita e di (18); di fronte a una (17); di lì a poco (17); di un sistema di (17); di un vero e (17); degli immigrati italiani nel (15); di un processo di (15); di una vera e (14); di un secolo fa (13); della seconda metà del (20); degli anni cinquanta e (18); di volta in volta (44); prima della seconda guerra (14);

- la preposizione IN:

nel mondo del lavoro (25); nel paese d'origine (26); in tutto il mondo (29); nel mercato del lavoro (36); in via di sviluppo (21); nei confronti degli italiani (18); nell'ambito della ricerca (15); in cerca di lavoro (14); in un secondo tempo (13); nel paese di arrivo (14); nel settore della ristorazione (17); nel corso del novecento (14); nel corso dell'Ottocento (18); in entrambi i casi (35);

nella maggior parte dei (62); negli Stati Uniti e (37); in una sorta di (23); nella prima metà del (26); in questo senso la (26); nella seconda metà del (35); nel corso del tempo (31); nei primi decenni del (30); in questo caso è (19); in questo caso il (20); in questo caso l' (20); in questo caso la (18); in questo modo si (19); nel tempo e nello (18); in un momento in (17); in base al quale (16); in tutta la sua (15); in primo luogo la (14); in una serie di (14); in Italia e in (13); in presenza di un (13); in questo caso di (13); in tal senso la (13); negli Stati Uniti in (13); negli Stati Uniti si (13); negli anni cinquanta e (18); negli anni sessanta e (16); nel corso degli ultimi (16); nel corso di una (16); nel corso degli anni (74); negli anni venti e (21); nella seconda metà degli (21); in un certo senso (27);

- la preposizione DA:

dal punto di vista (203); da parte degli immigrati (14); da parte del governo (14); da parte delle autorità (14); da parte delle istituzioni (14); da questo punto di (60); da un punto di (39); da parte di un (29); da parte di chi (15); dai primi anni del (15); dalla fine degli anni (31); da una parte e (16); da una serie di (17); da un lato la (14); da un lato e (30); dalla metà degli anni (18);

- la preposizione A:

all'interno di un (43); all'interno di una (34); all'interno della società (15); al paese d'origine (18); a causa della guerra (15); al momento della partenza (16); all'inizio del novecento (18); all'interno della città (13); all'interno della famiglia (21); alla fine del

secolo (19); alla fine dell'Ottocento (26); alla prima guerra mondiale (13); all'interno della comunità (38); all'interno dell'unione (23); al mercato del lavoro (18);

all'interno del quale (24); a causa di una (16); a fronte di una (16); all'interno di questo (16); a causa della loro (15); a differenza di quanto (15); a New York e (15); alla ricerca di una (15); a fronte di un (13); all'inizio degli anni (62); alla fine degli anni (62); alla meta degli anni (18); agli inizi degli anni (21); a tutti gli effetti (27);

al di la di (27); al di la delle (36); al di la della (26); al di la del (31); al di la dell' (18); al di la dei (22); al di fuori della (26); al di fuori del (22); al di fuori del (15); al di fuori dell' (14); al di fuori di (14); fino a quel momento (26);

- la preposizione PER:

per gli Stati uniti (14); per tutti gli anni (14); per la prima volta (102);

per il fatto che (17); per la maggior parte (51); per cento della popolazione (30); per il rinnovo del (22); per cento degli italiani (15); per la creazione di (13); per la realizzazione di (13);

- la preposizione SU:

sul mercato del lavoro (45); sulla base di una (16); su un totale di (14);

- la preposizione TRA:

tra i due paesi (24); tra le due guerre (26); tra gli anni cinquanta (15); tra la fine del (27); tra le altre cose (16); fra i due paesi (17); tra l'Italia e (15);

- la preposizione CON:

con l'intento di (33); con l'eccezione di (20); con l'intenzione di (18); con una serie di (18); con il compito di (15); con il paese di (13); con la terra di (13); con lo scopo di (13);

- la preposizione VERSO:

verso gli Stati uniti (17); verso la fine degli (20);

- la preposizione DOPO:

dopo la seconda guerra (28); dopo la caduta del (15); dopo la morte del (13); dopo la morte di (13);

- la preposizione DURANTE:

durante la seconda guerra (21); durante il regno di (14);

b. Frammenti di sintagma preposizionale con *che*

che a sua volta (24); che in alcuni casi (20); che in molti casi (13);  
di origine italiana che (15); dal momento che la (20);

c. Marcatore del discorso + frammento di sintagma preposizionale

anche in questo caso (67); anche dal punto di (13); e allo stesso tempo (22); e in modo particolare (16); e al tempo stesso (15); e per certi versi (17); e più in generale (33); ma allo stesso tempo (16); ben al di là (16); come nel caso della (15); come nel caso di (15);

d. Frammento di dipendenti relative

nel caso in cui (16); nel momento in cui (74); negli anni in cui (13);

e. Congiunzione + sintagma preposizionale

se da un lato (34).

*III.3.1.1.1.2 Lexical bundles che sono sintagmi nominali o incorporano frammenti di sintagmi nominali*

a. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione DI:

questo punto di vista (62); un punto di vista (51); maggior parte dei casi (56); maggior parte di loro (17); la storia dell'emigrazione (26); i figli di immigrati (24); storia dell'immigrazione italiana (23); la storia delle migrazioni (17); americani di origine italiana (13); commissariato generale dell'emigrazione (13); discendenti degli immigrati italiani (13); i figli degli immigrati (13); la terra d'origine (13); legame con l'Italia (13); giovani di origine italiana (43); il più delle volte (27); il fatto di essere (13); la caduta del muro (23); interno dell'unione europea (23); il rinnovo del parlamento (18); ministero degli affari esteri (35); il ministero del lavoro (16); il ministero degli esteri (17); gli uffici del lavoro (16); uffici provinciali del lavoro (16); i datori di lavoro (13); il permesso di soggiorno (13); il processo di integrazione (13); il senso di appartenenza (13); le condizioni di vita (14); il diritto di voto (21); diritti di cittadinanza politica (15); ruolo di primo piano (21); i cittadini dell'unione (19); il tasso di disoccupazione (15); forme di partecipazione politica (16); il paese d'origine (14), il paese di provenienza (14); Stati Uniti d'America (14); un titolo di studio (18); la qualità della vita (16);

fine degli anni sessanta (36); meta del XIX secolo (31); fine della seconda guerra (22); fine degli anni settanta (21); fine del XIX secolo (20); seconda meta del secolo (20); seconda meta dell'Ottocento (20); meta degli anni sessanta (17); fine degli anni cinquanta (16); meta degli anni ottanta (16); meta degli anni settanta (19); primi anni del novecento (18); primi decenni del novecento (18); inizio degli anni sessanta (17); fine degli anni venti (14); prima meta del novecento (14); anni venti del Novecento (13);

inizio degli anni novanta (13); inizio degli anni ottanta (13); inizio del XX secolo (13); meta del VI secolo (13); meta del XII secolo (13); la fine della guerra (14); la fine degli anni (36); seconda meta degli anni (34);

punto di vista dell' (17); punto di vista della (19); punto di vista la (15); la maggior parte degli (41); la maggior parte delle (39); la maggior parte dei (36); la maggior parte di (36); il maggior numero di (15); l'esistenza di un (23); l'esistenza di una (33); l'idea di una (23); l'idea di un (17); la necessità di una (15); la possibilità di un (14); il numero degli italiani (14); il risultato di una (14); un certo numero di (16); l'assenza di un (13); un gran numero di (14); uno degli aspetti più (14); caduta del muro di (20); la presenza di un (21); voto degli italiani all' (19); la creazione di una (20); la creazione di un (19); la costruzione di una (17); un ruolo di primo (19); paesi in via di (16); vita e di lavoro (13); mercato del lavoro e (32); le seconde generazioni di (13);

b. Sintagma nominale con frammento di una clausola preposizionale che comincia con una preposizione diversa da DI

- preposizione A

comunità italiane all'estero (16); italiani residenti all'estero (30); gli italiani all'estero (27); un ritorno alle radici (14); italiani all'estero e (20);

- preposizione IN

comunità italiana in Germania (47); immigrazione italiana in Germania (27); italiani negli Stati Uniti (31); immigrati italiani in Germania (18); italiano negli Stati Uniti (15); l'entrata in guerra (14); dati in nostro possesso (16); la presenza italiana in (16); la comunità italiana in (17); italiani e in particolare (15); tempo e nello spazio (13);

- preposizione CON

i conti con la (17); i legami con la (14);

- preposizione PER

-

una volta per tutte (13);

c. Sintagma nominale con complemento frammentato

il fatto che le (19); il fatto che la (18); il fatto che l' (13);

d. Sintagma nominale complesso

la seconda guerra mondiale (76); anni sessanta e settanta (30); la repubblica federale tedesca (19); seconda e terza generazione (17), la prima guerra mondiale (18); prima e seconda generazione (14); voto attivo e passivo (15); un anno e mezzo (14); vecchie e nuove migrazioni (13); anni cinquanta e sessanta (35); anni venti e trenta (21);

gli anni cinquanta e (20); prima guerra mondiale e (14); gli italiani e i (18); un vero e proprio (66); una vera e propria (86); gli anni sessanta e (14);

tutta una serie di (28); italiana negli Stati Uniti (24); entrata in vigore delle (14);

e. Coordinatore + sintagma nominale frammentato

e i loro discendenti (14); e i processi di (15); e la capacità di (14); e l'inizio del (13); e le condizioni di (13); come una sorta di (13);

f. Nome + proposizione relativa frammentata

il modo in cui (23); momento in cui la (28); un momento in cui (20); momento in cui si (15); il motivo per cui (17); il luogo in cui (14); anni in cui la (14); un periodo in cui (13).

### *III.3.1.1.1.3 Lexical bundles che incorporano frammenti di frasi verbali*

a. Essere + frase nominale o preposizionale frammentata

non è un caso (23); è il fatto che (20); è un caso che (20); è in grado di (25); sono in grado di (16); è il caso di (26); è possibile affermare che (14);

b. *Lexical bundles* con il verbo alla forma impersonale

si tratta di un (105); si tratta di una (54); si trattava di un (27); si trattava di una (23); si può dire che (25); vale a dire la (13);

c. *Bundles* complessi

partire dalla fine degli (25); partire dagli anni settanta (19); partire dagli anni Novanta (13); fare i conti con (25);

d. Preposizione + sintagma verbale frammentato

a che fare con (29); a partire dagli anni (88); a partire dalla fine (46); a partire dalla meta (22); a partire dai primi (14); a fare i conti (15); a giocare un ruolo (13); di origine italiana è (24); per quanto riguarda l' (52); per quanto riguarda la (40); per quanto riguarda le (22); per quanto riguarda i (17); per quanto riguarda il (36); per ciò che riguarda (20); per quel che riguarda (26); per quanto concerne la (18);

e. Congiunzione + frase verbale frammentata

se è vero che (21);

f. Frase passiva frammentata

come si è visto (37); come si è detto (28); si è deciso di (28);

g. Frase relativa frammentata

di cui si è (15); in cui si sono (15).

#### *III.3.1.1.1.4 Lexical bundles che sono congiunzioni*

nella misura in cui (18); in modo tale che (13); in modo tale da (13).

A questo punto possiamo procedere con i risultati emersi dall'analisi strutturale.

#### *III.3.1.1.2 I risultati emersi dall'analisi strutturale*

Come si può vedere dalla tabella 13, i *lexical bundles* con struttura nominale o preposizionale sono più usati dei *lexical bundles* verbali. I dati emersi dall'analisi supportano le ipotesi tradizionali sulla struttura prevalentemente nominale del linguaggio scritto e delle microlingue in particolare.

Lo spoglio ha mostrato che i *bundles* preposizionali e nominali rappresentano la maggioranza dei *clusters* identificati, coprendo l'89,2% di tutte le sequenze di quattro parole identificate. I *lexical bundles* preposizionali individuati coprono il 51,4% di tutti i *clusters* emersi dall'estrazione automatica del *corpus*, mentre i *clusters* con la struttura nominale il 37,8%. Le sequenze di più parole con una struttura verbale costituiscono il 9,8% di tutti i *bundles*, quelle congiunzionali solo l'1% di tutti i *lexical bundles* identificati.

I risultati coincidono con le ricerche svolte da Cortes (2004) sulla lingua inglese nell'ambito scientifico della storia. La studiosa riporta che i *bundles* identificati appartengono a due categorie strutturali: nominali e preposizionali.

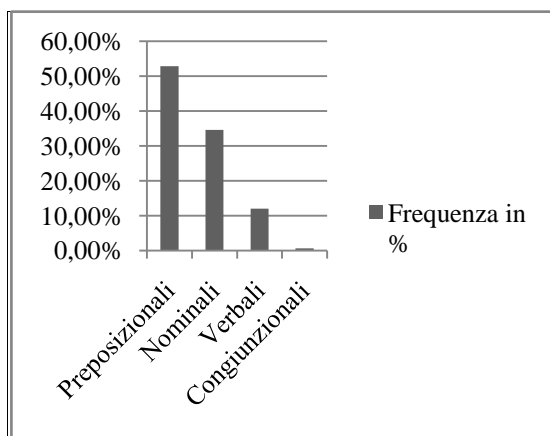
Tabella 13: Distribuzione strutturale dei *lexical bundles* di 4 parole in sequenza nel *corpus* di storia

<i>Lexical bundles</i> di 4 parole	
<b>Preposizionali</b>	204
<b>Nominali</b>	153
<b>Verbali</b>	39
<b>Congiunzionali</b>	3
<b>Totale</b>	<b>399</b>

Come già più volte accennato, nell'analisi dell'uso delle strutture formulaiche bisogna distinguere tra la numerosità dei *bundles* e la loro effettiva occorrenza o frequenza d'uso. Come si vede nel Grafico 3, nella microlingua della storia l'uso dei *bundles* diversi e la loro frequenza di occorrenza mostrano una tendenza uguale: i *bundles* nominali e preposizionali sono infatti usati più frequentemente dei *clusters* verbali.

Le sequenze nominali e preposizionali coprono rispettivamente il 34,6% e il 52,8% di tutte le frequenze nel *corpus*, mentre la frequenza dei *clusters* verbali costituisce il 12% di tutte le frequenze d'uso.

Grafico 3: Frequenza d'uso dei *lexical bundles* di 4 parole in sequenza nel *corpus* di storia



Il fattore che meglio spiega l'uso massiccio dei *lexical bundles* preposizionali e nominali è da un lato la particolare frequenza del processo di nominalizzazione diffuso nelle microlingue scientifico-professionali, dall'altro invece la spersonalizzazione o la necessità di minimizzare l'intervento individuale dello scienziato, concentrando la comunicazione sul messaggio del testo o sul risultato di una ricerca (Balboni, 2000).



La nominalizzazione consiste nella trasformazione di un sintagma verbale in un sintagma nominale. L'attenzione viene spostata dall'autore dell'azione al messaggio stesso, al fatto o all'azione, mettendo in ombra il complemento d'agente. La nominalizzazione provoca la riduzione dell'uso del verbo, evitando in questo modo l'uso del pronome, il quale a volta causa ambiguità e un'estrema sinteticità, ovvero la tendenza ad esprimere le nozioni nel modo più breve possibile (Gotti, 1991). Il processo di nominalizzazione inserisce enunciati chiari e precisi, senza dover ricorrere a un complesso meccanismo di concordanza fra i tempi verbali (Balboni, 2000). D'accordo con Balboni (2000), sosteniamo che ciò avviene in nome della non-ambiguità, della chiarezza e della semplicità espressiva.

Ciò provoca al livello morfosintattico una perdita di importanza del verbo e una riduzione delle forme verbali, delle persone e dei modi adoperati (Sobrero, 1993). Viene utilizzata la forma più elementare del verbo – la copula – e i verbi maggiormente impiegati sono difatti i verbi copulativi (Sobrero, 1993). Difatti, il verbo maggiormente rappresentato nel *corpus* di storia è il verbo *essere*. A mo' di esempio possiamo indicare i seguenti *bundles*: *non è un caso, è il fatto che, è in grado di, è possibile affermare*.

Un altro elemento distintivo delle microlingue è il ricorso alla forma passiva e alla forma impersonale, che si presentano molto più frequentemente che nel linguaggio comune. Il passivo, quindi, mette in luce l'informazione più importante di un enunciato, riducendo al minimo l'importanza del complemento d'agente o dell'autore. La forma passiva viene "[...] comunemente utilizzata per tematizzare il processo, il fatto o l'azione di cui si è trattato nella frase precedente, garantendo quindi una più funzionale e naturale progressione del flusso delle informazioni" (Gotti, 1991: 97). Anche questo tratto è riscontrabile nella struttura dei *lexical bundles*: per la forma passiva possiamo indicare *si è deciso di, come si è visto*; mentre la forma impersonale è osservabile in *bundles si tratta di un, si tratta di una, si trattava di un, si trattava di una*.

È interessante notare che tra i 39 *lexical bundles* verbali, 2 si trovano tra i top 10: al secondo posto troviamo una forma impersonale – *si tratta di un*, con una frequenza di 105 volte su 1.605.974 parole o 65 volte su un milione di parole e in 54 articoli scientifici differenti; al quarto posto è collocato il *bundle a partire dagli anni*, con una frequenza di 88 volte su 1.605.974 *tokens* o 54 volte su un milione di parole.

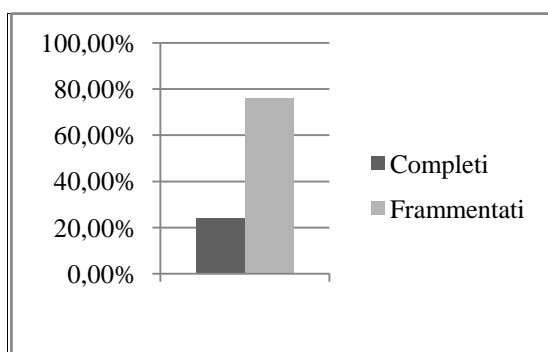
Come già abbiamo segnalato, la maggioranza dei *lexical bundles* può essere guardata come composta da frasi frammentate – cosa riscontrata anche da Biber (1999), nel cui studio meno del 5% dei *bundles* era composto da unità strutturali complete.

La frammentarietà e completezza dei *bundles* identificati nel *corpus* di storia è visibile nella tabella 14. Come si evince dal grafico 4, il 76,2% di tutte le sequenze di 4 parole ha una forma frammentata (come per esempio *a causa di un, all'interno di un, interno dell'unione europea*). In questo dato, la distanza tra i risultati sulla lingua inglese e quelli sul *corpus* specialistico in lingua italiana si fa più accentuata. Le clausole complete in lingua italiana – come ad esempio *il ministero del lavoro, il diritto di voto, il paese d'origine* – coprono il 23,8% di tutti i *bundles*, mettendo in luce un nucleo di parole abbastanza diverso come riscontrato nella lingua inglese. È interessante notare che tra i primi dieci *bundles* ce ne sono 2 che presentano una forma completa – *per la prima volta*, con l'occorrenza di 102 volte su 1.605.974 parole, e *la seconda guerra mondiale*, con l'occorrenza 76 su 1.605.974 parole.

Tabella 14: Distribuzione strutturale dei *lexical bundles* di 4 parole – completi e frammentati – nel *corpus* di storia

<b>Lexical bundles di 4 parole</b>	<b>Complet</b>	<b>Frammentati</b>
<b>Preposizionali</b>	62	142
<b>Nominali</b>	33	120
<b>Verbali</b>	0	39
<b>Congiunzionali</b>	0	3
<b>Totale</b>	<b>95</b>	<b>304</b>

Grafico 4: Percentuale delle forme complete e frammentate dei *lexical bundles* nel *corpus* di storia



Procediamo a questo punto con l'analisi del *corpus* di filosofia.

### III.3.1.2 Il *corpus* di filosofia

Come si evince dalla tabella 15, il *corpus* di filosofia si compone di 780.989 *tokens* o parole e di 41.727 *types* o parole distinte. L'analisi si basa su tre riviste scientifiche liberamente consultabili online, ovvero *Noema*, *Philosophical News* e *Segni e comprensione*. Anche in questa parte della ricerca si è preferito optare per le annate più recenti, per poter offrire un panorama attuale del linguaggio prefabbricato. Sono state pertanto scelte riviste dall'anno 2007 all'anno 2012. Lo spoglio include 100 articoli scientifici e 100 autori differenti – per minimizzare gli effetti idiosincratici relativi ai singoli scrittori.

Tabella 15: Dati statistici del *corpus* di filosofia

<b>Dati statistici del <i>corpus</i> di filosofia</b>	
Tokens	780.989
Types	41.727
Numero articoli scientifici	100
<i>Lexical bundles</i> diversi – occorrenza minima 7 volte	328
Frequenza dei <i>bundles</i>	4.089
% delle parole in <i>bundles</i>	2,1%

Si noterà che il *corpus* di filosofia consiste di 328 *lexical bundles* di quattro parole con un'occorrenza minima di 7 su 780.989 *tokens*, ovvero 10 volte su un milione di parole. Per il calcolo della frequenza di occorrenza, come già indicato in precedenza, si è seguita la regola della normalizzazione. Oltre a questo criterio, lo spoglio include solo i *lexical bundles* presenti in almeno quattro articoli scientifici diversi. Osservando le prime 100 posizioni (vedi Appendice 3) troviamo espressioni che in media appaiono in 13 articoli scientifici differenti. Questo fatto è riscontrabile anche nella lingua inglese e presenta una caratteristica attesa dei *lexical bundles*.

I *lexical bundles* di quattro parole sono il nucleo primario della presente analisi. Lo spoglio dei *clusters* del *corpus* di filosofia ha confermato le osservazioni emerse dall'analisi della lingua inglese e, inoltre, dall'analisi dell'ambito scientifico di storia.

Come riportato nell'appendice 7, i *bundles* di 4 parole incorporano nella struttura i *clusters* di tre parole e presentano delle estensioni delle sequenze di cinque parole. Riportiamo qui a titolo di esempio il *lexical bundle* di quattro segmenti in sequenza *dal punto di vista*, con una frequenza di occorrenza 206 su 780.989 parole o 26 volte su un milione di parole, e in 28 articoli scientifici diversi. I *bundles punto di vista* e *dal punto di*, con un'occorrenza 309 e 107 rispettivamente, sono incorporati nella struttura dei *clusters* di quattro parole; da contro, la sequenza estesa rappresenta un *bundle* di cinque parole – *da un punto di vista*.

Inoltre, è interessante notare che nella prima posizione del *corpus* di storia e filosofia troviamo lo stesso *lexical bundle*. Oltre a ciò, è importante notare che nelle prime dieci posizioni i *bundles* compaiono in media 62 su 780.989 di parole, ossia 79 volte su un milione di parole. In media, le espressioni di quattro parole nelle prime dieci posizioni sono presenti in 20 articoli scientifici differenti.

È interessante inoltre notare che nelle prime dieci posizioni di entrambi i *corpora* – il *corpus* di storia e di filosofia – troviamo quattro *lexical bundles* identici: *dal punto di vista, nel momento in cui, una vera e propria* e *si tratta di un*. La differenza riguarda il numero di occorrenze nel *corpus*.

Va oltre a ciò menzionato che tra i 63 *lexical bundles* verbali, 2 si trovano tra i top 10: al sesto posto troviamo frammento di una forma impersonale – *a che fare con*, con una frequenza di 51 volte su 789.989 parole o 65 volte su un milione di parole e in 23 articoli scientifici differenti; all'ottavo posto si colloca la forma impersonale *si tratta di un*, con una frequenza di 46 volte su 789.989 *tokens* o 58 volte su un milione di parole.

Possiamo constatare che i due *bundles* presentano nella struttura la forma impersonale, uno dei tratti maggiormente riscontrati nel linguaggio scientifico.

La tabella sotto riportata illustra la pervasività delle sequenze di più parole nell'ambito scientifico di filosofia.

Tabella 16: Il numero di *lexical bundles* differenti e la loro frequenza d'occorrenza nel *corpus* di filosofia

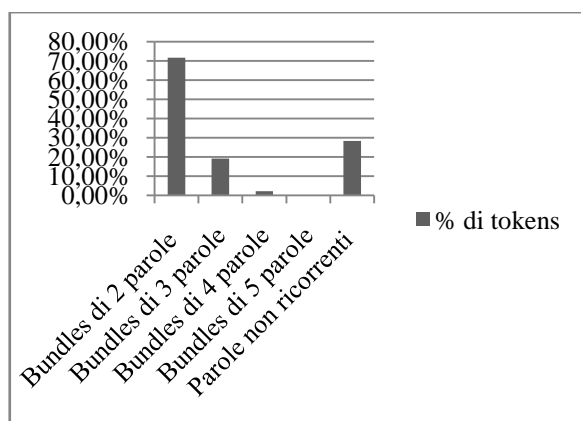
<i>Lexical bundles</i>	<i>Bundles diversi</i>	Occorrenze
<b>2 parole</b>	12.606	279.432
<b>3 parole</b>	3.488	50.065

<b>4 parole</b>	328	4.089
<b>5 parole</b>	26	310

Come vedremo nello svolgimento della presente ricerca, l'alta frequenza dei *lexical bundles* è un dato atteso, un fenomeno già verificatosi nella lingua inglese. I *lexical bundles* di tre e quattro parole figurano in una percentuale del 19,2% e del 2,1% di tutte le parole sottoposte allo spoglio. Sommando le sequenze di tre e quattro parole il *corpus* consiste di circa 4000 *bundles*. I *clusters* di tre parole hanno un'occorrenza di 64.000 volte su un milione di parole, quelli di quattro parole 5.200 volte su un milione di parole. I dati concordano sia con i risultati sulla lingua inglese (Conrad e Biber, 2004; Hyland, 2008a, 2008b), sia con i dati emersi dallo spoglio del *corpus* di storia.

Va inoltre menzionato che il *bundle* di quattro segmenti in sequenza più frequente nel *corpus* di filosofia – *dal punto di vista* – si presenta con una frequenza di 263 su un milione di parole; la sequenza più frequente di tre parole è di 395 volte su un milione di parole.

Grafico 5: Percentuale delle parole prefabbricate



Come già discusso in precedenza, il confronto relativo alle fasce di copertura dei *lexical bundles* mette in luce la loro pervasività.

Le espressioni di 3 segmenti in successione coprono una percentuale del 19,2% fra tutti i *tokens* analizzati, con le sequenze che compaiono più di 60.000 volte su un milione di parole, mentre le sequenze di 4 parole ricorrono più di 5.000 volte su un milione di parole, coprendo una fascia del 2,1% di tutti i *tokens*.

I dati circa la pervasività dei *lexical bundles* e la loro frequenza d'occorrenza sono dati attesi e, inoltre, già riscontrati nella lingua inglese.

Nella sezione seguente analizzeremo la struttura dei *lexical bundles* identificati nel *corpus* di filosofia.

### III.3.1.2.1 La struttura dei *lexical bundles* nel *corpus* di filosofia

Come già precedentemente accennato, i *lexical bundles* sono classificati rispettando il primo segmento dell'espressione. Nell'analisi del *corpus* di filosofia sono state rilevate cinque categorie principali: a) *clusters* preposizionali; b) *bundles* che incorporano sintagmi nominali; c) espressioni di più parole che contengono frammenti verbali; d) *bundles* congiunzionali e infine – e) *bundles* avverbiali.

Il criterio della classificazione decrescente dei gruppi è stato il criterio di quantità di *bundles* diversi emersi dall'analisi. Il primo gruppo include i *lexical bundles* preposizionali – in questa categoria sono state classificate le espressioni che cominciano con la preposizione DI (*dell'opera d'arte, della Critica di ragion*), la preposizione A (*a favore di un, al di là della*), la preposizione DA (*dal punto di vista, da un certo punto*), la preposizione IN (*in un certo senso, in entrambi i casi*), la preposizione PER (*per la prima volta*), SU (*sulla base di un*), TRA (*tra Ragione e fede*), CON (*con le parole di*) e infine la preposizione SOTTO (*sotto il nome di*). Nel secondo sottogruppo sono state classificate le espressioni che includono *che*, come per esempio *che in questo caso, che a sua volta*; quelle che iniziano con un marcatore del discorso – *e allo stesso tempo, ma non per questo*; frammenti di frasi relative (*al modo in cui, nel modo in cui*) e le sequenze che hanno una congiunzione come primo elemento (*se da un lato*).

La seconda categoria identificata è il gruppo dei sintagmi nominali. Sono state pertanto individuate espressioni che incorporano la preposizione DI (*un punto di vista, tradizione del pensiero occidentale*); PER (*una volta per tutte*) SU (*la domanda sulla verità*) e infine TRA (*la differenza tra la*). Il secondo sottogruppo rappresenta sintagmi nominali che includono un complemento frammentato, come per esempio *l'idea che il, il fatto che la*; il terzo include sintagmi nominali complessi (*l'uomo e il, la sua vita e*); nel quarto sottogruppo vengono elencate espressioni che iniziano con un coordinatore, come ad esempio *come una sorta di, e una sorta di*; e infine le espressioni che incorporano frammenti di frasi relative – *il modo in cui, una situazione in cui*.

La terza categoria include frammenti di sintagmi verbali e viene suddivisa in espressioni che hanno il verbo *essere* come copula (*è anche vero che*); che incorporano un verbo impersonale (*si tratta dunque di*); un verbo diverso da *essere* (*ci troviamo di fronte*); quelle che iniziano con un infinito (*fare i conti con*); una congiunzione (*se è per sé*); che sono frammenti di frasi passive (*come si è detto*); che iniziano con un pronome dimostrativo (*ciò che non è*); e infine con un pronome relativo (*cui non si può*).

Negli ultimi due gruppi possiamo riscontrare un numero abbastanza esiguo di espressioni – i *lexical bundles* che sono congiunzioni (*nella misura in cui*) o avverbi (*ma anche e soprattutto*).

Procediamo a questo punto con la classificazione strutturale dei *lexical bundles* di filosofia.

#### *III.3.1.2.1.1 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi preposizionali*

##### a. *Lexical bundles* preposizionali completi o frammentati

###### - la preposizione DI:

del mondo della vita (13), dell'opera d'arte (13); di tutte le cose (12); della critica della ragion (9); della fenomenologia dello spirito (8); di tutti gli uomini (7);

di una teoria della (12); di una vera e (12); di una sorta di (12); dell'uomo e della (9); di ciò che è (14); di tutto ciò che (12); del fatto che la (12); di un soggetto che (10); di per se non (9); di un processo di (9); della vita e della (8); dei bisogni e dei (7); della scienza e della (7); della verità e della (7); di una filosofia della (7); di un io che (7); di ciò che si (7); di volta in volta (39); di fronte ad un (19); di fronte a un (10); di fronte ad una (9);

###### - la preposizione A:

ad un certo punto (19); a un certo punto (16); al mondo delle cose (7); ad una sorta di (14); al tempo stesso la (10); a se stesso e (7); a sua volta un (7); a favore di un (8); a favore di una (9); al di fuori di (17); al di fuori del (21); al di fuori dell' (8); al di la della (21); al di la del (16); al di la di (16); al di la dell' (11); al di sopra di (15); alla stregua di un (11); all'interno del quale (9); all'interno di una (37); all'interno di un (33); all'interno di questa (8); alla fine degli anni (7); fino a che punto (9);

###### - la preposizione DA:

dal punto di vista (106); da un punto di (64); da questo punto di (25); da un lato la (15); da un certo punto (8); da un lato il (8); da una sorta di (8); dal momento che il (11); dal momento che la (9); da parte di un (7); da un lato l' (7); da una parte e (7); dal nostro punto di (7);

- la preposizione IN:

in un certo senso (28); in un certo modo (8); in entrambi i casi (12); in carne ed ossa (16); nella critica della ragione (11); in un primo tempo (8); in un primo momento (7); nel corso degli anni (7); in fin dei conti (7);

in questo senso la (21); in questo senso l' (18); in questo senso il (15); nel senso che la (13); nel senso che una (13); nel senso che non (8); nel senso che l' (7); non nel senso che (7); in questo senso si (12); in questo modo la (14); in questo modo si (8); in tal modo si (8); in tal modo l' (8); in tal modo la (10); in questo modo l' (9); in tutta la sua (17); in tutte le sue (15); in questo caso l' (13); in questo caso il (9); in questo caso la (8); in una situazione di (8); in vista di una (8); in una sorta di (9); nei termini di un (14); in essere e tempo (13); in sé e per (8); in quanto tale e (7); in questo caso si (7); in questo modo il (7); in questo senso l' (7); in un rapporto di (7); nella ricerca di un (7); nell'ambito di un (8); in altri termini il (7); in contrasto con la (7); in modo del tutto (7); non in grado di (8); in base alla quale (12); in base al quale (8); in relazione ad un (7); in virtù di un (7); nel senso di un (8); in base ad un (7);

- la preposizione PER:

per la prima volta (29); per una filosofia dell' (10);

- la preposizione SU:

sulla base di una (15); sulla base di un (12); su di un piano (8);

- la preposizione TRA:

-  
tra ragione e fede (7); tra l'uomo e (10); tra il soggetto e (8);

- la preposizione CON:

con le parole di (7);

- la preposizione SOTTO:

sotto il nome di (13);

b. Frammenti di sintagma preposizionale con che



che in questo caso (8); che a sua volta (20); che in qualche modo (9); per il fatto che (18);

c. Marcatore del discorso + frammento di clausola preposizionale

anche in questo caso (28);

e d'altra parte (9); e a sua volta (8); e per questo che (8); e in quanto tale (7), e in questo modo (7); e in tal modo (7); e in un certo (7); e proprio per questo (11); e in questo senso (11); e allo stesso tempo (13); e nello stesso tempo (16); e al tempo stesso (18); come nel caso dell' (7); ma non per questo (12); e in qualche modo (7); solo nella misura in (8); sia da un punto (7);

d. Frammenti di frasi relative

al modo in cui (8); nel modo in cui (15); in cui ci si (9); 7); nel momento in cui (79); in cui l'uomo (10); in cui il soggetto (8); per il quale il (10); per il quale la (7); nel senso in cui (8); in cui essa si (7); in cui la filosofia (7);

e. Congiunzione + frase preposizionale frammentata

se da un lato (20); se per un verso (8).

*III.3.1.2.1.2 Lexical bundles che sono sintagmi nominali o incorporano frammenti di sintagmi nominali*

a. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione DI:

un punto di vista (84); il punto di vista (20); questo punto di vista (30); punto di vista ontologico (8); il punto di partenza (25); un punto di partenza (7); il mondo delle cose (40); il mondo della vita (17); il fatto di essere (7); il dato di fatto (8); un dato di fatto (10); il concetto di verità (14); critica della ragione pura (24); pensiero di Merleau Ponty (9); la fenomenologia di Husserl (8); tradizione del pensiero occidentale (8); la storia della filosofia (7); la filosofia del linguaggio (13); la presa di coscienza (8); le cose del mondo (8); il modo di essere (11); il senso dell'essere (7); il senso di essere (9); principio di non contraddizione (11); la condizione di possibilità (8); le condizioni di possibilità (7); il problema del rapporto (10); una teoria della conoscenza (11);

suo punto di vista (11); punto di vista della (29); punto di vista del (17); punto di vista di (13); punto di vista dell' (12); punto di vista la (9); punto di vista il (7); punto di vista che (9); punto di partenza e (11); punto di partenza per (7); nostro punto di vista (8); punto di partenza di (9); l'idea di una (16); l'idea di un (15); mondo della vita e (11); il fatto che il (15); la possibilità di un (11); la possibilità di una (9); il risultato di un (11); il risultato di una (9); il rapporto con l' (7); problema del rapporto tra (8); tutte le forme

di (11); tutta una serie di (10); l'esistenza di un (11); l'esistenza di una (7); l'esigenza di una (9); la presenza di un (9); il principio di non (7);

b. Sintagma nominale con frammento di una clausola preposizionale che comincia con una preposizione diversa da DI

- la preposizione PER

una volta per tutte (20),

- la preposizione SU

la domanda sulla verità (7);

- la preposizione TRA

la differenza tra la (8);

c. Sintagma nominale frammentato con complemento frammentato

il fatto che la (21); il fatto che in (8); dato di fatto che (7); l'idea che il (7); l'idea che l' (10); l'idea che la (10);

- *che* + sintagma nominale frammentato

che il concetto di (9);

d. Sintagmi nominali complessi

l'io e il (8); l'uno e il (8); l'uomo e il (8); l'uomo e l' (8); l'uomo e la (8); la ragione e la (8); la sua vita e (7); una vera e propria (47); un vero e proprio (20); bisogni e dei desideri (7);

e. Marcatore del discorso + sintagma nominale frammentato

come una sorta di (11); come punto di partenza (7); ancora una volta la (7); e lo stesso che (7); e una sorta di (7); e il fatto che (9);

f. Sintagma nominale + proposizione relativa frammentata

momento in cui la (14); il modo in cui (24); una situazione in cui (7); momento in cui il (7); un mondo in cui (8); momento in cui si (16); modo in cui la (10); stesso modo in cui (10); modo in cui si (7); modo in cui si (7); modo in cui la (10); stesso modo in cui (10); momento in cui si (16); momento in cui il (7); una situazione in cui (7); un mondo in cui (8).

### III.3.1.2.1.3 Lexical bundles che incorporano frammenti di sintagmi verbali

#### a. Il verbo essere come copula

- essere + frase preposizionale frammentata

è in grado di (28); siamo in grado di (7); sia in grado di (7); essere in grado di (10);

- essere + frase nominale frammentata  
è il problema della (10);

- essere + aggettivo + clausola dipendente

è anche vero che (9);

#### b. *Lexical bundles* alla forma impersonale

si tratta dunque di (7); si tratta di un (46); si tratta di una (35); non si tratta di (26); vale a dire il (14); vale a dire la (10); vale a dire un (7); si pone il problema (7); si può dire che (9);

#### c. Frammento verbale con un verbo altro da essere

si trova di fronte (12); ci troviamo di fronte (14); abbiamo a che fare (7); consiste nel fatto che (9); ha nulla a che (7); non ha nulla a (7); ha a che fare (21); ha il compito di (12); non ha bisogno di (12);

#### d. Infinito + locuzione preposizionale / sintagma nominale

fare a meno di (21); fare i conti con (13);

#### e. Congiunzione + clausola verbale frammentata

se è per sé (7); se è solo se (7); se non ci fosse (9); se è vero che (15);

#### f. Clausola passiva frammentata

come si è visto (9); come si è detto (7);

#### g. *che* + clausola verbale frammentata

che fare con la (8); che fare con una (8); che non è mai (8); che non si può (8); che non può essere (10); che si tratta di (11); che come abbiamo visto (8); che ci permette di (7); che fa sì che (7); che fare con il (7); che non si lascia (7); che si tratti di (7); che a partire da (9);

#### h. Interrogativo + frase verbale frammentata

che cos'è la (11);

i. Pronome dimostrativo + frase verbale frammentata

ciò che non è (10); questo non significa che (10); tutto ciò che si (9); tutto ciò che è (8);

l. Pronome relativo + frase verbale frammentata

cui non si può (8); su cui si fonda (8); di cui non si sa (9);

m. Preposizioni + sintagma verbale frammentato

a che vedere con (9); a che fare con (51); da un lato è (11); da ciò che è (8);  
a partire da un (14); a partire da una (17); a partire dal quale (9); in grado di fornire (7);  
in grado di rendere (7); in grado di capire (8); per quanto riguarda il (15); per quanto  
riguarda la (9).

#### *III.3.1.2.1.4 Lexical bundles che sono congiunzioni o locuzioni congiuntive*

nella misura in cui (67); in modo tale che (19); a meno che non (18).

#### *III.3.1.2.1.5 Lexical bundles che sono avverbi*

ma anche e soprattutto (11).

Si procede ora con i risultati emersi dall'analisi strutturale.

#### *III.3.1.2.2 I risultati emersi dall'analisi strutturale*

La lettura empirica dei dati manifestatisi nel *corpus* di filosofia ci permette di trarre qualche osservazione. L'analisi ha mostrato che la maggioranza dei *bundles* appartiene a due categorie principali – i *bundles* preposizionali e nominali. Queste due categorie coprono in totale il 79,6% di tutte le forme precostituite nel *corpus*. Come già riscontrato nell'ambito scientifico di storia, il terzo posto è riservato ai *clusters* verbali. Queste combinazioni di parole rappresentano il 19,2 % di tutti i *lexical bundles* identificati nel *corpus*. Questi dati risultano particolarmente importanti perché consentono di applicare al *corpus* di filosofia la suddivisione interna del *corpus* di storia, poiché le percentuali della partizione dei *bundles* preposizionali e nominali

risultano concordanti. Questi dati confermano le nostre ipotesi, concordando pienamente con i risultati ottenuti per le scienze sociali della lingua inglese. Come indicato da Cortes (2004), i *lexical bundles* identificati nel linguaggio scientifico delle scienze sociali possono essere classificati in due categorie principali – le sequenze preposizionali e quelle nominali.

I risultati del confronto relativo alle percentuali indicano che il ricorso ai diversi tipi di *lexical bundles* di entrambe le discipline scientifiche risulta quasi congruente. Su questo dato si può notare una minima divergenza nell'uso dei *lexical bundles* verbali, l'uso dei quali si dimostra un po' più alto nell'ambito di filosofia. Infatti l'utilizzo dei *clusters* verbali corrisponde al 9,8% in storia e al 19,2% in filosofia.

Tabella 17: Distribuzione strutturale dei *lexical bundles* di 4 parole in sequenza nel *corpus* di filosofia

<b><i>Lexical bundles</i> di 4 parole</b>	<b><i>Bundles</i> diversi</b>	<b>Percentuale</b>
<b>Preposizionali</b>	161	49,1 %
<b>Nominali</b>	100	30,4 %
<b>Verbali</b>	63	19,3%
<b>Congiunzionali</b>	3	0,9 %
<b>Avverbiali</b>	1	0,3%
<b>Totale</b>	<b>328</b>	<b>100 %</b>

Come già più volte accennato, queste combinazioni ricorrenti di parole non fanno parte della grammatica tradizionale, poiché nella maggior parte dei casi non dispongono di una struttura completa. Dal computo sono state evidenziate 54 forme complete, costituendo una percentuale del 16,5%. La generalità dei *lexical bundles* riscontrati risulta avere una struttura frammentata – un dato concordante con le ricerche sui *lexical bundles* nella prosa accademica della lingua inglese (Biber, 1999). Dobbiamo precisare che la frammentarietà dei *lexical bundles* risulta in lingua inglese un po' più bassa, aggirandosi intorno al 5%.

Tabella 18: Distribuzione strutturale dei *lexical bundles* di 4 parole completi e frammentati nel *corpus* di filosofia

<i>Lexical bundles</i> di 4 parole	Completi	Frammentati
<b>Preposizionali</b>	31	130
<b>Nominali</b>	23	77
<b>Verbali</b>	0	63
<b>Congiunzionali</b>	0	3
<b>Avverbiali</b>	0	1
Totale	<b>54</b>	<b>274</b>

Qui di seguito esamineremo il *corpus* di medicina del lavoro secondo la prospettiva della struttura dei *lexical bundles*; in un secondo luogo passeremo all'esame di tutti e tre gli ambiti scientifici secondo la struttura delle sequenze di quattro segmenti. I dati saranno infine comparati con i risultati ottenuti dall'analisi del *corpus* della lingua parlata (LIP).

### *III.3.1.3 I lexical bundles nel corpus di medicina del lavoro*

Il *corpus* di medicina del lavoro comprende 1.047.442 *tokens* o parole e 38.365 *types* o parole diverse. La reperibilità del materiale online ha condizionato la scelta della rivista scientifica di medicina del lavoro – *Giornale di medicina del lavoro ed ergonomia*. Sono state scelte per la ricerca le annate più recenti, poiché si è voluto presentare lo stato attuale delle sequenze precostituite. La ricerca include 496 articoli scientifici differenti, scritti da 2415 studiosi – un dato interessante che tocca il nucleo della disciplina, la quale incrementa la collaborazione tra pari.

Tabella 19: Dati statistici del *corpus* di medicina del lavoro

<b>Dati statistici del <i>corpus</i> di medicina del lavoro</b>	
Tokens	1.047.442
Types	38.365
Numero articoli scientifici	496
<i>Lexical bundles</i> diversi – occorrenza minima 10 volte	890
Frequenza dei <i>bundles</i>	18.454
% delle parole in <i>bundles</i>	7%

Come si vede dalla tabella 19, il *corpus* di medicina del lavoro presenta 890 *lexical bundles* diversi, con una frequenza di 18.454 volte su circa un milione di parole. Lo spoglio incorpora solo le sequenze comparse in almeno cinque articoli scientifici diversi, con un'occorrenza di almeno 10 volte su un milione di parole. Esaminando la fascia fino a 100 sequenze di quattro parole si può notare che i *lexical bundles* figurano in media in 25 articoli scientifici diversi, manifestandosi da in un massimo di 43 a in un minimo di 7 testi diversi (vedi Appendice 4).

L'analisi dei *clusters* del *corpus* di medicina del lavoro ha riconfermato il fatto che i *clusters* di quattro parole sono delle estensioni dei *clusters* di tre parole, mentre le sequenze di cinque parole incorporano quelle di quattro parole. Riportiamo a titolo d'esempio la sequenza più frequente nel *corpus* di medicina – *nei luoghi di lavoro*, con l'occorrenza di 215 su 1.047.442 parole o 205 su un milione di parole. La sua versione più breve è *luoghi di lavoro*, la quale occorre 274 volte su un milione di parole, mentre la sua variante più lunga *prevenzione nei luoghi di lavoro* si presenta con una frequenza di 51 volte su un milione di parole.

Inoltre, è interessante notare che il *corpus* di medicina del lavoro presenta nel numero dei *lexical bundles* identificati una prima divergenza nelle tre microlingue analizzate. Dai dati emerge che il numero dei *clusters* nella scienza applicata è del 55% più alto che nell'ambito di storia e del 64% più alto che in filosofia. Di conseguenza, la più alta numerosità dei *lexical bundles* si riflette in una percentuale più alta di sequenze di quattro parole nel *corpus* analizzato. Su questo dato si nota una divergenza tra le discipline umanistiche da un lato e la scienza applicata dall'altro, sulla quale ci soffermeremo in seguito.

Ulteriori discrepanze quantitative riguardano la media in cui i *bundles* compaiono nella fascia fino a 10. Le prime dieci posizioni vedono confluire i tre tipi di *bundles* identificati – preposizionali, nominali e verbali – con una frequenza che varia da 215 a 94 su 1.047.442 parole, ovvero da 205 a 89 su un milione di parole. La loro occorrenza varia da 43 fino a 22 articoli scientifici, in media in 35 testi diversi. Si può notare che la frequenza di occorrenza dei primi dieci *lexical bundles* del *corpus* di medicina è assai più alta che l'occorrenza dei *bundles* nelle discipline umanistiche.

La pervasività dei *lexical bundles* nel *corpus* di medicina risulta abbastanza chiara se osserviamo la tabella 20.

Tabella 20: I *lexical bundles* differenti e la loro frequenza d'occorrenza

<i>Lexical bundles</i>	Diversi	Occorrenze
<b>2 parole</b>	12.177	431.246
<b>3 parole</b>	4.147	95.690
<b>4 parole</b>	890	18.454
<b>5 parole</b>	255	4.462

La tabella mette in luce un nucleo abbastanza ampio di *lexical bundles* di tre e quattro segmenti in sequenza – su 38.365 parole distinte riscontriamo quasi 5000 *clusters* diversi, nei termini della proporzione i *lexical bundles* di 3 parole in successione coprono una percentuale del 27,4% di tutte le parole, mentre i *bundles* di 4 parole una percentuale pari al 2,2%. I *bundles* di tre parole si presentano con una frequenza di 95.690 volte su un milione di parole; la frequenza delle espressioni di quattro parole è di circa 18.454 volte su un milione di parole. Una prima osservazione che si impone addentrandoci nel *corpus* è la conferma delle ipotesi poste all'inizio di questa ricerca sulla pervasività dei *lexical bundles* nel linguaggio scritto, ovvero nelle microlingue scientifico-professionali. Simili dati sono emersi nelle ricerche sulla lingua inglese.

### III.3.1.3.1 La struttura dei *lexical bundles* nel *corpus* di medicina del lavoro

Nel *corpus* di medicina del lavoro sono state identificate tre categorie principali: a) *lexical bundles* che sono sintagmi nominali o incorporano frammenti di sintagmi nominali; b) *bundles* con la struttura preposizionale; e infine c) *clusters* verbali.

Il primo gruppo maggiormente rappresentato è quello dei *lexical bundles* con la forma nominale. In questa categoria sono stati inclusi: i *clusters* che incorporano la preposizione DI (come ad esempio *il datore di lavoro, il medico del lavoro*), A (ad esempio *il lavoro a turni, effetti a lungo termine*), IN (*la prevenzione nei luoghi,*



sicurezza nei luoghi di), DA (uno studio condotto da, i soggetti affetti da), PER (interesse per il lavoro, soddisfazione per il lavoro), TRA (i rapporti tra lo, età compresa tra i) e la preposizione SU (gli infortuni sul lavoro, gli effetti sulla salute). I frammenti di clausole nominali includono inoltre i *clusters* con un complemento frammentato (*datore di lavoro che*); i sintagmi complessi, come ad esempio *la salute e sicurezza, la sicurezza e la*; le espressioni che iniziano con un coordinatore, come nel caso di *e sicurezza sul lavoro, e la capacità di*; quelle che cominciano con il pronome *che*, come ad esempio *che il medico competente, che la maggior parte*; e infine le sequenze che incorporano un pronome relativo – *malattie per le quali, malattie professionali di cui*.

La seconda categoria è rappresentata dai sintagmi preposizionali, i quali sono stati suddivisi in: gruppo che comincia con la preposizione DI (*della medicina del lavoro, dei lavoratori per la*), IN (*nei luoghi di lavoro, nel luogo di lavoro*), PER (*per la prima volta, per la salute e*), DA (*sul luogo del lavoro, su un campione di*), A (*al datore di lavoro, a maggior rischio di*), SU (*sul luogo del lavoro, su un campione di*), TRA (*tra i due gruppi*), CON (*con il datore di*), e il gruppo che inizia con tutte le altre preposizioni, come *secondo i principi della, contro gli infortuni sul, durante il periodo di*. In questa categoria troviamo anche le sequenze che cominciano con un coordinatore (*anche in questo caso, sia in termini di*); che sono frammenti di frasi relative (*nel caso in cui, di cui all'articolo*); quelle che iniziano con una congiunzione, come per esempio *se da un lato*; e infine le strutture complesse – *quali ad esempio la*.

Il terzo gruppo costituisce frammenti di sintagmi verbali. Dall'analisi sono emerse espressioni che includono il verbo *essere* come copula, come ad esempio *è in grado di, non è stato possibile*; *bundles* con un verbo alla forma impersonale, come ad esempio *si tratta di un*; *clusters* con un verbo diverso dal verbo *essere* – *ha il compito di, deve riguardare tutti*; espressioni, il cui primo segmento è un coordinatore (*sia per quanto riguarda*) o il pronome *che* (*che si tratta di*); *bundles* che sono frammenti di frasi passive, ad esempio *è stato dimostrato che, sono stati sottoposti a*, *bundles* che incorporano un pronome relativo (*in cui sono stati*), un infinito (*ridurre il rischio di*); *bundles* che cominciano con una preposizione o locuzione preposizionale (*al fine di verificare; per quanto riguarda la*), con un participio (*sottoposti a sorveglianza*

sanitaria); e infine i *bundles* che cominciano con un sostantivo (*medicina del lavoro dovrà*).

Passiamo a questo punto alla classificazione strutturale dei *lexical bundles*.

### III.3.1.3.1.1 *Lexical bundles che sono sintagmi nominali o incorporano frammenti di sintagmi nominali*

- a. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione DI:

il datore di lavoro (184); i datori di lavoro (40); datori di lavoro (27); il medico del lavoro (128); medico del lavoro competente (41); parte del medico competente (34); attività del medico competente (17); collaborazione del medico competente (14); i medici del lavoro (25); figura del medico competente (16); medico di medicina generale (16); medici di medicina generale (16); la medicina del lavoro (62); il ruolo del medico (15); direzione provinciale del lavoro (16); organizzazione mondiale di sanità (15); istituto superiore di sanità (18); il responsabile del servizio (15); attività della vita quotidiana (14); il ministero del lavoro (13); l'organizzazione del lavoro (13); la sindrome delle apnee (13); tutela della salute (81); la tutela della salute (47); la promozione della salute (39); la salute dei lavoratori (63); lo stato di salute (29); promozione della salute (21); la sicurezza dei lavoratori (19); le condizioni di salute (16); il gruppo di lavoro (30); un gruppo di lavoro (24); un gruppo di controllo (14); gli ambienti del lavoro (13); il carico di lavoro (16); la valutazione del rischio (109); la valutazione dei rischi (63); valutazione del rischio stress (38); i fattori di rischio (63); un fattore di rischio (30); altri fattori di rischio (27); principali fattori di rischio (17); fattori di rischio lavorativi (15); fattori di rischio professionali (17); la sicurezza del lavoro (17); sicurezza del lavoro ISPELSL (14); un aumento del rischio (17); la stima del rischio (15); i dispositivi di protezione (18); movimentazione manuale dei carichi (67); la prevenzione degli infortuni (16); dispositivi di protezione individuale (51); di protezione individuale DPI (15); movimentazione manuale di carichi (32); la qualità della vita (34); società italiana di medicina (29); tabella delle malattie professionali (22); periodo massimo di indennizzabilità (23); sindrome del tunnel carpale (23); attività di sorveglianza sanitaria (22); programma di sorveglianza sanitaria (15); risultati della sorveglianza sanitaria (22); sorveglianza sanitaria dei lavoratori (24); i risultati della sorveglianza (14); le misure di prevenzione (21); norme di buona tecnica (24); soggetti di sesso maschile (24); il miglioramento delle condizioni (21); organizzazione del lavoro (25); di sicurezza e salute (15); di specializzazione in medicina (15); di tumore del polmone (15); di vigilanza in materia (15); movimentazione manuale dei pazienti (17); gli organi di vigilanza (16); lavoratori di sesso maschile (15); i valori di riferimento (14); il processo di valutazione (14); un sistema di gestione (17); scopo del presente lavoro (21); valutazione del rischio chimico (20); valutazione dello stress lavoro (20); la gestione del rischio (20); la valutazione dello stress (20); la raccolta dei dati (19); sindrome delle apnee ostruttive (17); province autonome di Trento (17); indice di massa corporea (17); lo stile di vita (19); patologie degli arti superiori (18); tabelle

delle malattie professionali (18); nuova tabella della malattia (14); denuncia di malattia professionale (15); nuove tabelle delle malattie (15); maggior parte dei soggetti (12); soggetti di sesso femminile (12); un ambiente di lavoro (12); valori limite di esposizione (12); il giudizio di idoneità (48); un giudizio di idoneità (14); la SS degli OS (12); un punto di vista (49); risultati di questo studio (13); ruolo del medico competente (13); sovraccarico biomeccanico degli arti (13); gruppi di lavoratori esposti (12), i luoghi di lavoro (12); la figura del medico (12); la percezione del rischio (12); maggior parte dei casi (19); una revisione della letteratura (11); fattori di rischio biomeccanico (10); gli orari di lavoro (10); il coefficiente di correlazione (10); il livello di esposizione (10); il rischio di esposizione (10); il turno del mattino (10); il turno di notte (10); l'analisi dei dati (10); la collaborazione del medico (10); la formulazione del giudizio (10); i giorni di riposo (10); la valutazione dell'efficacia (10); le strategie di coping (10); lo scopo di fornire (10); lo svolgimento di attività (10); modello di regressione logistica (10); punto di vista clinico (10); un documento di consenso (10); un punto di riferimento (10); fattori di rischio cardiovascolare (11); fattori di rischio presente (11); i principi della medicina (11); il fumo della sigaretta (11); il servizio di prevenzione (11); la facilità di accesso (11); la percentuale di soggetti (11); la riduzione dei rischi (11); medicina preventiva dei lavoratori (11); programmi volontari di promozione (11); somministrazione di un questionario (11); kg di peso corporeo (11);

datore di lavoro e (74); datore di lavoro di (21); datore di lavoro deve (22); medico del lavoro e (23); medico del lavoro in (17); medicina del lavoro e (91); medicina del lavoro ed (31); ministero del lavoro e (16); principi della medicina del (17); tutela della salute dei (30); materia di tutela della (21); salute dei lavoratori e (18); gestione della salute e (17); promozione della salute nei (16); valutazione del rischio e (37); valutazione del rischio da (30); valutazione dei rischi e (34); fattori di rischio per (32); fattori di rischio e (27); fattore di rischio per (26); un aumentato rischio di (19); servizio di prevenzione e (55); misure di prevenzione e (33); servizi di prevenzione e (20); materia di salute e (49); cartella sanitaria e di (34); la presenza di un (34); la presenza di una (16); miglioramento delle condizioni di (33); sistema di gestione della (22); condizioni di salute e (23); condizioni di lavoro e (16); ambienti di lavoro e (16); stato di salute e (17); processo di valutazione dei (23); fumo di sigaretta e (16); i valori medi di (15); giudizio di idoneità alla (31); documento di valutazione dei (27); parte del datore di (22); formulazione del giudizio di (25); responsabile del servizio di (25); luoghi di lavoro e (24); la sorveglianza sanitaria dei (16); attività di vigilanza in (15); sicurezza degli ambienti di (15); un elevato grado di (15); un elevato numero di (15); ambito della medicina del (14); datore di lavoro in (14); i principali fattori di (14); la necessità di un (14); la società italiana di (14); i soggetti di sesso (12); la maggior parte dei (56); la maggior parte delle (28); la maggior parte degli (19); la maggior parte di (16); la maggior parte del (15); un maggior numero di (16); i risultati di questo (17); luogo di lavoro e (15); lavoro della salute e (18); medicina del lavoro con (18); ministero del lavoro della (14); rappresentanti dei lavori per (14); riduzione del rischio di (14); stato di salute dei (14); datore di lavoro a (13); documento di valutazione del (13); medici del lavoro e (12); ministero della salute e (12); rischio di esposizione a (12); carico del datore di (11); consumo di alcol e (11); datore di lavoro o (11); fattori di rischio di (11); prevenzione degli infortuni e (13); aumento del rischio di (12); condizioni di salute del (12); il maggior numero di (12); punto di vista della (16); la necessità del rischio (12); tema di salute e (11); un elevato rischio di (11); un numero limitato di (11); valutazione

del rischio per (11); valutazione di tutti i (11); volontari di promozione della (11); ambienti di vita e (10); condizioni di vita e (10); datore di lavoro per (10); giudizio di idoneità al (10); il test T di (10); la movimentazione manuale dei (10); la possibilità di un (10), lo scopo del presente (10); preventiva dei lavoratori della (10); medicina del lavoro sia (10); prevenzione degli infortuni sul (10); processo di valutazione del (10); programmi di promozione della (10); servizi di medicina del (10); statistica dei dati e (10); T di student per (10); test T di student (10); un gran numero di (10); una vasta gamma di (10); valutazione dei rischi che (10); valutazione dei rischi in (10); luoghi di lavoro delle (13); medicina del lavoro di (11); medicina del lavoro in (11); medicina del lavoro la (11); medicina del lavoro deve (11); riduzione del numero di (11); presenza o meno di (13); la concentrazione media di (11),

b. Sintagma nominale con frammento di una clausola preposizionale che comincia con una preposizione diversa da DI

- la preposizione A

idoneità alla mansione specifica (36); il lavoro a turni (27); effetti a lungo termine (14); vibrazioni trasmesse al corpo (13); esposizione ad agenti biologici (15); lavoro a turni e (29); lavoro al fine di (14); i dati relativi alle (12); i dati relativi ai (10); i lavoratori esposti a (10); abitudine al fumo di (16); i dati relativi agli (11);

- la preposizione IN

la prevenzione nei luoghi (20); sicurezza nei luoghi di (60); prevenzione nei luoghi di (55), salute nei luoghi di (34); specialisti in medicina del (15); specializzazione in medicina del (15); vigilanza in materia di (15); normativa in materia di (10);

- la preposizione DA

patologie da sovraccarico biomeccanico (21); rischio da stress lavoro (21); uno studio condotto da (17); i soggetti affetti da (16);

- la preposizione PER

rischi per la salute (77); rischio per la salute (44); rischi per la sicurezza (32); lavoratori per la sicurezza (23); interesse per il lavoro (22); soddisfazione per il lavoro (12); misure per la tutela (11); patto per la salute (10); tutti i rischi per (29); i rischi per la (39); linee guida per la (20); istituto superiore per la (14); sistema informativo nazionale per (12); un rischio per la (11); il rischio per la (11);

- la preposizione TRA

rapporti tra lo stato (14); età compresa tra i (28); i rapporti tra lo (14);

- la preposizione SU

gli infortuni sul lavoro (31); gli effetti sulla salute (11); effetti negativi sulla salute (10); infortuni sul lavoro INAIL (10); sicurezza sui luoghi di (15); uno studio condotto su (14); infortuni sul lavoro e (13);

c. Sintagma nominale con complemento frammentato

datore di lavoro che (17); medico del lavoro che (14);

d. Sintagma nominale complesso

la salute e sicurezza (49); sicurezza e la salute (47); rischio stress lavoro correlato (35); regioni e province autonome (25); piccole e medie imprese (20); lo stress lavoro correlato (17); indirizzi scientifici più avanzati (16); tutto il territorio nazionale (16); il servizio sanitario nazionale (13); una differenza statisticamente significativa (15); salute e la sicurezza (38); sostanze stupefacenti e psicotrope (14); fumatori e non fumatori (10); tutte le figure professionali (10); una correlazione statisticamente significativa (10); lo stato le regioni (10); elevato rischio di infortuni (10);

un vero e proprio (11); salute e sicurezza dei (38); salute e sicurezza sui (13), correlazione statisticamente significativa tra (11); una differenza significativa tra (10); valutazione e gestione del (10); prevenzione e sicurezza nei (10); origine lavorativa e di (10); lavoratrici e dei lavoratori (11), lavoro e della previdenza (11); la prima volta in (10); stress lavoro correlato in (10); stress lavoro correlato e (28); la sicurezza e la (48); la salute e la (40); il medico competente e (22); la prevenzione e la (24); salute e sicurezza e (14); la sorveglianza sanitaria e (13); la presenza o meno (10); il medico competente non (10); il medico competente si (10); gli indirizzi scientifici più (10); stato le regioni e (11); salute e sicurezza sul (79); salute e della sicurezza (70); sanitaria e di rischio (34), vita e di lavoro (24); infortuni e delle malattie (18); salute e sicurezza nei (18); informazione e formazione dei (12); cartelle sanitarie e di (13); Trento e di Bolzano (12); commissione consultiva permanente per (15); conferenza permanente per i (15); lavoro e quello di (15); regioni e delle province (11); regioni e le province (17); le regioni e le (17); salute e la sicurezza (38); prevenzione e la sicurezza (17); lavoro e igiene industriale (16); lavoro ed igiene industriale (10), prevenzione e protezione e (16); la durata e la (12);

autonome di Trento e (17); italiana di medicina del (30); nazionale per la prevenzione (17); biomeccanico degli arti superiori (13); superiore per la prevenzione (14); permanente per i rapporti (15); permanente per la salute (14); mondiale della sanità OMS (13); consultiva permanente per la (15); sanitarie e di rischio (15); collegati allo stress lavoro (11);

E.S.M.I. (55); G ITAL MED LAV (14); ITAL MED LAV ERG (14);

e. Marcatore del discorso + clausola nominale frammentata

e sicurezza sul lavoro (77); e sicurezza dei lavoratori (35); e la salute dei (31); e sicurezza nei luoghi (31), e la necessità di (25); e la sicurezza del (15); e la valutazione delle (15); e sicurezza degli ambienti (15); e gestione del rischio (17); e formazione dei lavoratori (14); e sicurezza del lavoro (14); e la capacità di (12); e la possibilità di (12);

e i livelli di (13); e sicurezza sui luoghi (13); e la mancanza di (10); e lo sviluppo di (10); e fattori di rischio (11); e la promozione delle (11); e la presenza di (16), e la sicurezza dei (16); e le province autonome (17); e il medico competente (10);

f. *che* + clausola nominale frammentata

che il medico competente (39); che la maggior parte (39); che la sorveglianza sanitaria (11); che le linee guida (12);  
che il datore di (26); che l'esposizione a (20); che la prevalenza di (11); che il medico del (16); che i livelli di (14); che la valutazione dei (10);

g. Sintagma nominale + pronome relativo

malattie per le quali (11); malattie professionali di cui (14); malattie la cui origine (17); caso in cui il (12); la cui origine lavorativa (17).

*III.3.1.3.1.2 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi preposizionali*

a. La preposizione DI:

della medicina del lavoro (94); dei medici del lavoro (38); del medico del lavoro (92); dei luoghi di lavoro (14); del datore di lavoro (112); dei datori di lavoro (36); dei lavoratori della sanità (19); della salute e sicurezza (63); della salute dei lavoratori (43); della sicurezza dei lavoratori (22); delle condizioni di salute (26); della valutazione dei rischi (40); della valutazione del rischio (35); del rischio stress lavoro (25); del rischio da stress (23); del lavoro della salute (18); del lavoro e igiene (16); del lavoro ed ergonomia (15); del posto di lavoro (15); della tutela della salute (25); dello stato di salute (42); della salute nei luoghi (22); dell'organizzazione del lavoro (38); del giudizio di idoneità (60); dei fattori di rischio (55); dello stress lavoro correlato (55); degli ambienti di lavoro (39); delle misure di prevenzione (27); del servizio di prevenzione (34); dei servizi di prevenzione (23); delle attività di prevenzione (26); della prevenzione nei luoghi (14); degli infortuni sul lavoro (27); del numero di infortuni (11); delle condizioni di lavoro (31); del rapporto di lavoro (30); del ministero della salute (23); del ministero del lavoro (20); dei dispositivi di protezione (22); del gruppo di lavoro (20); della sicurezza nei luoghi (18); delle attività di vigilanza (18); del documento di valutazione (16); della qualità della vita (16); del lavoro ed ergonomia (15); del lavoro della salute (18); della cuffia dei rotatori (18); delle attività della vigilanza (18); del lavoro ed igiene (16); dei giudizi di idoneità (15); dei valori di riferimento (14); dei livelli di esposizione (11); del sistema nervoso centrale (11); di assenza dal lavoro (11); di gestione del rischio (11); dei diritti di lavoratori (10); dei diritti dei lavoratori (10); del lavoro ed igiene (10); del mondo del lavoro (10); del servizio sanitario nazionale (10); della normativa in materia (10); degli organi di vigilanza (12); del mercato del lavoro (12); della promozione della salute (12);

di medicina del lavoro (163); di salute e sicurezza (94); di salute del lavoratore (22); di valutazione del rischio (89); di valutazione dei rischi (84); di stress lavoro correlato (25); di promozione della salute (70); di tutela della salute (54); di salute dei lavoratori

(25); di gestione della salute (13); di formazione e informazione (13); di idoneità alla mansione (36); di esposizione ad amianto (14); di una malattia professionale (14); di volta in volta (16); di tutti i rischi (18); di tutti i soggetti (18); di fattori di rischio (17); di prevenzione e sicurezza (26); di prevenzione e protezione (92); di infortuni sul lavoro (27); di sorveglianza sanitaria (20); di disturbi del sonno (20); di igiene e sicurezza (10); di inizio e fine (10); di massa corporea BMI (10); di sensibilità e specificità (10); di sicurezza sul lavoro (10); di valutare il rischio (10); di assunzione di sostanze (12); di gestione della sicurezza (12); di formazione ed informazione (12); di sesso maschile (12);

dei lavoratori per la (27); degli arti superiori a (26); del presente lavoro e (23); dei lavoratori e dei (22); del nostro studio e (16); dei lavoratori disposti a (14); dei rischi per la (14); della cartella sanitaria e (14); del lavoro e delle (13); del presente studio e (11); della sicurezza e della (11); delle regioni e delle (11); dei lavoratori e delle (10); del lavoro e del (10); del lavoro e di (10); del rischio e la (10); della società italiana di (10); della letteratura scientifica e (10); del medico competente alla (11); degli infortuni e delle (20); del rischio da parte (11); delle malattie professionali di (11); del lavoro e della (19); dell'arto superiore a (18); del medico competente di (12); della sorveglianza sanitaria e (12); delle lavoratrici e dei (12); di lavoro e del (12), di prevenzione e di (12); di salute e di (12); di Trento e di (12); del medico competente e (27); del medico competente in (16); della salute e della (82);

di vita e di (26); di un gruppo di (25); di lavoro e di (22); di un sistema di (17); di riferimento per la (18); di lavoro e dei (13); di esposizione e di (11); di questo studio e (11); di questo tipo di (11); di sostanze psicotrope e (11); di lavoro ai fini (10); di lavoro e la (10); di questo lavoro e (10); di una condizione di (10); di uno o più (10); di una condizione di (10); di medicina preventiva dei (10); di lavoratori esposti a (16); di rischio per la (25); di età compresa tra (38); di fronte ad una (11);

b. La preposizione IN:

nei luoghi di lavoro (215); nel luogo di lavoro (21); negli ambienti di lavoro (76); nell'ambiente di lavoro (20); nella valutazione del rischio (20); nella valutazione del rischio (20); nel mondo del lavoro (18); nel processo di valutazione (16); nella maggioranza dei casi (16); nel corso degli anni (16); nella valutazione dei rischi (12); negli ambienti di vita (12); nel corso della sorveglianza (13); nella fascia di età (13); nel documento di valutazione (13); nell'ambito della medicina (14); nella tabella I sono (11); nei confronti dei lavoratori (10);

nella maggior parte dei (24); nella maggior parte delle (11); nella popolazione generale e (21); negli ultimi anni e (19); nei lavoratori esposti a (10);

in ambiente di lavoro (23); in medicina del lavoro (94); in materia di salute (53); in materia di sicurezza (22); in materia di tutela (22); in tema di salute (14); in tutto il mondo (11); in tutti i casi (23); in uno studio condotto (22); in materia di prevenzione (18); in questi ultimi anni (16); in un recente studio (10); in via di sviluppo (10); in entrambi i sessi (10); in tema di sicurezza (10); in sede di conferenza (10);

in particolare per quanto (20); in un gruppo di (21); in questo caso il (18); in questo caso e (12); in presenza di una (15); in particolare per la (14); in questo caso la (11); in un ottica di (11); nel corso di un (10); in linea con il (13); in presenza di un (13); in questo caso si (13); in un campione di (10); in un contesto di (10); in collaborazione con gli (10); in collaborazione con l' (10); in collaborazione con il (12); in base a quanto (10); in accordo con la (11); in accordo con i (11); in accordo con quanto (14); in maniera tale da (11);

c. La preposizione PER:

per la prima volta (40); per la formazione continua (35); per la popolazione generale (29); per la sorveglianza sanitaria (21); per il medico competente (28); per gli operatori sanitari (17); per la salute umana (22); per la maggior parte (13); per i lavoratori esposti (10);

per la salute e (81); per la salute dei (24);); per la sicurezza e (50); per la valutazione del (45); per la valutazione dei (22); per la diagnosi di (31); per la tutela della (30); per la prevenzione e (25); per lo sviluppo di (21); per la determinazione di (19); per la presenza di (18); per la produzione di (18); per il datore di (18); per il medico del (22); per un totale di (33); per i rapporti tra (14); per la gestione del (14); per la realizzazione di (14); per la valutazione delle (14); per la valutazione della (14); per la prevenzione nei (13); per la raccolta dei (13); per il miglioramento delle (12); per una serie di (12); per la medicina del (10); per la prevenzione degli (10); per la prevenzione della (10); per la promozione della (11); per la SS degli (11); per la valutazione dell' (15);

d. La preposizione DA:

dal punto di vista (152); da parte del medico (53); da parte dei lavoratori (42); da parte del datore (22); da parte degli operatori (18); da parte del lavoratore (18); dal ministero della salute (13); da parte dei medici (12); dai medici del lavoro (11); dalla commissione consultiva permanente (10); dalla cessazione della lavorazione (17); dal datore di lavoro (33); dal gruppo di lavoro (15), da stress lavoro correlato (27);

da un minimo di (15); da un punto di (48); da parte di un (19); da una serie di (12); da un lato e (10);

e. La preposizione A:

al datore di lavoro (61); ai datori di lavoro (15); alla valutazione dei rischi (39); alla valutazione del rischio (38); ai fattori di rischio (23); al fumo di sigaretta (35); al fumo di tabacco (13); ai sensi dell'art. (29); allo stress lavoro correlato (23); al medico del lavoro (22); a livelli di rumore (18); a tutti i lavoratori (18); alla tutela della salute (18); alla promozione della salute (17); a fattori di rischio (18); al sistema mano braccio (18); alla promozione della salute (17); all'ambiente di lavoro (13); alla luce di quanto (13); a turni e notturno (10); alla medicina del lavoro (10); alle differenze di genere (11); a tutti i livelli (12);

a maggior rischio di (16); ad alto rischio di (14); ad elevato rischio di (10); ad un massimo di (14); ad un gruppo di (13); ad una condizione di (11); ad una serie di (10); alla movimentazione manuale dei (12); ai sensi del d (12);



a fronte di un (20); a fronte di una (20); al di sotto dei (16); al di sotto di (11); al di sotto del (12); al di sotto della (12); ai fini di una (10); a causa di una (11); all'interno di un (12); ai fini della valutazione (14); a carico degli arti (11); a carico del datore (11); a carico del sistema (11); rispetto alla popolazione generale (22);

f. La preposizione SU:

sul luogo del lavoro (52); sui luoghi di lavoro (34); sul posto di lavoro (28); sullo stato di salute (26); su tutto il territorio (18); sulla base dei dati (16); sulla valutazione del rischio (14); sulla salute dei lavoratori (12); sulle condizioni di salute (11); sulle condizioni di lavoro (13);

su un campione di (30); su un gruppo di (18); sulla base di quanto (11); sulla base di un (11); sui rischi per la (11); sulla base di queste (10);

g. La preposizione TRA:

tra i due gruppi (16);

tra lo stato le (11); tra i livelli di (12);

h. La preposizione CON:

con il medico competente (11);

con il datore di (33); con l'obiettivo di (17); con particolare riferimento alla (14); con particolare riferimento alle (12); con età media di (10); con lo scopo di (36);

i. Altre preposizioni:

secondo i principi della (20);

contro gli infortuni sul (17);

durante il periodo di (11);

l. Coordinatore + frammento di frase preposizionale

anche in questo caso (50); anche al fine di (12); ma anche per la (11);

sia in termini di (25);

e della sicurezza dei (21); e delle malattie professionali (15); e delle province autonome (14); e della previdenza sociale (12); e delle politiche sociali (12); e della sicurezza nei (10);

m. Frammento di frasi relative

nel caso in cui (38); nei casi in cui (25); nel momento in cui (22); nel momento in cui (22);

per i quali e (22); per le quali e (22); per le quali non (12); per i quali non (16);

di cui all'articolo (30); di cui all'art. (21); di cui agli artt. (11); di lavoro in cui (10);

n. Congiunzione + clausola preposizionale frammentata

se da un lato (14);

o. Strutture complesse

quali ad esempio la (11).

### *III.3.1.3.1.3 Lexical bundles che incorporano frammenti di sintagmi verbali*

a. Essere + sintagma preposizionale frammentato

è in grado di (65); sono in grado di (55); essere in grado di (30); sia in grado di (18); è stato quello di (34); è di rischio e (11); è più in generale (10),

b. Essere + aggettivo

non è stato possibile (16); è obbligatoria la denuncia (10);

c. Essere + nome

è stato oggetto di (12); i dati sono stati (21); i risultati sono stati (13); la ricerca è stata (13); il questionario è stato (14); i lavoratori sono stati (15); i soggetti sono stati (11); lo studio è stato (10);

d. *Lexical bundles* alla forma impersonale

si tratta di un (32); si tratta di una (17); si ritiene che la (13);

e. Verbo diverso da *essere*

ha il compito di (16); deve riguardare tutti (11); hanno la possibilità di (10); ha lo scopo di (14); ha evidenziato che la (11); si può affermare che (10); conto del fatto che (10); il medico competente può (15); la tabella I riporta (10); è risultato pari a (16); è risultata pari a (11);

f. Coordinatore + sintagma verbale frammentato

sia per quanto riguarda (27); più in grado di (11);

g. Clausola verbale frammentata con *che*

che si tratta di (11); medici competenti che operano (11);

h. Clausola verbale frammentata + complemento frammentato

è emerso che il (11);

i. Clausola passiva frammentata

è stato dimostrato che (30); sono stati sottoposti a (21); è stato utilizzato il (20); sono stati presi in (17); sono stati sottoposti ad (17); è stata effettuata una (16); è stata condotta una (12); è stata osservata una (12); è stata utilizzata la (12); sono stati suddivisi in (12); stati presi in considerazione (14); è stata valutata la (13); è stato somministrato un (11); è stata rilevata in (11), è stata rilevata una (11); si è ritenuto opportuno (10); sono stati analizzati (10); stato somministrato un questionario (10); sono stati confrontati con (10);

l. Pronome relativo + clausola verbale frammentata

le quali è obbligatoria (10); quali è obbligatoria la (10); cui origine lavorativa è (17); in cui sono stati (13);

m. Clausola verbale frammentata con l'infinito

ridurre il rischio di (10);

n. Preposizione + sintagma verbale frammentato

al fine di verificare (16); al fine evitare (18); al fine di poter (14); al fine di garantire (10); al fine di ridurre (11); al fine di ottenere (11); al fine di identificare (10); al fine di individuare (12); al fine di valutare (12); allo scopo di individuare (10); in grado di determinare (35); in grado di influenzare (29); in grado di fornire (19); in grado di indurre (14); in grado di garantire (10); in grado di modificare (10); grado di determinare un (11);

per quanto riguarda la (107); per quanto riguarda il (73); per quanto riguarda le (50); per quanto riguarda i (39); per quanto riguarda gli (36); per quanto riguarda l' (36); per quello che riguarda (15); per quanto concerne la (14); per quanto concerne le (13); per quanto concerne il (20); particolare per quanto riguarda (15); quanto riguarda la valutazione (10); soprattutto per quanto riguarda la (18);

tra coloro che hanno (14);

o. Sintagma verbale frammentato che comincia con participio passato

sottoposti a sorveglianza sanitaria (24); entrata in vigore del (18); esposti a rischio biologico (12); messa in atto di (10); trasmesse al sistema mano (15); trasmesse al corpo umano (14);

p. Sintagma verbale frammentato che comincia con un sintagma nominale

medicina del lavoro dovrà (11).

Ora possiamo procedere con l'esposizione dei risultati emersi dall'analisi strutturale.

### III.3.1.3.2 I risultati emersi dall'analisi strutturale

Dalla presa in esame della lista dei *lexical bundles* emersa dall'estrazione automatica del *corpus* di medicina, si distinguono tre categorie di *bundles*. La maggioranza delle sequenze, ovvero il 47,6%, risulta avere la struttura nominale mentre al secondo posto si trovano i *lexical bundles* preposizionali che coprono il 41,4% di tutti i *clusters* individuati. Da ultimo appaiono le sequenze verbali, per una percentuale dell'11%.

La struttura nominale del *corpus* conferma quindi le ipotesi tradizionali sulla struttura nominale del linguaggio scritto, in particolare del linguaggio scientifico.

Tabella 21: Distribuzione strutturale dei *lexical bundles* di 4 parole in sequenza nel *corpus* di medicina del lavoro

<i>Lexical bundles di 4 parole</i>	<i>Bundles diversi</i>
<b>Nominali</b>	424
<b>Preposizionali</b>	368
<b>Verbali</b>	98
Totale	<b>890</b>

Come già più volte segnalato, nello spoglio è necessario distinguere tra due tipi di misura – la quantità dei *lexical bundles* diversi e la loro ricorrenza nel *corpus*. Il confronto relativo ai due tipi di misurazione nel *corpus* di medicina non si manifesta in un diverso comportamento statistico, bensì risulta avere una fascia di autocopertura. I *bundles* nominali e preposizionali sono difatti usati più frequentemente dei *clusters* verbali. Le sequenze nominali e preposizionali coprono rispettivamente il 46,3% e il 44% di tutte le frequenze nel *corpus*; da contro, la frequenza dei *clusters* verbali costituisce il 9,7% di tutte le frequenze d'uso.

La misura della frequenza di occorrenza consente di confermare il dato atteso circa il carattere prevalentemente nominale dello scritto rispetto al parlato.

Tabella 22: Frequenza d'uso dei *lexical bundles* di 4 parole in sequenza nel *corpus* di medicina del lavoro

<b>Lexical bundles di 4 parole</b>	<b>Frequenza</b>
<b>Nominali</b>	8.531
<b>Preposizionali</b>	8.140
<b>Verbali</b>	1.770
<b>Totale</b>	<b>18.454</b>

Come già detto in precedenza, le combinazioni prese in esame non dispongono di una struttura completa. Dal computo sono state evidenziate 281 forme complete, che costituiscono una percentuale del 31%. La generalità dei *lexical bundles* riscontrati risulta avere una struttura frammentata, pur presentando una divergenza nella proporzione rispetto alla lingua inglese. Precedenti ricerche per la lingua inglese hanno rilevato che solo il 5% della prosa accademica in lingua inglese dispone di una struttura completa (Biber ed altri, 1999). Questo aspetto sarà approfondito più avanti quando si presenterà un esame complessivo dei tre ambiti scientifici.

Tabella 23: Distribuzione strutturale dei *lexical bundles* di 4 parole completi e frammentati nel *corpus* di medicina

<b>Lexical bundles di 4 parole</b>	<b>Completi</b>	<b>Frammentati</b>
<b>Nominali</b>	75	349
<b>Preposizionali</b>	187	181
<b>Verbali</b>	19	79
<b>Totale</b>	<b>281</b>	<b>609</b>

### **III.3.2 La struttura dei *lexical bundles* nella prosa accademica**

Qui di seguito esamineremo il *corpus* specialistico – di storia, filosofia e medicina del lavoro – facendoci guidare in un primo momento dalla quantità e dalla frequenza dei *bundles*; poi passeremo all'esame dei *bundles* secondo la prospettiva della loro struttura.

Il *corpus* specialistico si compone di 3.405.346 *tokens* e di 102.724 *types* o parole distinte. Nell'analisi sono stati inclusi 756 articoli scientifici e 2806 autori diversi.

Come illustra la tabella 24, nel *corpus* della prosa accademica sono stati individuati 220 *lexical bundles* di 4 parole diversi, con una frequenza di occorrenza di circa 4.000 volte su un milione di parole.

Tabella 24: Dati statistici del *corpus* specialistico

<b>Dati statistici del <i>corpus</i> specialistico</b>	
Tokens	3.405.346
Types	102.724
Numero articoli scientifici	756
<i>Lexical bundles</i> diversi – occorrenza minima 35 volte	220
Frequenza dei <i>bundles</i>	13.536
% delle parole in <i>bundles</i>	1,6%

Per cercare di identificare le linee di tendenza che caratterizzano le microlingue dalla prospettiva dei *lexical bundles*, si ritiene necessario in primo luogo proporre una lista di 22 *bundles* più frequenti nel *corpus* specialistico di 3.434.405 *tokens* – la storia e la filosofia come rappresentanti di discipline umanistiche – con un *corpus* di 2.386.963 – e di una scienza applicata, la medicina del lavoro, con un *corpus* basato su 1.047.442 parole. Ci si è limitati ai primi 22 *lexical bundles* poiché dall'estrazione automatica sono emersi solo 22 *clusters* di 5 parole in sequenza.

Dallo spoglio dell'intero *corpus* specialistico ci si è limitati alle espressioni con un'occorrenza minima di 10 volte su un milione di parole o 34 su 3.434.405 parole – come suggerito dalle ricerche precedenti svolte per la lingua inglese.

Tabella 25: Lista dei primi 22 *lexical bundles* di 3, 4 e 5 parole in sequenza nel *corpus* specialistico

N	Lexical bundles di 3 parole		Lexical bundles di 4 parole		Lexical bundles di 5 parole	
	Word	Freq.	Word	Freq.	Word	Freq.
1	IN GRADO DI	907	DAL PUNTO DI VISTA	460	DA UN PUNTO DI VISTA	151
2	PUNTO DI VISTA	887	NEI LUOGHI DI LAVORO	216	DA QUESTO PUNTO DI VISTA	92
3	PER QUANTO RIGUARDA	658	IL DATORE DI LAVORO	191	NELLA MAGGIOR PARTE DEI CASI	76
4	SI TRATTA DI	574	UN PUNTO DI VISTA	184	DELLA SALUTE E DELLA SICUREZZA	70
5	AL FINE DI	550	SI TRATTA DI UN	179	SALUTE E SICUREZZA SUL LAVORO	70
6	LA POSSIBILITÀ DI	508	NEL MOMENTO IN CUI	175	SICUREZZA NEI LUOGHI DI LAVORO	60
7	IN CUI SI	490	PER LA PRIMA VOLTA	171	DAL PUNTO DI VISTA DELLA	58
8	IN QUESTO CASO	480	DI MEDICINA DEL LAVORO	158	PREVENZIONE NEI LUOGHI DI LAVORO	54
9	UNA SERIE DI	473	PER QUANTO RIGUARDA LA	155	IN MATERIA DI SALUTE E	49
10	DATORE DI LAVORO	465	DA UN PUNTO DI	151	MATERIA DI SALUTE E SICUREZZA	49
11	DAL PUNTO DI	463	ANCHE IN QUESTO CASO	144	SERVIZIO DI PREVENZIONE E PROTEZIONE	47
12	LA PRESENZA DI	459	UNA VERA E PROPRIA	142	TUTELA DELLA SALUTE E DELLA	47
13	MEDICINA DEL LAVORO	454	IL MEDICO DEL LAVORO	128	LA SICUREZZA E LA SALUTE	45
14	DA UN LATO	442	PER QUANTO RIGUARDA IL	124	PER LA SALUTE E SICUREZZA	37
15	UNA SORTA DI	432	DEL DATORE DI LAVORO	119	DI MEDICINA DEL LAVORO E	36
16	IN TERMINI DI	430	È IN GRADO DI	118	LA SALUTE E LA SICUREZZA	36
17	IL FATTO CHE	422	LA VALUTAZIONE DEL RISCHIO	109	DI VITA E DI LAVORO	35
18	NEGLI STATI UNITI	407	SI TRATTA DI UNA	106	CARTELLA SANITARIA E DI RISCHIO	34
19	LA MAGGIOR PARTE	399	QUESTO PUNTO DI VISTA	101	DA PARTE DEL MEDICO COMPETENTE	34
20	DA PARTE DI	397	A PARTIRE DAGLI ANNI	100	PER LA VALUTAZIONE DEL RISCHIO	34
21	LA NECESSITÀ DI	375	DI VOLTA IN VOLTA	99	SALUTE E SICUREZZA DEI LAVORATORI	34
22	DEL LAVORO E	367	LA MAGGIOR PARTE DEI	98	SALUTE NEI LUOGHI DI LAVORO	34

La lista mette in luce il fatto che i *lexical bundles* sono delle espressioni in effetti molto frequenti. Questo elenco di *lexical bundles* sembra confermare da un lato i dati rilevati dalle prime analisi parziali di questa ricerca, dall'altro concorda pienamente con le osservazioni di Cortes (2004) sulla struttura dei *lexical bundles* nella prosa accademica della lingua inglese – molte delle espressioni di 4 o 5 segmenti in successione incorporano nella loro struttura *bundles* di 3 parole. Tra i primi 22 *bundles* ci sono 19 *lexical bundles* di 4 parole in sequenza, e tutti i 22 *lexical bundles* di 5 parole in successione, che nella propria struttura includono un *cluster* di 3 parole.

La prima prova circa l'importanza dei *bundles* nel *corpus* specialistico, che non può essere trascurata, viene dalla loro frequenza di occorrenza. La lista mette in risalto che i *bundles* di quattro parole nelle prime dieci posizioni occorrono in media con una frequenza di 204 volte su un milione di parole e in media in 85 articoli scientifici differenti. Il *bundle* più frequente – *dal punto di vista* – occorre 460 su 3.434.405 *tokens*, ovvero 134 volte su un milione di parole, manifestandosi in 158 articoli scientifici.

In media i primi 50 *clusters* occorrono 60 volte su un milione di parole e sono riscontrabili in almeno 50 articoli scientifici differenti. La lista indica che il *bundle dal punto di vista* è la sequenza più utilizzata, presentando una frequenza di occorrenza del 50% più alta rispetto al secondo classificato. Per quanto riguarda i *bundles* di tre parole, indichiamo a mo' di esempio il *lexical bundle* più frequente *in grado di*, che appare in 231 articoli scientifici con una frequenza di occorrenza di 907 volte su 3.434.405 parole o 264 volte su un milione di parole.

Ulteriori dati quantitativi riguardano l'occorrenza di tutti i *bundles* di quattro parole identificati nel *corpus* specialistico – tutte le sequenze di quattro parole occorrono più di 13.000 volte nell'intero *corpus*, ovvero di circa 4.000 volte su un milione di parole, mentre le sequenze di tre parole hanno una frequenza notevolmente più alta, di più di 135.000 volte nel *corpus*, ossia di circa 40.000 su un milione di parole.

Sommando la frequenza dei *bundles* di tre e quattro segmenti in successione si raggiunge un'occorrenza di circa 44.000 volte su un milione di parole.

I dati appena esposti concordano con le ricerche svolte per la lingua inglese, le quali indicano che i *bundles* più frequenti nella prosa accademica presentano la stessa



frequenza di occorrenza, ovvero da 200 a 400 volte su un milione di parole. I *bundles* di tre parole occorrono più di 60.000 volte nel *corpus* della scrittura accademica in lingua inglese, quelli di quattro parole più di 5.000 volte su un milione di parole (Conrad, Biber, 2004). Questi risultati concordano con i risultati emersi dall'analisi della scrittura accademica in lingua inglese, presentando però una frequenza minore in lingua italiana – del 33% e del 24% rispettivamente. La differenza nella proporzione quantitativa potrebbe essere attribuita alla diversa tipologia linguistica – tuttavia ulteriori ricerche sono necessarie per poter confermare questa osservazione.

Questo primo elenco dei *lexical bundles* sembra inoltre confermare i dati attesi rispetto alla proporzione testuale che i *bundles* ricoprono nel testo, coincidendo con i dati sulla lingua inglese. Il conteggio delle frequenze d'occorrenze dei *bundles* ha rivelato che il 12% di tutte le parole occorre in segmenti di tre parole in sequenza, mentre l'1,6% in segmenti di quattro parole in successione. Paragonando questo dato con il risultato in lingua inglese, si può notare una proporzione un po' più alta in lingua inglese, la quale si aggira intorno al 20%.

I risultati emersi concordano con l'osservazione di Liu (2012) sul bisogno di controllare la distribuzione dei *lexical bundles* per ogni disciplina scientifica, vista la discrepanza fra i risultati sulla frequenza dei *clusters* nella scrittura accademica generale e in ciascun ambito scientifico.

Prima di individuare il rapporto tra la lingua parlata e scritta nei termini di *lexical bundles*, può essere interessante presentare la loro struttura prevalente.

Tra i top 50 dei *lexical bundles* della medicina ci sono 43 *bundles* che sono connessi alla disciplina e non sono presenti in nessun'altra disciplina; tra i primi 50 della storia ce ne sono 11 che sono propriamente legati alla disciplina, e altri 10 *lexical bundles* che esprimono una collocazione temporale e non sono presenti in nessun'altra disciplina; nell'ambito della filosofia ce ne sono solo 3 che sono propriamente legati alla disciplina e altri 10 che non si trovano in nessun'altra disciplina – questi sono nella maggior parte dei casi orientati al testo.

In tutte e tre le discipline viene riscontrato solo 1 *lexical bundle* in comune – *dal punto di vista*. La medicina del lavoro e la storia hanno 3 *lexical bundles* in comune – *la maggior parte dei, per quanto riguarda il/la, anche in questo caso*. Due di questi riguardano l'organizzazione testuale. La medicina del lavoro e la filosofia hanno solo 1

*bundle* in comune – è in grado di. Si tratta anche in questo caso di un'espressione che concerne l'organizzazione testuale. La storia e la filosofia, due discipline umanistiche, hanno tra i primi 50 *lexical bundles* 17 *clusters* in comune. Tra questi ci sono soprattutto sequenze di parole che si riferiscono alla diversificazione dell'argomentare e quelle legate all'organizzazione testuale. Il risultato sopporta le osservazioni di Hyland (2008a) sull'affinità dei campi simili e quelle di Cortes (2004) sulla diversità dei *lexical bundles* in ambiti scientifici diversi.

L'analisi ha rilevato che la generalità dei *bundles* nella prosa accademica analizzata presenta una struttura nominale o preposizionale. I dati emersi dall'analisi del *corpus* coincidono in maniera quantitativa con i dati sulla prosa accademica in lingua inglese.

Proponiamo di seguito il raggruppamento strutturale dei *lexical bundles* settoriali.

### ***III.3.2.1 Lexical bundles che sono sintagmi nominali o incorporano frammenti di sintagmi nominali***

- a. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione DI:

il datore di lavoro (191); un punto di vista (184); il medico del lavoro (128); questo punto di vista (101); la maggior parte dei (98); medicina del lavoro e (91); tutela della salute e (81); maggior parte dei casi (78); datore di lavoro e (74); la maggior parte delle (73); la maggior parte degli (66); movimentazione manuale dei carichi (67); la presenza di un (64); punto di vista un (64); la valutazione dei rischi (63); i fattori di rischio (62); la medicina del lavoro (62); la salute dei lavoratori (62); servizio di prevenzione e (55); la maggior parte di (54); i datori di lavoro (53); la qualità della vita (50); dispositivi di protezione individuale (49); materia di salute e (49); il giudizio di idoneità (48); la tutela della salute (48); giovani di origine italiana (43); il punto di partenza (42); medico del lavoro competente (41); la promozione della salute (39); fine degli anni sessanta (38); la fine degli anni (38); seconda metà degli anni (38); valutazione del rischio stress (38); il punto di vista (37); l'esistenza di un (37); mercato del lavoro e (37); miglioramento delle condizioni di (37); salute e sicurezza dei (37); valutazione del rischio e (37); il gruppo di lavoro (36); l'esistenza di una (36); ministero degli affari esteri (35); parte del medico competente (34); punto di vista del (34); sanitaria e di rischio (34); valutazione dei rischi e (34); vita e di lavoro (35); cartella sanitaria e di (35); salute e della sicurezza (70); tutta una serie di (42); misure di prevenzione e (34);

- b. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione PER:

rischi per la salute (79); rischio per la salute (44); i rischi per la (42); una volta per tutte (35);

- c. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione SU:

salute e sicurezza sul (78); gli infortuni sul lavoro (34);

- d. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione IN:

sicurezza nei luoghi di (60); prevenzione nei luoghi di (55); salute nei luoghi di (34); comunità italiana in Germania (47);

- e. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione A:

idoneità alla mansione specifica (36);

- f. Sintagma nominale con frammento di una clausola preposizionale che comincia con la preposizione TRA:

età compresa tra i (35);

- g. Sintagma nominale complesso

una vera e propria (142); un vero e proprio (97); E.S.M.I. (55); la sicurezza e la (49); la seconda guerra mondiale (77); sicurezza e la salute (46); la salute e la (42); salute e la sicurezza (38); anni cinquanta e sessanta (35); rischio stress lavoro correlato (35); la salute e sicurezza (48);

- h. Coordinatore + sintagma nominale

e sicurezza dei lavoratori (34); e sicurezza sul lavoro (77);

- i. Sintagma nominale con complemento frammentato

il fatto che la (47);

- l. Nome + proposizione relativa frammentata

il modo in cui (55); momento in cui la (45); momento in cui si (36);

- m. *Che* + sintagma nominale

che la maggior parte (49); che il medico competente (39).

### *III.3.2.2 Lexical bundles che sono clausole preposizionali o frammenti di sintagmi preposizionali*

#### a. La preposizione DI:

di medicina del lavoro (158); del datore di lavoro (119); di volta in volta (99); della medicina del lavoro (94); di salute e sicurezza (94); della seconda guerra mondiale (93); di prevenzione e protezione (92); del medico del lavoro (90); di valutazione del rischio (88); di valutazione dei rischi (84); della salute e della (82); del mercato del lavoro (74); di promozione della salute (70); della salute e sicurezza (63); del giudizio di idoneità (60); di una serie di (58); di fronte a un (57); dei fattori di rischio (55); dello stress lavoro correlato (55); di tutela della salute (55); di un gruppo di (53); del lavoro e della (46); di vita e di (46); degli italiani all'estero (43); dei datori di lavoro (43); della salute dei lavoratori (43); dello stato di salute (43); di lavoro e di (42); della valutazione dei rischi (40); degli ambienti di lavoro (39); dell'organizzazione del lavoro (39); di un sistema di (39); dei medici del lavoro (38); della valutazione del rischio (36); di idoneità alla mansione (36); del servizio di prevenzione (34); delle condizioni di lavoro (34); di età compresa tra (45);

#### b. La preposizione IN:

nei luoghi di lavoro (216); nel corso degli anni (96); in medicina del lavoro (93); nella maggior parte dei (98); negli ambienti di lavoro (75); in entrambi i casi (56); in un certo senso (55); in materia di salute (53); in questo senso la (53); in questo caso il (47); in tutto il mondo (45); nel mondo del lavoro (43); negli Stati uniti e (41); nella seconda metà del (39); in questo caso la (37); in una sorta di (37); nel mercato del lavoro (37); in questo modo si (34); nei primi decenni del (34); nel corso del tempo (34); in grado di determinare (37);

nel momento in cui (175); nel caso in cui (60); nei casi in cui (35);

#### c. La preposizione DA:

dal punto di vista (460); da un punto di (151); da questo punto di (92); da parte di un (54); da un lato e (51); da parte dei lavoratori (43); dalla fine degli anni (37); dal datore di lavoro (36); da parte del medico (53); da un lato la (34);

#### d. La preposizione PER:

per la prima volta (171); per la salute e (81); per la maggior parte (66); per la sicurezza e (50); per la valutazione del (45); per un totale di (45); per il fatto che (35); per la formazione continua (35);

#### e. La preposizione A:

alla fine degli anni (77); all'interno di un (68); al datore di lavoro (64); al di là della (54); all'interno di una (54); al di là delle (51); al di là di (51); al di là del (50); al di

fuori del (47); all'inizio degli anni (45); alla valutazione dei rischi (39); al di fuori di (38); alla valutazione del rischio (38); a fronte di una (37); al fumo di sigaretta (35);

f. La preposizione SU:

sul luogo di lavoro (56); sul mercato del lavoro (45); su un campione di (42); sulla base di una (39); sul posto di lavoro (36); sui luoghi di lavoro (34);

g. La preposizione CON:

con lo scopo di (53); con l'intento di (45);

h. Marcatore del discorso + frammento del sintagma preposizionale

anche in questo caso (144);

i. Congiunzione + frammento del sintagma preposizionale

se da un lato (68); e più in generale (47); e allo stesso tempo (42); e al tempo stesso (37);

l. Frammenti di sintagma preposizione con *che*

che a sua volta (49).

### **III.3.2.3 Lexical bundles verbali**

a. *Lexical bundles* con il verbo alla forma impersonale

si tratta di un (179); si tratta di una (106); non si tratta di (41); si può dire che (37);

b. Essere + frammento nominale o preposizionale

è stato quello di (35); è in grado di (118); sono in grado di (77); essere in grado di (44); è il caso di (37); è la necessità di (37);

c. Frammento passivo

come si è visto (52); come si è detto (37);

d. Preposizione + sintagma verbale frammentato

per quanto riguarda il (124); per quanto riguarda l' (81); per quanto riguarda le (74); per quanto riguarda i (56); per quanto riguarda gli (48); per quel che riguarda (45); per quanto riguarda la (155); a partire dagli anni (100); a partire dalla fine (54); a che fare con (100);

e. Congiunzione + sintagma verbale frammentato

se è vero che (38);

f. Infinito + sintagma verbale frammentato

fare i conti con (40).

### III.3.2.4 *Lexical bundles che sono congiunzioni*

nella misura in cui (89).

Dallo spoglio è emerso che i *lexical bundles* preposizionali e nominali coprono l'87,2% di tutte le sequenze di quattro parole identificate – i *lexical bundles* con struttura preposizionale sono stati individuati nel 49,6% dei casi, quelli con struttura nominale nel 37,7% di tutte le espressioni di quattro parole. La forma verbale è stata identificata nel 12,3% di tutte le sequenze.

Tabella 26: La struttura dei *lexical bundles* di 4 parole nella prosa accademica

<b><i>Lexical bundles</i> di 4 parole</b>	<b><i>Bundles</i> diversi</b>
<b>Preposizionali</b>	109
<b>Nominali</b>	83
<b>Verbali</b>	27
<b>Congiunzionali</b>	1
<b>Totale</b>	<b>220</b>

Interessante si rivela un'altra differenziazione tra le discipline, con il *corpus* della medicina del lavoro che adopera i *bundles* passivi – come ad esempio *è stato dimostrato che, sono stati sottoposti a, è stato utilizzato il* – in maniera più significativa rispetto alle altre discipline. In tutti i casi riscontrati il passivo viene usato per stabilire una relazione di priorità (Balboni, 2000), ponendo al centro d'interesse l'evento, l'azione o l'argomento – una caratteristica della microlingua.

Nell'analisi sono emerse altre interessanti divergenze tra le discipline scientifiche. La prima divergenza si mostra nella numerosità dei *clusters* nella scienza applicata, la quale si presenta del 55% più alta che nell'ambito di storia e del 64% più alta che in

quello di filosofia. Di conseguenza, la più alta numerosità dei *lexical bundles* si rispecchia in una percentuale più alta di sequenze di quattro parole nel *corpus* analizzato. Da contro, nel *corpus* di filosofia è stato individuato il numero minimo di *lexical bundles* diversi, la minima frequenza di occorrenza e la minima percentuale di tutte le parole del *corpus* che si presentano nella forma dei *lexical bundle*. La storia e la filosofia, due discipline umanistiche, presentano una percentuale congruente rispetto alla proporzione della parole dell'intero *corpus* in forma di *lexical bundles*. La tabella 27 rispecchia questa differenza.

Tabella 27: La frequenza dei *bundles* nelle tre discipline

<b>Disciplina scientifica</b>	<b>Bundles diversi</b>	<b>Numero di occorrenze</b>	<b>% delle parole in bundles</b>
<b>Medicina del lavoro</b>	890	18.454	7%
<b>Storia</b>	399	9.182	2,3%
<b>Filosofia</b>	328	4.089	2,1%

La vasta scelta delle formule precostituite potrebbe essere spiegata in base alla natura della disciplina stessa: la medicina è una scienza che si fonda su tassonomie (Serianni, 2005) ed ha per questo bisogno di una terminologia più ampia e diversificata. Poiché la medicina trasmette contenuti tecnici, essa necessita di un vasto e differenziato uso di espressioni precostituite. La densità dei *lexical bundles* potrebbe essere spiegata in termini di rappresentazione visuale e grafica della disciplina, del suo denso contenuto tecnico (Hyland, 2008a). La concezione scientifica del ragionamento istituzionalizzato, legato all'osservazione scientifica, e di conseguenza la presentazione dei dati in forma visuale e grafica – che sembra imporre dei termini specifici per riferirsi a quel contesto – indica un impiego più vasto dei *bundles*. Pertanto, la maggioranza dei *bundles* trovati nel *corpus* di medicina del lavoro non è stata individuata nelle due discipline umanistiche, poiché legata al contenuto della disciplina stessa.

Da contro, il ragionamento istituzionalizzato delle discipline umanistiche, quali la storia e la filosofia, è più legato all'interpretazione e alla manipolazione o alla discussione critica dei dati, confidando di più in un modo di presentazione sofisticato e individuale. Per questo motivo, uno storico o un filosofo non confida nell'uso delle formule precostituite, ma piuttosto usa argomenti molteplici e variegati, per poter

diversificare il proprio discorso in modo che, seppur esprimendo uno stesso concetto, il linguaggio possa essere abbellito e ornato con scelte stilistiche variegiate.

Cerchiamo a questo punto di interpretare i dati fin qui presentati.

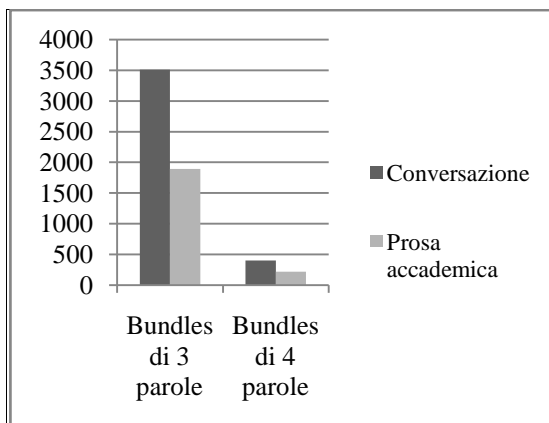
### III.3.3 Comparazione del parlato e dello scritto

Come questa ricerca ha illustrato, la pervasività dei *lexical bundles* può essere misurata – tramite la numerosità dei diversi *lexical bundles*, la frequenza d'occorrenza dei *bundles* e la proporzione delle parole di un *corpus* rappresentate in forma di *lexical bundles*.

Sommando i *bundles* di tre e quattro segmenti in sequenza del *corpus* del parlato analizzato, si nota che ci sono quasi 4000 *lexical bundles* diversi, con una frequenza di 39.000 volte su circa 500.000 parole, ovvero 78.000 occorrenza su un milione di parole. Seguendo la regola di normalizzazione, i *bundles* di tre parole occorrono 75.000 volte su un milione di parole, i *bundles* di quattro segmenti hanno una frequenza di 6.000 volte su un milione di parole.

Già nella sezione precedente sono stati evidenziati i dati quantitativi circa la numerosità e frequenza di occorrenza dei *lexical bundles* nella scrittura accademica. Dai dati si evince come la frequenza di occorrenza di *lexical bundles* di 4 parole in tutti e due i registri – conversazione e discorso accademico – è alta. Nella conversazione risulta tuttavia più alta.

Grafico 6 – Numero di *lexical bundles* nella conversazione e nella prosa accademica





Per rendere possibile il confronto tra lo scritto e il parlato si deve considerare anche la frequenza dei singoli *lexical bundles*. Si può constatare che il linguaggio parlato presenta solo alcune sequenze con una frequenza abbastanza alta – cosa riscontrata già in lingua inglese (Conrad, Biber, 2004). Da contro, nella scrittura accademica riscontriamo nelle prime cinquanta posizioni *bundles* che sono molto frequenti.

A mo' di esempio indichiamo il *bundle* di tre parole *non lo so*, che presenta un'occorrenza di 356 su un milione di parole, il più frequente *lexical bundle* di quattro parole – *va bene va bene* – ha una frequenza di 92 volte su un milione di parole. D'altro canto, le sequenze di quattro parole più frequenti nella prosa accademica occorrono da 200 a 400 volte su un milione di parole.

Ulteriori dati quantitativi riguardano la proporzione delle parole presenti in forma di *lexical bundles* di 3 e 4 parole in sequenza. La proporzione risulta più alta nella conversazione – le sequenze di 3 parole coprono una percentuale del 24%, quelle di 4 parole in sequenza una percentuale del 2,8%. La percentuale nella scrittura accademica si rivela alquanto più bassa – i *lexical bundles* di 3 parole presentano una copertura del 12% di tutte le parole, quelli di 4 parole in successione una copertura dell'1,6% – osservabile nel grafico 7 e 8.

Grafico 7 – Percentuale delle parole in *lexical bundles* in conversazione

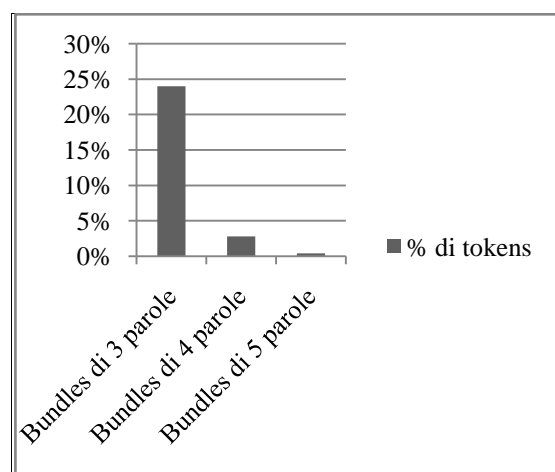
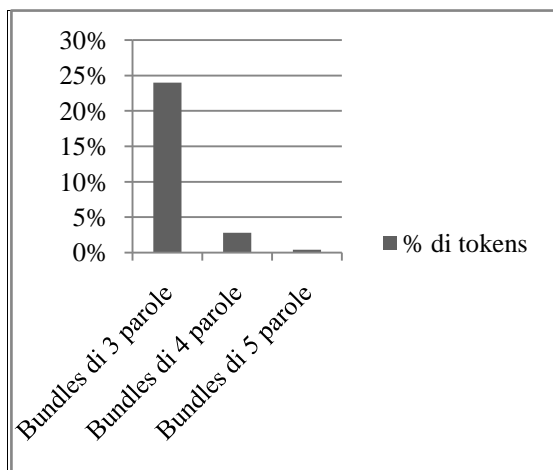


Grafico 8 – Percentuale delle parole in *lexical bundles* nella scrittura accademica



La necessità di un confronto tra il parlato e lo scritto non deve portare a dimenticare l'arbitrarietà del criterio imposto all'inizio della presente ricerca. Il limite dell'occorrenza dei *lexical bundles* stabilito su 10 volte su un milione di parole impone un'alta percentuale delle parole precostituite di 3 e 4 parole in sequenza. Da questa prospettiva, il loro uso ricorrente non sembra essere un puro caso (Conrad, Biber, 2004).

Nell'effettuare il confronto si è inoltre partiti dalla struttura grammaticale dei *lexical bundles*. Lo studio ha mostrato che la generalità dei *clusters* nel discorso accademico dispone di una struttura frammentata, con una tendenza verso i *bundles* nominali e preposizionali.

Come riportato da Biber ed altri (1999), solo il 15% dei *lexical bundles* nella conversazione può essere guardato come composto da frasi o clausole complete, mentre il resto è composto da frasi frammentate o clausole con nuovi frammenti incorporati.

Dallo spoglio svolto sul *corpus* LIP è emerso che il 18,4% delle espressioni di quattro parole è composto da sintagmi completi, mentre nella prosa accademica il 35% di tutti i *lexical bundles* rappresenta sintagmi completi. Questo dato illustra la divergenza tra il discorso accademico in lingua inglese e in lingua italiana – la scrittura scientifica inglese dispone del 5% di clausole complete. La diversa proporzione potrebbe essere attribuita alla diversa tipologia linguistica.

Un altro insieme di elementi interessanti per individuare la differenza tra il parlato e lo scritto è la struttura dei *lexical bundles* di quattro parole in tutti e due i *corpora* analizzati.

Il 62% di tutte le sequenze di quattro parole identificate nel *corpus* del parlato include un elemento verbale, da contro la struttura verbale è presente solo in una

proporzione equivalente al 12,3 % nel *corpus* specialistico. Piuttosto, l'87% di tutte le espressioni di quattro parole nella scrittura accademica risulta avere una struttura nominale o preposizionale.

I risultati possono essere comparati con quelli ottenuti nella ricerca sulla lingua inglese, mostrando un carattere concorde.

Ci sembra importante mettere in luce un'ulteriore divergenza tra il parlato e lo scritto: nel *corpus* del parlato è stato riscontrato un gruppo di *lexical bundles* avente struttura clausale. Questo gruppo copre il 29,5% di tutti i *bundles* di quattro parole nel *corpus* del parlato, da contro, nella scrittura accademica la stessa forma strutturale presenta solo il 7,3% delle forme dei *bundles* nel *corpus*. Il carattere clausale della conversazione è una caratteristica tipica del parlato inglese – paragonando i risultati tra il parlato e lo scritto in lingua italiana questa struttura viene preferita nella lingua parlata.

Tracy-Ventura ed altri (2007) hanno paragonato i modelli di utilizzo nelle tipologie e nelle distribuzioni per la lingua spagnola. Lo studio svolto sulla prosa accademica e sulle interviste sociolinguistiche ha mostrato la prevalenza dei *bundles* nominali e preposizionali nella prosa accademica, mentre il parlato dispone di più *lexical bundles* verbali. La lingua italiana si presenta da questa prospettiva vicina alla lingua spagnola, mostrando una tendenza verso i *bundles* preposizionali e nominali nel discorso scientifico.

Inoltre, le ricerche di Tracy-Ventura ed altri (2007) hanno mostrato che lo spagnolo accademico – paragonato alle interviste sociolinguistiche – dispone di una varietà più alta di *lexical bundles*. Da contro, gli studi di Butler (1997, in Tracy-Ventura, 2007) hanno rilevato un più alto numero di sequenze ripetute di quattro parole nei registri parlati. Gli studi svolti sulla lingua inglese (Biber ed altri, 1999, 2006) indicano una maggiore densità dei *lexical bundles* nel registro parlato.

I risultati delle ricerche sulla lingua italiana sono congruenti con i dati sulla lingua inglese – i *lexical bundles* nel registro parlato sono più numerosi. Quest'osservazione mette in luce la variazione della quantità dei *lexical bundles* attraverso i registri – un altro fattore che non deve essere dimenticato è la loro variazione quantitativa attraverso le discipline scientifiche. Come hanno rivelato le ricerche parziali sulla lingua italiana,

la numerosità di *lexical bundles* varia attraverso le discipline scientifiche, mostrando una maggiore densità dei *lexical bundles* nella scienza applicata.

Sulla base di quanto premesso possiamo concludere che i risultati di questa parte della ricerca coincidono con le ipotesi formulate all'inizio dello studio. Anche se i *lexical bundles* nella maggioranza dei casi presentano delle strutture grammaticalmente incomplete, possono essere classificati strutturalmente, confermando le ipotesi tradizionali sul carattere prevalentemente verbale del parlato e sul carattere nominale dello scritto. Considerando questo fatto, i *lexical bundles* sembrano essere dei blocchi importanti nella costruzione del discorso (Biber ed altri, 1999).

Questa parte della ricerca ha voluto mostrare la diversa struttura dei *lexical bundles* in registri diversi, senza addentrarci nelle funzioni discorsive delle sequenze di quattro parole, che saranno approfondite nella sezione seguente e riguardano solo il discorso accademico.

### **III.3.4 Variazione funzionale dei *lexical bundles* nel discorso accademico – microlingua della storia, della filosofia e della medicina del lavoro**

Lo studio si concentra sulle somiglianze e sulle divergenze dei *pattern* lessicali tra le discipline scientifiche in oggetto. L'esposizione dei *lexical bundles* si basa da un lato sui *pattern* che delineano le peculiarità di diversi ambiti scientifici, e dall'altro si addentra in una discussione particolareggiata delle somiglianze e divergenze funzionali di tre ambiti scientifici.

Lo scopo principale è analizzare la misura in cui i *lexical bundles* di 4 parole contribuiscono alla creazione del discorso accademico.

Il presente studio mostra come queste combinazioni di parole siano meccanismi discorsivi con delle funzioni riconoscibili, anche se il maggior numero di esse non rappresenta delle unità strutturali complete.

Dopo l'approccio numerico o automatico al *corpus*, si passa ora a un'analisi concentrata sulla funzione discorsiva dei *bundles* o sul contesto di occorrenza, un aspetto legato all'interpretazione umana. Analogamente a quanto era già stato fatto per

l'inglese, i segmenti ripetuti sono stati analizzati nel contesto d'uso in cui compaiono nel *corpus*.

Come vedremo nella sezione seguente, i dati statistici forniti dai *corpora* presi in esame vengono considerati alla luce di un'analisi funzionale delle occorrenze dei *lexical bundles*, contribuendo a definire ancor meglio la loro importanza. La ricerca inizia con l'analisi delle concordanze, volta ad esaminare le funzioni discorsive dei *lexical bundles*. Le concordanze sono uno strumento di analisi che integra l'approccio quantitativo con l'interpretazione funzionale di tali segmenti.

Ogni *bundle* è stato esaminato nel suo contesto originale, adoperando la funzione *concordance listing* del programma WordSmith Tools. Il programma permette la visualizzazione dei *bundles* all'interno del loro contesto d'uso, in base al quale è possibile definire l'uso linguistico. Ciascuna sequenza è stata, quindi, esaminata nel suo contesto originale e classificata nelle categorie. In alcuni casi, la funzione discorsiva dei *lexical bundles* risulta chiara anche fuori dal contesto – come nel caso dei *clusters* che esprimono una collocazione temporale (*al momento della partenza, fine degli anni Settanta*). Tuttavia, la funzione discorsiva della maggioranza delle sequenze è stata definita in base al loro contesto originale.

Alcuni *bundles* presentano una funzione molteplice, per esempio *al di fuori di, nel momento in cui*. In questi casi si è optato per il gruppo che presentava la maggiore frequenza d'uso – per esempio nel caso del *bundle – al di fuori del –*, che può avere sia la funzione di un delimitatore del discorso che di una collocazione locale, è stato scelto il sottogruppo dei segnali di cornice. Il *bundle*, infatti, appare innanzitutto nella funzione di un delimitatore del discorso.

I *lexical bundles* sono stati classificati in base alle funzioni discorsive. Come già precedentemente accennato, la tassonomia funzionale proposta per la lingua inglese è stata adottata anche per la lingua italiana. La scelta è stata condizionata dall'ipotesi proposta da Conrad e Biber (2004) secondo la quale la tassonomia elaborata dal gruppo di ricerca americano (Biber ed altri, 1999) può essere adoperata sia per diversi registri linguistici che per diverse lingue.

Gli studi svolti per la lingua inglese hanno mostrato che i *lexical bundles* sono estremamente comuni. Nella comunicazione, sia parlata che scritta, essi vengono

adoperati come sequenze non analizzate e svolgono delle funzioni discorsive ben identificabili (Biber, Conrad, 2004).

L'estrazione automatica e la successiva analisi dei *clusters* sostiene le osservazioni elaborate per la lingua inglese (Biber ed altri, 1999, 2004; Biber, 2006; Cortes, 2002, 2004; Hyland, 2008a, 2008b): i *bundles* identificati non sono accidentali, rappresentano, piuttosto, sequenze ad alta frequenza che richiedono una spiegazione. Anche nella lingua italiana le unità linguistiche identificate rappresentano dei meccanismi linguistici che svolgono delle funzioni linguistiche fondamentali e caratteristiche per ciascuna disciplina scientifica analizzata. Ogni ambito scientifico adopera un insieme dei *bundles* che contraddistingue la disciplina in questione. Le funzioni discorsive dei *pattern* analizzati dimostrano che i modelli ad alta frequenza rispecchiano lo status formulaico del linguaggio (Biber, 2006).

Proponiamo di seguito la classificazione funzionale dei *lexical bundles* di tre discipline scientifiche analizzate. Le prime due sezioni sono dedicate ai *lexical bundles* di due discipline umanistiche, la storia e la filosofia, l'ultima parte si concentra sull'analisi funzionale dei *bundles* della medicina del lavoro.

### **III.3.4.1 La funzione dei lexical bundles nel corpus di storia**

I *lexical bundles* emersi dall'analisi computazionale sono stati in un primo momento strutturati in categorie funzionali. Dopo la classificazione discorsiva si è passati alla presentazione dei risultati.

#### **III.3.4.1.1 La classificazione dei bundles in categorie funzionali**

I *bundles* di storia sono stati raggruppati in tre categorie principali: *bundles* orientati alla ricerca, al testo e al partecipante. I *clusters* sono stati ordinati in senso decrescente in base alla loro frequenza d'occorrenza.

I *bundles* orientati alla ricerca aiutano lo scrivente ad organizzare la propria attività e la propria esperienza del mondo reale e includono: a) *bundles* che esprimono

una collocazione temporale – *la seconda guerra mondiale, della seconda guerra mondiale, alla fine degli anni*; b) *bundles* che si riferiscono all'argomento principale della ricerca o agli argomenti inerenti alla ricerca, come ad esempio *la storia dell'immigrazione, giovani di origine italiana, i figli di immigrati*; c) espressioni che indicano una collocazione locale, come ad esempio *nel paese di origine, in tutto il mondo*; c) *clusters* che delimitano l'argomento della ricerca – *del muro di Berlino, dei diritti dell'uomo*; d) quantificatori – *bundles* che nella loro forma incorporano un determinante che esprime la quantità – *nella maggior parte dei, il più delle volte*.

Il secondo gruppo maggiormente rappresentato è quello dei *lexical bundles* concernenti l'organizzazione del testo a livello micro e macro. Questa categoria incorpora: a) segnali di cornice – questi *bundles* accentuano connessioni, specificano casi, indicano limitazioni o condizioni per l'interpretazione del discorso successivo o precedente, come ad esempio *dal punto di vista, in base al quale, anche in questo caso*; mentre al microlivello testuale, ovvero riferiti alle connessioni interne ai paragrafi o tra due frasi successive, queste sequenze stabiliscono collegamenti addizionali di tipo causativo, contrastivo o consecutivo tra due elementi o ne enfatizzano alcuni (*l'idea di un', e il fatto che*); b) segnali di passaggio discorsivo – collegamenti additivi o contrastivi tra gli elementi del testo (*se da un lato, in primo luogo la*); c) *lexical bundles* che indicano un risultato – *il risultato di una* –; d) *lexical bundles* che si riferiscono alla struttura del testo – *nell'ambito della ricerca*.

L'ultima categoria riguarda i *bundles* orientati al partecipante, sia allo scrivente sia al lettore. Il primo sottogruppo incorpora le sequenze di quattro parole che si concentrano sullo scrivente, esprimendo la sua opinione o quella della comunità scientifica – *in questo senso la, si è deciso di*; i *bundles* del secondo sottogruppo, invece, cercano di coinvolgere il lettore, come nel caso di *come si è visto, come si è detto*.

#### ***III.3.4.1.1.1 I lexical bundles orientati alla ricerca – aiutano lo scrivente a strutturare la propria attività e la propria esperienza del mondo reale***

a. Collocazione temporale:

della seconda guerra mondiale (86); della prima guerra mondiale (27); di un secolo fa (13); di volta in volta (44); nel corso del novecento (14); nel corso dell'Ottocento (18); nel corso del tempo (31); all'inizio del novecento (18); alla fine del secolo (19); alla fine dell'Ottocento (26); al momento della partenza (16); alla prima guerra mondiale (12); fino a quel momento (26); per tutti gli anni (14); per la prima volta (102); tra le due guerre (26); tra gli anni cinquanta (15); dopo la seconda guerra (28); durante la seconda guerra (21); la seconda guerra mondiale (76); anni sessanta e settanta (30); un anno e mezzo (14); anni cinquanta e sessanta (35); anni venti e trenta (21); la prima guerra mondiale (18); l'entrata in guerra (14); meta del XIX secolo (31); fine degli anni sessanta (36); fine della seconda guerra (22); fine degli anni settanta (21); fine del XIX secolo (20); seconda meta del secolo (20); seconda meta dell'Ottocento (20); meta degli anni sessanta (17); fine degli anni cinquanta (16); meta degli anni ottanta (16); meta degli anni settanta (19); primi anni del novecento (18); primi decenni del novecento (18); inizio degli anni sessanta (17); fine degli anni venti (14); prima meta del novecento (14); anni venti del Novecento (13); inizio degli anni novanta (13); inizio degli anni ottanta (13); inizio del XX secolo (13); meta del VI secolo (13); meta del XII secolo (13); la fine della guerra (14); prima della seconda guerra (14); nel corso degli anni (74); della prima meta del (13); degli anni settanta e (23); della seconda meta del (20); degli anni cinquanta e (18); nella prima meta del (26); nella seconda meta del (35); nei primi decenni del (30); negli anni cinquanta e (18); negli anni sessanta e (16); nel corso degli ultimi (16); negli anni venti e (21); nella seconda meta degli (21); in un momento in (17); dai primi anni del (15); dalla fine degli anni (31); dalla meta degli anni (18); alla fine degli anni (62); alla meta degli anni (18); agli inizi degli anni (21); all'inizio degli anni (62); tra la fine del (27); dopo la caduta del (15); dopo la morte del (13); dopo la morte di (13); durante il regno di (14); verso la fine degli (20); negli anni in cui (13); la fine degli anni (36); seconda meta degli anni (34); gli anni cinquanta e (20); prima guerra mondiale e (14); gli anni sessanta e (14); partire dalla fine degli (25); partire dagli anni settanta (19); partire dagli anni Novanta (13); a partire dagli anni (88); a partire dalla fine (46); a partire dalla meta (22); a partire dai primi (14); e l'inizio del (13); alla prima guerra mondiale (1); nel momento in cui (74); momento in cui la (28); un momento in cui (20); momento in cui si (15);

b. Argomento principale e argomenti inerenti alla ricerca:

la storia dell'emigrazione (26); i figli di immigrati (24); storia dell'immigrazione italiana (23); la storia delle migrazioni (17); americani di origine italiana (13); commissariato generale dell'emigrazione (13); discendenti degli immigrati italiani (13); i figli degli immigrati (13); giovani di origine italiana (43); la caduta del muro (23); legame con l'Italia (13); il rinnovo del parlamento (18); il processo di integrazione (13); il senso di appartenenza (13); le condizioni di vita (14); il diritto di voto (21); diritti di cittadinanza politica (15); italiani all'estero e (20); il tasso di disoccupazione (15); forme di partecipazione politica (16); le seconde generazioni di (13); caduta del muro di (20); voto degli italiani all' (19); la repubblica federale tedesca (19); seconda e terza generazione (17), prima e seconda generazione (14); voto attivo e passivo (15); vecchie e nuove migrazioni (13); la presenza italiana in (16); la comunità italiana in (17); e i loro discendenti (14); un ritorno alle radici (14); ministero degli affari esteri (35); il ministero del lavoro (16); il ministero degli esteri (17); gli uffici del lavoro (16); uffici



provinciali del lavoro (16); i datori di lavoro (13); il permesso di soggiorno (13); un titolo di studio (18); la qualità della vita (16); italiani e in particolare (15); gli italiani e i (18);

c. Collocazione locale:

nel paese d'origine (26); in tutto il mondo (29); nel paese di arrivo (14); all'interno della città (13); fra i due paesi (17); verso gli Stati Uniti (16); il paese d'origine (14); il paese di provenienza (14); Stati Uniti d'America (14); la terra d'origine (13); all'interno della società (15); del paese d'origine (20); all'interno della famiglia (21); all'interno della comunità (34); all'interno dell'unione (23); verso gli Stati Uniti (1); negli Stati Uniti e (37); in Italia e in (13); negli Stati Uniti in (13); negli Stati Uniti si (13); a New York e (15); tra l'Italia e (15); tra i due paesi (24); degli italiani all'estero (67); degli italiani a Berlino (24); degli italiani in America (23); degli italiani in Germania (30); degli italiani nel mondo (19); degli italiani in Francia (17); del paese di accoglienza (14); interno dell'unione europea (23); all'interno di questo (16); comunità italiana in Germania (47); immigrazione italiana in Germania (27); italiani negli Stati Uniti (31); immigrati italiani in Germania (18); italiano negli Stati Uniti (15); italiani residenti all'estero (30); gli italiani all'estero (27); comunità italiane all'estero (16); italiana negli Stati Uniti (24); nel mondo del lavoro (25); nel mercato del lavoro (36); nel settore della ristorazione (17); al paese d'origine (18); per gli Stati Uniti (14);

d. Delimitatori dell'argomento:

del muro di Berlino (20); dell'immigrazione di massa (14); di Anna Maria Martellone (14); dei diritti dell'uomo (13); della seconda generazione di (13); di prigionieri di guerra (13); della comunità italiana in (33); della comunità italiana di (22); dell'immigrazione italiana in (32); degli immigrati italiani nel (15); degli immigrati italiani in (23); da parte degli immigrati (14); del diritto di voto (15); delle condizioni di vita (14); del mercato del lavoro (62); di origine italiana che (15); di origine italiana è (24); a causa della guerra (15); al mercato del lavoro (18); sul mercato del lavoro (45); i cittadini dell'unione (19); mercato del lavoro e (32); nei confronti degli italiani (18); in cerca di lavoro (14); da parte del governo (14); da parte delle autorità (14); da parte delle istituzioni (14); in via di sviluppo (21); il fatto di essere (13);

f. Quantificatori:

nella maggior parte dei (62); in tutta la sua (15); per la maggior parte (51); per cento della popolazione (30); per cento degli italiani (15); maggior parte dei casi (56); maggior parte di loro (17); il più delle volte (27); la maggior parte degli (41); la maggior parte delle (39); la maggior parte dei (36); la maggior parte di (36); il maggior numero di (15); un gran numero di (14); un certo numero di (16).

*III.3.4.1.1.2 I lexical bundles orientati al testo – riguardanti l'organizzazione del testo e il suo significato in quanto portatore dell'argomento già precedentemente discusso*

**Al macrolivello**

**a. Segnali di cornice – accentuano connessioni, specificano casi, indicano limitazioni condizioni per l'interpretazione del discorso successivo o quello in precedenza:**

in base al quale (16); in presenza di un (13); sulla base di una (16); con l'eccezione di (20); che in alcuni casi (20); che in molti casi (13); come nel caso della (15); come nel caso di (15); anche in questo caso (67); nel caso in cui (16); l'assenza di un (13); di fronte a un (45); di fronte a una (17); punto di vista dell' (17); punto di vista della (19); punto di vista la (15); uno degli aspetti più (14); un punto di vista (51); dal punto di vista (203); da questo punto di (60); da un punto di (39); in questo caso è (19); in questo caso il (20); in questo caso l' (20); in questo caso la (18); in questo caso di (13); anche dal punto di (13); questo punto di vista (62); in entrambi i casi (35); in una sorta di (23); per quanto riguarda l' (52); per quanto riguarda la (40); per quanto riguarda le (22); per quanto riguarda i (17); per quanto riguarda il (36); per ciò che riguarda (20); per quel che riguarda (26); per quanto concerne la (18); al di la del (29); al di la di (26); al di la delle (35); al di la dell' (13); al di la della (26); al di la dei (18); al di fuori della (21); al di fuori del (16); al di fuori dell' (9); al di fuori di (14); da una serie di (17); di una serie di (44); in una serie di (14); nel momento in cui (74); momento in cui la (28); un momento in cui (20); momento in cui si (15); al di fuori della (5); al di fuori del (8); al di fuori dell' (5); tutta una serie di (28); con una serie di (18); a fronte di una (16); a fronte di un (13); dal momento che la (20); a causa di una (16); a causa della loro (15); con lo scopo di (13); di lì a poco (17); a tutti gli effetti (27); ben al di la (16); nella misura in cui (18); in modo tale che (13); in modo tale da (13); come una sorta di (13); da parte di un (29); da parte di chi (15);

**b. Segnali di passaggio discorsivo – collegamenti tra gli elementi del testo:**

in un secondo tempo (13); se da un lato (34); da una parte e (16); in primo luogo la (14); da un lato la (14); da un lato e (30);

**c. Segnali di risultato:** il risultato di una (14); dati in nostro possesso (16);

**d. Segnali strutturali – riflessione sul testo:** nell'ambito della ricerca (15);

**Al microlivello**

**Connessioni interne ai paragrafi o tra due frasi successive; stabiliscono collegamenti aggiuntivi di tipo causativo, contrastivo o consecutivo tra due elementi; enfatizzano elementi adiacenti:**

l'idea di una (23); l'idea di un' (17); la necessità di una (15); la possibilità di un (14); la presenza di un (21); di un vero e (17); di una vera e (14); anni in cui la (14); un periodo in cui (13); la creazione di una (20); la creazione di un (19); la costruzione di una (17); per il rinnovo del (22); per la creazione di (13); per la realizzazione di (13); alla ricerca di una (15); l'esistenza di un (23); l'esistenza di una (33); i legami con la (14); con il compito di (15); e la capacità di (14); e le condizioni di (13); e i processi di (15); di origine e di (20); del lavoro e della (27); di lavoro e di (20); di vita e di (18); un gruppo di (27); di un sistema di (17); di un processo di (15); ruolo di primo (19); ruolo di primo piano (21); il luogo in cui (14); il fatto che le (19); il fatto che la (18); il fatto che l' (13); con l'intento di (33); con l'intenzione di (18); nel corso di una (16); con il paese di (13); con la terra di (13); per il fatto che (17); e in modo particolare (16); su un totale di (14); un vero e proprio (66); una vera e propria (86); è il fatto che (20); è un caso che (20); è il caso di (26); il modo in cui (23); il motivo per cui (17); non è un caso (23); a giocare un ruolo (13); una volta per tutte (13); che a sua volta (24); nel tempo e nello (18); i conti con la (17); fare i conti con (25); a che fare con (29); a fare i conti (15); è in grado di (25); sono in grado di (16); paesi in via di (16); vita e di lavoro (13); tempo e nello spazio (13); di cui si è (15); in cui si sono (15).

#### *III.3.4.1.1.3 I lexical bundles orientati al partecipante- si concentrano sullo scrivente e sul lettore del testo*

##### **a. Opinione dell'autore o della comunità:**

in questo senso la (26); in questo modo si (19); in tal senso la (13); in un certo senso (27); a differenza di quanto (15); e per certi versi (17); si è deciso di (28); è possibile affermare che (14); se è vero che (21); si tratta di un (105); si tratta di una (54); si trattava di un (27); si trattava di una (23); si può dire che (25); vale a dire la (13);

**b. Coinvolgimento del lettore nel discorso:** come si è visto (37); come si è detto (28).

#### III.3.4.1.2 Risultati emersi dalla classificazione funzionale

La stragrande maggioranza dei *lexical bundles* identificati nel *corpus* di storia rispecchia una connessione intrinseca all'essenza della disciplina stessa.

Il termine storia deriva dal greco e significa nella sua accezione primaria "ricerca, indagine, cognizione<sup>42</sup>". La storia è la disciplina umanistica che ha per oggetto lo studio del passato, configurato in un'esposizione sistematica di fatti e avvenimenti umani. I fatti vengono interpretati alla luce di un'indagine critica, basata sulla ricerca sui fatti

<sup>42</sup> Definizione tratta dall'Enciclopedia Treccani della versione online – [www.treccani.it](http://www.treccani.it).

passati, sulla loro veridicità, sulle connessioni di causa ed effetto che li determinano. La narrazione storica si fonda sull'oggettività e sulla razionalità dello storico, sul principio dell'osservazione analitica, critica, speculativa ed etica del passato. La storia può essere interpretata come una disciplina umanistica o accademica dello studio dell'uomo e della condizione umana.

La narrazione storica è caratterizzata da un lato dalle osservazioni del mondo reale e dalla loro presentazione. Questo tipo di presentazione induce ad adoperare più *lexical bundles* orientati alla presentazione del periodo o dell'anno in cui gli eventi storici hanno avuto luogo, e dall'altro alla segnalazione del luogo dell'avvenimento storico. Tali sequenze rappresentano il 35,6% di tutti i *bundles* identificati.

A titolo d'esempio sono stati scelti i *lexical bundles* – *della Seconda guerra mondiale e per la prima volta* – che indicano una collocazione temporale, e il *bundle* – *in tutto il mondo* – che denota una collocazione locale. Tutti e tre i *bundles* presentano la più grande frequenza d'uso, nell'intero *corpus*, rispettivamente 102, 86 e 29 volte.

Altreitalie, 2008

Ciò si rifletté sul problema degli internamenti e delle persecuzioni contro gli italiani all'epoca *della Seconda guerra mondiale*.

Altreitalie, 2006

Inoltre, *per la prima volta* nella storia dell'emigrazione gli italiani residenti all'estero, grazie alla nuova legge elettorale, hanno avuto la possibilità di eleggere propri rappresentanti al Parlamento italiano, essi stessi provenienti prevalentemente dal mondo dell'emigrazione.

Altreitalie, 2005

Si tratta peraltro di atteggiamenti riscontrabili nelle comunità italiane emigrate *in tutto il mondo*.

Inoltre, gli articoli di ricerca sono contrassegnati dai *lexical bundles* che rispondono alla domanda – *Chi/Che cosa/Perché?* –, riferendosi al tema della ricerca stessa. Tali *bundles* contengono nella propria struttura parole che rispecchiano gli argomenti trattati negli articoli scientifici. Sono stati identificati 72 *bundles*, che rappresentano il 18% di tutte le sequenze di quattro parole. Indichiamo di seguito le sequenze con l'occorrenza più alta.

Altreitalie, 2006

Questa relazione è articolata in tre punti che riflettono le diverse fasi di un percorso migratorio: la prima parte è dedicata alla partenza cioè, ricordando brevemente *la storia dell'emigrazione italiana in Argentina*, tenta di dare una definizione essenzialmente giuridica dello statuto di "italoargentino" e di illustrare le procedure necessarie per tornare in Italia.

*La caduta del muro* di Berlino e le radicali trasformazioni politiche nell'Europa dell'Est segnarono una svolta nella provenienza regionale, nella tipologia e nei profili professionali della nuova immigrazione.

Altreitalie, 2008

Gli sviluppi *del mercato del lavoro*, la deindustrializzazione, lo spostamento delle risorse verso il terziario e la flessibilità occupazionale inizialmente avevano fatto sperare in un aumento della mobilità all'interno dell'Unione Europea.

Nonostante presenti un numero abbastanza esiguo, ci pare rilevante menzionare il gruppo dei *lexical bundles* con funzione molteplice. A mo' di esempio si è optato per le espressioni più ricorrenti. Il *bundle – al di fuori del –* può indicare sia una collocazione locale che porre una cornice al discorso precedente; mentre al *bundle – nel momento in cui –* può essere attribuita sia la funzione temporale che la funzione delimitativa di un discorso.

Altreitalie, 2004

Vorrei poi chiedervi se è pertinente parlare di «petites italties» *al di fuori del Nord America e dell'Australia* e, nel caso, quando si è iniziato a parlarne in Francia. Esistono ancora quartieri etnici italiani nel Paese?

Colui il quale non può rappresentare un amico è *al di fuori del* tessuto sociale. Coloro i quali violano la fides, intesa come buona fides fondamento del traffico giuridico, sono dichiarati infami, cioè non in grado di postulare pro alios.

Altreitalie, 2005

Per Halbwachs (1985) la storia in generale inizia *nel momento in cui* la tradizione orale termina, *nel momento in cui* la memoria sociale si estingue, perché i detentori di tale memoria scompaiono.

Quanto a sapere perché oggi la formula ha tanto successo *nel momento in cui* l'emigrazione è un fenomeno compiuto e che le poche "petites italties" che esistono sono delle reliquie commerciali più o meno ricostituite, non possiamo che azzardare delle ipotesi.

Lo spoglio ha mostrato che i *lexical bundles* rivolti alla ricerca coprono una percentuale del 57,4% di tutti i *lexical bundles* – nell'ambito dei *bundles* orientati alla ricerca, il 41,9% sono *bundles* che esprimono una collocazione temporale. A questi

seguono le sequenze che si riferiscono agli argomenti trattati negli articoli scientifici, con una percentuale del 31,4%, e i segmenti che esprimono una collocazione locale, i quali coprono il 20%.

L'alto uso dei *bundles* che rispondono alle domande – *Quando?*, *Dove?*, *Chi/Che cosa/Perché?* – rivela da un lato il pensiero narrativo di uno storico, dall'altro va attribuito alla necessità dello scrivente di convincere il lettore e dirigerlo verso una certa interpretazione dei fatti.

Hyland (2005) sostiene che un esperto di una scienza umanistica deve lavorare di più per poter instaurare un rapporto con il lettore, poiché le sue osservazioni non si basano su una ricerca scientifica, ovvero su un'esplicazione dei dati. Questo fatto induce gli esperti di discipline umanistiche ad adoperare più costrutti linguistici rivolti all'organizzazione discorsiva e alla valutazione dei fatti. L'esposizione narrativa delle vicende storiche risulta in un discorso accademico in cui le idee vengono presentate in un modo interpretativo, fondato sull'interpretazione personale dello storico.

L'osservazione proposta si riflette nella seconda categoria maggiormente rappresentata, ovvero nei *lexical bundles* rivolti al testo. Il gruppo copre il 38,3% di tutti i *bundles* identificati.

La prevalenza dei *lexical bundles* orientati al testo ha la funzione di situare o porre gli argomenti discussi in una cornice, delimitandoli sia specificando i casi singoli sia evidenziando delle connessioni o specificando delle condizioni sotto le quali un argomento può essere accettato. Questi tipi di *bundles* hanno la funzione di elaborare, comparare e accentuare aspetti di un argomento precedentemente discusso.

Altreitalie, 2003

Solo dopo parecchio tempo la famiglia si sarebbe resa conto che le scatole contenevano un'abbondante fornitura di cereali per la prima colazione, cibo "sano", anzi moralmente consigliabile *dal punto di vista* degli Americanizers, alimento totalmente sconosciuto per le famiglie italiane.

Il reclutamento avveniva, *anche in questo caso*, secondo il sistema delle catene parentali e amicali e veniva organizzato all'estero in squadre di lavoro, costituite essenzialmente da reti paesane.

Altreitalie, 2006

Intorno agli anni settanta, nel pieno del boom dell'immigrazione, il Parlamento tedesco, a differenza di quello francese o statunitense, aveva votato un orientamento *in base al quale* lo

stato tedesco veniva riconosciuto non come paese di immigrazione ma come nazione di soggiorno temporaneo e prolungato per i lavoratori stranieri.

Il secondo gruppo che presenta il maggior numero di *lexical bundles* è quello che si riferisce alle connessioni interne ai paragrafi o tra due frasi successive. I *bundles* appartenenti a questo gruppo stabiliscono collegamenti addizionali di tipo causativo, contrastivo o consecutivo tra due elementi o enfatizzano elementi al microlivello.

Altreitalie, 2008

Nel raccontare la storia familiare alcuni giovani italoamericani si soffermano sulle difficoltà incontrate dai propri avi quando giunsero negli Stati Uniti. In questo caso *l'idea di un'America* come luogo dove tutto è possibile, che molti emigranti serbavano prima di prendere il mare, lascia il posto a una realtà ben più cruda in cui essere immigrato significava lottare quotidianamente tra gli innumerevoli ostacoli disseminati lungo il cammino dell'integrazione sociale (lavoro, casa, imparare l'inglese).

Altreitalie, 2008

Di certo non dovette essere così in tutta Italia *e il fatto che* le nostre scarse testimonianze provengano dall'Italia del nord può essere un caso, ma sappiamo che il settentrione della penisola, in cui si concentrò la resistenza gotica, che vide il formarsi di dominazioni franche e che per primo subì la guerra longobarda [...]

L'ultima categoria che copre una percentuale del 4,3% di tutti i *bundles* emersi dall'estrazione automatica sono i *clusters* orientati al partecipante, ossia allo scrivente o al lettore. A titolo d'esempio indichiamo la sequenza più frequente – *si è deciso di*.

Altreitalie, 2004

A questo scopo *si è deciso di* restringere le osservazioni sulle tornate elettorali referendarie e nazionali svoltesi a partire dalle elezioni europee del 2004.

Possiamo concludere che le proporzioni dei *lexical bundles* identificati riflettono l'ideologia della disciplina umanistica – per la maggior parte i *lexical bundles* esprimono una collocazione temporale o locale o si riferiscono ai temi della ricerca. Inoltre, essi sono volti ad organizzare il testo in quanto portatore del messaggio o argomento della ricerca. Gli articoli scientifici di storia sono, a loro volta, caratterizzati da una narrazione storica che si svolge in un contesto accademico discorsivo e valutativo.

### III.3.4.2 La funzione dei *lexical bundles* nel corpus di filosofia

Procediamo a questo punto con l'analisi funzionale dei *lexical bundles* emersi dal corpus di filosofia.

#### III.3.4.2.1 La classificazione dei *bundles* in categorie funzionali

I *lexical bundles* di filosofia sono stati ordinati in quattro categorie principali – il gruppo orientato al testo, alla ricerca, al partecipante e infine sono stati elencati *lexical bundles* senza una funzione discorsiva identificabile.

I gruppi sono proposti in senso decrescente considerando la frequenza d'occorrenza. Alla prima categoria, ai *lexical bundles* orientati al testo, appartengono: a) *bundles* volti a inquadrare il testo in quanto portatore del messaggio o dell'argomento – queste espressioni delimitano il tema al macro- e al microlivello, come per esempio nel caso dei *bundles* – *dal punto di vista, nella misura in cui, sulla base di un, nei termini di un*; e al microlivello indicano connessioni interne ai paragrafi o tra due frasi successive, stabiliscono collegamenti addizionali di tipo causativo, contrastivo o consecutivo tra due elementi o accentuano segmenti adiacenti – come ad esempio *per una filosofia dell', una volta per tutte*; b) segnali di passaggio discorsivo che stabiliscono connessioni additive o contrastive tra elementi – *se da un lato, da un lato la*; c) *bundles* che indicano un risultato (*il risultato di un, il risultato di una*); d) *clusters* che strutturano il testo – *nella ricerca di un, nell'ambito di un*.

Il secondo gruppo propone i *lexical bundles* volti al partecipante e viene a sua volta diviso in due sottogruppi: a) *bundles* che esprimono l'opinione, l'atteggiamento o la valutazione dello scrivente – *in un certo senso, ci troviamo di fronte, che ci permette di*; b) *bundles* che concernono in maniera diretta il lettore – *come abbiamo visto, come si è detto*.

Il terzo gruppo rappresenta i *lexical bundles* orientati alla ricerca e include i *bundles* che si riferiscono agli argomenti della ricerca – *il mondo delle cose, il mondo della vita* –, quelli che la delimitano – *della Critica della ragion, punto di vista*



*ontologico* – o esprimono una collocazione temporale (*di volta in volta, per la prima volta*) e infine quelli che delimitano una quantità (*di tutte le cose, in tutte le sue*).

Nell'ultima tipologia sono stati classificati i *lexical bundles* senza una funzione discorsiva individuabile – come nel caso dei *bundles in cui ci si, per il quale il*. Tale tipologia risulta solamente nel 2,1% di tutti i *bundles* emersi e pertanto non viene esplicitata in prosieguo.

Proponiamo di seguito la classificazione funzionale.

### *III.3.4.2.1.1 I lexical bundles orientati al testo – riguardanti l'organizzazione del testo e il suo significato in quanto messaggio o argomento*

#### **Al macrolivello**

##### **a. Segnali di cornice:**

fino a che punto (9); a se stesso e (7); al di fuori di (17); al di fuori del (21); al di fuori dell' (8); al di la della (21); al di la del (16); al di la di (16); al di la dell' (11); al di sopra di (15); anche in questo caso (28); all'interno del quale (9); all'interno di una (37); all'interno di un (33); all'interno di questa (8); sulla base di una (15); sulla base di un (12); in base alla quale (12); in base al quale (8); in base ad un (7); su di un piano (8); solo nella misura in (8); di fronte ad un (2); di fronte a un (1); di fronte ad una (2); che in questo caso (8); come nel caso dell' (7); dal punto di vista (106); da un punto di (64); da questo punto di (25); da un certo punto (8); nei termini di un (14); in una sorta di (9); un punto di vista (84); il punto di vista (20); questo punto di vista (30); suo punto di vista (11); punto di vista della (29); punto di vista del (17); punto di vista di (13); punto di vista dell' (12); punto di vista la (9); punto di vista il (7); punto di vista che (9); in entrambi i casi (12); da una sorta di (8); in questo senso la (21); in questo senso l' (18); in questo caso l' (13); in questo caso il (9); in questo caso la (8); in vista di una (8); in questo caso si (7); in un rapporto di (7); in altri termini il (7); in modo del tutto (7); di fronte ad un (19); di fronte a un (10); di fronte ad una (9); ad una sorta di (14); a partire da un (14); a partire da una (17); a partire dal quale (9); per quanto riguarda il (15); per quanto riguarda la (9); come una sorta di (11); come punto di partenza (7); e una sorta di (7); su cui si fonda (8); in relazione ad un (7); a favore di un (8); a favore di una (9); in virtù di un (7); di una sorta di (12); da parte di un (7); al tempo stesso la (10); nella misura in cui (67); in modo tale che (19); modo in cui la (10); stesso modo in cui (10); modo in cui si (7); modo in cui si (7); modo in cui la (10); stesso modo in cui (10); momento in cui si (16); momento in cui il (7); al modo in cui (8); dal momento che la (9); alla stregua di un (11); e allo stesso tempo (13); e nello stesso tempo (16); e al tempo stesso (18); il punto di partenza (25); un punto di partenza (7); punto di partenza e (11); punto di partenza per (7); punto di partenza di (9); che a partire da (9); sotto il nome di (13); dal momento che il (11); a meno che non (18); il modo in cui (24);

### **Al microlivello**

di una filosofia della (7); di una teoria della (12); in carne ed ossa (12); una volta per tutte (20); e a sua volta (8); a sua volta un (7); in quanto tale e (7); l'idea di una (16); l'idea di un (15); la possibilità di un (11); la possibilità di una (9); in una situazione di (8); l'idea che il (7); l'idea che l' (10); l'idea che la (10); la differenza tra la (8); il rapporto con l' (7); l'esistenza di un (11); fare a meno di (21); una situazione in cui (7); dell'uomo e della (9); di un soggetto che (10); di un processo di (9); della vita e della (8); dei bisogni e dei (7); della scienza e della (7); della verità e della (7); del fatto che la (12); che a sua volta (20); il fatto di essere (7); tra l'uomo e (10); tra il soggetto e (8); il problema del rapporto (10); di un io che (7); di una vera e (12); l'esistenza di una (7); l'esigenza di una (9); la presenza di un (9); il principio di non (7); un mondo in cui (8); il fatto che il (15); di ciò che si (7); mondo della vita e (11); l'io e il (8); l'uno e il (8); l'uomo e il (8); l'uomo e l' (8); l'uomo e la (8); la ragione e la (8); la sua vita e (7); in cui l'uomo (10); in cui il soggetto (8); è il problema della (10); in cui la filosofia (7); il fatto che la (21); il fatto che in (8); non in grado di (8); una vera e propria (47); un vero e proprio (20); è in grado di (28); sia in grado di (7); essere in grado di (10); in grado di fornire (7); in grado di rendere (7); in grado di capire (8); l'uno all'altro (9); che il concetto di (9); fare i conti con (13); problema del rapporto tra (8); e il fatto che (9); e in quanto tale (7); di ciò che è (14); un mondo in cui (8); e lo stesso che (7); che fare con la (8); che fare con una (8);

#### **b. Segnali di passaggio discorsivo:**

se da un lato (20); e d'altra parte (9); da un lato la (15); da un lato il (8); da una parte e (7); da un lato è (11); in un primo momento (7); da un lato l' (7); in contrasto con la (7); sia da un punto (7); in un primo tempo (8);

**c. Segnali di risultato:** il risultato di un (11); il risultato di una (9);

**d. Segnali strutturalivi:** nella ricerca di un (7); nell'ambito di un (8).

### ***III.3.4.2.1.2 I lexical bundles orientati al partecipante – si concentrano sullo scrivente e sul lettore del testo e includono:***

#### **a. Opinione dell'autore o della comunità:**

dal nostro punto di (7); in un certo senso (28); in un certo modo (8); in fin dei conti (7); che in qualche modo (9); nostro punto di vista (8); si tratta dunque di (7); si tratta di un (46); si tratta di una (35); non si tratta di (26); vale a dire il (14); vale a dire la (10); vale a dire un (7); si pone il problema (7); si può dire che (9); si trova di fronte (12); ci troviamo di fronte (14); abbiamo a che fare (7); consiste nel fatto che (9); ha nulla a che (7); non ha nulla a (7); ha a che fare (21); ha il compito di (12); non ha bisogno di (12); se per un verso (8); questo non significa che (10); siamo in grado di (7); è anche vero che (9); che si tratta di (11); che ci permette di (7); che si tratti di (7); se è vero che (15); e in questo senso (11); in questo senso il (15); nel senso che la (13); nel senso che una

(13); nel senso che non (8); nel senso che l' (7); non nel senso che (7); in questo senso si (12); in questo modo la (14); in questo modo si (8); in tal modo si (8); in tal modo l' (8); in tal modo la (10); in questo modo l' (9); in questo modo il (7); in questo senso l' (7); nel senso di un (8); nel modo in cui (15); nel senso in cui (8); a che vedere con (9); a che fare con (51); e in questo modo (7); e in tal modo (7); e in un certo (7); e proprio per questo (11); ma non per questo (12); e in qualche modo (7); se è solo se (7); ma anche e soprattutto (11); e per questo che (8); per il fatto che (18); se non ci fosse (9); che non è mai (8); che non può essere (10); che non si lascia (7); ciò che non è (10); cui non si può (8); di cui non si sa (9); da ciò che è (8); ancora una volta la (7);

**b. Coinvolgimento del lettore nel discorso:** come si è visto (9); come si è detto (7); che come abbiamo visto (8); con le parole di (7).

### *III.3.4.2.1.3 I lexical bundles orientati alla ricerca – aiutano lo scrivente a strutturare la propria attività e la propria esperienza del mondo reale*

#### **a. Argomento di ricerca:**

tra ragione e fede (7); il mondo delle cose (40); il mondo della vita (17); il concetto di verità (14); Critica della ragione pura (24); pensiero di Merleau Ponty (9); la fenomenologia di Husserl (8); tradizione del pensiero occidentale (8); la storia della filosofia (7); la filosofia del linguaggio (13); la presa di coscienza (8); le cose del mondo (8); il modo di essere (11); il senso dell'essere (7); il senso di essere (9); una teoria della conoscenza (11); la domanda sulla verità (7); principio di non contraddizione (11);

#### **b. Delimitatori dell'argomento:**

della Critica della ragione (9); della Fenomenologia dello spirito (8); nella Critica della ragione (11); in Essere e tempo (13); il dato di fatto (8); un dato di fatto (10); dato di fatto che (7); punto di vista ontologico (8); del mondo della vita (13); al mondo delle cose (7); la condizione di possibilità (8); le condizioni di possibilità (7); bisogni e dei desideri (7); per una filosofia dell' (10); dell'opera d'arte (13);

#### **c. Collocazione temporale:**

di volta in volta (39); alla fine degli anni (7); nel corso degli anni (7); ad un certo punto (19); a un certo punto (16); per la prima volta (29); nel momento in cui (79); momento in cui la (14); momento in cui il (7); momento in cui si (16);

#### **d. Quantificatori:**

di tutte le cose (12); di tutti gli uomini (7); di tutto ciò che (12); in tutta la sua (17); in tutte le sue (15); tutte le forme di (11); tutta una serie di (10); tutto ciò che si (9); tutto ciò che è (8).

#### III.3.4.2.1.4 Altri bundles

in cui ci si (9); in cui essa si (7); per il quale il (10); per il quale la (7); che non si può (8); che fa sì che (7); che fare con il (7).

#### III.3.4.2.2 Risultati emersi dalla classificazione funzionale

Un passo maggiore verso il discorso persuasivo lo si fa nel linguaggio di filosofia, dato che esso non si basa su regole o sui dati di fatto. Il ragionamento in filosofia è all'insegna delle asserzioni, delle descrizioni e della valutazione – tutti elementi devono essere argomentati e giustificati. L'enunciato che ne deriva delimita un pensiero, un'idea, un'interpretazione o una supposizione, per pervenire ad una conclusione.

Il termine filosofia deriva dal greco e indica "desiderio di sapere". La filosofia è "attività di pensiero che attinge ciò che è costante e uniforme al di là del variare dei fenomeni, con l'ambizione di definire le strutture permanenti della realtà e di indicare norme universali di comportamento<sup>43</sup>".

Il discorso filosofico, basandosi solo sull'attività di pensiero, deve avere una trattazione degli argomenti notevolmente convincente – come tale si propone di convogliare il pensiero del lettore e di dirigerlo verso una certa deduzione o opinione. Il ragionamento è teorizzato per difendere o giustificare una tesi, un'opinione, un argomento. La riflessione di un filosofo che, a sua volta, va motivata e argomentata, induce ad una struttura logico-epistemologica con delle peculiari strategie discorsive e argomentative. L'intenzione del filosofo non è soltanto di fare delle asserzioni, delle descrizioni o delle valutazioni o qualsiasi altro tipo di discorso, bensì di argomentarle, disputarle, difenderle, sostenerle e motivarle con dei giudizi, delle spiegazioni e dimostrazioni (Berti, 1997).

Da questo punto di vista ne consegue, pertanto, l'uso massiccio dei *lexical bundles* orientati al testo. Difatti, lo spoglio ha mostrato che i *bundles* orientati al testo coprono una percentuale del 59,8% di tutte le sequenze – sono stati identificati 196 *clusters*. Da un lato, le loro funzioni discorsive prevalenti accentuano connessioni, specificano casi, indicano limitazioni o condizioni per l'interpretazione del discorso successivo o quello in precedenza, dall'altro lato stabiliscono collegamenti addizionali al

---

<sup>43</sup> Definizione tratta dall'Enciclopedia Treccani della versione online – [www.treccani.it](http://www.treccani.it).

microlivello testuale o accentuano caratteristiche dei segmenti adiacenti o delle asserzioni precedenti.

Indichiamo a titolo di esempio i *bundles* con l'occorrenza più alta – i *bundles sulla base di un, nei termini di un'* delimitano o inquadrano il discorso precedente; il *bundle – una volta per tutte* – accentua l'asserzione precedente.

Segni e comprensione, 2007

Il primo l'abbiamo detto: e che le procedure di accostamento al verbale devono essere non solo stabilite *sulla base di un* modello di segno in generale, cioè un modello semiotico, e quindi, come tale, non coincidente con quello di segno verbale, ma capace anche di rendere conto dei segni più refrattari rispetto a categorie semiotiche fonocentriche (come tali, pseudosemiotiche).

Segni e comprensione, 2004

Il tempo interno alla scrittura musicale non può scandirsi come una semplice successione di istanti, poiché il movimento dello sviluppo beethoveniano si pone *nei termini di un'* articolazione di frammenti: il tempo musicale si configura, piuttosto, come *kairos* (occasione).

Segni e comprensione, 2001

Ci sembra pertanto che un Campo Semantico Globale possa poter sussistere per intero, benché forse non *una volta per tutte*, viste le sedimentazioni di senso a cui è sottoposto; ciò che si nega è che esso sia descrivibile come sistema di relazioni binarie successive [...]

Va ricordato, inoltre, il gruppo dei *lexical bundles* con doppia funzione, sebbene presenti solo un numero limitato di *clusters*. Come è già stato indicato, i *bundles* con funzione molteplice sono stati classificati nel gruppo della loro funzione predominante. Nel caso della sequenza *su di un piano* si è optato per la sua funzione di un segnale di cornice, ovvero di una delimitazione o specificazione del discorso precedente e non della sua funzione come collocazione locale. La proporzione delle due funzioni è 7 verso 1. La funzione delimitante è altresì predominante tra le funzioni del *bundle – all'interno di un*. Tale funzione costituisce il 60% di tutte le funzioni, mentre il resto va attribuito alla funzione locale.

Segni e comprensione, 2001

L'intuizione basilare che soggiace a questo esperimento può essere esposta come segue: se il peso effettivo di un corpo può essere diminuito collocandolo *su di un piano* inclinato, allora anche la sua velocità di scivolamento diminuirà proporzionalmente.

Segni e comprensione, 2001

Egli pone il problema dell'essere, annunciato da Platone, sul piano dell'esperienza ossia *su di un piano* "realistico" dove la filosofia si coniuga al sapere umano.

Segni e comprensione, 2001

Immaginiamo un individuo, Carlo, *all'interno di un* cinema. Oltre ad entità quali un individuo (Carlo), la visione del film (evento) e la sua curiosità (proprietà), se ne possono enumerare tante altre che rendono difficile, se non impossibile rispondere ad una domanda quale "cosa è un cinema?"

Segni e comprensione, 2001

Pertanto e soprattutto –o anche– la delusione causata dalla sordità dell'intelligenza, tutta rinchiusa *all'interno di un* pensare inteso come luogo aristocratico, separato, che porta la Arendt a costruire la teoria che assegna il primato ad un pensiero che pensa non le cose, ma l'azione: [...]

Dalle 328 sequenze differenti si ricava il 23,2% di *bundles* che raffigurano l'opinione dello scrivente, le sue valutazioni e i suoi atteggiamenti, o si concentrano propriamente sul lettore, cercando di coinvolgerlo. La lettura funzionale di *bundles* di questa categoria consente di stabilire che il 94,7% dei *bundles* indica un atteggiamento posizionale dello scrivente, mentre il 5,3% cerca di convogliare il lettore. Dalla tipologia presente si può desumere il carattere argomentativo del testo filosofico. L'argomentatività di esso si riflette da un lato nella rappresentazione, esposizione, enunciazione del pensiero dello scrivente, mentre allo stesso tempo lo scrittore cerca di argomentare, mostrare o spiegare il concetto esposto.

A mo' di esempio indichiamo i *bundles* con la frequenza più alta – *si può dire che* indica un'esposizione del pensiero dello scrivente, da contro i *bundles* – *ci troviamo di fronte, che si tratta di* – specificano la coerenza interna al discorso precedente, con l'ausilio dei quali si cerca di chiarire quanto in precedenza discusso.

Segni e comprensione, 2010

*Si può dire che* l'inconscio costituisce l'archetipo del gioco del rocchetto: esso rappresenta il gioco simbolico fondamentale per Freud e per la psicoanalisi.

La conclusione dell'intero discorso riguardante la coscienza in tutte le sue caratteristiche, e quella per cui, anche in questo caso, *ci troviamo di fronte* ad un problema il cui studio esige la chiara consapevolezza *che si tratta di* un problema le cui soluzioni sono perseguibili soltanto attraverso delle indagini epistemologiche [...]

Il terzo gruppo maggiormente rappresentato è la tipologia dei *bundles* orientati alla ricerca. Tali sequenze coprono una percentuale del 14,9% di tutte le espressioni

individuate. La maggior parte dei *bundles* si orienta verso l'essenza della ricerca o cerca in qualche modo di delimitarla – si tratta del 63,5% di tutte le sequenze nella tipologia presente. Il resto si rispecchia in un determinante rivolto ad esprimere una quantità – *di tutte le cose* – o in una collocazione temporale – *di volta in volta*.

A titolo d'esempio elenchiamo le sequenze più ricorrenti – il *bundle – il mondo delle cose* indica l'argomento della ricerca, da contro i *bundles – per una filosofia dell', punto di vista ontologico* – inquadrano un aspetto della ricerca.

Segni e comprensione, 2010

Il soggetto della conoscenza – o il soggetto che “pone” *il mondo delle cose*, fino a caricarsi di un potere conoscitivo superiore, quello dell’“intelletto intuitivo” o dell’“intelletto archetipo” kantianamente inteso [...]

Noi siamo *per una filosofia dell'ago* e non della paglia.

Segni e comprensione, 2001

Una versione "forte" della sopravvenienza per esempio intende quest'ultima come una relazione che, dal *punto di vista ontologico*, vale in tutti i mondi possibili.

Segni e comprensione, 2007

Livelli che non possono non essere stabiliti in relazione ad un dato tempo e ad un dato luogo, ma che *di volta in volta* assumono quei caratteri strutturali che ne definiscono l'etnicità come eco-sistema.

[...] a tal proposito si muovevano spesso sulle sue labbra le parole del Credo: "Credo in Dio, creatore *di tutte le cose*, visibili e invisibili", riconoscendo Dio come il Creatore dell'Universo.

Possiamo concludere che i *lexical bundles* possono essere riconosciuti come un tratto linguistico con delle funzioni discorsive ben identificabili.

Il ragionamento filosofico, ovvero l'argomentare, il discutere, l'approfondire, l'analizzare e il giustificare una verità condivisa per venire ad un consenso che si basa su fatti e ragionamenti, si rispecchia in un pensiero discorsivo e persuasivo. Come tale utilizza in più ampia misura i *lexical bundles* orientati al testo e al partecipante.

Nella sezione seguente si cercherà di classificare, a seconda della loro funzione nel testo, i *lexical bundles* emersi dall'estrazione automatica del *corpus* di medicina del lavoro.

### III.3.4.3 La funzione dei *lexical bundles* nel corpus di medicina del lavoro

Prima di passare all'analisi dei risultati proponiamo la classificazione discorsiva dei *lexical bundles* del corpus di medicina del lavoro.

#### III.3.4.3.1 La classificazione dei *bundles* in categorie funzionali

I *lexical bundles* sono stati raggruppati in quattro categorie principali – *bundles* orientati alla ricerca, al testo, al partecipante e "altri *bundles*", ovvero espressioni senza una funzione contestuale.

La categoria delle espressioni di quattro parole orientate alla ricerca è stata suddivisa in sette sottogruppi: a) *lexical bundles* che si riferiscono all'argomento di ricerca o agli aspetti inerenti alla ricerca; b) *bundles* che descrivono il contesto di ricerca; c) *clusters* che delineano la procedura di ricerca, mostrando i modi in cui uno studio è stato svolto; d) *bundles* che esprimono una quantità; e) *lexical bundles* che specificano il luogo della ricerca; f) *lexical bundles* che specificano il periodo della ricerca; g) *lexical bundles* che esprimono un paragone. A questo punto dobbiamo menzionare che l'ultima classe incorpora solo un *bundle* – si è optato per questa scelta per presentare al meglio le funzioni delle espressioni analizzate nel contesto.

Dobbiamo inoltre puntualizzare che nel gruppo delle espressioni che indicano l'argomento di ricerca (gruppo a) sono state incluse tutte le espressioni che riguardano il tema centrale della ricerca e gli aspetti relativi allo studio. Per aspetti si intendono i temi correlati all'argomento principale – nel caso del tema centrale "Il rischio stress lavoro correlato", è stato come argomento principale indicato il *bundle* – *rischio stress lavoro correlato* –, come quelli interdipendenti invece i *lexical bundles*: *la tutela della salute, la promozione della salute, la salute dei lavoratori*.

In base alla frequenza d'occorrenza sono stati come seconda categoria classificati i *lexical bundles* concernenti l'organizzazione del testo e il suo significato, ovvero i segnali di coerenza testuale, in quanto portatori del messaggio. Questa categoria incorpora *bundles* che pongono una restrizione, una delimitazione



all'argomento di ricerca inquadrandolo con i margini, specificando le connessioni interne al testo, i risultati dello studio, e la sua organizzazione.

I *bundles* che esprimono l'opinione del ricercatore o della comunità scientifica e quelli che coinvolgono il lettore sono stati classificati come terza categoria.

All'ultimo gruppo appartengono i *clusters* senza una funzione discorsiva riconoscibile nel testo, come ad esempio *per i quali e, le quali è obbligatoria*.

#### *III.3.4.3.1.1 I lexical bundles orientati alla ricerca – aiutano lo scrivente a strutturare la propria attività e la propria esperienza del mondo reale*

##### **a. argomento principale e argomenti inerenti alla ricerca:**

la sindrome delle apnee (13); tutela della salute (81); la tutela della salute (47); la promozione della salute (39); la salute dei lavoratori (63); lo stato di salute (29); promozione della salute (21); la sicurezza dei lavoratori (19); le condizioni di salute (16); la sicurezza del lavoro (17); sicurezza del lavoro ISPELSL (14); la prevenzione degli infortuni (16); sindrome del tunnel carpale (23); la qualità della vita (34); il miglioramento delle condizioni (21); organizzazione del lavoro (25); di sicurezza e salute (15); di specializzazione in medicina (15); di tumore del polmone (15); sindrome delle apnee ostruttive (17); lo stile di vita (19); patologie degli arti superiori (18); denuncia di malattia professionale (15); il giudizio di idoneità (48); un giudizio di idoneità (14); la SS degli OS (12); ruolo del medico competente (13); la figura del medico (12); la percezione del rischio (12); fattori di rischio biomeccanico (10); gli orari di lavoro (10); il rischio di esposizione (10); il turno del mattino (10); il turno di notte (10); la collaborazione del medico (10); i giorni di riposo (10); fattori di rischio cardiovascolare (11); fattori di rischio presente (11); i principi della medicina (11); il fumo della sigaretta (11); la facilità di accesso (11); la riduzione dei rischi (11); medicina preventiva dei lavoratori (11); tutela della salute dei (30); materia di tutela della (21); salute dei lavoratori e (18); gestione della salute e (17); promozione della salute nei (16); servizio di prevenzione e (55); i fattori di rischio (63); un fattore di rischio (30); altri fattori di rischio (27); principali fattori di rischio (17), fattori di rischio lavorativi (15); fattori di rischio professionali (17); fattori di rischio per (32); fattori di rischio e (27); fattore di rischio per (26); la gestione del rischio (20); la valutazione dello stress (20); la valutazione dell'efficacia (10); valutazione del rischio e (37); valutazione del rischio da (30); valutazione dei rischi e (34); la formulazione del giudizio (10); sistema informativo nazionale per (12); E.S.M.I. (55); G ITAL MED LAV (14); ITAL MED LAV ERG (14); italiana di medicina del (30); nazionale per la prevenzione (17); patto per la salute (10); gli infortuni sul lavoro (31); infortuni sul lavoro INAIL (10); sicurezza sui luoghi di (15); infortuni sul lavoro e (13); infortuni e delle malattie (18); salute e sicurezza nei (18); della medicina del lavoro (94); della salute e sicurezza (63); della salute dei lavoratori (43); della sicurezza dei lavoratori

(22); delle condizioni di salute (26); della valutazione dei rischi (40); della valutazione del rischio (35); del rischio stress lavoro (25); del rischio da stress (23); del lavoro della salute (18); del lavoro e igiene (16); del lavoro ed ergonomia (15); del posto di lavoro (15); della tutela della salute (25); dello stato di salute (42); della salute nei luoghi (22); dell'organizzazione del lavoro (38); del giudizio di idoneità (60); dei fattori di rischio (55); dello stress lavoro correlato (55); degli ambienti di lavoro (39); responsabile del servizio di (25); stato di salute e (17); fumo di sigaretta e (16); giudizio di idoneità alla (31); parte del datore di (22); formulazione del giudizio di (25); la sorveglianza sanitaria dei (16); attività di vigilanza in (15); sicurezza degli ambienti di (15); ambito della medicina del (14); i principali fattori di (14); condizioni di vita e (10); giudizio di idoneità al (10); preventiva dei lavoratori della (10); prevenzione degli infortuni sul (10); processo di valutazione del (10); servizi di medicina del (10); medicina del lavoro di (11); medicina del lavoro in (11); medicina del lavoro la (11); medicina del lavoro deve (11); riduzione del numero di (11); effetti negativi sulla salute (10); patologie da sovraccarico biomeccanico (21); rischio da stress lavoro (21); gli effetti sulla salute (11); malattie per le quali (11); malattie professionali di cui (14); malattie la cui origine (17); materia di salute e (49); miglioramento delle condizioni di (33); condizioni di salute e (23); condizioni di lavoro e (16); lavoro della salute e (18); medicina del lavoro con (18); riduzione del rischio di (14); stato di salute dei (14); rischio di esposizione a (12); carico del datore di (11); consumo di alcol e (11); fattori di rischio di (11); prevenzione degli infortuni e (13); aumento del rischio di (12); condizioni di salute del (12); tema di salute e (11); idoneità alla mansione specifica (36); il lavoro a turni (27); effetti a lungo termine (14); lavoro a turni e (29); lavoro al fine di (14); abitudine al fumo di (16); la prevenzione nei luoghi (20); sicurezza nei luoghi di (60); prevenzione nei luoghi di (55), salute nei luoghi di (34); specializzazione in medicina del (15); rischi per la salute (77); rischio per la salute (44); rischi per la sicurezza (32); lavoratori per la sicurezza (23); interesse per il lavoro (22); soddisfazione per il lavoro (12); i rischi per la (39); un rischio per la (11); il rischio per la (11); lo stress lavoro correlato (17); il servizio sanitario nazionale (13); salute e la sicurezza (38); sostanze stupefacenti e psicotrope (14); salute e sicurezza dei (38); salute e sicurezza sui (13), valutazione e gestione del (10); prevenzione e sicurezza nei (10); origine lavorativa e di (10); lavoro e della previdenza (11); stress lavoro correlato in (10); stress lavoro correlato e (28); la sicurezza e la (48); la salute e la (40); la prevenzione e la (24); salute e sicurezza e (14); la sorveglianza sanitaria e (13); la presenza o meno (10); stato le regioni e (11); salute e sicurezza sul (79); salute e della sicurezza (70); sanitaria e di rischio (34), vita e di lavoro (24); informazione e formazione dei (12); lavoro e quello di (15); salute e la sicurezza (38); prevenzione e la sicurezza (17); lavoro e igiene industriale (16); lavoro ed igiene industriale (10), prevenzione e protezione e (16); biomeccanico degli arti superiori (13); permanente per i rapporti (15); permanente per la salute (14); consultiva permanente per la (15); sanitarie e di rischio (15); collegati allo stress lavoro (11); e sicurezza sul lavoro (77); e sicurezza dei lavoratori (35); e la salute dei (31); e sicurezza nei luoghi (31), e sicurezza degli ambienti (15); e gestione del rischio (17); e formazione dei lavoratori (14); e sicurezza del lavoro (14); e i livelli di (13); e sicurezza sui luoghi (13); e la mancanza di (10); e lo sviluppo di (10); e fattori di rischio (11); e la promozione delle (11); e la sicurezza dei (16); che l'esposizione a (20); che la prevalenza di (11); che i livelli di (14); che la valutazione dei (10); della prevenzione nei luoghi (14); degli infortuni sul lavoro (27); del numero di infortuni (11); delle condizioni di lavoro (31); del rapporto di lavoro (30); della sicurezza nei luoghi (18);

della qualità della vita (16); del lavoro ed ergonomia (15); del lavoro della salute (18); del lavoro ed igiene (16); dei giudizi di idoneità (15); del sistema nervoso centrale (11); di assenza dal lavoro (11); di gestione del rischio (11); dei diritti di lavoratori (10); dei diritti dei lavoratori (10); del lavoro ed igiene (10); del mondo del lavoro (10); del servizio sanitario nazionale (10); del mercato del lavoro (12); della promozione della salute (12); di medicina del lavoro (163); di salute e sicurezza (94); di salute del lavoratore (22); la durata e la (12); di stress lavoro correlato (25); di promozione della salute (70); di tutela della salute (54); di salute dei lavoratori (25); di gestione della salute (13); di formazione e informazione (13); di idoneità alla mansione (36); di esposizione ad amianto (14); di una malattia professionale (14); di fattori di rischio (17); di prevenzione e sicurezza (26); di prevenzione e protezione (92); di infortuni sul lavoro (27); di sorveglianza sanitaria (20); di disturbi del sonno (20); di igiene e sicurezza (10); di sensibilità e specificità (10); di sicurezza sul lavoro (10); di assunzione di sostanze (12); di formazione ed informazione (12); dei rischi per la (14); della cartella sanitaria e (14); del lavoro e delle (13); della sicurezza e della (11); delle regioni e delle (11); del lavoro e del (10); del lavoro e di (10); del rischio e la (10); degli infortuni e delle (20); del rischio da parte (11); delle malattie professionali di (11); del lavoro e della (19); dell'arto superiore a (18); di lavoro e del (12); di prevenzione e di (12); di salute e di (12); della salute e della (82); di vita e di (26); di lavoro e di (22); di lavoro e dei (13); di esposizione e di (11); di sostanze psicotrope e (11); di lavoro ai fini (10); di lavoro e la (10); di questo lavoro e (10); di una condizione di (10); di medicina preventiva dei (10); di rischio per la (25); nella valutazione del rischio (20); nella valutazione del rischio (20); nel mondo del lavoro (18); nel processo di valutazione (16); nella valutazione dei rischi (12); nell'ambito della medicina (14); nei confronti dei lavoratori (10); in medicina del lavoro (94); in materia di salute (53); in materia di sicurezza (22); in materia di tutela (22); in tema di salute (14); in materia di prevenzione (18); in via di sviluppo (10); in tema di sicurezza (10); per la formazione continua (35); per la popolazione generale (29); per la sorveglianza sanitaria (21); per il medico competente (28); per gli operatori sanitari (17); per la salute umana (22); per i lavoratori esposti (10); per la SS degli (11); contro gli infortuni sul (17); e della sicurezza dei (21); e delle malattie professionali (15); e delle province autonome (14); e della previdenza sociale (12); e delle politiche sociali (12); e della sicurezza nei (10); medicina del lavoro dovrà (11); dalla cessazione della lavorazione (17); da stress lavoro correlato (27); alla valutazione dei rischi (39); alla valutazione del rischio (38); ai fattori di rischio (23); al fumo di sigaretta (35); al fumo di tabacco (13); allo stress lavoro correlato (23); a livelli di rumore (18); alla tutela della salute (18); alla promozione della salute (17); a fattori di rischio (18); alla promozione della salute (17); all'ambiente di lavoro (13); alla luce di quanto (13); a turni e notturno (10); alla medicina del lavoro (10); ai fini della valutazione (14); a carico degli arti (11); a carico del datore (11); a carico del sistema (11); ridurre il rischio di (10); sulla valutazione del rischio (14); sullo stato di salute (26); sulla salute dei lavoratori (12); sulle condizioni di salute (11); sulle condizioni di lavoro (13); rischio stress lavoro correlato (35); medicina del lavoro e (91); medicina del lavoro ed (31); rapporti tra lo stato (14);

#### **b. descrizione del contesto di ricerca:**

gli organi di vigilanza (16); società italiana di medicina (29); lavoratori di sesso maschile (15); soggetti di sesso femminile (12); gruppi di lavoratori esposti (12); datore

di lavoro e (74); datore di lavoro di (21); datore di lavoro deve (22); medico del lavoro e (23); medico del lavoro in (17); ministero del lavoro e (16); soggetti di sesso maschile (24); il datore di lavoro (184); i datori di lavoro (40); datori di lavoro (27); il medico del lavoro (128); medico del lavoro competente (41); parte del medico competente (34); attività del medico competente (17); collaborazione del medico competente (14); i medici del lavoro (25); figura del medico competente (16); medico di medicina generale (16); medici di medicina generale (16); la medicina del lavoro (62); il ruolo del medico (15); direzione provinciale del lavoro (16); organizzazione mondiale di sanità (15); istituto superiore di sanità (18); il responsabile del servizio (15); attività della vita quotidiana (14); il ministero del lavoro (13); l'organizzazione del lavoro (13); medici del lavoro e (12); ministero della salute e (12); ministero del lavoro della (14); rappresentanti dei lavori per (14); datore di lavoro a (13); datore di lavoro in (14); la società italiana di (14); i soggetti di sesso (12); datore di lavoro o (11); datore di lavoro per (10); datore di lavoro che (17); medico del lavoro che (14); istituto superiore per la (14); lavoratrici e dei lavoratori (11); età compresa tra i (28); piccole e medie imprese (20); il medico competente non (10); il medico competente si (10); tutte le figure professionali (10); lo stato le regioni (10); i soggetti affetti da (16); commissione consultiva permanente per (15); fumatori e non fumatori (10); mondiale della sanità OMS (13); dei medici del lavoro (38); del medico del lavoro (92); del datore di lavoro (112); dei datori di lavoro (36); dei lavoratori della sanità (19); che il medico competente (39); medicina del lavoro sia (10); di sesso maschile (12); del ministero della salute (23); del ministero del lavoro (20); nella fascia di età (13); in un gruppo di (21); in entrambi i sessi (10); nella popolazione generale e (21); nei lavoratori esposti a (10); i lavoratori sono stati (15); i soggetti sono stati (11); volontari di promozione della (11); i lavoratori esposti a (10); che la sorveglianza sanitaria (11); che il datore di (26); specialisti in medicina del (15); vigilanza in materia di (15); il medico competente e (22); e il medico competente (10); la cui origine lavorativa (17); che il medico del (16); del gruppo di lavoro (20); del servizio di prevenzione (34); dei servizi di prevenzione (23); il servizio di prevenzione (11); servizi di prevenzione e (20); degli organi di vigilanza (12); dei lavoratori e dei (22); dei lavoratori per la (27); dei lavoratori disposti a (14); dei lavoratori e delle (10); della società italiana di (10); del medico competente alla (11); del medico competente di (12); della sorveglianza sanitaria e (12); delle lavoratrici e dei (12); del medico competente e (27); del medico competente in (16); di un gruppo di (25); di questo tipo di (11); di lavoratori esposti a (16); per il datore di (18); per il medico del (22); da parte del medico (53); da parte dei lavoratori (42); da parte del datore (22); da parte degli operatori (18); da parte del lavoratore (18); dal ministero della salute (13); da parte dei medici (12); dai medici del lavoro (11); dalla commissione consultiva permanente (10); dal datore di lavoro (33); dal gruppo di lavoro (15); al datore di lavoro (61); ai datori di lavoro (15); tra i due gruppi (16); con il medico competente (11); con il datore di (33); con età media di (10); ad un gruppo di (13); esposti a rischio biologico (12); tra coloro che hanno (14); sottoposti a sorveglianza sanitaria (24); medici competenti che operano (11); il medico competente può (15); al medico del lavoro (22); a tutti i lavoratori (18); cui origine lavorativa è (17); il gruppo di lavoro (30); un gruppo di lavoro (24); un gruppo di controllo (14); i dispositivi di protezione (18); dispositivi di protezione individuale (51); movimentazione manuale dei carichi (67); movimentazione manuale di carichi (32); movimentazione manuale dei pazienti (17); sovraccarico biomeccanico degli arti (13); kg di peso corporeo (11); il carico di lavoro (16); la movimentazione manuale dei (10);

vibrazioni trasmesse al corpo (13); esposizione ad agenti biologici (15); della cuffia dei rotatori (18); dei dispositivi di protezione (22); al sistema mano braccio (18); alla movimentazione manuale dei (12);

### **c. procedura di ricerca:**

il processo di valutazione (14); la stima del rischio (15); la raccolta dei dati (19); tabelle delle malattie professionali (18); nuova tabella della malattia (14); nuove tabelle delle malattie (15); il coefficiente di correlazione (10); processo di valutazione dei (23); T di student per (10); test T di student (10); normativa in materia di (10); protezione individuale (DPI) (15); un sistema di gestione (17); la valutazione del rischio (109); la valutazione dei rischi (63); valutazione del rischio stress (38); tabella delle malattie professionali (22); programma di sorveglianza sanitaria (15); le misure di prevenzione (21); valutazione del rischio chimico (20); valutazione dello stress lavoro (20); somministrazione di un questionario (11); lo svolgimento di attività (10); le strategie di coping (10); modello di regressione logistica (10); programmi volontari di promozione (11); i valori di riferimento (14); valori limite di esposizione (12); un documento di consenso (10); i valori medi di (15); cartella sanitaria e di (34); sistema di gestione della (22); sorveglianza sanitaria dei lavoratori (24); attività di sorveglianza sanitaria (22); indice di massa corporea (17); il livello di esposizione (10); principi della medicina del (17); misure di prevenzione e (33); norme di buona tecnica (24); documento di valutazione del (13); il test T di (10); linee guida per la (20); misure per la tutela (11); programmi di promozione della (10); indirizzi scientifici più avanzati (16); gli indirizzi scientifici più (10); del documento di valutazione (16); che le linee guida (12); cartelle sanitarie e di (13); nel documento di valutazione (13); delle misure di prevenzione (27); documento di valutazione dei (27); valutazione del rischio per (11); valutazione di tutti i (11); valutazione dei rischi che (10); valutazione dei rischi in (10); dei valori di riferimento (14); dei livelli di esposizione (11); delle attività di vigilanza (18); delle attività della vigilanza (18); delle attività di prevenzione (26); della normativa in materia (10); di valutazione del rischio (89); di valutazione dei rischi (84); di valutare il rischio (10); di gestione della sicurezza (12); della letteratura scientifica e (10); ai sensi dell'art. (29); trasmesse al sistema mano (15); trasmesse al corpo umano (14); entrata in vigore del (18); messa in atto di (10); di cui all'articolo (30); di cui all'art. (21); di cui agli artt. (11); sono stati analizzati (10); stato somministrato un questionario (10); sono stati confrontati con (10); alle differenze di genere (11); ai sensi del d (12); sono stati presi in (17); sono stati sottoposti ad (17); è stata effettuata una (16); è stata condotta una (12); è stata osservata una (12); è stata utilizzata la (12); sono stati suddivisi in (12); stati presi in considerazione (14); è stata valutata la (13); è stato somministrato un (11); è stata rilevata in (11), è stata rilevata una (11); il questionario è stato (14); sono stati sottoposti a (21); è stato utilizzato il (20); di vigilanza in materia (15); è obbligatoria la denuncia (10);

### **d. quantificatori:**

maggior parte dei soggetti (12); maggior parte dei casi (19); la percentuale di soggetti (11); un elevato grado di (15); un elevato numero di (15); la maggior parte dei (56); la maggior parte delle (28); la maggior parte degli (19); la maggior parte di (16); la maggior parte del (15); un maggior numero di (16); il maggior numero di (12); un

numero limitato di (11); un gran numero di (10); una vasta gamma di (10); un aumentato rischio di (19); un aumento del rischio (17); periodo massimo di indennizzabilità (23); un elevato rischio di (11); che la maggior parte (39); di tutti i rischi (18); di tutti i soggetti (18); nella maggioranza dei casi (16); nella maggior parte dei (24); nella maggior parte delle (11); di uno o più (10); di una condizione di (10); per la maggior parte (13); da un minimo di (15); a maggior rischio di (16); ad alto rischio di (14); ad elevato rischio di (10); ad un massimo di (14); a tutti i livelli (12); la concentrazione media di (11); tutti i rischi per (29); elevato rischio di infortuni (10); superiore per la prevenzione (14); in tutti i casi (23);

**e. collocazione locale:**

province autonome di Trento (17); ambienti di lavoro e (16); gli ambienti del lavoro (13); un ambiente di lavoro (12); i luoghi di lavoro (12); luogo di lavoro e (15); luoghi di lavoro e (24); luoghi di lavoro delle (13); la salute e sicurezza (49); sicurezza e la salute (47); regioni e province autonome (25); tutto il territorio nazionale (16); Trento e di Bolzano (12); regioni e delle province (11); regioni e le province (17); le regioni e le (17); nei luoghi di lavoro (215); nel luogo di lavoro (21); negli ambienti di lavoro (76); nell'ambiente di lavoro (20); dei luoghi di lavoro (14); autonome di Trento e (17); e le province autonome (17); di Trento e di (12); in ambiente di lavoro (23); sul luogo del lavoro (52); sui luoghi di lavoro (34); sul posto di lavoro (28); su tutto il territorio (18); ambienti di vita e (10); negli ambienti di vita (12); in tutto il mondo (11);

**f. collocazione temporale:**

la prima volta in (10); nel corso degli anni (16); in questi ultimi anni (16); negli ultimi anni e (19); per la prima volta (40); durante il periodo di (11); di volta in volta (16); di inizio e fine (10); nel corso della sorveglianza (13); nel momento in cui (10);

**g. paragone:** rispetto alla popolazione generale (22).

*III.3.4.3.1.2 I lexical bundles orientati al testo – riguardanti l'organizzazione del testo e il suo significato*

**Al macrolivello**

**a. segnali di cornice:**

un punto di vista (49); punto di vista clinico (10); un punto di riferimento (10); la necessità di un (14); punto di vista della (16); la possibilità di un (10); presenza o meno di (13); per quanto riguarda la (107); per quanto riguarda il (73); per quanto riguarda le (50); per quanto riguarda i (39); per quanto riguarda gli (36); per quanto riguarda l' (36); per quello che riguarda (15); per quanto concerne la (14); per quanto concerne le (13); per quanto concerne il (20); particolare per quanto riguarda (15); quanto riguarda la valutazione (10); soprattutto per quanto riguarda la (18); di fronte ad una (11); a

fronte di un (20); a fronte di una (20); al di sotto dei (16); al di sotto di (11); al di sotto del (12); al di sotto della (12); in questo caso il (18); in questo caso e (12); in presenza di una (15); in particolare per la (14); in questo caso la (11); in un ottica di (11); nel corso di un (10); in linea con il (13); in presenza di un (13); in questo caso si (13); in particolare per quanto (20); in base a quanto (10); in un campione di (10); in un contesto di (10); dal punto di vista (152); all'interno di un (12); da un punto di (48); da una serie di (12); anche in questo caso (50); di questo tipo di (11); per la presenza di (18); per una serie di (12); ad una serie di (10); sulla base di un (11); con particolare riferimento alla (14); con particolare riferimento alle (12); sia in termini di (25); nel caso in cui (38); nei casi in cui (25); sulla base di quanto (11); sia per quanto riguarda (27); sulla base di queste (10); tra i livelli di (12); su un campione di (30); su un gruppo di (18); di un sistema di (17); ad una condizione di (11); in accordo con la (11); in accordo con i (11); in accordo con quanto (14); in maniera tale da (11); in collaborazione con gli (10); in collaborazione con l' (10); in collaborazione con il (12); ai fini di una (10); a causa di una (11); da parte di un (19);

**b. segnali di passaggio discorsivo:** da un lato e (10); se da un lato (14);

**c. segnali di risultato:**

risultati della sorveglianza sanitaria (22); i risultati della sorveglianza (14); risultati di questo studio (13); l'analisi dei dati (10); una revisione della letteratura (11); i risultati di questo (17); statistica dei dati e (10); i dati relativi alle (12); i dati relativi ai (10); i dati relativi agli (11); una differenza statisticamente significativa (15); una correlazione statisticamente significativa (10); nella tabella I sono (11); sulla base dei dati (16); i dati sono stati (21); i risultati sono stati (13); è emerso che il (11); correlazione statisticamente significativa tra (11); una differenza significativa tra (10); quali ad esempio la (11); ha evidenziato che la (11); è risultato pari a (16); è risultata pari a (11);

**d. segnali strutturativi:**

uno studio condotto da (17); uno studio condotto su (14); conferenza permanente per i (15); in uno studio condotto (22); in un recente studio (10); la ricerca è stata (13); del presente lavoro e (23); del nostro studio e (16); del presente studio e (11); di questo studio e (11); in sede di conferenza (10); lo studio è stato (10); la tabella I riporta (10);

### **Al microlivello**

al fine di verificare (16); al fine di evitare (18); al fine di poter (14); al fine di garantire (10); al fine di ridurre (11); al fine di ottenere (11); al fine di identificare (10); al fine di individuare (12); al fine di valutare (12); allo scopo di individuare (10); in grado di determinare (35); in grado di influenzare (29); in grado di fornire (19); in grado di indurre (14); in grado di garantire (10); in grado di modificare (10); grado di determinare un (11); con lo scopo di (36); di riferimento per la (18); con l'obiettivo di (17); anche al fine di (12); ma anche per la (11); secondo i principi della (20); la presenza di un (34); la presenza di una (16); un vero e proprio (11); i rapporti tra lo (14); e la capacità di (12); e la possibilità di (12); e la necessità di (25); e la sicurezza del (15); e la valutazione delle (15); e la presenza di (16); caso in cui il (12); per la

valutazione dell' (15); per lo sviluppo di (21); per la determinazione di (19); per la produzione di (18); per i rapporti tra (14); per la gestione del (14); per la realizzazione di (14); per la valutazione delle (14); per la valutazione della (14); per la prevenzione nei (13); per la raccolta dei (13); per il miglioramento delle (12); per la medicina del (10); per la prevenzione degli (10); per la prevenzione della (10); per la promozione della (11); per la salute e (81); per la salute dei (24); per la sicurezza e (50); per la prevenzione e (25); per la tutela della (30); per la valutazione del (45); per la valutazione dei (22); per la diagnosi di (31); sui rischi per la (11); di lavoro in cui (10); è in grado di (65); sono in grado di (55); essere in grado di (30); sia in grado di (18); è stato quello di (34); è di rischio e (11); è più in generale (10); ha il compito di (16); deve riguardare tutti (11); hanno la possibilità di (10); conto del fatto che (10); più in grado di (11).

#### *III.3.4.3.1.3 I lexical bundles orientati al partecipante – si concentrano sullo scrivente e sul lettore del testo*

##### **a. caratteristiche posizionali:**

che si tratta di (11); si tratta di un (32); si tratta di una (17); si ritiene che la (13); non è stato possibile (16); si è ritenuto opportuno (10); si può affermare che (10);

##### **b. caratteristiche di coinvolgimento:**

scopo del presente lavoro (21); lo scopo di fornire (10); lo scopo del presente (10); è stato dimostrato che (30); ha lo scopo di (14).

#### *III.3.4.3.1.4 Altri bundles*

per i quali e (22); per le quali e (22); per le quali non (12); per i quali non (16); in cui sono stati (13); le quali è obbligatoria (10); quali è obbligatoria la (10).

#### *III.3.4.3.2 I risultati emersi dall'analisi funzionale*

Dalla classificazione dei *lexical bundles* del linguaggio della medicina del lavoro si può desumere il legame del discorso scientifico ai dati e alle relazioni tra di loro, ai risultati, al mondo reale che si presenta come osservabile ed oggettivo. Appunto per questo, si può riscontrare la stragrande maggioranza dei *lexical bundles* connessi allo



svolgimento di ricerca, alla presentazione dell'argomento di ricerca e ai limiti posti nell'ambito di essa, e infine all'analisi di dati.

La medicina del lavoro è una "branca della medicina che si occupa dello studio, della prevenzione e della terapia di malattie derivanti dalle attività lavorative. È strettamente collegata, per gli aspetti normativi e prescrittivi, al diritto del lavoro. Il tradizionale campo di interesse della medicina del lavoro è lo studio degli effetti sulla salute di fattori (chimici, fisici, biologici) cui i lavoratori sono esposti nell'ambiente dove svolgono la loro attività<sup>44</sup>". La medicina del lavoro è una scienza applicativa, una branca della medicina clinica che può essere comparata ad una scienza naturale.

Le scienze naturali non sono tanto legate all'argomentare, quanto piuttosto alla presentazione dei fatti, dei dati e di conseguenza alla dimostrazione di nuove formulazioni o di ipotesi da verificare o giustificare (Hyland, 2008a).

La formulazione di nuove conclusioni e la dimostrazione empirica impongono un ragionamento deduttivo, dominato dalla realtà sperimentata, dalle conoscenze attuali, nel quale non c'è posto per un argomentare persuasivo. In questo contesto sia il ricercatore che l'uditore vengono messi in ombra.

Il linguaggio è preciso e molto condensato, orientato ai fatti e ai dati, oggettivo, caratterizzato da un'economia del linguaggio, espressa nei termini di un'eliminazione di elementi pleonastici sul piano sintattico (Serianni, 2005).

Nel linguaggio della medicina del lavoro analizzato emergono esempi di *lexical bundles* che rispecchiano l'ideologia scientifica, ovvero il ragionamento deduttivo e dimostrativo legato alle situazioni specifiche.

Lo svolgimento della ricerca e la sua presentazione dimostrano una predominanza di *lexical bundles* orientati allo svolgimento di ricerca, alla descrizione dell'oggetto di ricerca o del contesto specifico, illustrando aspetti inerenti ai modelli di utilizzo, ai materiali adoperati e all'ambiente lavorativo. Questo gruppo di *lexical bundles* include innanzitutto espressioni di quattro parole a struttura nominale. Lo stile nominale assume il ruolo della condensazione informativa dello studio in oggetto. In termini percentuali, questo gruppo copre il 77,1% di tutti i *lexical bundles* emersi dall'estrazione automatica del *corpus*.

---

<sup>44</sup> Definizione tratta dall'Enciclopedia Treccani della versione online – [www.treccani.it](http://www.treccani.it).

Il gruppo maggiormente rappresentato è quello dei *lexical bundles* orientati all'ambito della ricerca, all'argomento centrale della ricerca e agli aspetti inerenti ad essa.

I *lexical bundles* che appartengono a questa classe coprono il 38,9% di tutti i *bundles* identificati all'interno di *lexical bundles* orientati alla ricerca, segnalando in questo modo l'importanza attribuita all'argomento analizzato.

In questa categoria sono stati rilevati *clusters* esclusivamente nominali e preposizionali.

Seguendo il criterio dell'occorrenza nel *corpus* indichiamo a mo' di esempio alcuni *bundles* appartenenti a questa categoria. Come argomento principale è stato scelto il tema "Il rischio stress lavoro correlato" – i *bundles* che sono emersi dall'estrazione automatica del *corpus* sono:

Giornale di medicina del lavoro ed ergonomia, 2010

La natura del *rischio stress lavoro-correlato*, diverso da altri rischi occupazionali (es. rumore) affrontabili con metodi e strumenti di misura standardizzati, richiede l'applicazione di strumenti condivisi da tutti i soggetti aziendali [...]

La valutazione dello stress costituisce parte imprescindibile per *la tutela della salute* in ambito lavorativo.

I datori di lavoro devono garantire attivamente il miglioramento della salute e del benessere dei lavoratori in quanto *la promozione della salute* nei luoghi di lavoro [...]

Accanto ai tradizionali rischi per *la salute dei lavoratori* da agenti chimici, fisici, meccanici e biologici, hanno acquisito peculiare rilievo nell'ambito della medicina del lavoro i cosiddetti rischi "trasversali", tra i quali sono inclusi i "rischi di tipo organizzativo" che possono causare "distress e disagio lavorativo".

Riportiamo inoltre i due *lexical bundles* appartenenti a questa categoria che presentano la massima occorrenza nel *corpus*, superiore a 100 volte.

Per ogni lavoratore sottoposto a sorveglianza sanitaria si decidono le eventuali altre analisi da effettuare, alla luce delle più recenti conoscenze disponibili in materia *di medicina del lavoro*.

Proprio per questo *la valutazione del rischio* deve tener conto del fatto che il tipo di riprogrammazione dell'espressione genica e chiaramente in funzione della dose dello xenobiotico.

Il secondo sottogruppo maggiormente rappresentato è la categoria dei *bundles* che si riferisce alla descrizione del contesto di ricerca o dei suoi aspetti, dei soggetti sottoposti alla ricerca, dei materiali o degli strumenti utilizzati.

Elenchiamo di seguito alcuni esempi, tenendo presente dalla scelta il criterio della massima occorrenza d'uso. In questa categoria si possono riscontrare espressioni appartenenti al gruppo nominale e preposizionale.

Giornale di medicina del lavoro ed ergonomia, 2008

Altro momento fondamentale è il rapporto che il *Medico del Lavoro* deve instaurare con il medico curante del lavoratore [...]

Il successivo giudizio d'idoneità sarà quindi annotato sul documento sanitario personale, comunicato al lavoratore e trasmesso *al datore di lavoro* a seguito delle valutazioni indicate dal comma ora indicato.

Per quanto concerne i lavoratori essi sono stati suddivisi in base all'età, al sesso, ed alla anzianità lavorativa di comparto. Si sono valutati inoltre i *dispositivi di protezione individuale* in dotazione, l'articolazione degli orari di servizio, l'entità del ricambio del personale, gli indici di malattia e di infortunio [...].

I *bundles* che coprono quasi il 10,1% di tutti i *clusters* identificati e il 15% dei *bundles* orientati alla ricerca, descrivono la procedura tramite la quale una ricerca o un'analisi è stata svolta. Questo gruppo indica quanto valore viene attribuito, nell'ambito di una scienza applicata, allo svolgimento di ricerca.

Consideriamo a titolo d'esempio alcuni *lexical bundles* tratti dal *corpus*. Sono stati scelti i *lexical bundles* che presentano la maggior frequenza d'uso, variabile nella fascia da 120 a 50 volte nell'intero *corpus*. Si è optato per le espressioni con la massima frequenza in tre gruppi strutturali – nominale, preposizionale e verbale, anche se il gruppo nominale con la forma *nome + preposizione DI* copre la maggior parte dei *bundles* identificati.

Giornale di medicina del lavoro ed ergonomia, 2008

[...] tali modifiche sono state apportate sulla base delle informazioni ottenute dall'OECD attraverso *la valutazione del rischio* per la salute umana e per l'ambiente derivante dall'utilizzo di queste molecole in svariati settori.

Commissione Scientifica, istituita ai sensi dell'art. 10 del D.L. 38/2000, al termine dei lavori e dopo una fase di confronto con le parti sociali, ha elaborato la nuova *tabella delle malattie professionali* in industria ed agricoltura di cui agli art. 3 e 211 del T.U [...]

Giornale di medicina del lavoro ed ergonomia, 2009

*Il coefficiente di correlazione* dei PMF nell'affaticamento (CCR) infine ha fornito valori promettenti, riconoscendo come massimali il 90% dei casi [...]

Giornale di medicina del lavoro ed ergonomia, 2009

Infatti durante questa fase è necessario individuare quelle lavorazioni che, anche dopo l'attuazione *delle misure di prevenzione*, mantengono un livello di rischio residuo tale da compromettere la salute dei lavoratori.

I campioni *sono stati analizzati* presso l'ARPA di Reggio Emilia.

L'uso massiccio dei *lexical bundles* orientati alla ricerca indica l'essenza della ideologia scientifica della disciplina analizzata. La medicina del lavoro, una scienza applicata, con delle caratteristiche inerenti alla scienza naturale, enfatizza la ricerca empirica, ponendo come obiettivo primario il progresso della conoscenza, trasformando i risultati dello studio in prodotti e servizi.

La seconda categoria principale maggiormente rappresentata è il gruppo dei *bundles* orientati al testo. In termini di percentuale riveste questo gruppo il 20,7% di tutti i *bundles* identificati. Tale risultato tende a enfatizzare il lato empirico della scienza analizzata. I *bundles* rappresentati in maggior misura sono le sequenze che inquadrano o delimitano la ricerca, seguite da quelle che indicano dei passaggi discorsivi nel testo. La presentazione dello sbocco pratico, ovvero dell'utilizzo dei risultati, presenta *bundles* che delimitano l'ambito nel quale uno studio viene presentato, comparato con gli studi precedenti, e *bundles* che indicano i risultati ottenuti dalla ricerca.

Seguendo il criterio di massima occorrenza d'uso, elenchiamo di seguito i due *bundles* più frequenti, la cui frequenza è superiore a 150 e 100 rispettivamente. Tutti e due i *bundles* delimitano la ricerca o una parte di essa.

Giornale di medicina del lavoro ed ergonomia, 2008

Gli effetti, sia nel caso di un'inattivazione sia di una sovraespressione di miRNA, possono essere importanti *dal punto di vista* della patologia.

*Per quanto riguarda la tossicità e la carcinogenicità a livello epatico, l'attivazione del recettore PPAR. nel modello animale da parte di PFOS e PFOA (per il quale si rimanda al capitolo successivo) [...]*

Il 99% dei *lexical bundles* è da un lato contrassegnato dalla neutralità verso il destinatario, dall'altro il ricercatore viene messo in ombra. Il linguaggio adoperato assume in questo senso un modo di argomentare particolare, tendendo ad una struttura dimostrativa, concentrandosi sulle ricerche, sui fatti, sugli esperimenti e sui risultati raggiunti dalle analisi.

Si presuppone, inoltre, l'appartenenza dell'uditorio ad un gruppo ristretto, agli addetti al lavoro, mettendo pertanto in risalto l'elemento empirico e minimizzando l'importanza sia dell'uditore che del ricercatore, una caratteristica propria del linguaggio scientifico.

Ora possiamo mettere a confronto le funzioni discorsive dei *lexical bundles* delle tre discipline scientifiche analizzate.

### **III.3.5 Comparazione dei *lexical bundles* di tre ambiti scientifici – della storia, filosofia e medicina del lavoro**

A questo punto la ricerca si addentra in una discussione particolareggiata delle somiglianze e divergenze funzionali dei tre ambiti scientifici presi in considerazione, mostrando in che modo i *lexical bundles* di 4 parole contribuiscono alla creazione del discorso accademico. La ricerca mostra come queste sequenze di parole siano elementi discorsivi con delle funzioni identificabili, pur non rappresentando delle unità linguistiche complete.

La prima divergenza riguarda il numero dei *lexical bundles* identificati. Dallo spoglio è emerso che il maggior numero di sequenze di 4 parole è stato rilevato nel corpus di medicina del lavoro, con quasi il 50% in più di *bundles*.

Hyland (2008a) suggerisce che la densità dei *lexical bundles* nelle scienze naturali potrebbe riflettere la dipendenza della retorica dalla rappresentazione visuale e grafica – i *lexical bundles* identificati nelle scienze naturali non sono stati riscontrati in nessun'altra disciplina.

Cortes (2004) osserva che il minor numero di *bundles* nelle discipline umanistiche potrebbe indicare da un lato la confidenza degli autori in una scelta diversificata e più elaborata delle strutture linguistiche, dall'altro, invece, una preferenza per le sequenze più brevi di quattro parole. Come indicato da Hyland (2004), questo risultato riflette le tendenze più discorsive e valutative dell'argomentazione nelle scienze umanistiche, laddove la persuasione è più interpretativa e meno basata sull'empirismo.

Questa supposizione potrebbe essere in parte confermata dal fatto che le prime 50 sequenze di 3 parole riscontrate nel *corpus* di storia compaiono in media più di 180 volte su un milione di parole. Dobbiamo inoltre puntualizzare che tutte le espressioni di 3 parole nel *corpus* di storia hanno una frequenza d'occorrenza superiore a 90.000 su un milione di parole. Lo stesso risultato è riscontrabile nel *corpus* di filosofia, il quale mette alla luce 3488 *lexical bundles* di 3 parole diversi, con 50.065 occorrenze, ovvero circa 64.000 volte su un milione di parole. Il *corpus* di medicina del lavoro si presenta dal punto di vista dell'occorrenza dei *lexical bundles* di 3 parole simile al *corpus* di storia, presentando la stessa frequenza d'occorrenza su un milione di parole, bensì una quantità di *clusters* leggermente più alta.

Questi dati concordano con le osservazioni di Cortes (2004) – la linguista suggerisce che gli autori di storia fanno uso di singoli elementi di vocabolario o di altre espressioni per esprimere la stessa funzione.

Siamo consapevoli che i risultati numerici rappresentano a questo punto solo un'indicazione dell'uso reale degli esperti in materia – ulteriori studi che riguardano sequenze di 3 parole sono necessari per poter giungere a conclusioni definitive.

La seconda discrepanza riguarda la concentrazione di diverse tipologie di *lexical bundles* in tre discipline. Lo spoglio ha mostrato che nel *corpus* di storia i *lexical bundles* rivolti alla ricerca coprono una percentuale del 57,4% di tutti i *lexical bundles*, il maggior numero dei quali può essere attribuito alla classe delle collocazioni temporali, di quelle locali e alle sequenze riferite agli argomenti trattati negli articoli scientifici. La narrazione storica si riverbera proprio nei *bundles* che specificano il periodo, il luogo dell'avvenimento storico o che si riferiscono al tema discusso.

La preferenza per i *lexical bundles* orientati alla ricerca è stata evidenziata anche dalla classificazione dei *lexical bundles* del *corpus* di medicina del lavoro, in cui il 77,1% appartiene a questa categoria. A differenza del *corpus* di storia, il maggior

numero dei *lexical bundles* riferiti alla ricerca, può essere ascritto al gruppo dei *clusters* che si riferiscono agli argomenti trattati nel discorso scientifico. Questo gruppo ricopre quasi il 40% di tutte le sequenze di 4 parole identificate nel *corpus* di medicina del lavoro, evidenziando quanto valore viene attribuito alla presentazione della ricerca stessa. I *lexical bundles* di questa tipologia accentuano i temi della ricerca, descrivono lo svolgimento di essa, gli oggetti adoperati e l'ambiente lavorativo. Inoltre, chiariscono i modelli di utilizzo e lo svolgimento dello studio e determinano particolari aspetti di esso. La preferenza della disciplina scientifica per l'esposizione del mondo empirico culmina nella presentazione di fatti concreti, dimostrabili e verificabili tramite ricerche. La ricerca pratica, concreta e tangibile si rileva nell'utilizzo dei meccanismi discorsivi che la rispecchiano – nel nostro caso dei *lexical bundles* orientati alla ricerca.

Il legame del discorso della medicina del lavoro coi dati e con le relazioni tra di loro, con i risultati e col mondo reale che si presenta come osservabile ed oggettivo si rispecchia nella scelta di *lexical bundles* connessi allo svolgimento di ricerca, alla presentazione dell'argomento di ricerca e ai limiti posti nell'ambito di essa, e infine all'analisi dei dati. Tale risultato tende a enfatizzare il lato empirico della medicina del lavoro. Hyland (2004) indica che nelle scienze naturali la scala di questo uso serve a attribuire alla scrittura un senso maggiore di conformità al reale e di maggiore importanza ai dati di laboratorio.

I dati esposti sono riassunti nella tabella seguente.

Tabella 28: Distribuzione delle funzioni dei *lexical bundles* attraverso le discipline

<b>Disciplina</b>	<b>Orientati alla ricerca</b>	<b>Al testo</b>	<b>Al partecipante</b>	<b>Altri</b>	<b>Totale %</b>
<b>Storia</b>	57,4	38,3	4,3	/	100
<b>Filosofia</b>	14,9	59,8	23,2	2,1	100
<b>Medicina</b>	77,1	20,7	1,4	0,8	100
<b>La media</b>	49,8	39,6	9,6	1	100

Come si evince dalla tabella, i *lexical bundles* orientati al testo trovano maggiore riscontro nelle discipline umanistiche – discipline con una rilevante importanza posta sull'argomentazione, mirante ad un'esplicazione che si basa sui fatti e ragionamenti.

Il testo filosofico, in particolare, si rivolge all'analisi di problemi etici e metafisici, ovvero problemi non incentrati sul mondo fisico, presentando argomenti, ragioni, motivi

e sostenendo una tesi. Questo ragionamento trova riscontro nei *lexical bundles* orientati al testo. L'argomentazione non indica semplicemente una lista di argomenti, ma la loro strutturazione in un ragionamento (Ruffaldi, Trombino, 2004). Il ragionamento argomentativo dell'autore o della comunità scientifica trova maggiore riscontro nei *lexical bundles* orientati al testo. Questi, a loro volta, si condensano particolarmente nei *clusters* che pongono un limite alla ricerca, indicano connessioni, limitazioni o condizioni per l'interpretazione del discorso successivo o precedente oppure specificano casi. Come si vede dalla tabella, il pensiero filosofico presenta il maggior numero di *lexical bundles* orientati al testo e al partecipante, un fatto che riflette il modo di ragionare filosofico.

L'argomentare su una verità condivisa, la quale a sua volta viene discussa, approfondita, analizzata, argomentata, per giungere a un consenso che si basa su fatti e ragionamenti, su una verità non oggettiva, si rispecchia nei *lexical bundles* che incorporano un'idea dell'autore o della comunità scientifica. Questa caratteristica del pensiero accademico è da attribuire principalmente all'ideologia filosofica, all'essenza della disciplina, in cui c'è maggiore spazio per l'espressione dell'opinione, delle valutazioni, di un giudizio da parte dello scrivente. Come già evidenziato dall'analisi delle funzioni discorsive dei *lexical bundles* in filosofia, quasi tutti i *lexical bundles* di questa categoria indicano un atteggiamento dello scrivente.

Da contro, la storia come disciplina umanistica risulta avere un numero abbastanza esiguo di *lexical bundles* orientati al partecipante. Questo fatto rivela la natura della narrazione storica, tutta volta alla puntualizzazione temporale, locale e tematica, cosa che porta a concentrarsi innanzitutto sulle risposte alle domande *Quando, Dove, Chi o Che cosa e Perché*. Lo storico vuole soprattutto assicurare una corretta comprensione dell'avvenimento storico, mettendo in secondo luogo il giudizio o l'atteggiamento dello scrivente.

Il secondo piano di lettura funzionale riguarda le caratteristiche dell'intero *corpus* specialistico. Sommando le percentuali dei *lexical bundles* si ricava che il *corpus* analizzato presenta per frequenza decrescente i *bundles* orientati alla ricerca, al testo, al partecipante e infine altri *bundles*, ovvero strutture senza una funzione identificabile.

Siamo consapevoli che tali dati devono essere analizzati con cautela viste le caratteristiche differenti di ambiti scientifici analizzati. Occorre sottolineare che i



risultati esposti sostengono in maniera puntuale le ricerche svolte per la lingua inglese (Cortes, 2004; Biber, 2006; Hyland 2008 a, 2008b).

Entrambe le lingue, sia l'inglese che l'italiano, mostrano notevoli variazioni nella frequenza dei *bundles*, nella loro struttura e nelle funzioni attraverso le discipline scientifiche del discorso scientifico.

Gli andamenti funzionali sembrano confermare i dati attesi e già riscontrati nella lingua inglese: i *lexical bundles* rappresentano delle unità linguistiche con delle funzioni discorsive ben identificabili. Come tali dimostrano che i modelli di frequenza alta rispecchiano lo status formulaico del linguaggio (Biber, 2006).

Da contro, siamo consapevoli che i dati devono essere presi in considerazione con una certa cautela perché i *corpora* analizzati non coincidono nel numero di parole o *tokens* e nel numero di testi presi in esame. Ciononostante, basandoci sugli studi precedenti e sui risultati emersi, possiamo confidare nel fatto che i risultati ci offrono un panorama delle differenze valide tra gli ambiti scientifici.

Nonostante le variazioni fra i *corpora* in termini di numero di *tokens* e di testi, i risultati indicano che i *lexical bundles* di 4 parole dovrebbero essere apprezzati come peculiarità linguistiche nella produzione del discorso scientifico.

Le loro strutture e la loro variazione attraverso gli ambiti scientifici diversi potrebbero aiutarci a definire meglio una microlingua e i modi in cui gli studiosi adoperano le strutture formulaiche. Inoltre, i risultati di quest'indagine possono contribuire alla comprensione del ruolo dei *lexical bundles* nella strutturazione del discorso accademico.

### **III.4 Conclusioni**

Il filone di ricerca sui *lexical bundles*, volto ad analizzare le combinazioni ripetute di parole contigue, o le sequenze di quattro parole nei registri scritti e parlati, si basa sull'analisi di grandi quantità di testi prodotti in maniera naturale. I *corpora* sono adoperati per individuare elementi comuni e divergenti ai vari registri e ai vari tipi di discorso. Le sequenze di 4 parole rilevate sono così frequenti che i dati prodotti dall'esplorazione di un *corpus* richiedono una spiegazione (Biber ed altri, 2004).

Anche se i *lexical bundles* non sono né espressioni idiomatiche né espressioni strutturalmente complete, bensì piuttosto delle unità frammentate, e come tali non riconosciute dalla grammatica tradizionale, sono blocchi importanti nella creazione del discorso. I *bundles* costituiscono dei blocchi costitutivi di due segmenti, di clausole o frasi più lunghe, manifestandosi come una cornice per le informazioni seguenti (Biber, 2009). I ricercatori (Biber ed altri, 2004) vedono i *lexical bundles* come "cornici" strutturali che dicono al lettore o all'ascoltatore come intendere le informazioni che seguono in termini di organizzazione del discorso e del mondo reale, dell'atteggiamento dello scrivente e del coinvolgimento del lettore. La prova più generale circa l'importanza dei *lexical bundles* proviene dalla loro frequenza d'uso e da ovvie funzioni del discorso.

Le analisi finora fatte e qui proposte rappresentano solo un avvio e un suggerimento per le ricerche future, sottolineando una caratteristica del linguaggio che si configura in una distribuzione diversa nei registri e nelle discipline scientifiche differenti. Questa diversa distribuzione fornisce una prima indicazione sull'importanza delle strutture precostituite in quanto elemento comunicativo.

Il presente studio rappresenta un primo passo nell'esplorazione dei *lexical bundles* in lingua italiana. Si è pertanto consapevoli del fatto che sono necessari ulteriori studi per poter meglio elaborare il raggruppamento di queste sequenze, sia dal punto di vista strutturale che da quello funzionale. L'analisi qui presente mette in risalto che i *lexical bundles* sono una caratteristica saliente del linguaggio, sia parlato che scritto. Questa salienza richiede comunque uno studio più approfondito e forse orientato all'analisi di sequenze di diversa lunghezza, dal momento che la formulaicità del linguaggio figura come una caratteristica complessa del linguaggio.

## Appendici

**Appendice 1: Lexical bundles di 4 parole in sequenza nel corpus LIP**

LIP - lexical bundles di 4 parole				
N	Lexical bundle	Frequenza	Articoli	%
1	VA BENE VA BENE	46	26	8.609271049
2	DAL PUNTO DI VISTA	43	28	9.271523476
3	NO NONONO	36	19	6.291390896
4	IO NON LO SO	35	22	7.284768105
5	È UNA COSA CHE	27	21	6.953642368
6	A UN CERTO PUNTO	25	22	7.284768105
7	NEL MOMENTO IN CUI	25	22	7.284768105
8	UN PUNTO DI VISTA	22	14	4.635761738
9	AL DI LÀ DI	21	16	5.29801321
10	SÌ SÌ SÌ SÌ	21	15	4.966887474
11	VA BE VA BE	21	13	4.304635525
12	DA UN PUNTO DI	20	13	4.304635525
13	UN GRUPPO DI LAVORO	19	4	1.324503303
14	CIAO CIAOCIAOCIAO	17	9	2.980132341
15	IL NUMERO DI TELEFONO	17	6	1.986755013
16	PER QUANTO RIGUARDA LA	17	14	4.635761738
17	AL CENTO PER CENTO	16	7	2.317880869
18	IN TUTTO IL MONDO	16	11	3.642384052
19	NON E NON E	16	15	4.966887474
20	ZZZ ZZZZZZZZZ	16	5	1.655629158
21	DA QUESTO PUNTO DI	15	9	2.980132341
22	È UN ALTRA COSA	15	14	4.635761738
23	IL DISEGNO DI LEGGE	15	3	0.993377507
24	IL PROBLEMA È CHE	15	11	3.642384052
25	IN UN CERTO SENSO	15	11	3.642384052
26	NON CE L'HO	15	12	3.973510027
27	QUESTO PUNTO DI VISTA	15	9	2.980132341

28	UN QUARTO D'ORA	15	8	2.649006605
29	DI QUELLO CHE È	14	11	3.642384052
30	E QUESTA È LA	14	13	4.304635525
31	HO CAPITO HO CAPITO	14	9	2.980132341
32	QUELLI CHE SONO I	14	11	3.642384052
33	DI DIDIDI	13	9	2.980132341
34	NON È DETTO CHE	13	10	3.311258316
35	QUELLE CHE SONO LE	13	10	3.311258316
36	UNA COSA DEL GENERE	13	11	3.642384052
37	CHE COSA VUOL DIRE	12	8	2.649006605
38	DI QUESTO TIPO DI	12	7	2.317880869
39	E ALLORA E ALLORA	12	5	1.655629158
40	È LA STESSA COSA	12	12	3.973510027
41	IO CE L'HO	12	10	3.311258316
42	IO GLI HO DETTO	12	8	2.649006605
43	NO NON LO SO	12	12	3.973510027
44	NON LO SO EH	12	11	3.642384052
45	NON LO SO SE	12	9	2.980132341
46	PER QUANTO RIGUARDA IL	12	11	3.642384052
47	BENE VA BENE VA	11	6	1.986755013
48	COME SI PUÒ DIRE	11	8	2.649006605
49	E QUESTO È UN	11	9	2.980132341
50	L'UNICA COSA CHE	11	11	3.642384052
51	LA PRIMA VOLTA CHE	11	9	2.980132341
52	MI HA DETTO CHE	11	10	3.311258316
53	NON È CHE NON	11	10	3.311258316
54	NON LO SO MA	11	8	2.649006605
55	QUESTA È UNA COSA	11	7	2.317880869
56	QUESTO È QUELLO CHE	11	9	2.980132341
57	TANT È VERO CHE	11	10	3.311258316
58	TUTTA UNA SERIE DI	11	9	2.980132341
59	AH NON LO SO	10	7	2.317880869

60	ALLA FINE DELL' ANNO	10	6	1.986755013
61	ALLE DUE E MEZZA	10	8	2.649006605
62	E NON LO SO	10	10	3.311258316
63	E UN PO' DI	10	9	2.980132341
64	IL CINQUANTA PER CENTO	10	7	2.317880869
65	IL DISCORSO È QUESTO	10	7	2.317880869
66	M'HA DETTO CHE	10	9	2.980132341
67	NEL MODO PIÙ ASSOLUTO	10	5	1.655629158
68	NON LO SO IO	10	8	2.649006605
69	PER QUANTO RIGUARDA L'	10	8	2.649006605
70	QUESTO È UN ALTRO	10	8	2.649006605
71	SÌ SÌ VA BENE	10	7	2.317880869
72	UN PO' DI PIÙ	10	10	3.311258316
73	AL MOMENTO IN CUI	9	3	0.993377507
74	CUI USCIAMO DA QUESTA	9	3	0.993377507
75	DA TUTTE LE PARTI	9	8	2.649006605
76	E CHE NE SO	9	7	2.317880869
77	E LA COSA PIÙ	9	7	2.317880869
78	E QUESTO È IL	9	7	2.317880869
79	È UN DISCORSO CHE	9	7	2.317880869
80	È UN PROBLEMA CHE	9	8	2.649006605
81	È UNA COSA MOLTO	9	9	2.980132341
82	HO CAPITO VA BENE	9	6	1.986755013
83	I GRUPPI DI LAVORO	9	3	0.993377507
84	IN CUI USCIAMO DA	9	3	0.993377507
85	LA COSA PIÙ IMPORTANTE	9	6	1.986755013
86	MA NON È CHE	9	9	2.980132341
87	MI RENDO CONTO CHE	9	7	2.317880869
88	MOMENTO IN CUI USCIAMO	9	3	0.993377507
89	NON È CHE SI	9	8	2.649006605
90	NON È UN PROBLEMA	9	9	2.980132341
91	NON È UNA COSA	9	6	1.986755013

92	PER LA PRIMA VOLTA	9	8	2.649006605
93	PER QUANTO RIGUARDA LE	9	8	2.649006605
94	PUNTO DI VISTA DELLA	9	8	2.649006605
95	QUELLO CHE È IL	9	8	2.649006605
96	QUESTO È UN PROBLEMA	9	7	2.317880869
97	SITUAZIONE IN MEDIO ORIENTE	9	3	0.993377507
98	UN PO' DI TEMPO	9	7	2.317880869
99	USCIAMO DA QUESTA SCUOLA	9	3	0.993377507
100	VA BENE CIAO CIAO	9	5	1.655629158
101	C'È UN PO' DI	9	8	2.649006605
102	ALL'INTERNO DI QUESTO	8	7	2.317880869
103	ALL'ORDINE DEL GIORNO	8	6	1.986755013
104	ALLORA VEDIAMO UN PO'	8	4	1.324503303
105	CHE CI SIA UN	8	7	2.317880869
106	CHE È UNA COSA	8	6	1.986755013
107	E NELLO STESSO TEMPO	8	8	2.649006605
108	È UN PROBLEMA DI	8	5	1.655629158
109	E VA BE'È	8	7	2.317880869
110	EH PER QUANTO RIGUARDA	8	6	1.986755013
111	IL SABATO E LA	8	3	0.993377507
112	IO NON L'HO	8	6	1.986755013
113	LO SO LO SO	8	7	2.317880869
114	NO NONO IO	8	8	2.649006605
115	NO NO VA BE	8	8	2.649006605
116	NON LO SO NON	8	8	2.649006605
117	PER QUELLO CHE RIGUARDA	8	7	2.317880869
118	QUELLO CHE È STATO	8	8	2.649006605
119	SÌ SÌ SÌÈ	8	8	2.649006605
120	SÌ SÌ SÌ VA	8	8	2.649006605
121	UN'ALTRA COSA CHE	8	8	2.649006605
122	UN MILIONE E MEZZO	8	6	1.986755013
123	UN MOMENTO IN CUI	8	7	2.317880869

124	VA BE MA NON	8	7	2.317880869
125	A CHE FARE CON	7	7	2.317880869
126	CHE A QUESTO PUNTO	7	6	1.986755013
127	CHE COS'È CHE	7	7	2.317880869
128	CHE COSA CHE COSA	7	4	1.324503303
129	CI TROVIAMO DI FRONTE	7	4	1.324503303
130	COSE DI QUESTO GENERE	7	4	1.324503303
131	DI QUELLA CHE È	7	6	1.986755013
132	DI QUELLI CHE SONO	7	6	1.986755013
133	DUE O TRE VOLTE	7	6	1.986755013
134	E CHE CI SONO	7	7	2.317880869
135	È LA PRIMA VOLTA	7	6	1.986755013
136	EH NON È CHE	7	5	1.655629158
137	EH NON LO SO	7	6	1.986755013
138	HA DETTO CHE NON	7	6	1.986755013
139	HO CAPITO VA BE	7	7	2.317880869
140	IN MODO TALE CHE	7	6	1.986755013
141	IO MI RENDO CONTO	7	6	1.986755013
142	IO NON SO SE	7	6	1.986755013
143	MA QUESTO È UN	7	7	2.317880869
144	MI SEMBRA CHE NON	7	7	2.317880869
145	NEI LUOGHI DI LAVORO	7	4	1.324503303
146	NO MA NON È	7	7	2.317880869
147	NO NONO NON	7	7	2.317880869
148	NON CE L'HA	7	5	1.655629158
149	NON È CHE CI	7	6	1.986755013
150	NON È SOLO UN	7	5	1.655629158
151	PER QUANTO RIGUARDA I	7	6	1.986755013
152	PER TUTTA LA VITA	7	6	1.986755013
153	PERCHÉ È UNA COSA	7	6	1.986755013
154	QUELLA CHE È LA	7	6	1.986755013
155	QUINDI QUESTO È UN	7	5	1.655629158



156	SABATO E LA DOMENICA	7	3	0.993377507
157	TUTTE LE VOLTE CHE	7	6	1.986755013
158	UN ANNO E MEZZO	7	5	1.655629158
159	UNA COSA UN PO'	7	7	2.317880869
160	VA BE MA IO	7	7	2.317880869
161	A MENO CHE NON	6	6	1.986755013
162	A PORTATA DI MANO	6	3	0.993377507
163	A QUELLA CHE È	6	6	1.986755013
164	A QUELLO CHE È	6	6	1.986755013
165	A UN CERTO MOMENTO	6	5	1.655629158
166	AH SÌ SÌ SÌ	6	6	1.986755013
167	ALL'INTERNO DI UN	6	5	1.655629158
168	ALLORA A QUESTO PUNTO	6	6	1.986755013
169	ANCHE SITUAZIONE IN MEDIO	6	3	0.993377507
170	CE L'HO IO	6	5	1.655629158
171	C'È ANCHE UN'ALTRA	6	6	1.986755013
172	C'È NIENTE DA FARE	6	6	1.986755013
173	CENTINAIA DI MIGLIAIA DI	6	3	0.993377507
174	CHE COS'È E	6	6	1.986755013
175	CHE NON È CHE	6	5	1.655629158
176	CHE NON È SOLO	6	4	1.324503303
177	CHE VOGLIONO ANDARE DI	6	3	0.993377507
178	CHI L'HA DETTO	6	4	1.324503303
179	CIOÉ NON È CHE	6	6	1.986755013
180	COM'È CHE SI	6	5	1.655629158
181	COSA SUCCEDA AL MOMENTO	6	3	0.993377507
182	D'ALTRA PARTE NON	6	5	1.655629158
183	DA PARTE DEI PROFESSORI	6	3	0.993377507
184	DA QUESTA SCUOLA PUÒ	6	3	0.993377507
185	DA UN'ALTRA PARTE	6	6	1.986755013
186	DEI GRUPPI DI LAVORO	6	3	0.993377507
187	DI CHE COSA SI	6	6	1.986755013

188	DI LAVORO E FORMATO	6	3	0.993377507
189	DI NOMI TIPO EH	6	3	0.993377507
190	DI UNA PERSONA CHE	6	4	1.324503303
191	DIREI CHE PER OGNI	6	3	0.993377507
192	E A QUESTO PUNTO	6	6	1.986755013
193	E ANCHE UN'ALTRA	6	6	1.986755013
194	È FORMATO DA SEI	6	3	0.993377507
195	È IL FATTO CHE	6	5	1.655629158
196	È IN GRADO DI	6	6	1.986755013
197	E NIENTE DA FARE	6	6	1.986755013
198	È PER QUELLO CHE	6	5	1.655629158
199	È SEMPRE STATO COSÌ	6	3	0.993377507
200	È UN CASO CHE	6	4	1.324503303
201	È UN È UN	6	6	1.986755013
202	È UNA COSA IMPORTANTE	6	5	1.655629158
203	È UNA COSA UN	6	6	1.986755013
204	E VA BE MA	6	6	1.986755013
205	È VERO È VERO	6	5	1.655629158
206	ECCO QUESTO È UN	6	6	1.986755013
207	FARA DEI GRUPPI DI	6	3	0.993377507
208	FAREI NEL PRIMO GIORNO	6	3	0.993377507
209	FATTO DA UN ALUNNO	6	3	0.993377507
210	FORMATO DA SEI PERSONE	6	3	0.993377507
211	GRUPPO DI LAVORO CHE	6	3	0.993377507
212	GRUPPO DI LAVORO È	6	3	0.993377507
213	GUARDARE CHI È ASSENTE	6	3	0.993377507
214	HA LE SUE IDEE	6	3	0.993377507
215	HA QUALCOSA DA DIRE	6	6	1.986755013
216	I VARI GRUPPI DI	6	3	0.993377507
217	IL FATTO CHE EH	6	5	1.655629158
218	IL FATTO CHE NON	6	6	1.986755013
219	IL PROBLEMA È QUESTO	6	6	1.986755013

220	IN UN ALTRO MODO	6	5	1.655629158
221	IO NON CREDO CHE	6	6	1.986755013
222	IO NON È CHE	6	6	1.986755013
223	IO NON MI SENTO	6	5	1.655629158
224	L'HO DETTO PRIMA	6	3	0.993377507
225	L'ULTIMA COSA CHE	6	6	1.986755013
226	LA POLITICA UN ALTRO	6	3	0.993377507
227	LA VIOLENZA NEGLI STADI	6	3	0.993377507
228	LAVORO È FORMATO DA	6	3	0.993377507
229	LISTA DI NOMI TIPO	6	3	0.993377507
230	ME L'HA DETTO	6	6	1.986755013
231	NO NO MA NON	6	6	1.986755013
232	NO NO NO È	6	5	1.655629158
233	NO NON L'HO	6	5	1.655629158
234	NON CE LA FA	6	3	0.993377507
235	NON È CHE IO	6	6	1.986755013
236	NON L'HO MAI	6	3	0.993377507
237	NON LO SO COMUNQUE	6	6	1.986755013
238	NON ME LO RICORDO	6	5	1.655629158
239	NON SI SA MAI	6	5	1.655629158
240	OGNUNO HA LE SUE	6	3	0.993377507
241	ORA SO NOVE ARGOMENTI	6	3	0.993377507
242	PER ORA SO NOVE	6	3	0.993377507
243	PERCHÉ SE NO È	6	6	1.986755013
244	PUNTO DI VISTA ECONOMICO	6	5	1.655629158
245	QUANDO SI PARLA DI	6	5	1.655629158
246	QUARTO D'ORA MA	6	3	0.993377507
247	QUELLI CHE VOGLIONO ANDARE	6	3	0.993377507
248	QUELLO CHE È SUCCESSO	6	5	1.655629158
249	QUELLO CHE PUÒ ESSERE	6	3	0.993377507
250	QUESTA È LA MIA	6	6	1.986755013
251	QUESTA È UN'ALTRA	6	5	1.655629158

252	QUESTA SCUOLA PUÒ ESSERE	6	3	0.993377507
253	QUESTO È UN PO'	6	6	1.986755013
254	QUINDI NON È CHE	6	6	1.986755013
255	SCUOLA PUÒ ESSERE UN	6	3	0.993377507
256	SE UN GRUPPO DI	6	3	0.993377507
257	SÌ CE L'HO	6	6	1.986755013
258	SI FARÀ DEI GRUPPI	6	3	0.993377507
259	SÌ GUARDA CHI È	6	3	0.993377507
260	SÌ SÌ NO NO	6	6	1.986755013
261	SI TRATTA DI UN	6	6	1.986755013
262	SUCCEDE AL MOMENTO IN	6	3	0.993377507
263	TE L'HO DETTO	6	6	1.986755013
264	TUTTE LE PERSONE CHE	6	4	1.324503303
265	UNA COSA CHE NON	6	6	1.986755013
266	UNA LISTA DI NOMI	6	3	0.993377507
267	UNO DUE TRE QUATTRO	6	5	1.655629158
268	VA FATTO DA UN	6	3	0.993377507
269	VA IN CLASSE NORMALMENTE	6	3	0.993377507
270	VARI GRUPPI DI LAVORO	6	3	0.993377507
271	A PARTE IL FATTO	5	4	1.324503303
272	A QUESTO PUNTO DI	5	5	1.655629158
273	A QUESTO PUNTO EH	5	5	1.655629158
274	A QUESTO PUNTO NON	5	5	1.655629158
275	AH E VA BE	5	3	0.993377507
276	AH LO SO MA	5	4	1.324503303
277	AH VA BE VA	5	4	1.324503303
278	AL DI LÀ DEL	5	5	1.655629158
279	ALL'UNA E MEZZA	5	3	0.993377507
280	ALLE PRIME DIECI PERSONE	5	4	1.324503303
281	ANCHE DAL PUNTO DI	5	4	1.324503303
282	BE E VA BE	5	4	1.324503303
283	BE MA NON C'È	5	5	1.655629158

284	BENE VA BENE CIAO	5	3	0.993377507
285	CAPITO SÌ SÌ SÌ	5	5	1.655629158
286	CHE È QUELLO CHE	5	5	1.655629158
287	CHE È STATO FATTO	5	5	1.655629158
288	CHE FANNO PARTE DELLA	5	5	1.655629158
289	CHE IN QUESTO MOMENTO	5	3	0.993377507
290	CHE NE SO IO	5	3	0.993377507
291	CHE NON È UN	5	5	1.655629158
292	CHE NON L'HO	5	5	1.655629158
293	CHE VEDETE IN SOVRIMPRESSIONE	5	3	0.993377507
294	CHI È CHI È	5	4	1.324503303
295	CI SONO ANCHE DELLE	5	3	0.993377507
296	COME IN QUESTO CASO	5	5	1.655629158
297	COME SI È DETTO	5	5	1.655629158
298	DI FRONTE A UN	5	4	1.324503303
299	DI FRONTE A UNA	5	5	1.655629158
300	DI LÀ DI QUESTO	5	4	1.324503303
301	DI QUELLO CHE SI	5	5	1.655629158
302	DI TUTTO QUELLO CHE	5	5	1.655629158
303	DI UN DI UN	5	3	0.993377507
304	DI UN MILIONE E	5	5	1.655629158
305	DI UNA COSA CHE	5	4	1.324503303
306	DI UNA SERIE DI	5	5	1.655629158
307	DUE UN MILIONE E	5	5	1.655629158
308	È ANCHE VERO CHE	5	3	0.993377507
309	E CHE SI CHIAMA	5	4	1.324503303
310	È IL CASO DI	5	5	1.655629158
311	E POI E POI	5	5	1.655629158
312	E QUESTA È UNA	5	4	1.324503303
313	E QUINDI QUESTO È	5	4	1.324503303
314	È STATO UN PO'	5	5	1.655629158
315	È UN' ALTRA UN'	5	5	1.655629158

316	È VERO CHE NON	5	4	1.324503303
317	EH SÌ SÌ SÌ	5	5	1.655629158
318	EH UNA SERIE DI	5	5	1.655629158
319	EH VEDIAMO UN PO'	5	4	1.324503303
320	GLI HO DETTO MA	5	5	1.655629158
321	GRAZIE CIAO CIAOCIAO	5	5	1.655629158
322	HO L'IMPRESSIONE CHE	5	5	1.655629158
323	IN FIN DEI CONTI	5	3	0.993377507
324	IN QUESTO CASO NON	5	5	1.655629158
325	IN TUTTI I MODI	5	5	1.655629158
326	IN UN MODO O	5	4	1.324503303
327	IN UNA SITUAZIONE DI	5	5	1.655629158
328	IO CREDO CHE IL	5	5	1.655629158
329	IO CREDO CHE QUESTO	5	5	1.655629158
330	IO PENSO CHE SIA	5	5	1.655629158
331	L'HO GIA DETTO	5	5	1.655629158
332	L'UNICA COSA È	5	5	1.655629158
333	LA MAGGIOR PARTE DEI	5	5	1.655629158
334	LA PAROLA AL PRESIDENTE	5	3	0.993377507
335	LA PRIMA COSA CHE	5	4	1.324503303
336	LE PERSONE CHE SONO	5	4	1.324503303
337	LE STESSE COSE E	5	5	1.655629158
338	MA AL DI LÀ	5	5	1.655629158
339	MA IO NON MI	5	5	1.655629158
340	MA NON È VERO	5	5	1.655629158
341	MA NON LO SO	5	5	1.655629158
342	MI HANNO DETTO CHE	5	5	1.655629158
343	MI PARE CHE SIA	5	4	1.324503303
344	NEL SENSO CHE EH	5	3	0.993377507
345	NEL SENSO CHE NON	5	4	1.324503303
346	NO NO VA BENE	5	4	1.324503303
347	NO PERCHÉ NON È	5	5	1.655629158

348	NOI NON È CHE	5	4	1.324503303
349	NON CI SONO PIÙ	5	4	1.324503303
350	NON CI STA NO	5	5	1.655629158
351	NON È CHE LO	5	4	1.324503303
352	NON È IL CASO	5	5	1.655629158
353	NON È RIUSCITO A	5	4	1.324503303
354	NON È SOLO LA	5	3	0.993377507
355	NON È UN CASO	5	3	0.993377507
356	NON HA NON HA	5	4	1.324503303
357	NON L'HO SENTITO	5	5	1.655629158
358	NON LO SO CHE	5	4	1.324503303
359	NON ME L'HA	5	5	1.655629158
360	NUMERI CHE VEDETE IN	5	4	1.324503303
361	PARTE IL FATTO CHE	5	4	1.324503303
362	PER FAR SÌ CHE	5	5	1.655629158
363	PER UN PO' DI	5	5	1.655629158
364	PER UNA SERIE DI	5	5	1.655629158
365	PERÒ A QUESTO PUNTO	5	5	1.655629158
366	PIÙ NON LO SO	5	5	1.655629158
367	POI CI SONO ANCHE	5	5	1.655629158
368	POI UN'ALTRA COSA	5	5	1.655629158
369	PUNTO DI VISTA DI	5	5	1.655629158
370	PUNTO DI VISTA È	5	4	1.324503303
371	QUELLA CHE È STATA	5	3	0.993377507
372	QUELLO CHE È L'	5	5	1.655629158
373	QUELLO CHE HA FATTO	5	4	1.324503303
374	QUELLO CHE HO DETTO	5	5	1.655629158
375	QUELLO CHE STA SUCCEDENDO	5	5	1.655629158
376	QUELLO CHE TI PARE	5	4	1.324503303
377	QUESTA È STATA LA	5	5	1.655629158
378	S'ERA DETTO CHE	5	5	1.655629158
379	SE NON RICORDO MALE	5	5	1.655629158

380	SÌ SÌ VA BE	5	5	1.655629158
381	SI TRATTA DI EH	5	5	1.655629158
382	SI TRATTA DI UNA	5	5	1.655629158
383	TE L'HA DETTO	5	5	1.655629158
384	TROVIAMO DI FRONTE A	5	3	0.993377507
385	UN'ALTRA COSA È	5	5	1.655629158
386	UN'ALTRA UN'ALTRA	5	5	1.655629158
387	UN CERTO NUMERO DI	5	3	0.993377507
388	UN CERTO TIPO DI	5	5	1.655629158
389	UN DATO DI FATTO	5	5	1.655629158
390	UN DISCORSO CHE RIGUARDA	5	3	0.993377507
391	UN PO' DI MUSICA	5	5	1.655629158
392	UN PO' UN PO'	5	4	1.324503303
393	UN PUNTO DI RIFERIMENTO	5	3	0.993377507
394	UNA COSA CHE È	5	5	1.655629158
395	UNA COSA CHE IO	5	4	1.324503303
396	UNA REGOLA DI CONDOTTA	5	3	0.993377507
397	UNA SERIE DI EH	5	5	1.655629158
398	VA BE E VA	5	4	1.324503303
399	VA BE VA BENE	5	5	1.655629158
400	VE L'HO DETTO	5	5	1.655629158
401	VEDIAMO UN PO DI	5	5	1.655629158
402	VENIAMO A CASA VOSTRA	5	3	0.993377507
403	VUOL DIRE VUOL DIRE	5	5	1.655629158
		3118		



**Appendice 2: Lexical bundles di 4 parole in sequenza nel corpus di storia**

Corpus di STORIA – lexical bundles di 4 parole, occorrenza 13				
N	Lexical bundle	Frequenza	Articoli	%
1	DAL PUNTO DI VISTA	203	78	48.44720459
2	SI TRATTA DI UN	105	54	33.54037094
3	PER LA PRIMA VOLTA	102	56	34.78260803
4	A PARTIRE DAGLI ANNI	88	40	24.84472084
5	DELLA SECONDA GUERRA MONDIALE	86	29	18.01242256
6	UNA VERA E PROPRIA	86	50	31.05590057
7	LA SECONDA GUERRA MONDIALE	76	24	14.9068327
8	NEL CORSO DEGLI ANNI	74	28	17.39130402
9	NEL MOMENTO IN CUI	74	41	25.46583939
10	ANCHE IN QUESTO CASO	67	35	21.73913002
11	DEGLI ITALIANI ALL'ESTERO	67	17	10.55900574
12	UN VERO E PROPRIO	66	44	27.32919312
13	ALL INIZIO DEGLI ANNI	62	30	18.63354111
14	ALLA FINE DEGLI ANNI	62	38	23.60248375
15	DEL MERCATO DEL LAVORO	62	21	13.04347801
16	NELLA MAGGIOR PARTE DEI	62	29	18.01242256
17	QUESTO PUNTO DI VISTA	62	34	21.11801147
18	DA QUESTO PUNTO DI	60	32	19.87577629
19	MAGGIOR PARTE DEI CASI	56	26	16.14906883
20	SI TRATTA DI UNA	54	42	26.08695602
21	PER QUANTO RIGUARDA L'	52	27	16.77018547
22	PER LA MAGGIOR PARTE	51	28	17.39130402
23	UN PUNTO DI VISTA	51	32	19.87577629
24	COMUNITA ITALIANA IN GERMANIA	47	6	3.726708174
25	A PARTIRE DALLA FINE	46	20	12.42236042
26	DI FRONTE A UN	45	27	16.77018547
27	SUL MERCATO DEL LAVORO	45	13	8.074534416

28	DI UNA SERIE DI	44	30	18.63354111
29	DI VOLTA IN VOLTA	44	30	18.63354111
30	ALL'INTERNO DI UN	43	30	18.63354111
31	GIOVANI DI ORIGINE ITALIANA	43	7	4.347826004
32	LA MAGGIOR PARTE DEGLI	41	30	18.63354111
33	PER QUANTO RIGUARDA LA	40	23	14.28571415
34	DA UN PUNTO DI	39	27	16.77018547
35	LA MAGGIOR PARTE DELLE	39	31	19.25465775
36	ALL'INTERNO DELLA COMUNITÀ	38	21	13.04347801
37	COME SI È VISTO	37	29	18.01242256
38	LA MAGGIOR PARTE DI	37	21	13.04347801
39	NEGLI STATI UNITI E	37	17	10.55900574
40	AL DI LÀ DELLE	36	24	14.9068327
41	FINE DEGLI ANNI SESSANTA	36	22	13.66459656
42	LA FINE DEGLI ANNI	36	22	13.66459656
43	LA MAGGIOR PARTE DEI	36	25	15.52795029
44	NEL MERCATO DEL LAVORO	36	16	9.937888145
45	PER QUANTO RIGUARDA IL	36	23	14.28571415
46	ANNI CINQUANTA E SESSANTA	35	15	9.316770554
47	IN ENTRAMBI I CASI	35	30	18.63354111
48	MINISTERO DEGLI AFFARI ESTERI	35	11	6.832298279
49	NELLA SECONDA METÀ DEL	35	25	15.52795029
50	ALL'INTERNO DI UNA	34	26	16.14906883
51	SE DA UN LATO	34	20	12.42236042
52	SECONDA METÀ DEGLI ANNI	34	24	14.9068327
53	CON L'INTENTO DI	33	19	11.80124187
54	DELLA COMUNITA ITALIANA IN	33	9	5.590062141
55	E PIÙ IN GENERALE	33	21	13.04347801
56	L'ESISTENZA DI UNA	33	22	13.66459656
57	DELL'EMIGRAZIONE ITALIANA IN	32	11	6.832298279
58	MERCATO DEL LAVORO E	32	14	8.695652008
59	AL DI LÀ DEL	31	27	16.77018547

60	DALLA FINE DEGLI ANNI	31	15	9.316770554
61	ITALIANI NEGLI STATI UNITI	31	16	9.937888145
62	METÀ DEL XIX SECOLO	31	17	10.55900574
63	NEL CORSO DEL TEMPO	31	16	9.937888145
64	ANNI SESSANTA E SETTANTA	30	17	10.55900574
65	DA UN LATO E	30	23	14.28571415
66	DEGLI ITALIANI IN GERMANIA	30	6	3.726708174
67	ITALIANI RESIDENTI ALL ESTERO	30	10	6.21118021
68	NEI PRIMI DECENNI DEL	30	17	10.55900574
69	PER CENTO DELLA POPOLAZIONE	30	14	8.695652008
70	A CHE FARE CON	29	18	11.18012428
71	DA PARTE DI UN	29	20	12.42236042
72	IN TUTTO IL MONDO	29	22	13.66459656
73	COME SI È DETTO	28	20	12.42236042
74	DOPO LA SECONDA GUERRA	28	13	8.074534416
75	MOMENTO IN CUI LA	28	14	8.695652008
76	TUTTA UNA SERIE DI	28	18	11.18012428
77	A TUTTI GLI EFFETTI	27	20	12.42236042
78	AL DI LÀ DI	27	22	13.66459656
79	DEL LAVORO E DELLA	27	9	5.590062141
80	DELLA PRIMA GUERRA MONDIALE	27	17	10.55900574
81	DI UN GRUPPO DI	27	21	13.04347801
82	GLI ITALIANI ALL ESTERO	27	11	6.832298279
83	IL PIÙ DELLE VOLTE	27	19	11.80124187
84	IMMIGRAZIONE ITALIANA IN GERMANIA	27	6	3.726708174
85	IN UN CERTO SENSO	27	17	10.55900574
86	SI TRATTAVA DI UN	27	20	12.42236042
87	TRA LA FINE DEL	27	20	12.42236042
88	AL DI FUORI DELLA	26	21	13.04347801
89	AL DI LÀ DELLA	26	19	11.80124187
90	ALLA FINE DELL OTTOCENTO	26	15	9.316770554
91	È IL CASO DI	26	21	13.04347801

92	FINO A QUEL MOMENTO	26	22	13.66459656
93	IN QUESTO SENSO LA	26	16	9.937888145
94	LA STORIA DELL'EMIGRAZIONE	26	9	5.590062141
95	NEL PAESE D ORIGINE	26	13	8.074534416
96	NELLA PRIMA METÀ DEL	26	23	14.28571415
97	PER QUEL CHE RIGUARDA	26	11	6.832298279
98	TRA LE DUE GUERRE	26	8	4.968944073
99	È IN GRADO DI	25	19	11.80124187
100	FARE I CONTI CON	25	18	11.18012428
101	NEL MONDO DEL LAVORO	25	13	8.074534416
102	PARTIRE DALLA FINE DEGLI	25	12	7.453416348
103	SI PUÒ DIRE CHE	25	22	13.66459656
104	ALL'INTERNO DEL QUALE	24	11	6.832298279
105	CHE A SUA VOLTA	24	18	11.18012428
106	DEGLI ITALIANI A BERLINO	24	6	3.726708174
107	DI ORIGINE ITALIANA E	24	16	9.937888145
108	I FIGLI DI IMMIGRATI	24	2	1.242236018
109	ITALIANA NEGLI STATI UNITI	24	9	5.590062141
110	TRA I DUE PAESI	24	11	6.832298279
111	ALL'INTERNO DELL UNIONE	23	7	4.347826004
112	DEGLI ANNI SETTANTA E	23	13	8.074534416
113	DEGLI IMMIGRATI ITALIANI IN	23	9	5.590062141
114	DEGLI ITALIANI D AMERICA	23	13	8.074534416
115	IL MODO IN CUI	23	15	9.316770554
116	IN UNA SORTA DI	23	15	9.316770554
117	INTERNO DELL UNIONE EUROPEA	23	7	4.347826004
118	L'ESISTENZA DI UN	23	16	9.937888145
119	L'IDEA DI UNA	23	16	9.937888145
120	LA CADUTA DEL MURO	23	8	4.968944073
121	NON È UN CASO	23	16	9.937888145
122	SI TRATTAVA DI UNA	23	13	8.074534416
123	STORIA DELL'EMIGRAZIONE ITALIANA	23	8	4.968944073

124	A PARTIRE DALLA META	22	16	9.937888145
125	AL DI FUORI DEL	22	19	11.80124187
126	AL DI LÀ DEI	22	20	12.42236042
127	DELLA COMUNITÀ ITALIANA DI	22	8	4.968944073
128	E ALLO STESSO TEMPO	22	15	9.316770554
129	FINE DELLA SECONDA GUERRA	22	16	9.937888145
130	PER IL RINNOVO DEL	22	7	4.347826004
131	PER QUANTO RIGUARDA LE	22	16	9.937888145
132	AGLI INIZI DEGLI ANNI	21	15	9.316770554
133	ALL'INTERNO DELLA FAMIGLIA	21	11	6.832298279
134	ANNI VENTI E TRENTA	21	17	10.55900574
135	DURANTE LA SECONDA GUERRA	21	12	7.453416348
136	FINE DEGLI ANNI SETTANTA	21	13	8.074534416
137	IL DIRITTO DI VOTO	21	9	5.590062141
138	IN VIA DI SVILUPPO	21	12	7.453416348
139	LA PRESENZA DI UN	21	13	8.074534416
140	NEGLI ANNI VENTI E	21	15	9.316770554
141	NELLA SECONDA METÀ DEGLI	21	16	9.937888145
142	RUOLO DI PRIMO PIANO	21	14	8.695652008
143	SE È VERO CHE	21	18	11.18012428
144	CADUTA DEL MURO DI	20	8	4.968944073
145	CHE IN ALCUNI CASI	20	14	8.695652008
146	CON L'ECCEZIONE DI	20	16	9.937888145
147	DAL MOMENTO CHE LA	20	16	9.937888145
148	DEL MURO DI BERLINO	20	8	4.968944073
149	DEL PAESE D'ORIGINE	20	10	6.21118021
150	DELLA SECONDA METÀ DEL	20	15	9.316770554
151	DI LAVORO E DI	20	13	8.074534416
152	DI ORIGINE E DI	20	12	7.453416348
153	E IL FATTO CHE	20	17	10.55900574
154	È UN CASO CHE	20	14	8.695652008
155	FINE DEL XIX SECOLO	20	15	9.316770554

156	GLI ANNI CINQUANTA E	20	9	5.590062141
157	IN QUESTO CASO IL	20	16	9.937888145
158	IN QUESTO CASO L'	20	14	8.695652008
159	ITALIANI ALL'ESTERO E	20	11	6.832298279
160	LA CREAZIONE DI UNA	20	11	6.832298279
161	PER CIÒ CHE RIGUARDA	20	13	8.074534416
162	SECONDA METÀ DEL SECOLO	20	15	9.316770554
163	SECONDA METÀ DELL'OTTOCENTO	20	12	7.453416348
164	UN MOMENTO IN CUI	20	9	5.590062141
165	VERSO LA FINE DEGLI	20	13	8.074534416
166	ALLA FINE DEL SECOLO	19	15	9.316770554
167	DEGLI ITALIANI NEL MONDO	19	12	7.453416348
168	I CITTADINI DELL'UNIONE	19	7	4.347826004
169	IL FATTO CHE LE	19	18	11.18012428
170	IN QUESTO CASO È	19	13	8.074534416
171	IN QUESTO MODO SI	19	11	6.832298279
172	LA CREAZIONE DI UN	19	16	9.937888145
173	LA REPUBBLICA FEDERALE TEDESCA	19	7	4.347826004
174	METÀ DEGLI ANNI SETTANTA	19	13	8.074534416
175	PARTIRE DAGLI ANNI SETTANTA	19	9	5.590062141
176	PUNTO DI VISTA DELLA	19	14	8.695652008
177	UN RUOLO DI PRIMO	19	15	9.316770554
178	VOTO DEGLI ITALIANI ALL'	19	9	5.590062141
179	AL DI LÀ DELL'	18	16	9.937888145
180	AL MERCATO DEL LAVORO	18	8	4.968944073
181	AL PAESE D ORIGINE	18	14	8.695652008
182	ALL INIZIO DEL NOVECENTO	18	12	7.453416348
183	ALLA METÀ DEGLI ANNI	18	12	7.453416348
184	CON L'INTENZIONE DI	18	15	9.316770554
185	CON UNA SERIE DI	18	12	7.453416348
186	DALLA METÀ DEGLI ANNI	18	12	7.453416348
187	DEGLI ANNI CINQUANTA E	18	15	9.316770554

188	DI VITA E DI	18	11	6.832298279
189	GLI ITALIANI E I	18	13	8.074534416
190	IL FATTO CHE LA	18	14	8.695652008
191	IL RINNOVO DEL PARLAMENTO	18	6	3.726708174
192	IMMIGRATI ITALIANI IN GERMANIA	18	6	3.726708174
193	IN QUESTO CASO LA	18	14	8.695652008
194	LA PRIMA GUERRA MONDIALE	18	13	8.074534416
195	NEGLI ANNI CINQUANTA E	18	11	6.832298279
196	NEI CONFRONTI DEGLI ITALIANI	18	13	8.074534416
197	NEL CORSO DELL'OTTOCENTO	18	12	7.453416348
198	NEL TEMPO E NELLO	18	18	11.18012428
199	NELLA MISURA IN CUI	18	17	10.55900574
200	PER QUANTO CONCERNE LA	18	18	11.18012428
201	PRIMI ANNI DEL NOVECENTO	18	8	4.968944073
202	PRIMI DECENNI DEL NOVECENTO	18	8	4.968944073
203	SI È DECISO DI	18	8	4.968944073
204	TEMPO E NELLO SPAZIO	18	18	11.18012428
205	UN TITOLO DI STUDIO	18	11	6.832298279
206	DA UNA SERIE DI	17	16	9.937888145
207	DEGLI ITALIANI IN FRANCIA	17	8	4.968944073
208	DI FRONTE A UNA	17	17	10.55900574
209	DI LÌ A POCO	17	14	8.695652008
210	DI UN SISTEMA DI	17	16	9.937888145
211	DI UN VERO E	17	11	6.832298279
212	E PER CERTI VERSI	17	11	6.832298279
213	FRA I DUE PAESI	17	8	4.968944073
214	I CONTI CON LA	17	13	8.074534416
215	IL MINISTERO DEGLI ESTERI	17	5	3.105590105
216	IL MOTIVO PER CUI	17	12	7.453416348
217	IN UN MOMENTO IN	17	9	5.590062141
218	INIZIO DEGLI ANNI SESSANTA	17	12	7.453416348
219	L'IDEA DI UN	17	13	8.074534416

220	LA COMUNITÀ ITALIANA IN	17	11	6.832298279
221	LA COSTRUZIONE DI UNA	17	16	9.937888145
222	LA STORIA DELLE MIGRAZIONI	17	10	6.21118021
223	MAGGIOR PARTE DI LORO	17	10	6.21118021
224	METÀ DEGLI ANNI SESSANTA	17	14	8.695652008
225	NEL SETTORE DELLA RISTORAZIONE	17	8	4.968944073
226	PER IL FATTO CHE	17	15	9.316770554
227	PER QUANTO RIGUARDA I	17	15	9.316770554
228	PUNTO DI VISTA DELL'	17	15	9.316770554
229	SECONDA E TERZA GENERAZIONE	17	9	5.590062141
230	VERSO GLI STATI UNITI	17	7	4.347826004
231	A CAUSA DI UNA	16	10	6.21118021
232	A FRONTE DI UNA	16	14	8.695652008
233	AL MOMENTO DELLA PARTENZA	16	9	5.590062141
234	ALL'INTERNO DI QUESTO	16	13	8.074534416
235	BEN AL DI LÀ	16	12	7.453416348
236	COMUNITÀ ITALIANE ALL ESTERO	16	7	4.347826004
237	DA UNA PARTE E	16	13	8.074534416
238	DATI IN NOSTRO POSSESSO	16	7	4.347826004
239	E IN MODO PARTICOLARE	16	9	5.590062141
240	FINE DEGLI ANNI CINQUANTA	16	13	8.074534416
241	FORME DI PARTECIPAZIONE POLITICA	16	10	6.21118021
242	GLI UFFICI DEL LAVORO	16	6	3.726708174
243	IL MINISTERO DEL LAVORO	16	5	3.105590105
244	IN BASE AL QUALE	16	14	8.695652008
245	LA PRESENZA ITALIANA IN	16	9	5.590062141
246	LA QUALITÀ DELLA VITA	16	7	4.347826004
247	MA ALLO STESSO TEMPO	16	15	9.316770554
248	META DEGLI ANNI OTTANTA	16	15	9.316770554
249	NEGLI ANNI SESSANTA E	16	9	5.590062141
250	NEL CASO IN CUI	16	13	8.074534416
251	NEL CORSO DEGLI ULTIMI	16	11	6.832298279



252	NEL CORSO DI UNA	16	14	8.695652008
253	PAESI IN VIA DI	16	11	6.832298279
254	SONO IN GRADO DI	16	14	8.695652008
255	SULLA BASE DI UNA	16	15	9.316770554
256	TRA LE ALTRE COSE	16	12	7.453416348
257	UFFICI PROVINCIALI DEL LAVORO	16	5	3.105590105
258	UN CERTO NUMERO DI	16	16	9.937888145
259	A CAUSA DELLA LORO	15	15	9.316770554
260	A DIFFERENZA DI QUANTO	15	15	9.316770554
261	A NEW YORK E	15	11	6.832298279
262	AL DI FUORI DEI	15	13	8.074534416
263	ALL'INTERNO DELLA SOCIETÀ	15	14	8.695652008
264	ALLA RICERCA DI UNA	15	11	6.832298279
265	COME NEL CASO DELLA	15	12	7.453416348
266	COME NEL CASO DI	15	13	8.074534416
267	CON IL COMPITO DI	15	11	6.832298279
268	DA PARTE DI CHI	15	14	8.695652008
269	DAI PRIMI ANNI DEL	15	5	3.105590105
270	DEGLI IMMIGRATI ITALIANI NEL	15	10	6.21118021
271	DEL DIRITTO DI VOTO	15	14	8.695652008
272	DI CUI SI È	15	12	7.453416348
273	DI ORIGINE ITALIANA CHE	15	8	4.968944073
274	DI UN PROCESSO DI	15	10	6.21118021
275	DIRITTI DI CITTADINANZA POLITICA	15	3	1.863354087
276	DOPO LA CADUTA DEL	15	6	3.726708174
277	E AL TEMPO STESSO	15	14	8.695652008
278	E I PROCESSI DI	15	8	4.968944073
279	IL MAGGIOR NUMERO DI	15	12	7.453416348
280	IL TASSO DI DISOCCUPAZIONE	15	5	3.105590105
281	IN TUTTA LA SUA	15	13	8.074534416
282	ITALIANI E IN PARTICOLARE	15	9	5.590062141
283	ITALIANO NEGLI STATI UNITI	15	7	4.347826004

284	LA NECESSITÀ DI UNA	15	15	9.316770554
285	MOMENTO IN CUI SI	15	13	8.074534416
286	NEGLI STATI UNITI IN	15	10	6.21118021
287	NELL'AMBITO DELLA RICERCA	15	12	7.453416348
288	PER CENTO DEGLI ITALIANI	15	6	3.726708174
289	PUNTO DI VISTA LA	15	11	6.832298279
290	TRA GLI ANNI CINQUANTA	15	6	3.726708174
291	TRA L'ITALIA E	15	9	5.590062141
292	VOTO ATTIVO E PASSIVO	15	6	3.726708174
293	A CAUSA DELLA GUERRA	14	11	6.832298279
294	A FARE I CONTI	14	11	6.832298279
295	A PARTIRE DAI PRIMI	14	8	4.968944073
296	AL DI FUORI DELL	14	13	8.074534416
297	AL DI FUORI DI	14	13	8.074534416
298	ANNI IN CUI LA	14	11	6.832298279
299	DA PARTE DEGLI IMMIGRATI	14	12	7.453416348
300	DA PARTE DEL GOVERNO	14	11	6.832298279
301	DA PARTE DELLE AUTORITA	14	11	6.832298279
302	DA PARTE DELLE ISTITUZIONI	14	11	6.832298279
303	DA UN LATO LA	14	10	6.21118021
304	DEL PAESE DI ACCOGLIENZA	14	8	4.968944073
305	DELL'IMMIGRAZIONE DI MASSA	14	11	6.832298279
306	DELLE CONDIZIONI DI VITA	14	13	8.074534416
307	DI ANNA MARIA MARTELLONE	14	4	2.484472036
308	DI UNA VERA E	14	14	8.695652008
309	DURANTE IL REGNO DI	14	5	3.105590105
310	E I LORO DISCENDENTI	14	10	6.21118021
311	E LA CAPACITÀ DI	14	8	4.968944073
312	È POSSIBILE AFFERMARE CHE	14	5	3.105590105
313	ENTRATA IN VIGORE DELLA	14	8	4.968944073
314	FINE DEGLI ANNI VENTI	14	12	7.453416348
315	GLI ANNI SESSANTA E	14	11	6.832298279

316	I LEGAMI CON LA	14	11	6.832298279
317	IL LUOGO IN CUI	14	13	8.074534416
318	IL NUMERO DEGLI ITALIANI	14	11	6.832298279
319	IL PAESE D'ORIGINE	14	8	4.968944073
320	IL PAESE DI PROVENIENZA	14	11	6.832298279
321	IL RISULTATO DI UNA	14	12	7.453416348
322	IN CERCA DI LAVORO	14	12	7.453416348
323	IN CUI SI SONO	14	9	5.590062141
324	IN PRIMO LUOGO LA	14	13	8.074534416
325	IN UNA SERIE DI	14	11	6.832298279
326	L ENTRATA IN GUERRA	14	8	4.968944073
327	LA FINE DELLA GUERRA	14	10	6.21118021
328	LA POSSIBILITÀ DI UN	14	14	8.695652008
329	LE CONDIZIONI DI VITA	14	14	8.695652008
330	NEL CORSO DEL NOVECENTO	14	14	8.695652008
331	NEL PAESE DI ARRIVO	14	11	6.832298279
332	PER GLI STATI UNITI	14	9	5.590062141
333	PER TUTTI GLI ANNI	14	8	4.968944073
334	PRIMA DELLA SECONDA GUERRA	14	11	6.832298279
335	PRIMA E SECONDA GENERAZIONE	14	14	8.695652008
336	PRIMA GUERRA MONDIALE E	14	11	6.832298279
337	PRIMA METÀ DEL NOVECENTO	14	12	7.453416348
338	STATI UNITI D'AMERICA	14	10	6.21118021
339	SU UN TOTALE DI	14	13	8.074534416
340	UN ANNO E MEZZO	14	10	6.21118021
341	UN GRAN NUMERO DI	14	13	8.074534416
342	UN RITORNO ALLE RADICI	14	5	3.105590105
343	UNO DEGLI ASPETTI PIU	14	10	6.21118021
344	A FRONTE DI UN	13	13	8.074534416
345	A GIOCARE UN RUOLO	13	9	5.590062141
346	ALL'INTERNO DELLA CITTA	13	8	4.968944073
347	ALLA PRIMA GUERRA MONDIALE	13	10	6.21118021

348	AMERICANI DI ORIGINE ITALIANA	13	6	3.726708174
349	ANCHE DAL PUNTO DI	13	10	6.21118021
350	ANNI VENTI DEL NOVECENTO	13	10	6.21118021
351	CHE IN MOLTI CASI	13	11	6.832298279
352	COME UNA SORTA DI	13	11	6.832298279
353	COMMISSARIATO GENERALE DELL'EMIGRAZIONE	13	7	4.347826004
354	CON IL PAESE DI	13	13	8.074534416
355	CON LA TERRA D	13	7	4.347826004
356	CON LO SCOPO DI	13	12	7.453416348
357	DEI DIRITTI DELL'UOMO	13	10	6.21118021
358	DELLA PRIMA METÀ DEL	13	10	6.21118021
359	DELLA SECONDA GENERAZIONE DI	13	6	3.726708174
360	DI PRIGIONIERI DI GUERRA	13	8	4.968944073
361	DI UN SECOLO FA	13	5	3.105590105
362	DISCENDENTI DEGLI IMMIGRATI ITALIANI	13	8	4.968944073
363	DOPO LA MORTE DEL	13	11	6.832298279
364	DOPO LA MORTE DI	13	10	6.21118021
365	E L'INIZIO DEL	13	10	6.21118021
366	E LE CONDIZIONI DI	13	8	4.968944073
367	I DATORI DI LAVORO	13	8	4.968944073
368	I FIGLI DEGLI IMMIGRATI	13	7	4.347826004
369	IL FATTO CHE L'	13	9	5.590062141
370	IL FATTO DI ESSERE	13	13	8.074534416
371	IL PERMESSO DI SOGGIORNO	13	6	3.726708174
372	IL PROCESSO DI INTEGRAZIONE	13	5	3.105590105
373	IL SENSO DI APPARTENENZA	13	10	6.21118021
374	IN ITALIA E IN	13	12	7.453416348
375	IN MODO TALE CHE	13	10	6.21118021
376	IN MODO TALE DA	13	8	4.968944073
377	IN PRESENZA DI UN	13	9	5.590062141
378	IN QUESTO CASO DI	13	10	6.21118021
379	IN TAL SENSO LA	13	10	6.21118021

380	IN UN SECONDO TEMPO	13	7	4.347826004
381	INIZIO DEGLI ANNI NOVANTA	13	13	8.074534416
382	INIZIO DEGLI ANNI OTTANTA	13	10	6.21118021
383	INIZIO DEL XX SECOLO	13	9	5.590062141
384	L'ASSENZA DI UN	13	12	7.453416348
385	LA TERRA D'ORIGINE	13	7	4.347826004
386	LE SECONDE GENERAZIONI DI	13	5	3.105590105
387	LEGAME CON L'ITALIA	13	5	3.105590105
388	METÀ DEL VI SECOLO	13	6	3.726708174
389	METÀ DEL XII SECOLO	13	6	3.726708174
390	NEGLI ANNI IN CUI	13	10	6.21118021
391	NEGLI STATI UNITI SI	13	10	6.21118021
392	PARTIRE DAGLI ANNI NOVANTA	13	7	4.347826004
393	PER LA CREAZIONE DI	13	7	4.347826004
394	PER LA REALIZZAZIONE DI	13	13	8.074534416
395	UN PERIODO IN CUI	13	12	7.453416348
396	UNA VOLTA PER TUTTE	13	12	7.453416348
397	VALE A DIRE LA	13	12	7.453416348
398	VECCHIE E NUOVE MIGRAZIONI	13	5	3.105590105
399	VITA E DI LAVORO	13	6	3.726708174
		9182		

**Appendice 3: Lexical bundles di 4 parole in sequenza nel corpus di filosofia**

Corpus di FILOSOFIA – lexical bundles di 4 parole, occorrenza 7				
N	Lexical bundle	Frequenza	Articoli	%
1	DAL PUNTO DI VISTA	106	28	71.79486847
2	UN PUNTO DI VISTA	84	23	58.9743576
3	NEL MOMENTO IN CUI	79	27	69.2307663
4	NELLA MISURA IN CUI	67	19	48.71794891
5	DA UN PUNTO DI	64	22	56.41025543
6	A CHE FARE CON	51	23	58.9743576
7	UNA VERA E PROPRIA	47	19	48.71794891
8	SI TRATTA DI UN	46	23	58.9743576
9	IL MONDO DELLE COSE	40	4	10.2564106
10	DI VOLTA IN VOLTA	39	17	43.58974457
11	ALL'INTERNO DI UNA	37	19	48.71794891
12	SI TRATTA DI UNA	35	20	51.28205109
13	ALL'INTERNO DI UN	33	20	51.28205109
14	QUESTO PUNTO DI VISTA	30	19	48.71794891
15	PER LA PRIMA VOLTA	29	16	41.0256424
16	PUNTO DI VISTA DELLA	29	18	46.15384674
17	ANCHE IN QUESTO CASO	28	14	35.89743423
18	È IN GRADO DI	28	14	35.89743423
19	IN UN CERTO SENSO	28	14	35.89743423
20	NON SI TRATTA DI	26	20	51.28205109
21	DA QUESTO PUNTO DI	25	17	43.58974457
22	IL PUNTO DI PARTENZA	25	19	48.71794891
23	CRITICA DELLA RAGION PURA	24	9	23.07692337
24	IL MODO IN CUI	24	17	43.58974457
25	AL DI FUORI DEL	21	14	35.89743423
26	AL DI LÀ DELLA	21	12	30.76922989
27	FARE A MENO DI	21	14	35.89743423

28	HA A CHE FARE	21	14	35.89743423
29	IL FATTO CHE LA	21	15	38.46154022
30	IN QUESTO SENSO LA	21	10	25.64102554
31	CHE A SUA VOLTA	20	14	35.89743423
32	IL PUNTO DI VISTA	20	12	30.76922989
33	IN LINEA DI PRINCIPIO	20	12	30.76922989
34	SE DA UN LATO	20	13	33.33333206
35	UN VERO E PROPRIO	20	14	35.89743423
36	UNA VOLTA PER TUTTE	20	13	33.33333206
37	AD UN CERTO PUNTO	19	7	17.94871712
38	DI FRONTE AD UN	19	15	38.46154022
39	IN MODO TALE CHE	19	13	33.33333206
40	NON È ALTRO CHE	19	11	28.20512772
41	A MENO CHE NON	18	11	28.20512772
42	E AL TEMPO STESSO	18	13	33.33333206
43	IN QUESTO SENSO L	18	10	25.64102554
44	PER IL FATTO CHE	18	12	30.76922989
45	A PARTIRE DA UNA	17	11	28.20512772
46	AL DI FUORI DI	17	14	35.89743423
47	IL MONDO DELLA VITA	17	7	17.94871712
48	IN TUTTA LA SUA	17	15	38.46154022
49	PUNTO DI VISTA DEL	17	13	33.33333206
50	A UN CERTO PUNTO	16	7	17.94871712
51	AL DI LÀ DEL	16	10	25.64102554
52	AL DI LÀ DI	16	9	23.07692337
53	E NELLO STESSO TEMPO	16	10	25.64102554
54	IN CARNE ED OSSA	16	10	25.64102554
55	L IDEA DI UNA	16	12	30.76922989
56	MOMENTO IN CUI SI	16	13	33.33333206
57	AL DI SOPRA DI	15	12	30.76922989
58	DA UN LATO LA	15	11	28.20512772
59	IL FATTO CHE IL	15	8	20.5128212

60	IN QUESTO SENSO IL	15	10	25.64102554
61	IN TUTTE LE SUE	15	12	30.76922989
62	L IDEA DI UN	15	11	28.20512772
63	NEL MODO IN CUI	15	9	23.07692337
64	PER QUANTO RIGUARDA IL	15	11	28.20512772
65	SE E VERO CHE	15	8	20.5128212
66	SULLA BASE DI UNA	15	9	23.07692337
67	A PARTIRE DA UN	14	10	25.64102554
68	AD UNA SORTA DI	14	11	28.20512772
69	CI TROVIAMO DI FRONTE	14	8	20.5128212
70	DI CIÒ CHE E	14	7	17.94871712
71	IL CONCETTO DI VERITÀ	14	4	10.2564106
72	IN QUESTO MODO LA	14	8	20.5128212
73	MOMENTO IN CUI LA	14	11	28.20512772
74	NEI TERMINI DI UN	14	6	15.38461494
75	VALE A DIRE IL	14	5	12.82051277
76	DEL MONDO DELLA VITA	13	10	25.64102554
77	DELL'OPERA D'ARTE	13	4	10.2564106
78	E ALLO STESSO TEMPO	13	7	17.94871712
79	FARE I CONTI CON	13	9	23.07692337
80	IN ESSERE E TEMPO	13	6	15.38461494
81	IN QUESTO CASO L'	13	11	28.20512772
82	LA FILOSOFIA DEL LINGUAGGIO	13	6	15.38461494
83	NEL SENSO CHE LA	13	5	12.82051277
84	NEL SENSO DI UNA	13	10	25.64102554
85	PUNTO DI VISTA DI	13	9	23.07692337
86	SOTTO IL NOME DI	13	8	20.5128212
87	DI TUTTE LE COSE	12	11	28.20512772
88	DI TUTTO CIÒ CHE	12	6	15.38461494
89	DI UNA TEORIA DELLA	12	6	15.38461494
90	DI UNA VERA E	12	10	25.64102554
91	HA IL COMPITO DI	12	10	25.64102554



92	IN BASE ALLA QUALE	12	9	23.07692337
93	IN ENTRAMBI I CASI	12	9	23.07692337
94	IN QUESTO SENSO SI	12	7	17.94871712
95	MA NON PER QUESTO	12	10	25.64102554
96	NON HA BISOGNO DI	12	9	23.07692337
97	PUNTO DI VISTA DELL'	12	8	20.5128212
98	SI TROVA DI FRONTE	12	7	17.94871712
99	SULLA BASE DI UN	12	8	20.5128212
100	AL DI LÀ DELLE	11	7	17.94871712
101	ALLA STREGUA DI UN	11	8	20.5128212
102	CHE COS'È LA	11	8	20.5128212
103	CHE SI TRATTA DI	11	9	23.07692337
104	COME UNA SORTA DI	11	8	20.5128212
105	DA UN LATO E	11	8	20.5128212
106	DAL MOMENTO CHE IL	11	7	17.94871712
107	DI UNA SORTA DI	11	7	17.94871712
108	E IN QUESTO SENSO	11	9	23.07692337
109	E PROPRIO PER QUESTO	11	7	17.94871712
110	IL MODO D'ESSERE	11	7	17.94871712
111	IL RISULTATO DI UN	11	9	23.07692337
112	L'ESISTENZA DI UN	11	9	23.07692337
113	LA POSSIBILITÀ DI UN	11	8	20.5128212
114	MA ANCHE E SOPRATTUTTO	11	9	23.07692337
115	MONDO DELLA VITA E	11	4	10.2564106
116	NELLA CRITICA DELLA RAGION	11	6	15.38461494
117	PRINCIPIO DI NON CONTRADDIZIONE	11	4	10.2564106
118	PUNTO DI PARTENZA E	11	9	23.07692337
119	SUO PUNTO DI VISTA	11	9	23.07692337
120	TUTTE LE FORME DI	11	5	12.82051277
121	UNA TEORIA DELLA CONOSCENZA	11	5	12.82051277
122	AL TEMPO STESSO LA	10	9	23.07692337
123	CHE NON PUÒ ESSERE	10	6	15.38461494

124	CIÒ CHE NON È	10	7	17.94871712
125	DEL FATTO CHE LA	10	9	23.07692337
126	DI FRONTE A UN	10	9	23.07692337
127	DI UN SOGGETTO CHE	10	6	15.38461494
128	E CIÒ CHE E	10	7	17.94871712
129	E IL PROBLEMA DELLA	10	4	10.2564106
130	ESSERE IN GRADO DI	10	10	25.64102554
131	IL PROBLEMA DEL RAPPORTO	10	6	15.38461494
132	IN CUI L'UOMO	10	8	20.5128212
133	IN TAL MODO LA	10	5	12.82051277
134	L'ESSERE E IL	10	9	23.07692337
135	L'IDEA CHE L'	10	10	25.64102554
136	L'IDEA CHE LA	10	7	17.94871712
137	MODO IN CUI LA	10	10	25.64102554
138	PER IL QUALE IL	10	9	23.07692337
139	PER UNA FILOSOFIA DELL'	10	4	10.2564106
140	QUESTO NON SIGNIFICA CHE	10	9	23.07692337
141	STESSO MODO IN CUI	10	8	20.5128212
142	TRA L'UOMO E	10	6	15.38461494
143	TUTTA UNA SERIE DI	10	7	17.94871712
144	UN DATO DI FATTO	10	6	15.38461494
145	VALE A DIRE LA	10	5	12.82051277
146	A CHE VEDERE CON	9	9	23.07692337
147	A FAVORE DI UNA	9	6	15.38461494
148	A PARTIRE DAL QUALE	9	7	17.94871712
149	ALL'INTERNO DEL QUALE	9	7	17.94871712
150	CHE A PARTIRE DA	9	7	17.94871712
151	CHE IL CONCETTO DI	9	6	15.38461494
152	CHE IN QUALCHE MODO	9	6	15.38461494
153	COME SI È VISTO	9	6	15.38461494
154	CONSISTE NEL FATTO CHE	9	9	23.07692337
155	DAL MOMENTO CHE LA	9	6	15.38461494

156	DELL'UOMO E DELLA	9	4	10.2564106
157	DELLA CRITICA DELLA RAGION	9	7	17.94871712
158	DI CUI NON SI	9	8	20.5128212
159	DI FRONTE AD UNA	9	8	20.5128212
160	DI PER SÉ NON	9	8	20.5128212
161	DI UN PROCESSO DI	9	9	23.07692337
162	È ANCHE VERO CHE	9	5	12.82051277
163	È IL FATTO CHE	9	7	17.94871712
164	FINO A CHE PUNTO	9	7	17.94871712
165	IL RISULTATO DI UNA	9	6	15.38461494
166	IL SENSO D'ESSERE	9	5	12.82051277
167	IN CUI CI SI	9	9	23.07692337
168	IN QUESTO CASO IL	9	9	23.07692337
169	IN QUESTO MODO L'	9	6	15.38461494
170	IN UNA SORTA DI	9	8	20.5128212
171	L'ESIGENZA DI UNA	9	8	20.5128212
172	L'UNO ALL'ALTRO	9	5	12.82051277
173	LA POSSIBILITÀ DI UNA	9	5	12.82051277
174	LA PRESENZA DI UN	9	8	20.5128212
175	NOSTRO PUNTO DI VISTA	9	8	20.5128212
176	PENSIERO DI MERLEAU PONTY	9	6	15.38461494
177	PER QUANTO RIGUARDA LA	9	7	17.94871712
178	PUNTO DI PARTENZA DI	9	6	15.38461494
179	PUNTO DI VISTA CHE	9	6	15.38461494
180	PUNTO DI VISTA LA	9	9	23.07692337
181	SE NON CI FOSSE	9	5	12.82051277
182	SI PUÒ DIRE CHE	9	8	20.5128212
183	TUTTO CIÒ CHE SI	9	4	10.2564106
184	A FAVORE DI UN	8	7	17.94871712
185	AL DI FUORI DELL'	8	4	10.2564106
186	AL MODO IN CUI	8	7	17.94871712
187	ALL'INTERNO DI QUESTA	8	7	17.94871712

188	CHE COME ABBIAMO VISTO	8	6	15.38461494
189	CHE FARE CON LA	8	6	15.38461494
190	CHE FARE CON UNA	8	7	17.94871712
191	CHE IN QUESTO CASO	8	7	17.94871712
192	CHE NON È MAI	8	7	17.94871712
193	CHE NON SI PUÒ	8	4	10.2564106
194	CUI NON SI PUÒ	8	7	17.94871712
195	DA CIÒ CHE È	8	6	15.38461494
196	DA UN CERTO PUNTO	8	4	10.2564106
197	DA UN LATO IL	8	7	17.94871712
198	DA UNA SORTA DI	8	8	20.5128212
199	DELLA FENOMENOLOGIA DELLO SPIRITO	8	4	10.2564106
200	DELLA VITA E DELLA	8	5	12.82051277
201	E A SUA VOLTA	8	7	17.94871712
202	E PER QUESTO CHE	8	5	12.82051277
203	IL DATO DI FATTO	8	6	15.38461494
204	IL FATTO CHE IN	8	5	12.82051277
205	IN BASE AL QUALE	8	6	15.38461494
206	IN CUI IL SOGGETTO	8	5	12.82051277
207	IN GRADO DI DIRE	8	6	15.38461494
208	IN QUESTO CASO LA	8	6	15.38461494
209	IN QUESTO MODO SI	8	7	17.94871712
210	IN SÉ E PER	8	7	17.94871712
211	IN TAL MODO L'	8	5	12.82051277
212	IN TAL MODO SI	8	7	17.94871712
213	IN UN CERTO MODO	8	7	17.94871712
214	IN UN PRIMO TEMPO	8	5	12.82051277
215	IN UNA SITUAZIONE DI	8	5	12.82051277
216	IN VISTA DI UNA	8	6	15.38461494
217	L'IO E IL	8	8	20.5128212
218	L'UNO E L'	8	7	17.94871712
219	L'UOMO E IL	8	7	17.94871712

220	L'UOMO E L'	8	5	12.82051277
221	L'UOMO E LA	8	5	12.82051277
222	LA CONDIZIONE DI POSSIBILITÀ	8	6	15.38461494
223	LA DIFFERENZA TRA LA	8	6	15.38461494
224	LA FENOMENOLOGIA DI HUSSERL	8	4	10.2564106
225	LA PRESA DI COSCIENZA	8	7	17.94871712
226	LA RAGIONE E LA	8	7	17.94871712
227	LE COSE DEL MONDO	8	4	10.2564106
228	NEL SENSO CHE NON	8	6	15.38461494
229	NEL SENSO DI UN	8	7	17.94871712
230	NEL SENSO IN CUI	8	6	15.38461494
231	NELL' AMBITO DI UN	8	8	20.5128212
232	NON IN GRADO DI	8	6	15.38461494
233	PROBLEMA DEL RAPPORTO TRA	8	5	12.82051277
234	PUNTO DI VISTA ONTOLOGICO	8	7	17.94871712
235	QUANDO SI TRATTA DI	8	4	10.2564106
236	SE PER UN VERSO	8	7	17.94871712
237	SOLO NELLA MISURA IN	8	7	17.94871712
238	SU CUI SI FONDA	8	7	17.94871712
239	SU DI UN PIANO	8	6	15.38461494
240	TRA IL SOGGETTO E	8	7	17.94871712
241	TRADIZIONE DEL PENSIERO OCCIDENTALE	8	5	12.82051277
242	TUTTO CIO CHE E	8	6	15.38461494
243	UN MONDO IN CUI	8	6	15.38461494
244	A SÉ STESSO E	7	6	15.38461494
245	A SUA VOLTA UN	7	6	15.38461494
246	ABBIAMO A CHE FARE	7	4	10.2564106
247	AL MONDO DELLE COSE	7	4	10.2564106
248	ALLA FINE DEGLI ANNI	7	7	17.94871712
249	ANCORA UNA VOLTA LA	7	6	15.38461494
250	BISOGNI E DEI DESIDERI	7	4	10.2564106
251	CHE CI PERMETTE DI	7	6	15.38461494

252	CHE FA SÌ CHE	7	5	12.82051277
253	CHE FARE CON IL	7	6	15.38461494
254	CHE NON SI LASCIA	7	5	12.82051277
255	CHE SI TRATTI DI	7	4	10.2564106
256	COME NEL CASO DELL'	7	7	17.94871712
257	COME PUNTO DI PARTENZA	7	6	15.38461494
258	COME SI È DETTO	7	4	10.2564106
259	CON LE PAROLE DI	7	6	15.38461494
260	DA PARTE DI UN	7	5	12.82051277
261	DA UN LATO L'	7	5	12.82051277
262	DA UNA PARTE È	7	6	15.38461494
263	DAL NOSTRO PUNTO DI	7	6	15.38461494
264	DATO DI FATTO CHE	7	5	12.82051277
265	DEI BISOGNI E DEI	7	4	10.2564106
266	DELLA SCIENZA E DELLA	7	6	15.38461494
267	DELLA VERITÀ E DELLA	7	5	12.82051277
268	DI CIÒ CHE SI	7	7	17.94871712
269	DI TUTTI GLI UOMINI	7	5	12.82051277
270	DI UN IO CHE	7	5	12.82051277
271	DI UNA FILOSOFIA DELLA	7	4	10.2564106
272	E IN QUALCHE MODO	7	6	15.38461494
273	E IN QUANTO TALE	7	7	17.94871712
274	E IN QUESTO MODO	7	6	15.38461494
275	E IN TAL MODO	7	6	15.38461494
276	E IN UN CERTO	7	5	12.82051277
277	E LO STESSO CHE	7	4	10.2564106
278	E UNA SORTA DI	7	5	12.82051277
279	HA NULLA A CHE	7	6	15.38461494
280	IL FATTO DI ESSERE	7	7	17.94871712
281	IL PRINCIPIO DI NON	7	4	10.2564106
282	IL RAPPORTO CON L'	7	5	12.82051277
283	IL SENSO DELL'ESSERE	7	6	15.38461494

284	IN ALTRI TERMINI IL	7	6	15.38461494
285	IN BASE AD UN	7	5	12.82051277
286	IN CONTRASTO CON LA	7	5	12.82051277
287	IN CUI ESSA SI	7	4	10.2564106
288	IN CUI LA FILOSOFIA	7	7	17.94871712
289	IN FIN DEI CONTI	7	4	10.2564106
290	IN GRADO DI FORNIRE	7	6	15.38461494
291	IN GRADO DI RENDERE	7	5	12.82051277
292	IN MODO DEL TUTTO	7	6	15.38461494
293	IN QUANTO TALE E	7	5	12.82051277
294	IN QUESTO CASO SI	7	6	15.38461494
295	IN QUESTO MODO IL	7	5	12.82051277
296	IN QUESTO SENSO E	7	6	15.38461494
297	IN RELAZIONE AD UN	7	5	12.82051277
298	IN RELAZIONE CON LA	7	6	15.38461494
299	IN UN PRIMO MOMENTO	7	5	12.82051277
300	IN UN RAPPORTO DI	7	5	12.82051277
301	IN VIRTÙ DI UN	7	5	12.82051277
302	L'ESISTENZA DI UNA	7	7	17.94871712
303	L'IDEA CHE IL	7	6	15.38461494
304	LA DOMANDA SULLA VERITÀ	7	4	10.2564106
305	LA STORIA DELLA FILOSOFIA	7	6	15.38461494
306	LA SUA VITA E	7	5	12.82051277
307	LE CONDIZIONI DI POSSIBILITÀ	7	4	10.2564106
308	MODO IN CUI SI	7	6	15.38461494
309	MOMENTO IN CUI IL	7	5	12.82051277
310	NEL CORSO DEGLI ANNI	7	6	15.38461494
311	NEL SENSO CHE L'	7	5	12.82051277
312	NELLA RICERCA DI UN	7	5	12.82051277
313	NON HA NULLA A	7	6	15.38461494
314	NON NEL SENSO CHE	7	5	12.82051277
315	PER IL QUALE LA	7	7	17.94871712

316	PUNTO DI PARTENZA PER	7	6	15.38461494
317	PUNTO DI VISTA IL	7	6	15.38461494
318	SÉ E PER SÉ	7	6	15.38461494
319	SE E SOLO SE	7	5	12.82051277
320	SI PONE IL PROBLEMA	7	5	12.82051277
321	SI TRATTA DUNQUE DI	7	4	10.2564106
322	SIA DA UN PUNTO	7	4	10.2564106
323	SIA IN GRADO DI	7	6	15.38461494
324	SIAMO IN GRADO DI	7	6	15.38461494
325	TRA RAGIONE E FEDE	7	4	10.2564106
326	UN PUNTO DI PARTENZA	7	4	10.2564106
327	UNA SITUAZIONE IN CUI	7	6	15.38461494
328	VALE A DIRE UN	7	5	12.82051277
		4089		



**Appendice 4: Lexical bundles di 4 parole in sequenza nel corpus di medicina del lavoro**

Corpus di MEDICINA del lavoro – lexical bundles di 4 parole, occorrenza 10				
N	Lexical bundle	Frequenza	Articoli	%
1	NEI LUOGHI DI LAVORO	215	43	34.40000153
2	IL DATORE DI LAVORO	184	37	29.60000038
3	DI MEDICINA DEL LAVORO	163	34	27.20000076
4	DAL PUNTO DI VISTA	152	53	42.40000153
5	IL MEDICO DEL LAVORO	128	22	17.60000038
6	DEL DATORE DI LAVORO	112	26	20.79999924
7	LA VALUTAZIONE DEL RISCHIO	109	37	29.60000038
8	PER QUANTO RIGUARDA LA	107	41	32.79999924
9	DELLA MEDICINA DEL LAVORO	94	28	22.39999962
10	DI SALUTE E SICUREZZA	94	27	21.60000038
11	IN MEDICINA DEL LAVORO	94	18	14.39999962
12	DEL MEDICO DEL LAVORO	92	22	17.60000038
13	DI PREVENZIONE E PROTEZIONE	92	28	22.39999962
14	MEDICINA DEL LAVORO E	91	27	21.60000038
15	DI VALUTAZIONE DEL RISCHIO	89	33	26.39999962
16	DI VALUTAZIONE DEI RISCHI	84	27	21.60000038
17	DELLA SALUTE E DELLA	82	25	20
18	PER LA SALUTE E	81	34	27.20000076
19	TUTELA DELLA SALUTE E	81	26	20.79999924
20	SALUTE E SICUREZZA SUL	79	23	18.39999962
21	E SICUREZZA SUL LAVORO	77	23	18.39999962
22	RISCHI PER LA SALUTE	77	32	25.60000038
23	NEGLI AMBIENTI DI LAVORO	76	36	28.79999924
24	DATORE DI LAVORO E	74	25	20
25	PER QUANTO RIGUARDA IL	73	37	29.60000038
26	DI PROMOZIONE DELLA SALUTE	70	23	18.39999962
27	SALUTE E DELLA SICUREZZA	70	20	16

28	MOVIMENTAZIONE MANUALE DEI CARICHI	67	15	12
29	È IN GRADO DI	65	36	28.79999924
30	DELLA SALUTE E SICUREZZA	63	22	17.60000038
31	I FATTORI DI RISCHIO	63	28	22.39999962
32	LA SALUTE DEI LAVORATORI	63	29	23.20000076
33	LA VALUTAZIONE DEI RISCHI	63	25	20
34	LA MEDICINA DEL LAVORO	62	19	15.19999981
35	AL DATORE DI LAVORO	61	18	14.39999962
36	DEL GIUDIZIO DI IDONEITÀ	60	15	12
37	SICUREZZA NEI LUOGHI DI	60	19	15.19999981
38	LA MAGGIOR PARTE DEI	56	29	23.20000076
39	DEI FATTORI DI RISCHIO	55	20	16
40	DELLO STRESS LAVORO CORRELATO	55	11	8.800000191
41	E.S.M.I.	55	18	14.39999962
42	PREVENZIONE NEI LUOGHI DI	55	15	12
43	SERVIZIO DI PREVENZIONE E	55	20	16
44	SONO IN GRADO DI	55	30	24
45	DI TUTELA DELLA SALUTE	54	23	18.39999962
46	DA PARTE DEL MEDICO	53	21	16.79999924
47	IN MATERIA DI SALUTE	53	19	15.19999981
48	SUL LUOGO DI LAVORO	52	22	17.60000038
49	DISPOSITIVI DI PROTEZIONE INDIVIDUALE	51	18	14.39999962
50	ANCHE IN QUESTO CASO	50	22	17.60000038
51	PER LA SICUREZZA E	50	21	16.79999924
52	PER QUANTO RIGUARDA LE	50	21	16.79999924
53	LA SALUTE E SICUREZZA	49	20	16
54	MATERIA DI SALUTE E	49	16	12.80000019
55	UN PUNTO DI VISTA	49	28	22.39999962
56	DA UN PUNTO DI	48	27	21.60000038
57	IL GIUDIZIO DI IDONEITÀ	48	18	14.39999962
58	LA SICUREZZA E LA	48	26	20.79999924
59	LA TUTELA DELLA SALUTE	47	21	16.79999924

60	SICUREZZA E LA SALUTE	47	25	20
61	PER LA VALUTAZIONE DEL	45	31	24.79999924
62	RISCHIO PER LA SALUTE	44	23	18.39999962
63	DELLA SALUTE DEI LAVORATORI	43	20	16
64	DA PARTE DEI LAVORATORI	42	26	20.79999924
65	DELLO STATO DI SALUTE	42	21	16.79999924
66	MEDICO DEL LAVORO COMPETENTE	41	15	12
67	DELLA VALUTAZIONE DEI RISCHI	40	22	17.60000038
68	I DATORI DI LAVORO	40	17	13.60000038
69	LA SALUTE E LA	40	20	16
70	PER LA PRIMA VOLTA	40	24	19.20000076
71	ALLA VALUTAZIONE DEI RISCHI	39	17	13.60000038
72	CHE IL MEDICO COMPETENTE	39	15	12
73	CHE LA MAGGIOR PARTE	39	19	15.19999981
74	DEGLI AMBIENTI DI LAVORO	39	16	12.80000019
75	I RISCHI PER LA	39	19	15.19999981
76	LA PROMOZIONE DELLA SALUTE	39	17	13.60000038
77	PER QUANTO RIGUARDA I	39	18	14.39999962
78	ALLA VALUTAZIONE DEL RISCHIO	38	18	14.39999962
79	DEI MEDICI DEL LAVORO	38	13	10.39999962
80	DELL ORGANIZZAZIONE DEL LAVORO	38	10	8
81	DI ETA COMPRESA TRA	38	19	15.19999981
82	NEL CASO IN CUI	38	20	16
83	SALUTE E LA SICUREZZA	38	21	16.79999924
84	SALUTE E SICUREZZA DEI	38	20	16
85	VALUTAZIONE DEL RISCHIO STRESS	38	8	6.400000095
86	VALUTAZIONE DEL RISCHIO E	37	26	20.79999924
87	CON LO SCOPO DI	36	19	15.19999981
88	DEI DATORI DI LAVORO	36	18	14.39999962
89	DI IDONEITA ALLA MANSIONE	36	13	10.39999962
90	IDONEITA ALLA MANSIONE SPECIFICA	36	11	8.800000191
91	PER QUANTO RIGUARDA GLI	36	25	20

92	PER QUANTO RIGUARDA L'	36	15	12
93	AL FUMO DI SIGARETTA	35	13	10.39999962
94	DELLA VALUTAZIONE DEL RISCHIO	35	21	16.79999924
95	E SICUREZZA DEI LAVORATORI	35	19	15.19999981
96	IN GRADO DI DETERMINARE	35	18	14.39999962
97	PER LA FORMAZIONE CONTINUA	35	5	4
98	RISCHIO STRESS LAVORO CORRELATO	35	7	5.599999905
99	CARTELLA SANITARIA E DI	34	10	8
100	DEL SERVIZIO DI PREVENZIONE	34	13	10.39999962
101	È STATO QUELLO DI	34	16	12.80000019
102	LA PRESENZA DI UN	34	23	18.39999962
103	LA QUALITÀ DELLA VITA	34	18	14.39999962
104	PARTE DEL MEDICO COMPETENTE	34	16	12.80000019
105	SALUTE NEI LUOGHI DI	34	12	9.600000381
106	SANITARIA E DI RISCHIO	34	10	8
107	SUI LUOGHI DI LAVORO	34	22	17.60000038
108	VALUTAZIONE DEI RISCHI E	34	22	17.60000038
109	CON IL DATORE DI	33	13	10.39999962
110	DAL DATORE DI LAVORO	33	13	10.39999962
111	MIGLIORAMENTO DELLE CONDIZIONI DI	33	22	17.60000038
112	MISURE DI PREVENZIONE E	33	20	16
113	PER UN TOTALE DI	33	20	16
114	FATTORI DI RISCHIO PER	32	16	12.80000019
115	MOVIMENTAZIONE MANUALE DI CARICHI	32	13	10.39999962
116	RISCHI PER LA SICUREZZA	32	17	13.60000038
117	SI TRATTA DI UN	32	21	16.79999924
118	DELLE CONDIZIONI DI LAVORO	31	19	15.19999981
119	E LA SALUTE DEI	31	20	16
120	E SICUREZZA NEI LUOGHI	31	13	10.39999962
121	GIUDIZIO DI IDONEITÀ ALLA	31	13	10.39999962
122	GLI INFORTUNI SUL LAVORO	31	14	11.19999981
123	MEDICINA DEL LAVORO ED	31	9	7.199999809

124	PER LA DIAGNOSI DI	31	11	8.800000191
125	DEL RAPPORTO DI LAVORO	30	6	4.800000191
126	DI CUI ALL'ARTICOLO	30	7	5.599999905
127	E STATO DIMOSTRATO CHE	30	18	14.39999962
128	ESSERE IN GRADO DI	30	21	16.79999924
129	IL GRUPPO DI LAVORO	30	14	11.19999981
130	ITALIANA DI MEDICINA DEL	30	16	12.80000019
131	PER LA TUTELA DELLA	30	18	14.39999962
132	SU UN CAMPIONE DI	30	18	14.39999962
133	TUTELA DELLA SALUTE DEI	30	16	12.80000019
134	UN FATTORE DI RISCHIO	30	15	12
135	VALUTAZIONE DEL RISCHIO DA	30	17	13.60000038
136	AI SENSI DELL'ART	29	6	4.800000191
137	IN GRADO DI INFLUENZARE	29	17	13.60000038
138	LAVORO A TURNI E	29	9	7.199999809
139	LO STATO DI SALUTE	29	15	12
140	PER LA POPOLAZIONE GENERALE	29	16	12.80000019
141	SOCIETA ITALIANA DI MEDICINA	29	15	12
142	TUTTI I RISCHI PER	29	15	12
143	ETA COMPRESA TRA I	28	15	12
144	LA MAGGIOR PARTE DELLE	28	17	13.60000038
145	PER IL MEDICO COMPETENTE	28	12	9.600000381
146	STRESS LAVORO CORRELATO E	28	9	7.199999809
147	SUL POSTO DI LAVORO	28	20	16
148	ALTRI FATTORI DI RISCHIO	27	15	12
149	DA STRESS LAVORO CORRELATO	27	9	7.199999809
150	DATORI DI LAVORO E	27	17	13.60000038
151	DEGLI INFORTUNI SUL LAVORO	27	14	11.19999981
152	DEI LAVORATORI PER LA	27	17	13.60000038
153	DEL MEDICO COMPETENTE E	27	16	12.80000019
154	DELLE MISURE DI PREVENZIONE	27	19	15.19999981
155	DI INFORTUNI SUL LAVORO	27	16	12.80000019

156	DOCUMENTO DI VALUTAZIONE DEI	27	14	11.19999981
157	FATTORI DI RISCHIO E	27	12	9.600000381
158	IL LAVORO A TURNI	27	12	9.600000381
159	SIA PER QUANTO RIGUARDA	27	13	10.39999962
160	CHE IL DATORE DI	26	13	10.39999962
161	DEGLI ARTI SUPERIORI E	26	12	9.600000381
162	DELLE ATTIVITA DI PREVENZIONE	26	12	9.600000381
163	DELLE CONDIZIONI DI SALUTE	26	15	12
164	DI PREVENZIONE E SICUREZZA	26	12	9.600000381
165	DI VITA E DI	26	19	15.19999981
166	FATTORE DI RISCHIO PER	26	13	10.39999962
167	SULLO STATO DI SALUTE	26	19	15.19999981
168	DEL RISCHIO STRESS LAVORO	25	7	5.599999905
169	DELLA TUTELA DELLA SALUTE	25	15	12
170	DI RISCHIO PER LA	25	13	10.39999962
171	DI SALUTE DEI LAVORATORI	25	12	9.600000381
172	DI STRESS LAVORO CORRELATO	25	8	6.400000095
173	DI UN GRUPPO DI	25	18	14.39999962
174	E LA NECESSITÀ DI	25	17	13.60000038
175	FORMULAZIONE DEL GIUDIZIO DI	25	10	8
176	I MEDICI DEL LAVORO	25	13	10.39999962
177	NEI CASI IN CUI	25	16	12.80000019
178	ORGANIZZAZIONE DEL LAVORO E	25	13	10.39999962
179	PER LA PREVENZIONE E	25	11	8.800000191
180	REGIONI E PROVINCE AUTONOME	25	9	7.199999809
181	RESPONSABILE DEL SERVIZIO DI	25	10	8
182	SIA IN TERMINI DI	25	15	12
183	DEI VIGILI DEL FUOCO	24	7	5.599999905
184	LA PREVENZIONE E LA	24	13	10.39999962
185	LUOGHI DI LAVORO E	24	13	10.39999962
186	NELLA MAGGIOR PARTE DEI	24	19	15.19999981
187	NORME DI BUONA TECNICA	24	7	5.599999905

188	PER LA SALUTE DEI	24	16	12.80000019
189	SOGGETTI DI SESSO MASCHILE	24	12	9.600000381
190	SORVEGLIANZA SANITARIA DEI LAVORATORI	24	14	11.19999981
191	SOTTOPOSTI A SORVEGLIANZA SANITARIA	24	13	10.39999962
192	UN GRUPPO DI LAVORO	24	11	8.800000191
193	VITA E DI LAVORO	24	17	13.60000038
194	AI FATTORI DI RISCHIO	23	15	12
195	ALLO STRESS LAVORO CORRELATO	23	11	8.800000191
196	CONDIZIONI DI SALUTE E	23	16	12.80000019
197	DEI SERVIZI DI PREVENZIONE	23	11	8.800000191
198	DEL MINISTERO DELLA SALUTE	23	12	9.600000381
199	DEL PRESENTE LAVORO E	23	14	11.19999981
200	DEL RISCHIO DA STRESS	23	12	9.600000381
201	IN AMBIENTE DI LAVORO	23	15	12
202	IN TUTTI I CASI	23	14	11.19999981
203	LAVORATORI PER LA SICUREZZA	23	14	11.19999981
204	MEDICO DEL LAVORO E	23	15	12
205	PERIODO MASSIMO DI INDENNIZZABILITA	23	5	4
206	PROCESSO DI VALUTAZIONE DEI	23	13	10.39999962
207	SINDROME DEL TUNNEL CARPALE	23	12	9.600000381
208	AL MEDICO DEL LAVORO	22	13	10.39999962
209	ATTIVITA DI SORVEGLIANZA SANITARIA	22	13	10.39999962
210	DA PARTE DEL DATORE	22	13	10.39999962
211	DATORE DI LAVORO DEVE	22	10	8
212	DEI DISPOSITIVI DI PROTEZIONE	22	11	8.800000191
213	DEI LAVORATORI E DEI	22	16	12.80000019
214	DELLA SALUTE NEI LUOGHI	22	11	8.800000191
215	DELLA SICUREZZA DEI LAVORATORI	22	15	12
216	DI LAVORO E DI	22	15	12
217	DI SALUTE DEL LAVORATORE	22	10	8
218	IL MEDICO COMPETENTE E	22	11	8.800000191
219	IN MATERIA DI SICUREZZA	22	13	10.39999962

220	IN MATERIA DI TUTELA	22	11	8.800000191
221	IN UNO STUDIO CONDOTTO	22	17	13.60000038
222	INTERESSE PER IL LAVORO	22	5	4
223	NEL MOMENTO IN CUI	22	13	10.39999962
224	PARTE DEL DATORE DI	22	13	10.39999962
225	PER I QUALI E	22	14	11.19999981
226	PER IL MEDICO DEL	22	9	7.199999809
227	PER LA SALUTE UMANA	22	10	8
228	PER LA VALUTAZIONE DEI	22	19	15.19999981
229	PER LE QUALI E	22	11	8.800000191
230	RISPETTO ALLA POPOLAZIONE GENERALE	22	9	7.199999809
231	RISULTATI DELLA SORVEGLIANZA SANITARIA	22	12	9.600000381
232	SISTEMA DI GESTIONE DELLA	22	11	8.800000191
233	TABELLA DELLE MALATTIE PROFESSIONALI	22	7	5.599999905
234	DATORE DI LAVORO DI	21	12	9.600000381
235	DI CUI ALL'ART	21	6	4.800000191
236	DI INFORMAZIONE E FORMAZIONE	21	15	12
237	E DELLA SICUREZZA DEI	21	14	11.19999981
238	I DATI SONO STATI	21	14	11.19999981
239	IL MIGLIORAMENTO DELLE CONDIZIONI	21	16	12.80000019
240	IN UN GRUPPO DI	21	12	9.600000381
241	LE MISURE DI PREVENZIONE	21	13	10.39999962
242	MATERIA DI TUTELA DELLA	21	10	8
243	NEL LUOGO DI LAVORO	21	13	10.39999962
244	NELLA POPOLAZIONE GENERALE E	21	13	10.39999962
245	PATOLOGIE DA SOVRACCARICO BIOMECCANICO	21	9	7.199999809
246	PER LA SORVEGLIANZA SANITARIA	21	17	13.60000038
247	PER LO SVILUPPO DI	21	17	13.60000038
248	PROMOZIONE DELLA SALUTE E	21	9	7.199999809
249	RISCHIO DA STRESS LAVORO	21	9	7.199999809
250	SCOPO DEL PRESENTE LAVORO	21	13	10.39999962
251	SERVIZI DI PREVENZIONE E	21	13	10.39999962



252	SONO STATI SOTTOPOSTI A	21	8	6.400000095
253	VALUTAZIONE DEL RISCHIO CHIMICO	21	6	4.800000191
254	VALUTAZIONE DELLO STRESS LAVORO	21	8	6.400000095
255	A FRONTE DI UN	20	11	8.800000191
256	A FRONTE DI UNA	20	15	12
257	CHE L'ESPOSIZIONE A	20	7	5.599999905
258	DEGLI INFORTUNI E DELLE	20	10	8
259	DEL GRUPPO DI LAVORO	20	11	8.800000191
260	DEL MINISTERO DEL LAVORO	20	11	8.800000191
261	DI DISTURBI DEL SONNO	20	5	4
262	DI SORVEGLIANZA SANITARIA E	20	14	11.19999981
263	E STATO UTILIZZATO IL	20	13	10.39999962
264	IN PARTICOLARE PER QUANTO	20	13	10.39999962
265	LA GESTIONE DEL RISCHIO	20	10	8
266	LA PREVENZIONE NEI LUOGHI	20	7	5.599999905
267	LA VALUTAZIONE DELLO STRESS	20	7	5.599999905
268	LINEE GUIDA PER LA	20	13	10.39999962
269	NELL AMBIENTE DI LAVORO	20	5	4
270	NELLA VALUTAZIONE DEL RISCHIO	20	14	11.19999981
271	PER QUANTO CONCERNE IL	20	16	12.80000019
272	PICCOLE E MEDIE IMPRESE	20	10	8
273	SECONDO I PRINCIPI DELLA	20	14	11.19999981
274	DA PARTE DI UN	19	16	12.80000019
275	DEI LAVORATORI DELLA SANITA	19	7	5.599999905
276	DEL LAVORO E DELLA	19	12	9.600000381
277	IN GRADO DI FORNIRE	19	15	12
278	LA MAGGIOR PARTE DEGLI	19	15	12
279	LA RACCOLTA DEI DATI	19	15	12
280	LA SICUREZZA DEI LAVORATORI	19	13	10.39999962
281	LO STILE DI VITA	19	7	5.599999905
282	MAGGIOR PARTE DEI CASI	19	16	12.80000019
283	NEGLI ULTIMI ANNI E	19	10	8

284	PER LA DETERMINAZIONE DI	19	8	6.400000095
285	UN AUMENTATO RISCHIO DI	19	8	6.400000095
286	A LIVELLI DI RUMORE	18	6	4.800000191
287	A TUTTI I LAVORATORI	18	12	9.600000381
288	AL FINE DI EVITARE	18	13	10.39999962
289	ALLA TUTELA DELLA SALUTE	18	16	12.80000019
290	DA PARTE DEGLI OPERATORI	18	13	10.39999962
291	DEL LAVORO DELLA SALUTE	18	7	5.599999905
292	DELL ARTO SUPERIORE E	18	6	4.800000191
293	DELLA CUFFIA DEI ROTATORI	18	5	4
294	DELLA SICUREZZA NEI LUOGHI	18	9	7.199999809
295	DELLE ATTIVITA DI VIGILANZA	18	7	5.599999905
296	DI RIFERIMENTO PER LA	18	11	8.800000191
297	DI TUTTI I RISCHI	18	14	11.19999981
298	DI TUTTI I SOGGETTI	18	10	8
299	ENTRATA IN VIGORE DEL	18	7	5.599999905
300	I DISPOSITIVI DI PROTEZIONE	18	10	8
301	IN MATERIA DI PREVENZIONE	18	14	11.19999981
302	IN QUESTO CASO IL	18	14	11.19999981
303	INFORTUNI E DELLE MALATTIE	18	8	6.400000095
304	ISTITUTO SUPERIORE DI SANITA	18	8	6.400000095
305	LAVORO DELLA SALUTE E	18	7	5.599999905
306	MEDICINA DEL LAVORO CON	18	9	7.199999809
307	NEL MONDO DEL LAVORO	18	13	10.39999962
308	PATOLOGIE DEGLI ARTI SUPERIORI	18	5	4
309	PER LA PRESENZA DI	18	11	8.800000191
310	PER LA PRODUZIONE DI	18	9	7.199999809
311	SALUTE DEI LAVORATORI E	18	11	8.800000191
312	SALUTE E SICUREZZA NEI	18	9	7.199999809
313	SIA IN GRADO DI	18	13	10.39999962
314	SOPRATTUTTO PER QUANTO RIGUARDA	18	14	11.19999981
315	SU TUTTO IL TERRITORIO	18	11	8.800000191

316	SU UN GRUPPO DI	18	11	8.800000191
317	TABELLE DELLE MALATTIE PROFESSIONALI	18	6	4.800000191
318	A FATTORI DI RISCHIO	17	10	8
319	AL SISTEMA MANO BRACCIO	17	7	5.599999905
320	ALLA PROMOZIONE DELLA SALUTE	17	11	8.800000191
321	ATTIVITA DEL MEDICO COMPETENTE	17	8	6.400000095
322	AUTONOME DI TRENTO E	17	7	5.599999905
323	CON L OBIETTIVO DI	17	11	8.800000191
324	CONTRO GLI INFORTUNI SUL	17	8	6.400000095
325	DALLA CESSAZIONE DELLA LAVORAZIONE	17	5	4
326	DATORE DI LAVORO CHE	17	9	7.199999809
327	DI FATTORI DI RISCHIO	17	10	8
328	DI UN SISTEMA DI	17	12	9.600000381
329	E GESTIONE DEL RISCHIO	17	9	7.199999809
330	E LE PROVINCE AUTONOME	17	7	5.599999905
331	FATTORI DI RISCHIO PROFESSIONALI	17	11	8.800000191
332	GESTIONE DELLA SALUTE E	17	11	8.800000191
333	I RISULTATI DI QUESTO	17	10	8
334	INDICE DI MASSA CORPOREA	17	11	8.800000191
335	LA CUI ORIGINE LAVORATIVA	17	5	4
336	LA SICUREZZA DEL LAVORO	17	7	5.599999905
337	LE REGIONI E LE	17	7	5.599999905
338	LO STRESS LAVORO CORRELATO	17	8	6.400000095
339	MALATTIE LA CUI ORIGINE	17	5	4
340	MEDICO DEL LAVORO IN	17	11	8.800000191
341	MOVIMENTAZIONE MANUALE DEI PAZIENTI	17	5	4
342	NAZIONALE PER LA PREVENZIONE	17	11	8.800000191
343	PER GLI OPERATORI SANITARI	17	9	7.199999809
344	PER IL DATORE DI	17	11	8.800000191
345	PREVENZIONE E LA SICUREZZA	17	7	5.599999905
346	PRINCIPALI FATTORI DI RISCHIO	17	10	8
347	PRINCIPI DELLA MEDICINA DEL	17	8	6.400000095

348	PROVINCE AUTONOME DI TRENTO	17	7	5.599999905
349	REGIONI E LE PROVINCE	17	7	5.599999905
350	SANITARIE E DI RISCHIO	17	6	4.800000191
351	SI TRATTA DI UNA	17	13	10.39999962
352	SINDROME DELLE APNEE OSTRUTTIVE	17	7	5.599999905
353	SONO STATI PRESI IN	17	13	10.39999962
354	SONO STATI SOTTOPOSTI AD	17	8	6.400000095
355	STATO DI SALUTE E	17	13	10.39999962
356	UN AUMENTO DEL RISCHIO	17	12	9.600000381
357	UN SISTEMA DI GESTIONE	17	9	7.199999809
358	UNO STUDIO CONDOTTO DA	17	10	8
359	A MAGGIOR RISCHIO DI	16	9	7.199999809
360	ABITUDINE AL FUMO DI	16	11	8.800000191
361	AL DI SOTTO DEI	16	9	7.199999809
362	AMBIENTI DI LAVORO E	16	12	9.600000381
363	CHE IL MEDICO DEL	16	12	9.600000381
364	CONDIZIONI DI LAVORO E	16	11	8.800000191
365	CUI ORIGINE LAVORATIVA E	16	5	4
366	DA PARTE DEL LAVORATORE	16	11	8.800000191
367	DEL DOCUMENTO DI VALUTAZIONE	16	6	4.800000191
368	DEL LAVORO E IGIENE	16	12	9.600000381
369	DEL MEDICO COMPETENTE IN	16	7	5.599999905
370	DEL NOSTRO STUDIO E	16	9	7.199999809
371	DI LAVORATORI ESPOSTI A	16	13	10.39999962
372	DI VOLTA IN VOLTA	16	11	8.800000191
373	DIREZIONE PROVINCIALE DEL LAVORO	16	8	6.400000095
374	E LA PRESENZA DI	16	12	9.600000381
375	E LA SICUREZZA DEI	16	10	8
376	E RISULTATO PARI A	16	7	5.599999905
377	E STATA EFFETTUATA UNA	16	10	8
378	FIGURA DEL MEDICO COMPETENTE	16	7	5.599999905
379	FUMO DI SIGARETTA E	16	7	5.599999905

380	GLI ORGANI DI VIGILANZA	16	8	6.400000095
381	HA IL COMPITO DI	16	12	9.600000381
382	I SOGGETTI AFFETTI DA	16	7	5.599999905
383	IL CARICO DI LAVORO	16	14	11.19999981
384	IN QUESTI ULTIMI ANNI	16	11	8.800000191
385	INDIRIZZI SCIENTIFICI PIÙ AVANZATI	16	12	9.600000381
386	LA MAGGIOR PARTE DI	16	12	9.600000381
387	LA PRESENZA DI UNA	16	14	11.19999981
388	LA PREVENZIONE DEGLI INFORTUNI	16	11	8.800000191
389	LA SORVEGLIANZA SANITARIA DEI	16	12	9.600000381
390	LAVORO E IGIENE INDUSTRIALE	16	12	9.600000381
391	LE CONDIZIONI DI SALUTE	16	10	8
392	MEDICI DI MEDICINA GENERALE	16	10	8
393	MEDICO DI MEDICINA GENERALE	16	8	6.400000095
394	MINISTERO DEL LAVORO E	16	10	8
395	NEL CORSO DEGLI ANNI	16	15	12
396	NEL PROCESSO DI VALUTAZIONE	16	11	8.800000191
397	NELLA MAGGIORANZA DEI CASI	16	13	10.39999962
398	NON E STATO POSSIBILE	16	10	8
399	PER I QUALI NON	16	14	11.19999981
400	PREVENZIONE E PROTEZIONE E	16	11	8.800000191
401	PROMOZIONE DELLA SALUTE NEI	16	7	5.599999905
402	PUNTO DI VISTA DELLA	16	11	8.800000191
403	SULLA BASE DEI DATI	16	11	8.800000191
404	TRA I DUE GRUPPI	16	10	8
405	TUTTO IL TERRITORIO NAZIONALE	16	9	7.199999809
406	UN MAGGIOR NUMERO DI	16	9	7.199999809
407	AI DATORI DI LAVORO	15	11	8.800000191
408	AL FINE DI VERIFICARE	15	10	8
409	ATTIVITA DI VIGILANZA IN	15	5	4
410	COMMISSIONE CONSULTIVA PERMANENTE PER	15	9	7.199999809
411	CONFERENZA PERMANENTE PER I	15	7	5.599999905

412	CONSULTIVA PERMANENTE PER LA	15	9	7.199999809
413	DA UN MINIMO DI	15	9	7.199999809
414	DAL GRUPPO DI LAVORO	15	12	9.600000381
415	DEI GIUDIZI DI IDONEITA	15	10	8
416	DEL LAVORO ED ERGONOMIA	15	5	4
417	DEL POSTO DI LAVORO	15	9	7.199999809
418	DELLA QUALITÀ DELLA VITA	15	11	8.800000191
419	DENUNCIA DI MALATTIA PROFESSIONALE	15	6	4.800000191
420	DI PROTEZIONE INDIVIDUALE DPI	15	8	6.400000095
421	DI SICUREZZA E SALUTE	15	7	5.599999905
422	DI SPECIALIZZAZIONE IN MEDICINA	15	10	8
423	DI TUMORE DEL POLMONE	15	5	4
424	DI VIGILANZA IN MATERIA	15	5	4
425	E DELLE MALATTIE PROFESSIONALI	15	9	7.199999809
426	E LA SICUREZZA DEL	15	6	4.800000191
427	E LA VALUTAZIONE DELLE	15	8	6.400000095
428	E SICUREZZA DEGLI AMBIENTI	15	8	6.400000095
429	ESPOSIZIONE AD AGENTI BIOLOGICI	15	11	8.800000191
430	FATTORI DI RISCHIO LAVORATIVI	15	8	6.400000095
431	I LAVORATORI SONO STATI	15	7	5.599999905
432	I VALORI MEDI DI	15	8	6.400000095
433	IL MEDICO COMPETENTE PUO	15	8	6.400000095
434	IL RESPONSABILE DEL SERVIZIO	15	7	5.599999905
435	IL RUOLO DEL MEDICO	15	10	8
436	IN PRESENZA DI UNA	15	11	8.800000191
437	LA MAGGIOR PARTE DEL	15	9	7.199999809
438	LA STIMA DEL RISCHIO	15	7	5.599999905
439	LAVORATORI DI SESSO MASCHILE	15	6	4.800000191
440	LAVORO E QUELLO DI	15	12	9.600000381
441	LUOGO DI LAVORO E	15	12	9.600000381
442	NUOVE TABELLE DELLE MALATTIE	15	6	4.800000191
443	ORGANIZZAZIONE MONDIALE DELLA SANITÀ	15	10	8

444	PARTICOLARE PER QUANTO RIGUARDA	15	8	6.400000095
445	PER LA VALUTAZIONE DELL'	15	7	5.599999905
446	PER QUEL CHE RIGUARDA	15	11	8.800000191
447	PERMANENTE PER I RAPPORTI	15	7	5.599999905
448	PROGRAMMA DI SORVEGLIANZA SANITARIA	15	11	8.800000191
449	SICUREZZA DEGLI AMBIENTI DI	15	8	6.400000095
450	SICUREZZA SUI LUOGHI DI	15	12	9.600000381
451	SPECIALISTI IN MEDICINA DEL	15	6	4.800000191
452	SPECIALIZZAZIONE IN MEDICINA DEL	15	10	8
453	TRASMESSE AL SISTEMA MANO	15	7	5.599999905
454	UN ELEVATO GRADO DI	15	7	5.599999905
455	UN ELEVATO NUMERO DI	15	9	7.199999809
456	UNA DIFFERENZA STATISTICAMENTE SIGNIFICATIVA	15	11	8.800000191
457	VIGILANZA IN MATERIA DI	15	5	4
458	AD ALTO RISCHIO DI	14	10	8
459	AD UN MASSIMO DI	14	9	7.199999809
460	AI FINI DELLA VALUTAZIONE	14	10	8
461	AL FINE DI POTER	14	11	8.800000191
462	AMBITO DELLA MEDICINA DEL	14	7	5.599999905
463	ATTIVITA DELLA VITA QUOTIDIANA	14	7	5.599999905
464	CHE I LIVELLI DI	14	7	5.599999905
465	COLLABORAZIONE DEL MEDICO COMPETENTE	14	10	8
466	CON PARTICOLARE RIFERIMENTO ALLA	14	9	7.199999809
467	DATORE DI LAVORO IN	14	7	5.599999905
468	DEI LAVORATORI ESPOSTI A	14	10	8
469	DEI LUOGHI DI LAVORO	14	11	8.800000191
470	DEI RISCHI PER LA	14	11	8.800000191
471	DEI VALORI DI RIFERIMENTO	14	9	7.199999809
472	DELLA CARTELLA SANITARIA E	14	5	4
473	DELLA PREVENZIONE NEI LUOGHI	14	9	7.199999809
474	DI ESPOSIZIONE AD AMIANTO	14	6	4.800000191
475	DI UNA MALATTIA PROFESSIONALE	14	5	4

476	E DELLE PROVINCE AUTONOME	14	6	4.800000191
477	E FORMAZIONE DEI LAVORATORI	14	13	10.39999962
478	E SICUREZZA DEL LAVORO	14	6	4.800000191
479	EFFETTI A LUNGO TERMINE	14	11	8.800000191
480	G ITAL MED LAV	14	6	4.800000191
481	HA LO SCOPO DI	14	12	9.600000381
482	I PRINCIPALI FATTORI DI	14	9	7.199999809
483	I RAPPORTI TRA LO	14	6	4.800000191
484	I RISULTATI DELLA SORVEGLIANZA	14	11	8.800000191
485	I VALORI DI RIFERIMENTO	14	5	4
486	IL PROCESSO DI VALUTAZIONE	14	10	8
487	IL QUESTIONARIO E STATO	14	8	6.400000095
488	IN ACCORDO CON QUANTO	14	11	8.800000191
489	IN GRADO DI INDURRE	14	9	7.199999809
490	IN PARTICOLARE PER LA	14	11	8.800000191
491	IN TEMA DI SALUTE	14	14	11.19999981
492	ISTITUTO SUPERIORE PER LA	14	5	4
493	ITAL MED LAV ERG	14	6	4.800000191
494	LA NECESSITÀ DI UN	14	12	9.600000381
495	LA SOCIETÀ ITALIANA DI	14	5	4
496	LAVORO AL FINE DI	14	10	8
497	MALATTIE PROFESSIONALI DI CUI	14	5	4
498	MEDICO DEL LAVORO CHE	14	7	5.599999905
499	MINISTERO DEL LAVORO DELLA	14	7	5.599999905
500	NELL AMBITO DELLA MEDICINA	14	7	5.599999905
501	NUOVA TABELLA DELLE MALATTIE	14	5	4
502	PER I RAPPORTI TRA	14	6	4.800000191
503	PER LA GESTIONE DEL	14	12	9.600000381
504	PER LA REALIZZAZIONE DI	14	12	9.600000381
505	PER LA VALUTAZIONE DELLA	14	11	8.800000191
506	PER LA VALUTAZIONE DELLE	14	13	10.39999962
507	PER QUANTO ATTIENE LA	14	8	6.400000095



508	PER QUANTO CONCERNE LA	14	9	7.199999809
509	PERMANENTE PER LA SALUTE	14	8	6.400000095
510	RAPPORTI TRA LO STATO	14	6	4.800000191
511	RAPPRESENTANTI DEI LAVORATORI PER	14	9	7.199999809
512	RIDUZIONE DEL RISCHIO DI	14	6	4.800000191
513	SALUTE E SICUREZZA E	14	12	9.600000381
514	SE DA UN LATO	14	10	8
515	SICUREZZA DEL LAVORO ISPESL	14	5	4
516	SOSTANZE STUPEFACENTI E PSICOTROPE	14	5	4
517	STATI PRESI IN CONSIDERAZIONE	14	11	8.800000191
518	STATO DI SALUTE DEI	14	9	7.199999809
519	SULLA VALUTAZIONE DEL RISCHIO	14	7	5.599999905
520	SUPERIORE PER LA PREVENZIONE	14	5	4
521	TRA COLORO CHE HANNO	14	8	6.400000095
522	TRASMESSE AL CORPO INTERO	14	11	8.800000191
523	UN GIUDIZIO DI IDONEITÀ	14	10	8
524	UN GRUPPO DI CONTROLLO	14	9	7.199999809
525	UNO STUDIO CONDOTTO SU	14	13	10.39999962
526	AD UN GRUPPO DI	13	7	5.599999905
527	AL FUMO DI TABACCO	13	7	5.599999905
528	ALL AMBIENTE DI LAVORO	13	7	5.599999905
529	ALLA LUCE DI QUANTO	13	11	8.800000191
530	BIOMECCANICO DEGLI ARTI SUPERIORI	13	6	4.800000191
531	CARTELLE SANITARIE E DI	13	5	4
532	DAL MINISTERO DELLA SALUTE	13	11	8.800000191
533	DATORE DI LAVORO A	13	7	5.599999905
534	DEL LAVORO E DELLE	13	11	8.800000191
535	DI FORMAZIONE E INFORMAZIONE	13	10	8
536	DI GESTIONE DELLA SALUTE	13	9	7.199999809
537	DI LAVORO E DEI	13	11	8.800000191
538	DOCUMENTO DI VALUTAZIONE DEL	13	6	4.800000191
539	E I LIVELLI DI	13	10	8

540	E SICUREZZA SUI LUOGHI	13	10	8
541	È STATA VALUTATA LA	13	6	4.800000191
542	GLI AMBIENTI DI LAVORO	13	7	5.599999905
543	I RISULTATI SONO STATI	13	11	8.800000191
544	IL MINISTERO DEL LAVORO	13	7	5.599999905
545	IL SERVIZIO SANITARIO NAZIONALE	13	6	4.800000191
546	IN CUI SONO STATI	13	11	8.800000191
547	IN LINEA CON I	13	11	8.800000191
548	IN PRESENZA DI UN	13	9	7.199999809
549	IN QUESTO CASO SI	13	13	10.39999962
550	INFORTUNI SUL LAVORO E	13	8	6.400000095
551	L ORGANIZZAZIONE DEL LAVORO	13	7	5.599999905
552	LA RICERCA È STATA	13	12	9.600000381
553	LA SINDROME DELLE APNEE	13	6	4.800000191
554	LA SORVEGLIANZA SANITARIA E	13	9	7.199999809
555	LUOGHI DI LAVORO DELLE	13	6	4.800000191
556	MONDIALE DELLA SANITA OMS	13	10	8
557	NEL CORSO DELLA SORVEGLIANZA	13	10	8
558	NEL DOCUMENTO DI VALUTAZIONE	13	11	8.800000191
559	NELLA FASCIA DI ETA	13	10	8
560	PER LA MAGGIOR PARTE	13	8	6.400000095
561	PER LA PREVENZIONE NEI	13	6	4.800000191
562	PER LA RACCOLTA DEI	13	12	9.600000381
563	PER QUANTO CONCERNE LE	13	10	8
564	POPOLAZIONE PRESA IN ESAME	13	7	5.599999905
565	PRESENZA O MENO DI	13	11	8.800000191
566	PREVENZIONE DEGLI INFORTUNI E	13	7	5.599999905
567	RISULTATI DI QUESTO STUDIO	13	9	7.199999809
568	RUOLO DEL MEDICO COMPETENTE	13	9	7.199999809
569	SALUTE E SICUREZZA SUI	13	10	8
570	SI RITIENE CHE LA	13	11	8.800000191
571	SOVRACCARICO BIOMECCANICO DEGLI ARTI	13	6	4.800000191

572	SULLE CONDIZIONI DI LAVORO	13	9	7.199999809
573	VIBRAZIONI TRASMESSE AL CORPO	13	10	8
574	A TUTTI I LIVELLI	12	11	8.800000191
575	AI SENSI DEL D	12	11	8.800000191
576	AL DI SOTTO DEL	12	12	9.600000381
577	AL DI SOTTO DELLA	12	11	8.800000191
578	AL FINE DI INDIVIDUARE	12	9	7.199999809
579	AL FINE DI VALUTARE	12	9	7.199999809
580	ALL INTERNO DI UN	12	7	5.599999905
581	ALLA MOVIMENTAZIONE MANUALE DEI	12	6	4.800000191
582	ANCHE AL FINE DI	12	9	7.199999809
583	AUMENTO DEL RISCHIO DI	12	10	8
584	CASO IN CUI IL	12	6	4.800000191
585	CHE LE LINEE GUIDA	12	5	4
586	CON PARTICOLARE RIFERIMENTO ALLE	12	6	4.800000191
587	CONDIZIONI DI SALUTE DEL	12	7	5.599999905
588	DA PARTE DEI MEDICI	12	7	5.599999905
589	DA UNA SERIE DI	12	10	8
590	DEGLI ORGANI DI VIGILANZA	12	9	7.199999809
591	DEL MEDICO COMPETENTE DI	12	7	5.599999905
592	DEL MERCATO DEL LAVORO	12	7	5.599999905
593	DELLA PROMOZIONE DELLA SALUTE	12	8	6.400000095
594	DELLA SORVEGLIANZA SANITARIA E	12	10	8
595	DELLE LAVORATRICI E DEI	12	6	4.800000191
596	DI ASSUNZIONE DI SOSTANZE	12	5	4
597	DI FORMAZIONE ED INFORMAZIONE	12	11	8.800000191
598	DI GESTIONE DELLA SICUREZZA	12	7	5.599999905
599	DI LAVORO E DEL	12	9	7.199999809
600	DI PREVENZIONE E DI	12	8	6.400000095
601	DI SALUTE E DI	12	6	4.800000191
602	DI SESSO MASCHILE E	12	6	4.800000191
603	DI TRENTO E DI	12	6	4.800000191

604	E DELLA PREVIDENZA SOCIALE	12	5	4
605	E DELLE POLITICHE SOCIALI	12	10	8
606	E LA CAPACITA DI	12	11	8.800000191
607	E LA POSSIBILITA DI	12	8	6.400000095
608	È STATA CONDOTTA UNA	12	12	9.600000381
609	È STATA OGGETTO DI	12	9	7.199999809
610	È STATA OSSERVATA UNA	12	11	8.800000191
611	È STATA UTILIZZATA LA	12	12	9.600000381
612	ESPOSTI A RISCHIO BIOLOGICO	12	5	4
613	GRUPPI DI LAVORATORI ESPOSTI	12	10	8
614	I DATI RELATIVI ALLE	12	6	4.800000191
615	I LUOGHI DI LAVORO	12	8	6.400000095
616	I SOGGETTI DI SESSO	12	8	6.400000095
617	IL MAGGIOR NUMERO DI	12	9	7.199999809
618	IN COLLABORAZIONE CON IL	12	8	6.400000095
619	IN QUESTO CASO E	12	8	6.400000095
620	INFORMAZIONE E FORMAZIONE DEI	12	11	8.800000191
621	LA DURATA E LA	12	6	4.800000191
622	LA FIGURA DEL MEDICO	12	8	6.400000095
623	LA NECESSITÀ DI UNA	12	10	8
624	LA PERCEZIONE DEL RISCHIO	12	10	8
625	LA SS DEGLI OS	12	5	4
626	MAGGIOR PARTE DEI SOGGETTI	12	8	6.400000095
627	MEDICI DEL LAVORO E	12	9	7.199999809
628	MINISTERO DELLA SALUTE E	12	9	7.199999809
629	NEGLI AMBIENTI DI VITA	12	8	6.400000095
630	NELLA VALUTAZIONE DEI RISCHI	12	11	8.800000191
631	PER IL MIGLIORAMENTO DELLE	12	9	7.199999809
632	PER LE QUALI NON	12	8	6.400000095
633	PER UNA SERIE DI	12	11	8.800000191
634	RISCHIO DI ESPOSIZIONE A	12	9	7.199999809
635	SISTEMA INFORMATIVO NAZIONALE PER	12	6	4.800000191

636	SODDISFAZIONE PER IL LAVORO	12	5	4
637	SOGGETTI DI SESSO FEMMINILE	12	5	4
638	SONO STATI SUDDIVISI IN	12	8	6.400000095
639	SULLA SALUTE DEI LAVORATORI	12	11	8.800000191
640	TRA I LIVELLI DI	12	8	6.400000095
641	TRENTO E DI BOLZANO	12	6	4.800000191
642	UN AMBIENTE DI LAVORO	12	9	7.199999809
643	VALORI LIMITE DI ESPOSIZIONE	12	8	6.400000095
644	A CARICO DEGLI ARTI	11	8	6.400000095
645	A CARICO DEL DATORE	11	7	5.599999905
646	A CARICO DEL SISTEMA	11	9	7.199999809
647	A CAUSA DI UNA	11	11	8.800000191
648	AD UNA CONDIZIONE DI	11	8	6.400000095
649	AL DI SOTTO DI	11	9	7.199999809
650	AL FINE DI IDENTIFICARE	11	8	6.400000095
651	AL FINE DI OTTENERE	11	9	7.199999809
652	ALLE DIFFERENZE DI GENERE	11	11	8.800000191
653	CARICO DEL DATORE DI	11	7	5.599999905
654	CHE LA PREVALENZA DI	11	5	4
655	CHE LA SORVEGLIANZA SANITARIA	11	10	8
656	CHE SI TRATTA DI	11	7	5.599999905
657	COLLEGATI ALLO STRESS LAVORO	11	8	6.400000095
658	CON IL MEDICO COMPETENTE	11	9	7.199999809
659	CONSUMO DI ALCOL E	11	7	5.599999905
660	CORRELAZIONE STATISTICAMENTE SIGNIFICATIVA TRA	11	6	4.800000191
661	DAI MEDICI DEL LAVORO	11	8	6.400000095
662	DATORE DI LAVORO O	11	7	5.599999905
663	DEI LIVELLI DI ESPOSIZIONE	11	8	6.400000095
664	DEL MEDICO COMPETENTE ALLA	11	5	4
665	DEL NUMERO DI INFORTUNI	11	6	4.800000191
666	DEL PRESENTE STUDIO E	11	9	7.199999809
667	DEL RISCHIO DA PARTE	11	10	8

668	DEL SISTEMA NERVOSO CENTRALE	11	10	8
669	DELLA SICUREZZA E DELLA	11	8	6.400000095
670	DELLA SICUREZZA SUL LAVORO	11	9	7.199999809
671	DELLE MALATTIE PROFESSIONALI DI	11	5	4
672	DELLE REGIONI E DELLE	11	6	4.800000191
673	DEVE RIGUARDARE TUTTI I	11	9	7.199999809
674	DI ASSENZA DAL LAVORO	11	9	7.199999809
675	DI CUI AGLI ARTT	11	5	4
676	DI ESPOSIZIONE E DI	11	8	6.400000095
677	DI FRONTE AD UNA	11	11	8.800000191
678	DI GESTIONE DEL RISCHIO	11	7	5.599999905
679	DI QUESTO STUDIO E	11	6	4.800000191
680	DI QUESTO TIPO DI	11	11	8.800000191
681	DI SOSTANZE PSICOTROPE E	11	5	4
682	DURANTE IL PERIODO DI	11	7	5.599999905
683	È DI RISCHIO E	11	7	5.599999905
684	È EMERSO CHE IL	11	9	7.199999809
685	E FATTORI DI RISCHIO	11	8	6.400000095
686	E LA PROMOZIONE DELLA	11	10	8
687	È RISULTATA PARI A	11	8	6.400000095
688	È STATA RILEVATA IN	11	8	6.400000095
689	È STATA RILEVATA UNA	11	10	8
690	È STATO SOMMINISTRATO UN	11	7	5.599999905
691	FATTORI DI RISCHIO CARDIOVASCOLARE	11	6	4.800000191
692	FATTORI DI RISCHIO DI	11	8	6.400000095
693	FATTORI DI RISCHIO PRESENTI	11	7	5.599999905
694	GLI EFFETTI SULLA SALUTE	11	10	8
695	GRADO DI DETERMINARE UN	11	9	7.199999809
696	HA EVIDENZIATO CHE LA	11	10	8
697	I DATI RELATIVI AGLI	11	10	8
698	I PRINCIPI DELLA MEDICINA	11	8	6.400000095

699	I SOGGETTI SONO STATI	11	5	4
700	IL FUMO DI SIGARETTA	11	5	4
701	IL RISCHIO PER LA	11	9	7.199999809
702	IL SERVIZIO DI PREVENZIONE	11	7	5.599999905
703	IN ACCORDO CON I	11	11	8.800000191
704	IN ACCORDO CON LA	11	9	7.199999809
705	IN MANIERA TALE DA	11	10	8
706	IN QUESTO CASO LA	11	9	7.199999809
707	IN TUTTO IL MONDO	11	10	8
708	IN UN OTTICA DI	11	10	8
709	KG DI PESO CORPOREO	11	7	5.599999905
710	LA CONCENTRAZIONE MEDIA DI	11	7	5.599999905
711	LA FACILITÀ DI ACCESSO	11	5	4
712	LA PERCENTUALE DI SOGGETTI	11	8	6.400000095
713	LA RIDUZIONE DEI RISCHI	11	10	8
714	LAVORATRICI E DEI LAVORATORI	11	5	4
715	LAVORO E DELLA PREVIDENZA	11	5	4
716	LE QUALI È OBBLIGATORIA	11	7	5.599999905
717	LO STATO LE REGIONI	11	6	4.800000191
718	MA ANCHE PER LA	11	11	8.800000191
719	MALATTIE PER LE QUALI	11	8	6.400000095
720	MEDICI COMPETENTI CHE OPERANO	11	5	4
721	MEDICINA DEL LAVORO DI	11	5	4
722	MEDICINA DEL LAVORO IN	11	9	7.199999809
723	MEDICINA DEL LAVORO LA	11	8	6.400000095
724	MEDICINA PREVENTIVA DEI LAVORATORI	11	7	5.599999905
725	MEDICO DEL LAVORO DEVE	11	6	4.800000191
726	MEDICO DEL LAVORO DOVRA	11	5	4
727	MISURE PER LA TUTELA	11	9	7.199999809
728	NELLA MAGGIOR PARTE DELLE	11	10	8
729	NELLA TABELLA I SONO	11	11	8.800000191
730	ORIGINE LAVORATIVA E DI	11	5	4

731	PER LA PROMOZIONE DELLA	11	5	4
732	PER LA SS DEGLI	11	5	4
733	PIÙ IN GRADO DI	11	8	6.400000095
734	PREVENZIONE E SICUREZZA NEI	11	8	6.400000095
735	PROGRAMMI VOLONTARI DI PROMOZIONE	11	9	7.199999809
736	QUALI AD ESEMPIO LA	11	11	8.800000191
737	QUALI È OBBLIGATORIA LA	11	7	5.599999905
738	REGIONI E DELLE PROVINCE	11	6	4.800000191
739	RIDUZIONE DEL NUMERO DI	11	6	4.800000191
740	RIGUARDARE TUTTI I RISCHI	11	9	7.199999809
741	SI PUÒ AFFERMARE CHE	11	8	6.400000095
742	SOMMINISTRAZIONE DI UN QUESTIONARIO	11	9	7.199999809
743	SONO STATI CONFRONTATI CON	11	6	4.800000191
744	STATO LE REGIONI E	11	6	4.800000191
745	STRESS LAVORO CORRELATO IN	11	6	4.800000191
746	SUI RISCHI PER LA	11	8	6.400000095
747	SULLA BASE DI QUANTO	11	7	5.599999905
748	SULLA BASE DI UN	11	10	8
749	SULLE CONDIZIONI DI SALUTE	11	8	6.400000095
750	TEMA DI SALUTE E	11	11	8.800000191
751	TRA LO STATO LE	11	6	4.800000191
752	UN ELEVATO RISCHIO DI	11	7	5.599999905
753	UN MAGGIOR RISCHIO DI	11	9	7.199999809
754	UN NUMERO LIMITATO DI	11	8	6.400000095
755	UN RISCHIO PER LA	11	8	6.400000095
756	UN VERO E PROPRIO	11	10	8
757	UNA REVISIONE DELLA LETTERATURA	11	8	6.400000095
758	VALUTAZIONE DEL RISCHIO PER	11	9	7.199999809
759	VALUTAZIONE DI TUTTI I	11	8	6.400000095
760	VOLONTARI DI PROMOZIONE DELLA	11	9	7.199999809
761	A TURNI E NOTTURNO	10	5	4
762	AD ELEVATO RISCHIO DI	10	6	4.800000191



763	AD UNA SERIE DI	10	9	7.199999809
764	AI FINI DI UNA	10	8	6.400000095
765	AL FINE DI GARANTIRE	10	5	4
766	AL FINE DI RIDURRE	10	9	7.199999809
767	ALLA MEDICINA DEL LAVORO	10	7	5.599999905
768	ALLO SCOPO DI INDIVIDUARE	10	7	5.599999905
769	AMBIENTI DI VITA E	10	7	5.599999905
770	CHE LA VALUTAZIONE DEI	10	8	6.400000095
771	CON ETÀ MEDIA DI	10	8	6.400000095
772	CONDIZIONI DI VITA E	10	9	7.199999809
773	CONTO DEL FATTO CHE	10	9	7.199999809
774	DA UN LATO E	10	10	8
775	DALLA COMMISSIONE CONSULTIVA PERMANENTE	10	7	5.599999905
776	DATORE DI LAVORO PER	10	5	4
777	DEI DIRITTI DEI LAVORATORI	10	7	5.599999905
778	DEI LAVORATORI E DELLE	10	7	5.599999905
779	DEL LAVORO E DEL	10	9	7.199999809
780	DEL LAVORO E DI	10	8	6.400000095
781	DEL LAVORO ED IGIENE	10	7	5.599999905
782	DEL MONDO DEL LAVORO	10	8	6.400000095
783	DEL RISCHIO E LA	10	9	7.199999809
784	DEL SERVIZIO SANITARIO NAZIONALE	10	5	4
785	DELLA LETTERATURA SCIENTIFICA E	10	7	5.599999905
786	DELLA NORMATIVA IN MATERIA	10	5	4
787	DELLA SOCIETÀ ITALIANA DI	10	8	6.400000095
788	DI IGIENE E SICUREZZA	10	8	6.400000095
789	DI INIZIO E FINE	10	6	4.800000191
790	DI LAVORO AI FINI	10	9	7.199999809
791	DI LAVORO E LA	10	8	6.400000095
792	DI LAVORO IN CUI	10	7	5.599999905
793	DI MASSA CORPOREA BMI	10	8	6.400000095
794	DI MEDICINA PREVENTIVA DEI	10	6	4.800000191

795	DI QUESTO LAVORO E	10	8	6.400000095
796	DI SENSIBILITÀ E SPECIFICITÀ	10	7	5.599999905
797	DI SICUREZZA SUL LAVORO	10	8	6.400000095
798	DI UNA CONDIZIONE DI	10	8	6.400000095
799	DI UNO O PIÙ	10	9	7.199999809
800	DI VALUTARE IL RISCHIO	10	8	6.400000095
801	E DELLA SICUREZZA NEI	10	6	4.800000191
802	E IL MEDICO COMPETENTE	10	7	5.599999905
803	E LA MANCANZA DI	10	9	7.199999809
804	E LO SVILUPPO DI	10	9	7.199999809
805	È OBBLIGATORIA LA DENUNCIA	10	6	4.800000191
806	E PIÙ IN GENERALE	10	8	6.400000095
807	EFFETTI NEGATIVI SULLA SALUTE	10	6	4.800000191
808	ELEVATO RISCHIO DI INFORTUNI	10	7	5.599999905
809	FATTORI DI RISCHIO BIOMECCANICO	10	5	4
810	FUMATORI E NON FUMATORI	10	6	4.800000191
811	GIUDIZIO DI IDONEITÀ AL	10	5	4
812	GLI INDIRIZZI SCIENTIFICI PIÙ	10	8	6.400000095
813	GLI ORARI DI LAVORO	10	7	5.599999905
814	HANNO LA POSSIBILITÀ DI	10	5	4
815	I DATI RELATIVI AI	10	9	7.199999809
816	I GIORNI DI RIPOSO	10	5	4
817	I LAVORATORI ESPOSTI A	10	6	4.800000191
818	IL COEFFICIENTE DI CORRELAZIONE	10	5	4
819	IL LIVELLO DI ESPOSIZIONE	10	7	5.599999905
820	IL MEDICO COMPETENTE NON	10	6	4.800000191
821	IL MEDICO COMPETENTE SI	10	8	6.400000095
822	IL RISCHIO DI ESPOSIZIONE	10	6	4.800000191
823	IL TEST T DI	10	5	4
824	IL TURNO DEL MATTINO	10	5	4
825	IL TURNO DI NOTTE	10	5	4
826	IN BASE A QUANTO	10	7	5.599999905

827	IN COLLABORAZIONE CON GLI	10	6	4.800000191
828	IN COLLABORAZIONE CON L'	10	10	8
829	IN ENTRAMBI I SESSI	10	8	6.400000095
830	IN GRADO DI GARANTIRE	10	7	5.599999905
831	IN GRADO DI MODIFICARE	10	7	5.599999905
832	IN SEDE DI CONFERENZA	10	6	4.800000191
833	IN TEMA DI SICUREZZA	10	5	4
834	IN UN CAMPIONE DI	10	7	5.599999905
835	IN UN CONTESTO DI	10	10	8
836	IN UN RECENTE STUDIO	10	9	7.199999809
837	IN VIA DI SVILUPPO	10	8	6.400000095
838	INFORTUNI SUL LAVORO INAIL	10	6	4.800000191
839	L'ANALISI DEI DATI	10	6	4.800000191
840	LA COLLABORAZIONE DEL MEDICO	10	9	7.199999809
841	LA FORMULAZIONE DEL GIUDIZIO	10	6	4.800000191
842	LA MOVIMENTAZIONE MANUALE DEI	10	8	6.400000095
843	LA POSSIBILITÀ DI UN	10	7	5.599999905
844	LA PRESENZA O MENO	10	8	6.400000095
845	LA PRIMA VOLTA IN	10	9	7.199999809
846	LA TABELLA I RIPORTA	10	9	7.199999809
847	LA VALUTAZIONE DELL'EFFICACIA	10	6	4.800000191
848	LAVORO ED IGIENE INDUSTRIALE	10	7	5.599999905
849	LE STRATEGIE DI COPING	10	8	6.400000095
850	LO SCOPO DEL PRESENTE	10	9	7.199999809
851	LO SCOPO DI FORNIRE	10	7	5.599999905
852	LO STUDIO E STATO	10	8	6.400000095
853	LO SVOLGIMENTO DI ATTIVITÀ	10	9	7.199999809
854	MEDICINA DEL LAVORO SIA	10	7	5.599999905
855	MESSA IN ATTO DI	10	8	6.400000095
856	MODELLO DI REGRESSIONE LOGISTICA	10	5	4
857	NEI CONFRONTI DEI LAVORATORI	10	7	5.599999905
858	NEI LAVORATORI ESPOSTI A	10	6	4.800000191

859	NEL CORSO DI UN	10	7	5.599999905
860	NORMATIVA IN MATERIA DI	10	5	4
861	PATTO PER LA SALUTE	10	5	4
862	PER I LAVORATORI ESPOSTI	10	7	5.599999905
863	PER LA MEDICINA DEL	10	7	5.599999905
864	PER LA PREVENZIONE DEGLI	10	6	4.800000191
865	PER LA PREVENZIONE DELLA	10	6	4.800000191
866	PREVENTIVA DEI LAVORATORI DELLA	10	6	4.800000191
867	PREVENZIONE DEGLI INFORTUNI SUL	10	5	4
868	PROCESSO DI VALUTAZIONE DEL	10	8	6.400000095
869	PROGRAMMI DI PROMOZIONE DELLA	10	7	5.599999905
870	PUNTO DI VISTA CLINICO	10	9	7.199999809
871	QUANTO RIGUARDA LA VALUTAZIONE	10	7	5.599999905
872	RIDURRE IL RISCHIO DI	10	8	6.400000095
873	SERVIZI DI MEDICINA DEL	10	6	4.800000191
874	SI E RITENUTO OPPORTUNO	10	9	7.199999809
875	SONO STATI ANALIZZATI I	10	9	7.199999809
876	STATISTICA DEI DATI E	10	9	7.199999809
877	STATO SOMMINISTRATO UN QUESTIONARIO	10	7	5.599999905
878	SULLA BASE DI QUESTE	10	9	7.199999809
879	T DI STUDENT PER	10	5	4
880	TEST T DI STUDENT	10	5	4
881	TUTTE LE FIGURE PROFESSIONALI	10	8	6.400000095
882	UN DOCUMENTO DI CONSENSO	10	7	5.599999905
883	UN GRAN NUMERO DI	10	9	7.199999809
884	UN PUNTO DI RIFERIMENTO	10	8	6.400000095
885	UNA CORRELAZIONE STATISTICAMENTE SIGNIFICATIVA	10	6	4.800000191
886	UNA DIFFERENZA SIGNIFICATIVA TRA	10	6	4.800000191
887	UNA VASTA GAMMA DI	10	8	6.400000095
888	VALUTAZIONE DEI RISCHI CHE	10	8	6.400000095
889	VALUTAZIONE DEI RISCHI IN	10	9	7.199999809
890	VALUTAZIONE E GESTIONE DEL	10	6	4.800000191

		13536		
--	--	-------	--	--

**Appendice 5: Lexical bundles di 4 parole in sequenza nel corpus specialistico (storia, filosofia e medicina del lavoro)**

Corpus specialistico - lexical bundles di 4 parole				
N	Lexical bundle	Frequenza	Articoli	%
1	DAL PUNTO DI VISTA	460	158	48.76543045
2	NEI LUOGHI DI LAVORO	216	44	13.58024693
3	IL DATORE DI LAVORO	191	41	12.65432072
4	UN PUNTO DI VISTA	184	83	25.61728477
5	SI TRATTA DI UN	179	98	30.24691391
6	NEL MOMENTO IN CUI	175	81	25
7	PER LA PRIMA VOLTA	171	96	29.62962914
8	DI MEDICINA DEL LAVORO	158	33	10.18518543
9	PER QUANTO RIGUARDA LA	155	70	21.60493851
10	DA UN PUNTO DI	151	76	23.45679092
11	ANCHE IN QUESTO CASO	144	70	21.60493851
12	UNA VERA E PROPRIA	142	77	23.76543236
13	IL MEDICO DEL LAVORO	128	22	6.790123463
14	PER QUANTO RIGUARDA IL	124	71	21.91357994
15	DEL DATORE DI LAVORO	119	30	9.259259224
16	È IN GRADO DI	118	69	21.29629707
17	LA VALUTAZIONE DEL RISCHIO	109	37	11.41975307
18	SI TRATTA DI UNA	106	75	23.14814758
19	QUESTO PUNTO DI VISTA	101	62	19.13580322
20	A PARTIRE DAGLI ANNI	100	49	15.12345695
21	DI VOLTA IN VOLTA	99	58	17.90123367
22	LA MAGGIOR PARTE DEI	98	58	17.90123367
23	UN VERO E PROPRIO	97	68	20.98765373
24	NEL CORSO DEGLI ANNI	96	48	14.81481457
25	DELLA MEDICINA DEL LAVORO	94	28	8.641975403
26	DI SALUTE E SICUREZZA	94	27	8.333333015
27	DELLA SECONDA GUERRA MONDIALE	93	35	10.80246925

28	IN MEDICINA DEL LAVORO	93	17	5.246913433
29	DA QUESTO PUNTO DI	92	56	17.28395081
30	DI PREVENZIONE E PROTEZIONE	92	28	8.641975403
31	MEDICINA DEL LAVORO E	91	27	8.333333015
32	DEL MEDICO DEL LAVORO	90	21	6.481481552
33	NELLA MAGGIOR PARTE DEI	89	51	15.74074078
34	NELLA MISURA IN CUI	89	40	12.34567928
35	DI VALUTAZIONE DEL RISCHIO	88	32	9.876543045
36	A CHE FARE CON	87	47	14.50617313
37	DI VALUTAZIONE DEI RISCHI	84	27	8.333333015
38	DELLA SALUTE E DELLA	82	25	7.716049194
39	PER LA SALUTE E	81	34	10.49382687
40	PER QUANTO RIGUARDA L'	81	38	11.72839546
41	TUTELA DELLA SALUTE E	81	26	8.024691582
42	RISCHI PER LA SALUTE	79	32	9.876543045
43	MAGGIOR PARTE DEI CASI	78	45	13.88888931
44	SALUTE E SICUREZZA SUL	78	22	6.790123463
45	ALLA FINE DEGLI ANNI	77	51	15.74074078
46	LA SECONDA GUERRA MONDIALE	77	25	7.716049194
47	SONO IN GRADO DI	77	48	14.81481457
48	E SICUREZZA SUL LAVORO	76	22	6.790123463
49	NEGLI AMBIENTI DI LAVORO	75	35	10.80246925
50	DATORE DI LAVORO E	74	25	7.716049194
51	DEL MERCATO DEL LAVORO	74	28	8.641975403
52	PER QUANTO RIGUARDA LE	74	39	12.0370369
53	LA MAGGIOR PARTE DELLE	73	53	16.3580246
54	DI PROMOZIONE DELLA SALUTE	70	23	7.098765373
55	SALUTE E DELLA SICUREZZA	70	20	6.172839642
56	ALL'INTERNO DI UN	68	43	13.27160454
57	SE DA UN LATO	68	43	13.27160454
58	MOVIMENTAZIONE MANUALE DEI CARICHI	67	15	4.629629612
59	LA MAGGIOR PARTE DEGLI	66	51	15.74074078

60	PER LA MAGGIOR PARTE	66	38	11.72839546
61	AL DATORE DI LAVORO	64	21	6.481481552
62	LA PRESENZA DI UN	64	44	13.58024693
63	PUNTO DI VISTA DELLA	64	43	13.27160454
64	DELLA SALUTE E SICUREZZA	63	22	6.790123463
65	LA VALUTAZIONE DEI RISCHI	63	25	7.716049194
66	I FATTORI DI RISCHIO	62	27	8.333333015
67	LA MEDICINA DEL LAVORO	62	19	5.864197731
68	LA SALUTE DEI LAVORATORI	62	28	8.641975403
69	DEL GIUDIZIO DI IDONEITÀ	60	15	4.629629612
70	NEL CASO IN CUI	60	38	11.72839546
71	SICUREZZA NEI LUOGHI DI	60	19	5.864197731
72	DI UNA SERIE DI	58	42	12.9629631
73	DI FRONTE A UN	57	38	11.72839546
74	IN ENTRAMBI I CASI	56	48	14.81481457
75	PER QUANTO RIGUARDA I	56	33	10.18518543
76	SUL LUOGO DI LAVORO	56	26	8.024691582
77	DEI FATTORI DI RISCHIO	55	20	6.172839642
78	DELLO STRESS LAVORO CORRELATO	55	11	3.395061731
79	DI TUTELA DELLA SALUTE	55	24	7.407407284
80	E.S.M.I.	55	18	5.55555344
81	IL MODO IN CUI	55	38	11.72839546
82	IN UN CERTO SENSO	55	31	9.567901611
83	PREVENZIONE NEI LUOGHI DI	55	15	4.629629612
84	SERVIZIO DI PREVENZIONE E	55	20	6.172839642
85	A PARTIRE DALLA FINE	54	27	8.333333015
86	AL DI LÀ DELLA	54	38	11.72839546
87	ALL'INTERNO DI UNA	54	36	11.11111069
88	DA PARTE DI UN	54	40	12.34567928
89	LA MAGGIOR PARTE DI	54	34	10.49382687
90	CON LO SCOPO DI	53	34	10.49382687
91	DA PARTE DEL MEDICO	53	21	6.481481552



92	DI UN GRUPPO DI	53	40	12.34567928
93	I DATORI DI LAVORO	53	25	7.716049194
94	IN MATERIA DI SALUTE	53	19	5.864197731
95	IN QUESTO SENSO LA	53	31	9.567901611
96	COME SI È VISTO	52	41	12.65432072
97	AL DI LÀ DELLE	51	34	10.49382687
98	AL DI LÀ DI	51	39	12.0370369
99	DA UN LATO E	51	41	12.65432072
100	AL DI LÀ DEL	50	40	12.34567928
101	LA QUALITÀ DELLA VITA	50	25	7.716049194
102	PER LA SICUREZZA E	50	21	6.481481552
103	CHE A SUA VOLTA	49	37	11.41975307
104	CHE LA MAGGIOR PARTE	49	29	8.950616837
105	DISPOSITIVI DI PROTEZIONE INDIVIDUALE	49	17	5.246913433
106	LA SICUREZZA E LA	49	27	8.333333015
107	MATERIA DI SALUTE E	49	16	4.938271523
108	IL GIUDIZIO DI IDONEITÀ	48	18	5.55555344
109	LA SALUTE E SICUREZZA	48	19	5.864197731
110	LA TUTELA DELLA SALUTE	48	22	6.790123463
111	PER QUANTO RIGUARDA GLI	48	37	11.41975307
112	AL DI FUORI DEL	47	36	11.11111069
113	COMUNITÀ ITALIANA IN GERMANIA	47	6	1.851851821
114	E PIÙ IN GENERALE	47	32	9.876543045
115	IL FATTO CHE LA	47	35	10.80246925
116	IN QUESTO CASO IL	47	39	12.0370369
117	DEL LAVORO E DELLA	46	21	6.481481552
118	DI VITA E DI	46	32	9.876543045
119	SICUREZZA E LA SALUTE	46	24	7.407407284
120	ALL'INIZIO DEGLI ANNI	45	25	7.716049194
121	CON L'INTENTO DI	45	26	8.024691582
122	DI ETÀ COMPRESA TRA	45	24	7.407407284
123	IN TUTTO IL MONDO	45	37	11.41975307

124	MOMENTO IN CUI LA	45	28	8.641975403
125	PER LA VALUTAZIONE DEL	45	31	9.567901611
126	PER QUEL CHE RIGUARDA	45	26	8.024691582
127	PER UN TOTALE DI	45	28	8.641975403
128	SUL MERCATO DEL LAVORO	45	13	4.012345791
129	ESSERE IN GRADO DI	44	35	10.80246925
130	RISCHIO PER LA SALUTE	44	23	7.098765373
131	DA PARTE DEI LAVORATORI	43	27	8.333333015
132	DEGLI ITALIANI ALL'ESTERO	43	11	3.395061731
133	DEI DATORI DI LAVORO	43	22	6.790123463
134	DELLA SALUTE DEI LAVORATORI	43	20	6.172839642
135	DELLO STATO DI SALUTE	43	22	6.790123463
136	GIOVANI DI ORIGINE ITALIANA	43	7	2.160493851
137	NEL MONDO DEL LAVORO	43	26	8.024691582
138	DI LAVORO E DI	42	28	8.641975403
139	E ALLO STESSO TEMPO	42	28	8.641975403
140	I RISCHI PER LA	42	20	6.172839642
141	IL PUNTO DI PARTENZA	42	35	10.80246925
142	LA SALUTE E LA	42	22	6.790123463
143	SU UN CAMPIONE DI	42	23	7.098765373
144	TUTTA UNA SERIE DI	42	29	8.950616837
145	MEDICO DEL LAVORO COMPETENTE	41	15	4.629629612
146	NEGLI STATI UNITI E	41	21	6.481481552
147	NON SI TRATTA DI	41	34	10.49382687
148	DELLA VALUTAZIONE DEI RISCHI	40	22	6.790123463
149	FARE I CONTI CON	40	29	8.950616837
150	ALLA VALUTAZIONE DEI RISCHI	39	17	5.246913433
151	CHE IL MEDICO COMPETENTE	39	15	4.629629612
152	DEGLI AMBIENTI DI LAVORO	39	16	4.938271523
153	DELL ORGANIZZAZIONE DEL LAVORO	39	11	3.395061731
154	DI UN SISTEMA DI	39	32	9.876543045
155	LA PROMOZIONE DELLA SALUTE	39	17	5.246913433

156	NELLA SECONDA METÀ DEL	39	29	8.950616837
157	SIA PER QUANTO RIGUARDA	39	24	7.407407284
158	SULLA BASE DI UNA	39	30	9.259259224
159	AL DI FUORI DI	38	33	10.18518543
160	ALLA VALUTAZIONE DEL RISCHIO	38	18	5.55555344
161	DEI MEDICI DEL LAVORO	38	13	4.012345791
162	FINE DEGLI ANNI SESSANTA	38	24	7.407407284
163	LA FINE DEGLI ANNI	38	24	7.407407284
164	SALUTE E LA SICUREZZA	38	21	6.481481552
165	SE È VERO CHE	38	28	8.641975403
166	SECONDA META DEGLI ANNI	38	27	8.333333015
167	VALUTAZIONE DEL RISCHIO STRESS	38	8	2.469135761
168	A FRONTE DI UNA	37	30	9.259259224
169	COME SI È DETTO	37	26	8.024691582
170	DALLA FINE DEGLI ANNI	37	19	5.864197731
171	E AL TEMPO STESSO	37	31	9.567901611
172	È IL CASO DI	37	31	9.567901611
173	È LA NECESSITÀ DI	37	27	8.333333015
174	IL PUNTO DI VISTA	37	28	8.641975403
175	IN GRADO DI DETERMINARE	37	20	6.172839642
176	IN QUESTO CASO È	37	27	8.333333015
177	IN QUESTO CASO LA	37	29	8.950616837
178	IN UNA SORTA DI	37	27	8.333333015
179	L'ESISTENZA DI UN	37	28	8.641975403
180	MERCATO DEL LAVORO E	37	18	5.55555344
181	MIGLIORAMENTO DELLE CONDIZIONI DI	37	26	8.024691582
182	NEL MERCATO DEL LAVORO	37	17	5.246913433
183	SALUTE E SICUREZZA DEI	37	19	5.864197731
184	SI PUÒ DIRE CHE	37	33	10.18518543
185	VALUTAZIONE DEL RISCHIO E	37	26	8.024691582
186	DAL DATORE DI LAVORO	36	16	4.938271523
187	DELLA VALUTAZIONE DEL RISCHIO	36	22	6.790123463

188	DI IDONEITÀ ALLA MANSIONE	36	13	4.012345791
189	IDONEITÀ ALLA MANSIONE SPECIFICA	36	11	3.395061731
190	IL GRUPPO DI LAVORO	36	17	5.246913433
191	L'ESISTENZA DI UNA	36	27	8.333333015
192	MOMENTO IN CUI SI	36	31	9.567901611
193	SUL POSTO DI LAVORO	36	24	7.407407284
194	AL FUMO DI SIGARETTA	35	13	4.012345791
195	ANNI CINQUANTA E SESSANTA	35	15	4.629629612
196	È STATO QUELLO DI	35	18	5.55555344
197	ETÀ COMPRESA TRA I	35	22	6.790123463
198	MINISTERO DEGLI AFFARI ESTERI	35	11	3.395061731
199	NEI CASI IN CUI	35	26	8.024691582
200	PER IL FATTO CHE	35	27	8.333333015
201	PER LA FORMAZIONE CONTINUA	35	5	1.54320991
202	RISCHIO STRESS LAVORO CORRELATO	35	7	2.160493851
203	UNA VOLTA PER TUTTE	35	27	8.333333015
204	VITA E DI LAVORO	35	22	6.790123463
205	CARTELLA SANITARIA E DI	34	10	3.086419821
206	DA UN LATO LA	34	26	8.024691582
207	DEL SERVIZIO DI PREVENZIONE	34	13	4.012345791
208	DELLE CONDIZIONI DI LAVORO	34	22	6.790123463
209	E SICUREZZA DEI LAVORATORI	34	18	5.55555344
210	GLI INFORTUNI SUL LAVORO	34	17	5.246913433
211	IN QUESTO MODO SI	34	25	7.716049194
212	MISURE DI PREVENZIONE E	34	21	6.481481552
213	NEI PRIMI DECENNI DEL	34	21	6.481481552
214	NEL CORSO DEL TEMPO	34	19	5.864197731
215	PARTE DEL MEDICO COMPETENTE	34	16	4.938271523
216	PUNTO DI VISTA DEL	34	29	8.950616837
217	SALUTE NEI LUOGHI DI	34	12	3.703703642
218	SANITARIA E DI RISCHIO	34	10	3.086419821
219	SUI LUOGHI DI LAVORO	34	22	6.790123463

220	VALUTAZIONE DEI RISCHI E	34	22	6.790123463
		13536		

**Appendice 6: Le prime 40 posizioni dei *lexical bundles* di 3, 4, 5 parole in sequenza nel *corpus* di storia**

Lexical bundles di 3 parole			Lexical bundles di 4 parole		Lexical bundles di 5 parole	
N	Lexical bundle	Freq.	Lexical bundle	Freq.	Lexical bundle	Freq.
1	NEGLI STATI UNITI	362	DAL PUNTO DI VISTA	203	DA QUESTO PUNTO DI VISTA	60
2	PUNTO DI VISTA	350	SI TRATTA DI UN	105	NELLA MAGGIOR PARTE DEI CASI	56
3	SI TRATTA DI	284	PER LA PRIMA VOLTA	102	DA UN PUNTO DI VISTA	39
4	UNA SERIE DI	274	A PARTIRE DAGLI ANNI	88	DOPO LA SECONDA GUERRA MONDIALE	26
5	DA UN LATO	239	DELLA SECONDA GUERRA MONDIALE	86	A PARTIRE DALLA FINE DEGLI	25
6	IN GRADO DI	238	UNA VERA E PROPRIA	86	PARTIRE DALLA FINE DEGLI ANNI	25
7	IN CUI SI	234	LA SECONDA GUERRA MONDIALE	76	DELLA COMUNITÀ ITALIANA IN GERMANIA	24
8	ALL'INTERNO DELLA	230	NEL CORSO DEGLI ANNI	74	ALL'INTERNO DELL UNIONE EUROPEA	23
9	UNA SORTA DI	230	NEL MOMENTO IN CUI	74	FINE DELLA SECONDA GUERRA MONDIALE	21
10	DA PARTE DI	229	ANCHE IN QUESTO CASO	67	NELLA SECONDA METÀ DEGLI ANNI	21
11	IN QUESTO CASO	229	DEGLI ITALIANI ALL'ESTERO	67	CADUTA DEL MURO DI BERLINO	20
12	LA MAGGIOR PARTE	221	UN VERO E PROPRIO	66	LA CADUTA DEL MURO DI	20
13	LA POSSIBILITÀ DI	220	ALL INIZIO DEGLI ANNI	62	NON È UN CASO CHE	20
14	DAL PUNTO DI	204	ALLA FINE DEGLI ANNI	62	VERSO LA FINE DEGLI ANNI	20
15	PER QUANTO RIGUARDA	201	DEL MERCATO DEL LAVORO	62	A PARTIRE DAGLI ANNI SETTANTA	19
16	ALL'INTERNO DI	197	NELLA MAGGIOR PARTE DEI	62	VOTO DEGLI ITALIANI ALL'ESTERO	19
17	IL FATTO CHE	196	QUESTO PUNTO DI VISTA	62	DAL PUNTO DI VISTA DELLA	18
18	MERCATO DEL LAVORO	194	DA QUESTO PUNTO DI	60	DURANTE LA SECONDA GUERRA MONDIALE	18
19	DI ORIGINE ITALIANA	191	MAGGIOR PARTE DEI CASI	56	NEL MOMENTO IN CUI LA	18
20	A PARTIRE DAL	187	SI TRATTA DI UNA	54	NEL TEMPO E NELLO SPAZIO	18
21	SECONDA GUERRA MONDIALE	180	PER QUANTO RIGUARDA L	52	PER IL RINNOVO DEL PARLAMENTO	18
22	AL DI LÀ	175	PER LA MAGGIOR PARTE	51	UN RUOLO DI PRIMO PIANO	18
23	PIU O MENO	167	UN PUNTO DI VISTA	51	DI UN VERO E PROPRIO	17
24	IN CUI LA	159	COMUNITÀ ITALIANA IN GERMANIA	47	IN UN MOMENTO IN CUI	17

25	IN PRIMO LUOGO	150	A PARTIRE DALLA FINE	46	LA MAGGIOR PARTE DI LORO	17
26	IN TERMINI DI	150	DI FRONTE A UN	45	NEGLI ANNI VENTI E TRENTA	17
27	E IN PARTICOLARE	149	SUL MERCATO DEL LAVORO	45	NEGLI ANNI SESSANTA E SETTANTA	16
28	ALL'INTERNO DEL	148	DI UNA SERIE DI	44	PAESI IN VIA DI SVILUPPO	16
29	DELLA COMUNITÀ ITALIANA	146	DI VOLTA IN VOLTA	44	FARE I CONTI CON LA	15
30	IN QUESTO SENSO	144	ALL'INTERNO DI UN	43	TRA GLI ANNI CINQUANTA E	15
31	NEL CORSO DEL	142	GIOVANI DI ORIGINE ITALIANA	43	ALLA FINE DEGLI ANNI SESSANTA	14
32	FINE DEGLI ANNI	139	LA MAGGIOR PARTE DEGLI	41	DAL PUNTO DI VISTA DELL	14
33	IN CUI IL	135	PER QUANTO RIGUARDA LA	40	DI UNA VERA E PROPRIA	14
34	A PARTIRE DALLA	133	DA UN PUNTO DI	39	LA STORIA DELL EMIGRAZIONE ITALIANA	14
35	LA NECESSITÀ DI	132	LA MAGGIOR PARTE DELLE	39	NEGLI ANNI CINQUANTA E SESSANTA	14
36	LA PRESENZA DI	132	ALL'INTERNO DELLA COMUNITÀ	38	NEI PRIMI DECENNI DEL NOVECENTO	14
37	ITALIANI ALL'ESTERO	131	COME SI È VISTO	37	QUESTO PUNTO DI VISTA LA	14
38	DA PARTE DEI	128	LA MAGGIOR PARTE DI	37	A FARE I CONTI CON	13
39	E COSÌ VIA	128	NEGLI STATI UNITI E	37	A PARTIRE DAGLI ANNI NOVANTA	13
40	DEGLI IMMIGRATI ITALIANI	126	AL DI LÀ DELLE	36	ANCHE DAL PUNTO DI VISTA	13

**Appendice 7: Le prime 40 posizioni dei *lexical bundles* di 3, 4, 5 parole in sequenza nel *corpus* di filosofia**

	Bundles di 3 parole		Bundles di 4 parole		Bundles di 5 parole	
	Lexical bundle	Freq.	Lexical bundle	Freq.	Lexical bundle	Freq.
1	PUNTO DI VISTA	309	DAL PUNTO DI VISTA	106	DA UN PUNTO DI VISTA	64
2	IN GRADO DI	210	UN PUNTO DI VISTA	84	DA QUESTO PUNTO DI VISTA	25
3	IN CUI SI	171	NEL MOMENTO IN CUI	79	DAL PUNTO DI VISTA DELLA	24
4	SI TRATTA DI	170	NELLA MISURA IN CUI	67	HA A CHE FARE CON	15
5	UNA SORTA DI	166	DA UN PUNTO DI	64	NEL MOMENTO IN CUI SI	14
6	IL FATTO CHE	161	A CHE FARE CON	51	DI UNA VERA E PROPRIA	12
7	IL CONCETTO DI	157	UNA VERA E PROPRIA	47	DAL PUNTO DI VISTA DI	11
8	IN QUESTO SENSO	138	SI TRATTA DI UN	46	DAL PUNTO DI VISTA DEL	9
9	VALE A DIRE	120	IL MONDO DELLE COSE	40	DAL PUNTO DI VISTA DELL	9
10	A SUA VOLTA	114	DI VOLTA IN VOLTA	39	NELLA CRITICA DELLA RAGION PURA	9
11	DA UN LATO	111	ALL'INTERNO DI UNA	37	A CHE FARE CON LA	8
12	IN CUI LA	111	SI TRATTA DI UNA	35	DELLA CRITICA DELLA RAGION PURA	8
13	ALL'INTERNO DI	108	ALL'INTERNO DI UN	33	L'UNO E L'ALTRO	8
14	DAL PUNTO DI	107	QUESTO PUNTO DI VISTA	30	NEL MOMENTO IN CUI LA	8
15	UN PUNTO DI	104	PER LA PRIMA VOLTA	29	NON SI TRATTA DI UNA	8
16	CHE NON SI	103	PUNTO DI VISTA DELLA	29	SOLO NELLA MISURA IN CUI	8
17	È LA SUA	103	ANCHE IN QUESTO CASO	28	A CHE FARE CON IL	7
18	MOMENTO IN CUI	101	È IN GRADO DI	28	A CHE FARE CON UNA	7
19	NEL SENSO CHE	101	IN UN CERTO SENSO	28	DAL NOSTRO PUNTO DI VISTA	7
20	IN QUESTO CASO	95	NON SI TRATTA DI	26	DEI BISOGNI E DEI DESIDERI	7



21	CIÒ CHE È	93	DA QUESTO PUNTO DI	25	IL PRINCIPIO DI NON CONTRADDIZIONE	7
22	MONDO DELLE COSE	93	IL PUNTO DI PARTENZA	25	IL PROBLEMA DEL RAPPORTO TRA	7
23	IN QUESTO MODO	91	CRITICA DELLA RAGION PURA	24	IL PUNTO DI VISTA DEL	7
24	CHE NON È	90	IL MODO IN CUI	24	IN SÉ E PER SÉ	7
25	IN CUI IL	90	AL DI FUORI DEL	21	NON HA NULLA A CHE	7
26	LA POSSIBILITÀ DI	89	AL DI LÀ DELLA	21	SIA DA UN PUNTO DI	7
27	NON C'È	88	FARE A MENO DI	21		
28	A PARTIRE DA	86	HA A CHE FARE	21		
29	AL DI LÀ	85	IL FATTO CHE LA	21		
30	IL PROBLEMA DELLA	82	IN QUESTO SENSO LA	21		
31	NEL MOMENTO IN	80	CHE A SUA VOLTA	20		
32	NON PUÒ ESSERE	79	IL PUNTO DI VISTA	20		
33	A QUESTO PUNTO	78	IN LINEA DI PRINCIPIO	20		
34	IN QUALCHE MODO	78	SE DA UN LATO	20		
35	ALL'INTERNO DELLA	77	UN VERO E PROPRIO	20		
36	IN CUI L'	77	UNA VOLTA PER TUTTE	20		
37	IN TAL MODO	75	AD UN CERTO PUNTO	19		
38	L'IDEA DI	71	DI FRONTE AD UN	19		
39	A CHE FARE	70	IN MODO TALE CHE	19		
40	È IL SUO	70	NON È ALTRO CHE	19		

**Appendice 8: Le prime 40 posizioni dei *lexical bundles* di 3, 4, 5 parole in sequenza nel *corpus* di medicina del lavoro**

	Lexical bundles di 3 parole		Lexical bundles di 4 parole		Lexical bundles di 5 parole	
N	Lexical bundle	Freq.	Lexical bundle	Freq.	Lexical bundle	Freq.
1	MEDICINA DEL LAVORO	463	NEI LUOGHI DI LAVORO	215	SALUTE E SICUREZZA SUL LAVORO	71
2	IN GRADO DI	460	IL DATORE DI LAVORO	184	DELLA SALUTE E DELLA SICUREZZA	70
3	DATORE DI LAVORO	437	DI MEDICINA DEL LAVORO	163	SICUREZZA NEI LUOGHI DI LAVORO	60
4	PER QUANTO RIGUARDA	417	DAL PUNTO DI VISTA	152	PREVENZIONE NEI LUOGHI DI LAVORO	54
5	AL FINE DI	382	IL MEDICO DEL LAVORO	128	IN MATERIA DI SALUTE E	49
6	FATTORI DI RISCHIO	368	DEL DATORE DI LAVORO	112	MATERIA DI SALUTE E SICUREZZA	49
7	VALUTAZIONE DEL RISCHIO	366	LA VALUTAZIONE DEL RISCHIO	109	DA UN PUNTO DI VISTA	48
8	LA PRESENZA DI	302	PER QUANTO RIGUARDA LA	107	SERVIZIO DI PREVENZIONE E PROTEZIONE	47
9	LUOGHI DI LAVORO	287	DELLA MEDICINA DEL LAVORO	94	TUTELA DELLA SALUTE E DELLA	47
10	DEL MEDICO COMPETENTE	283	DI SALUTE E SICUREZZA	94	LA SICUREZZA E LA SALUTE	46
11	DI LAVORO È	275	IN MEDICINA DEL LAVORO	94	PER LA SALUTE E SICUREZZA	38
12	MEDICO DEL LAVORO	275	DEL MEDICO DEL LAVORO	92	DI MEDICINA DEL LAVORO E	36
13	VALUTAZIONE DEI RISCHI	273	DI PREVENZIONE E PROTEZIONE	92	LA SALUTE E LA SICUREZZA	36
14	DEL LAVORO E	258	MEDICINA DEL LAVORO E	91	SALUTE E SICUREZZA DEI LAVORATORI	35
15	IL MEDICO COMPETENTE	258	DI VALUTAZIONE DEL RISCHIO	89	CARTELLA SANITARIA E DI RISCHIO	34
16	SALUTE E SICUREZZA	246	DI VALUTAZIONE DEI RISCHI	84	DA PARTE DEL MEDICO COMPETENTE	34
17	PER LA SALUTE	243	DELLA SALUTE E DELLA	82	PER LA VALUTAZIONE DEL RISCHIO	34
18	PUNTO DI VISTA	229	PER LA SALUTE E	81	SALUTE NEI LUOGHI DI LAVORO	34
19	IN TERMINI DI	216	TUTELA DELLA SALUTE E	81	CON IL DATORE DI LAVORO	33
20	NEI LUOGHI DI	216	SALUTE E SICUREZZA SUL	79	DI SALUTE E SICUREZZA SUL	33
21	STRESS LAVORO CORRELATO	212	E SICUREZZA SUL LAVORO	77	PER LA SICUREZZA E LA	33

22	LA NECESSITÀ DI	211	RISCHI PER LA SALUTE	77	E SICUREZZA NEI LUOGHI DI	31
23	IN CASO DI	210	NEGLI AMBIENTI DI LAVORO	76	GIUDIZIO DI IDONEITÀ ALLA MANSIONE	30
24	DELLA SALUTE E	207	DATORE DI LAVORO E	74	PER LA SALUTE E LA	30
25	LA POSSIBILITÀ DI	199	PER QUANTO RIGUARDA IL	73	DEL SERVIZIO DI PREVENZIONE E	29
26	DA PARTE DEL	186	DI PROMOZIONE DELLA SALUTE	70	DI TUTELA DELLA SALUTE E	29
27	IL DATORE DI	184	SALUTE E DELLA SICUREZZA	70	IL DATORE DI LAVORO E	29
28	I LIVELLI DI	179	MOVIMENTAZIONE MANUALE DEI CARICHI	67	ITALIANA DI MEDICINA DEL LAVORO	29
29	IN MATERIA DI	176	È IN GRADO DI	65	SOCIETÀ ITALIANA DI MEDICINA DEL	29
30	TUTELA DELLA SALUTE	170	DELLA SALUTE E SICUREZZA	63	DI IDONEITÀ ALLA MANSIONE SPECIFICA	28
31	LA SORVEGLIANZA SANITARIA	169	I FATTORI DI RISCHIO	63	PER LA TUTELA DELLA SALUTE	28
32	DI MEDICINA DEL	168	LA SALUTE DEI LAVORATORI	63	TUTTI I RISCHI PER LA	28
33	DI PREVENZIONE E	168	LA VALUTAZIONE DEI RISCHI	63	DOCUMENTO DI VALUTAZIONE DEI RISCHI	27
34	AMBIENTI DI LAVORO	167	LA MEDICINA DEL LAVORO	62	TUTELA DELLA SALUTE DEI LAVORATORI	27
35	PER LA VALUTAZIONE	165	AL DATORE DI LAVORO	61	CHE IL DATORE DI LAVORO	26
36	PROMOZIONE DELLA SALUTE	163	DEL GIUDIZIO DI IDONEITÀ	60	E LA SALUTE DEI LAVORATORI	26
37	GIUDIZIO DI IDONEITÀ	159	SICUREZZA NEI LUOGHI DI	60	MISURE DI PREVENZIONE E PROTEZIONE	26
38	IL RISCHIO DI	158	LA MAGGIOR PARTE DEI	56	SICUREZZA E LA SALUTE DEI	26
39	LA MAGGIOR PARTE	158	DEI FATTORI DI RISCHIO	55	TUTELA DELLA SALUTE E SICUREZZA	26
40	IN QUESTO CASO	157	DELLO STRESS LAVORO CORRELATO	55	DEL RISCHIO STRESS LAVORO CORRELATO	25

## Bibliografia

ALTENBERG B., 1998, "On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations", in COWIE A.P. (a cura di) *Phraseology. Theory, Analysis and Applications*, Oxford, Clarendon Press.

ALTENBERG B., EEG-OLOFSSON M., 1990, "Phraseology in Spoken English: Presentation of a Project", in AARTS J. e MEIJS W. (a cura di), *Theory and practice in corpus linguistics*, Amsterdam-Atlanta, Rodopi.

BALBONI P.E., 2000, *Le microlingue scientifico-professionali: natura e insegnamento*, Torino, UTET.

BALLY C., 1951, *Traité de stylistique française*, Ginevra, Georg & cie.

BERRUTO G., 1985, "Per una caratterizzazione del parlato: l'italiano parlato ha un'altragrammatica?", in HOLTUS G., RADTKE E. (a cura di) *Gesprochenes Italienisch in Geschichte und Gegenwart*, Tübingen, Narr.

BERTI E., 1997, "La classicità di un testo filosofico", in *Ars Interpretandi, Annuario di ermeneutica giuridica*.

BIBER D., 1995, *Dimensions of register variation, A cross-linguistic comparison*, Cambridge, Cambridge University Press.

BIBER D., 2006, *University Language: A Corpus-Based Study of Spoken and Written Registers*, Amsterdam and Philadelphia, John Benjamins.

BIBER D., 2009, "A corpus driven approach to formulaic language in English. Multi-word patterns in speech and writing", in *International Journal of Corpus Linguistic*, Amsterdam and Philadelphia, John Benjamins.

BIBER D., BARBIERI F., 2007, "Lexical bundles in university spoken and written Registers", in *English for Specific Purposes*, n. 26.

BIBER D., CONRAD S., 1999, "Lexical bundles in conversation and academic prose", in HASSELGÅRD H., OKSEFJELL S., *Out of Corpora: Studies in Honour of Stig Johansson*, Amsterdam, Rodopi.

BIBER D., CONRAD S., CORTES V., 2003, "Lexical bundles in speech and writing: an initial taxonomy", in WILSON A., RAYSON P., MCENERY T., *Corpus Linguistics by the Lune*, Frankfurt am Main, Peter Lang.

BIBER D., CONRAD S., CORTES V., 2004, "If you look at ... Lexical bundles in academic lectures and textbooks", in *Applied Linguistics*, n.25.

BIBER D. ET AL., 1999, *The Longman Grammar of Spoken and Written English*, Harlow, Longman.

BIBER D., TRACY-VENTURA N., 2007, "Dimensions of register variation in Spanish", in PARODI G. (a cura di), *Working with Spanish Corpora*, London/New York, Continuum.

BORTOLINI U., TAGLIAVINI C., ZAMPOLLI A., 1972, *Lessico di frequenza della lingua italiana contemporanea*, Milano, Garzanti.

BOWKER L., PEARSON J., 2002, *Working with Specialized Language. A practical guide tousing corpora*, London/New York, Routledge.

BUTLER C. S., 1997, "Repeated word combinations in spoken and written text: some implications for functional grammar", in BUTLER C.S. ET AL. (a cura di), *A Fund of Ideas: Recent Developments In Functional Grammar*, Amsterdam, IFOTT University of Amsterdam.

CANTARINI S., 2004, *Costrutti con verbo supporto*, Bologna, Pàtron.

CARDONA M., 2008, "La comprensione e produzione di idioms. Aspetti psicolinguistici e riflessioni glottodidattiche", in *Studi di glottodidattica*, n. 3.

CHIARI I., 2004, *Informatica e lingue naturali. Teorie e applicazioni computazionali per la ricerca sulle lingue*, Roma, Aracne.

CHIARI I., 2007, *Introduzione alla linguistica computazionale*, Roma, Laterza.

CONRAD S., BIBER D., 2004, "The frequency and Use of Lexical Bundles in Conversation and Academic Prose", in *Lexicographica*, n. 20.

CORTES V., 2002, "Lexical bundles in freshman composition", in REPPEN R., FITZMAURICES., BIBER D., *Using Corpora to Explore Linguistic Variation*, Amsterdam/Philadelphia, John Benjamins.

CORTES V., 2004, "Lexical bundles in published and student disciplinary writing: examples from history and biology", in *English for Specific Purposes*, n. 23.

CORTES V., 2008, "A comparative analysis of lexical bundles in academic history writing in English and Spanish", in *English for Specific Purposes*, n. 25.

COWIE A.P., 1991, "Multiwords units in newspaper language", in *Cahiers de l'Institut de Linguistique de Louvain*, n. 17(1-3).

COWIE A.P., 1999, *English dictionaries for foreign learners: a history*, Oxford, Oxford University Press.

CRESTI E., MONEGLIA M., 2005, *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphia, John Benjamins.

- CULPEPER J., and KYTÖ M., 2002, "Lexical bundles in Early Modern English dialogues", in FANEGO T., MENDEZ-NAYA N., SEONE E. (a cura di), *Sound, Words, Texts and Change*, Amsterdam/Philadelphia, John Benjamins.
- D'AGOSTINO E., 1998, "Il lessico di frequenza dell'italiano parlato", in *Quaderns d'Italia*, n. 3.
- DELMONTE R., 1997, "Rappresentazioni lessicali e linguistica computazionale" in DE MAURO T., LO CASCIO V. (a cura di), *Lessico e grammatica: Teorie linguistiche e Applicazioni lessicografiche. Atti del Convegno Interannuale della Società di Linguistica italiana (SLI)*, Roma, Bulzoni.
- DE MAURO T. ET AL., 1993, *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.
- DE SAUSSURE F., 1978, *Corso di linguistica generale, Introduzione, traduzione e commento di Tullio De Mauro*, Bari, Editori Laterza.
- ERMAN B., WARREN B., 2000, "The idiom principle and the open choice principle", in *Text*, n. 20(1).
- FARAGO LEONARDI M., 1982, *Proposte linguistiche e glottodidattiche di J.R. Firth*, Bergamo, Minerva Italica.
- FIRTH J. R., 1968, *Selected Papers of J. R. Firth, 1952-59. Edited by F. R. Palmer*, Bloomington, Indiana University Press.
- GIORDANO R., VOGHERA M., 2002, "Verb system and verb usage in spoken and written italian", in *Journées internationales d'Analyse statistique des Données Textuelles*, n.6.
- GOTTI M., 1991, *I linguaggi specialistici. Caratteristiche linguistiche e criteri pragmatici*, Firenze, La Nuova Italia.
- GRANGER S., 1998, "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae", in COWIE A.P. (a cura di) *Phraseology. Theory, Analysis and Applications*, Oxford, Clarendon Press.
- GRANGER S., 2011, "From phraseology to pedagogy: Challenges and prospects", in HERBST T., UHRIG P., SCHLLER S. (a cura di), *Chunks in the Description of Language. A tribute to John Sinclair*, Berlin/New York, Mouton de Gruyter.
- HALLIDAY M.A.K., 1989, *Spoken and written language*, Oxford, Oxford University Press.
- HALLIDAY M.A.K., HASSAN R., 1976, *Cohesion in English*, London, Longman.
- HOEY M., 1991, *Patterns of lexis in text*, Oxford, Oxford University Press.

- HOEY M., 2004, "Lexical priming and the properties of text", in PARTINGTON A., MORLEY J., HAARMAN L. (a cura di), *Corpora and discourse*, Bern/New York, Peter Lang.
- HOEY M., 2005, *Lexical priming: A new theory of words and language*, London, Routledge.
- HOWARTH P.A., 1996, *Phraseology in English Academic Writing, Some implications for language learning and dictionary making*, Tübingen, Max Niemeyer Verlag.
- HOWARTH P., 1998a, "Phraseology and Second Language Proficiency", in *Applied Linguistics*, n. 19/1.
- HOWARTH P., 1998b, "The Phraseology of Learners' Academic Writing", in COWIE A. P. (a cura di) *Phraseology. Theory, Analysis and Applications*, Oxford, Clarendon Press.
- HYLAND K., 2004, "Disciplinary interactions: Metadiscourse in L2 postgraduate writing" in *Journal of Second Language Writing*, n. 13.
- HYLAND K., 2005, "Stance and engagement: a model of interaction in academic discourse", in *Discourse studies*, n. 7(2).
- HYLAND K., 2008a, "As can be seen: Lexical bundles and disciplinary variation", in *English for Specific Purposes*, n. 27.
- HYLAND K., 2008b, "Academic clusters: text patterning in published and postgraduate Writing", in *International Journal of Applied Linguistics*, n. 18.
- JALALI H., GHAYOONI S., 2010, "A comparative qualitative study of lexical bundles in three academic genres of applied linguistics", in *The modern journal of applied linguistics*, n. 2:4.
- JALALI H., RASEKH A. E., RIZI M. T., 2008, "Lexical bundles and disciplinary variation: the case of applied linguistics", in *Iranian Journal of Language Studies (IJLS)*, n. 2(4).
- JESPERSEN O., 1924, *The Philosophy of Grammar*, London, George Allen and Unwin.
- KIM Y., 2009, "Korean lexical bundles in conversation and academic texts", in *Corpora*, n. 4(2).
- KJELLMER G., 1984, "Some thoughts on collocational distinctiveness", in AARTS J., MEIJS W. (a cura di), *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*, Amsterdam, Rodopi.
- KJELLMER G., 1994, *A dictionary of English collocations*, Oxford, Clarendon Press.
- LAUDANNA A., GAZZELLINI S., DE MARTINO M., 2004, "Representation of grammatical properties of Italian verbs in the mental lexicon", *Brain and Language*, n.90.

LAUDANNA A., VOGHERA M., GAZZELLINI S., 2002, "Lexical Representations of Written Nouns and Verbs in Italian", in *Brain and Language*, n. 81.

LAUDANNA A., VOGHERA M., 2002, "Nouns and verbs as grammatical categories in the lexicon", in *Rivista di linguistica*, n. 14.1.

LIU D., 2012, "The most frequently-used multi-word constructions in academic written English: A multi-corpus study", in *English for Specific Purposes*, n. 31.

LO CASCIO V., 1991, *Grammatica dell'argomentare*, Firenze, La Nuova Italia Editrice.

LO CASCIO V., 1994, "Ricchezza e povertà dell'italiano parlato in Italia e all'estero", in DE MAURO T. (a cura di), *Come parlano gli italiani*, Firenze, La Nuova Italia Editrice.

LO CASCIO V., 1997, "Semantica lessicale e i criteri di collocazione nei dizionari bilingui a stampa ed elettronici", IN DE MAURO T., LO CASCIO V. (a cura di), *Lessico e grammatica, Teorie linguistiche e applicazioni didattiche*, Roma, Bulzoni.

LEECH G., 1992, "Corpora and theories of linguistic performance", in STARTVIK J. (a cura di), *Directions in corpus linguistics, Proceedings of Nobel Symposium 82 Stockholm*, Berlin, Mouton de Gruyter.

MARELLO C., 1996, *Le parole dell'italiano, lessico e dizionari*, Bologna, Zanichelli.

MASINI F., 2009, "Combinazioni di parole e parole sintagmatiche", in Lombardi Vallauri, Edoardo, Lunella Mereu (a cura di), *Spazi linguistici. Studi in onore di Raffaele Simone*, Roma, Bulzoni.

MCENERY T., WILSON A., 2001, *Corpus linguistics*, Edinburgh, Edinburgh University Press.

MCENERY T., WILSON A., 2007, *Corpus linguistics: an introduction*, Edinburgh, Edinburgh University Press.

MCENERY T., XIAO R., TONO Y., 2006, *Corpus-Based Language Studies, An advanced resource book*, London/New York, Routledge.

NATTINGER J., DECARRICO J., 1992, *Lexical phrases and language teaching*, Oxford, Oxford University Press.

NEKRASOVA T. M., 2009, "English L1 and L2 Speakers' Knowledge of Lexical Bundles", in *Language Learning*, n. 59.

NESI H., BASTURKMEN H., 2006, "Lexical bundles and discourse signaling in academic lectures", in *International Journal of Corpus Linguistics*, n. 11.

NESSSELHAUF N., 2005, *Collocations in a learner corpus*, Philadelphia/Amsterdam, John Benjamins.



- NORRICK N. R., 1987, "Functions of repetition in conversation", in *Text*, n. 7(3).
- PALMER H. E., 1966, *Second interim report on English collocations, submitted to the Tenth Annual Conference of English Teachers, under the auspices of the Institute for Research in English Teaching, Dept. of Education, Tokyo, October 1933*, Tokyo, Kaitakusha.
- PARTINGTON A., MORLEY J., HAARMAN L., 2004, *Corpora and discourse*, Bern/New York, Peter Lang.
- PAWLEY A., SYDER F.H., 1983, "Two puzzles for linguistic theory: nativelike selection and nativelike fluency", in RICHARDS J.C., SCHMIDT R.W. (a cura di), *Language and Communication*, New York, Longman.
- PRAT ZAGREBELSKY M. T., 2002, "L'uso dei *corpora* nell'analisi contrastiva di saggi argomentativi di studenti universitari italiani e anglosassoni: I think versus I feel", in BECCARIA G. L., MARELLO C., *La parola al testo. Scritti per Bice Mortara Garavelli*, Alessandria, Dell'Orso.
- RENOUF A., 1992, "What do you think of that? A pilot study of the phraseology of the core words of English.", in LEITNER G. (a cura di), *New Directions in English Language Corpora: Methodology, Results, Software Developments*, Berlin, Mouton de Gruyter.
- RENOUF A., SINCLAIR J., 1991, "Collocational frameworks in English", in AJMER K., ALTENBERG B. (a cura di), *English Corpus Linguistics: Studies in the Honour of Jan Svartvik*, London, Longman.
- RÖMER U., 2009, "The inseparability of lexis and grammar, Corpus linguistic perspectives", in *Annual Review of Cognitive Linguistics*, n. 7.
- RUFFALDI E., TROMBINO M., 2004, *L'officina del pensiero. Insegnare e apprendere filosofia*, Milano, LED.
- SCOTT M., TRIBBLE C., 2006, *Textual Patterns: Key words and Corpus Analysis in Language Education*, Amsterdam/Philadelphia, John Benjamins.
- SERIANNI L., 2005, *Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente*, Milano, Garzanti Libri.
- SIMONE R., 1997, "Esistono verbi sintagmatici in italiano?", in DE MAURO T., LO CASCIO V., (a cura di), *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche*, Roma, Bulzoni.
- SIMONE R., 2006, "Classi di costruzioni", in GRANDI N. E IANNÀCCARO G. (a cura di), *Zhì. Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*, Cesena/Roma, Caissa Italia.

SIMONE R., 2007, "Constructions and categories in verbal and signed languages", in PIZZUTO E., PIETRADREA P., SIMONE R., *Verbal and signed languages: comparing structures, constructs and methodologies*, Berlin, Mouton de Gruyter.

SIMONE R., 2008, "Verbi sintagmatici come costruzione e come categoria", in CINI M. (a cura di), *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca*, Frankfurt am Main, Peter Lang.

SIMPSON-VLACH R., ELLIS N. C., 2010, "An Academic Formulas List: New Methods in Phraseology Research", in *Applied Linguistics*, n. 31/4.

SINCLAIR J., CARTER R., 2004, *Trust the text: Language, corpus and discourse*, London/New York, Routledge.

SINCLAIR J. ET AL., 2004, *English collocation studies, the OSTI report*, London/New York, Continuum.

SINCLAIR J. M., 1982, "Reflections on Computer Corpora in English Language Research", in JOHANNSON S. (a cura di), *Computer Corpora in English Language Research*, Bergen, Norwegian Computer Centre for the Humanities.

SINCLAIR J. M., 1991, *Corpus, Concordance, Collocation*, Oxford/New York, Oxford University Press.

SINCLAIR J. M., 2005, "Corpus and Text – Basic principles", in WYNNE M. (a cura di), *Developing Linguistic Corpora Guide to Good Practice*, Eynsham/Oxford, Information Press.

SINCLAIR J. M., JONES S., 1996, "English lexical collocations", in SINCLAIR J. M., FOLEY J. A., *J. M. Sinclair on lexis & lexicography*, Singapore, UniPress.

SOBRERO A., BENINCÀ P., BERRUTO G., 1993, *Introduzione all'italiano contemporaneo, La variazione e gli usi*, Roma, Laterza. Include:

BERRUTO G., "La varietà del repertorio", "Le varietà del repertorio", "Varietà diamesiche, diastratiche, diafasiche".

SOBRERO A., "Lingue speciali".

SPINA S., 2001, *Fare i conti con le parole. Introduzione alla linguistica dei corpora*, Perugia, Guerra Edizioni.

SPINA S., 2010, "The Dictionary of Italian Collocations: Design and Integration in an Online Learning Environment", in CALZOLARI N. ET AL., *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Malta, European Language Resources Association.

TOGNINI-BONELLI E., 2000, "Il corpus in classe: da una nuova concezione della lingua a una nuova concezione della didattica", in FAVRETTI R. R. (a cura di), *Linguistica e informatica. Corpora, multimedialità e percorsi di apprendimento*, Roma, Bulzoni.

- TOGNINI-BONELLI E., 2001, *Corpus linguistics at work*, Amsterdam/Philadelphia, John Benjamins.
- TRACY-VENTURA N., CORTES V., BIBER D., 2007, "Lexical bundles in speech and writing", in *Working with Spanish corpora*, PARODI G. (a cura di), London/New York, Continuum.
- TREMBLAY A., DERWING B., LIBBEN G., 2009, "Are Lexical Bundles Stored and Processed as Single Units", in *Working Papers of the Linguistics Circle of the University of Victoria*, n. 19.
- VOGHERA M., 1992, *Sintassi e intonazione nell'italiano parlato*, Bologna, Il Mulino.
- VOGHERA M., 1994, "Lessemi complessi: percorsi di lessicalizzazione a confronto", in *Lingua e stile*, Anno XXIX, n. 2.
- VOGHERA M., 2004, "Polirematiche", in GROSSMANN M., RAINER F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Niemeyer.
- VOGHERA M., 2005, "La misura delle categorie sintattiche", in DE MAURO T., CHIARI I. (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Bologna, Aracne.
- VOGHERA M., LAUDANNA A., 2003, "Proprietà categoriali e rappresentazione lessicale del verbo: una prospettiva interdisciplinare", in GIACOMO-MARCELLESI M. e ROCCHETTI A. (a cura di), *Il verbo italiano, Studi diacronici, sincronici, contrastivi, didattici, Atti del XXXV Congresso internazionale di studi, Parigi, 20-22 settembre 2001*, Roma, Bulzoni.
- WRAY A., 1999, "Formulaic language in learners and native speakers", in *Language Teaching*, n. 32(4).
- WRAY A., PERKINS M. R., 2000, "The functions of formulaic language: an integrated model", in *Language and Communication*, n. 20.
- ZGUSTA L., 1967, "Multiword lexical units", in *Linguistic studies presented to Andre' Martinet on the occasion of his sixtieth birthday*, New York, Linguistic Circle of New York.

## **Estratto per riassunto della tesi di dottorato**

**Studente:** Mojca Del Fabbro  
**Matricola:** 955639  
**Dottorato:** Scienze del linguaggio  
**Ciclo:** 24°

**Titolo della tesi:** *Lexical bundles* come tratto linguistico saliente

### **Abstract:**

Questa tesi si pone come obiettivo principale quello di presentare la pervasività dei *lexical bundles* nel parlato italiano, basato sul *corpus* del *Lessico di frequenza dell'italiano parlato*, e nel discorso accademico.

Partendo dalle prospettive teoriche dello studio sulle combinazioni di parole, la presente tesi offre una panoramica delle strutture grammaticali e delle funzioni discorsive dei *lexical bundles*, allo scopo di mostrare la loro importanza nella formulazione del discorso. Analizzando due diversi registri – la conversazione e la prosa accademica – e concentrandosi propriamente sull'analisi di tre microlingue, relative alle discipline di storia, filosofia e di medicina del lavoro, si intende mostrare come la prospettiva di registro e di disciplina scientifica risultino importanti nello studio dei *pattern* formulaici.

Prendendo atto della pervasività dei *lexical bundles* nelle microlingue scientifico-professionali e della loro variazione attraverso le aree scientifiche, la presente ricerca sostiene che i *lexical bundles* debbano essere integrati tra le caratteristiche sintattico-lessicali delle microlingue in oggetto.

This thesis places its focus on the pervasiveness of lexical bundles found in spoken Italian, based on the corpus *Lessico di frequenza dell'italiano parlato*, as well as on academic discourse.

Starting from the theoretical perspectives on the studies of word combinations, this thesis offers an overview of grammatical structures and discourse functions of the lexical bundles as a means to demonstrate their importance in formulating discourse. By analyzing the registers of colloquial speech and of academic prose along with an in-depth analysis of three languages for specific purposes in respect to history, philosophy and occupational medicine, this thesis intends to demonstrate how the register and discipline variation result essential in the study of formulaic patterns.

By acknowledging the predominance of the lexical bundles within languages for specific purposes and their variations throughout scientific spheres, this research serves to establish that the lexical bundles should be integrated within the lexical-syntactical characteristics of the languages for specific purposes aforementioned.