

# INDICATORI DI COMPLESSITÀ NEL PARLATO DEGLI INSEGNANTI DI ITALIANO L2: UN'ANALISI QUANTITATIVA

Lorena Salvati<sup>1</sup>, Irene Russo<sup>2</sup>

## 1. INTRODUZIONE

Sul parlato dell'insegnante di L2 (identificato anche come *teacher talk*) esiste una vasta letteratura, in particolare per quanto riguarda l'insegnamento dell'inglese come lingua seconda. Agli approcci di carattere più teorico-normativo si accompagnano, negli ultimi trenta anni, approcci che prendono spunto dalle tecniche di analisi della linguistica dei *corpora* e che si focalizzano su casi di studio specifici, andando a verificare su dati concreti le ipotesi formulate dalla teoria.

L'ipotesi di base di questo ambito di ricerca è che gli insegnanti compiono – in maniera non sempre consapevole e pianificata – degli adeguamenti nel loro modo di parlare di fronte ad una classe di apprendenti. Tali adeguamenti riguardano più livelli linguistici e variano di intensità a seconda del livello complessivo di competenza degli apprendenti.

Nel presente lavoro ci proponiamo di analizzare quantitativamente la complessità del parlato di insegnanti madrelingua di italiano L2 raccolto e trascritto durante lezioni appartenenti a due livelli del ), *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione* (QCER, Council of Europe, 2002 [2001]), A1 e B1<sup>3</sup>. Una parte delle trascrizioni riguarda lezioni svolte in classe (*corpus* ParInIt, *Parlato di Insegnanti di Italiano*), nella quale vi è compresenza fisica tra insegnante e apprendenti, un secondo *corpus* è invece composto da lezioni somministrate *online* in maniera asincrona, tramite un canale YouTube (*corpus* Oneworlditaliano).

Proponiamo una classificazione degli adeguamenti rispetto alla quale l'analisi quantitativa della complessità degli indicatori linguistici verificherà se è possibile distinguere sia tra il livello A1 e il livello B1 nel *corpus* raccolto in presenza e nel *corpus* relativo alle lezioni *on line*.

L'obiettivo finale è comprendere se un'analisi quantitativa dei dati possa aiutare ad individuare gli adeguamenti e le modifiche linguistiche attuate dai docenti per favorire una maggiore comprensibilità dell'*input* da parte degli apprendenti. Allo stesso tempo, si vuole investigare se tali variabili siano sufficientemente distintive da caratterizzare il parlato dell'insegnante di italiano L2 come adeguato rispetto ai due livelli del QCER presi in considerazione. Infine, l'analisi delle trascrizioni di lezioni svolte *online* in modalità asincrona aiuterà a capire se gli adeguamenti si verificano anche quando non vi è compresenza in un medesimo spazio fisico e l'insegnante non riceve un *feedback* immediato dagli studenti.

<sup>1</sup> Centro Provinciale per l'Istruzione degli Adulti (CPIA), Pistoia

<sup>2</sup> Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR), Pisa. Il contributo è frutto del lavoro congiunto delle due autrici che hanno concordato l'impianto generale e la suddivisione in parti. Tuttavia, Lorena Salvati ha curato i paragrafi 1, 3, 3.1, 4.2, 4.4 mentre Irene Russo ha curato i paragrafi 2, 3.2, 4, 4.1, 4.3, 5.

<sup>3</sup> Ulteriori indagini, anche di tipo qualitativo, sono state condotte nell'ambito della ricerca di dottorato svolta presso l'Università per Stranieri di Perugia: si rimanda a L. Salvati, *Il parlato degli insegnanti di italiano L2: un'indagine in classi di livello elementare e intermedio* (tesi di dottorato non pubblicata).

L'articolo è strutturato come segue: nel paragrafo 2 un breve stato dell'arte sul parlato degli insegnanti di inglese come L2 introduce alcuni degli aspetti che verranno esaminati; nel paragrafo 3 sono descritti i due *corpora* oggetto dell'indagine; nel paragrafo 4 si presenta l'analisi. Il paragrafo 4 è articolato in sottosezioni, ciascuna delle quali si focalizza su un livello di analisi. I risultati vengono discussi e sintetizzati nelle conclusioni (paragrafo 5).

## 2. CARATTERISTICHE DEL PARLATO DELL'INSEGNANTE

Il parlato dell'insegnante L2 è stato ampiamente investigato nella letteratura di ambito anglofono con il nome di *teacher talk*, uno specifico registro sociolinguistico differente dal *foreign talk*, sebbene per certi versi ad esso simile. Se il *foreign talk* è il parlato generico di un parlante nativo verso un parlante non nativo, il *teacher talk* è delimitato dal fine educativo dello scambio e dal fatto che avviene in un contesto di gruppo-classe. L'insegnante opera degli aggiustamenti di carattere temporaneo nella sua comunicazione con gli apprendenti non madrelingua che si realizzano con minore frequenza – fino a scomparire – con l'aumentare della competenza dell'apprendente.

Le ricerche condotte analizzando dati empirici indicano che il parlato dell'insegnante diretto ad apprendenti di lingua seconda è soggetto a modifiche sul piano fonologico, lessicale, sintattico e discorsivo volte a migliorare la comprensibilità dell'*input*, andando a costituire uno dei molteplici fattori che contribuiscono al processo di acquisizione. Non è tuttavia possibile verificare quanto la comprensibilità dell'*input* influenzi tale processo: per la sua complessità risulta difficile isolare la singola variabile “effetto del parlato dell'insegnante”. La semplificazione stessa dell'*input*, avvenendo su più livelli, non può essere usata come variabile unica da correlare con l'apprendimento di una lingua. Non si può infatti determinare, ad esempio, se gli adeguamenti sul piano fonologico siano più efficaci di quelli sul piano lessicale, dato il carattere olistico dell'interazione insegnante-studente.

Esistono molteplici caratteristiche linguistiche e discorsive che caratterizzano il parlato degli insegnanti di L2, riportate dalla letteratura. Un nutrito numero di studi su tali tematiche non è però in genere accompagnato da misure di significatività statistica e quindi alcuni dei risultati andrebbero ulteriormente verificati con l'analisi di nuovi dati. Scopo del presente stato dell'arte è riportare le conclusioni più rilevanti per i vari livelli di analisi linguistica, mettendo in luce quando misure di carattere quantitativo hanno permesso di verificare le ipotesi di partenza.

Da un punto di vista discorsivo, si suppone che la frequenza d'uso di ripetizioni e parafrasi aumenti la comprensibilità dell'*input*, dando agli apprendenti maggiori opportunità di processare l'informazione. Ad esempio, in uno studio condotto su insegnanti inglesi e francesi, Wesche e Ready (1985) hanno riscontrato un maggior uso di ripetizioni quando questi ultimi si rivolgevano ad apprendenti L2 rispetto a quando insegnavano a studenti L1. Tale caratteristica tende a diminuire con il passare del tempo: Ellis (1985) ha riscontrato che in un periodo di nove mesi lo stesso insegnante tende ad usare meno ripetizioni, percependo una maggiore competenza da parte degli studenti.

Da un punto di vista strettamente linguistico un aspetto importante delle modifiche apportate dall'insegnante riguarda gli aggiustamenti sintattici, che hanno però un limite nella problematicità di individuazione dell'unità minima di analisi quando si tratta del parlato trascritto in quanto la segmentazione in frasi del parlato è resa difficile dalla frammentarietà della comunicazione orale e non è eseguibile automaticamente. In termini generali, viene comunque riscontrato che il parlato diretto ad apprendenti di L2 contiene unità con un numero di parole minore, siano queste unità frasi, enunciati o unità di altro tipo. Osservando un insegnante in una stessa classe è stato riscontrato che la lunghezza

media delle frasi aumenta con il tempo (Hakansson, 1986), ad indicare che l'insegnante esercita un controllo su questa variabile.

Le modifiche sintattiche che faciliterebbero la comprensione dell'apprendente riguardano la lunghezza dell'enunciato, la distribuzione dei tipi di frase (frasi dichiarative, interrogative e imperative), l'uso di strutture sintattiche meno marcate (tempo presente), la grammaticalità degli enunciati e l'uso di frasi subordinate (Chaudron, 1988). Ad esempio, è stato riscontrato un uso maggiore dei verbi al tempo presente (Mizon, 1981) nel parlato degli insegnanti di L2, mentre Henzl (1979) riscontra minore complessità inflessionale, mancanza di costruzioni passive e dell'uso del condizionale. L'analisi quantitativa del livello sintattico è ovviamente problematica anche per le trascrizioni in italiano, ma la complessità morfologica può essere più agevolmente investigata attraverso un indicatore che utilizziamo nel paragrafo 4.1.

È inoltre possibile che gli insegnanti usino frasi/enunciati non grammaticali, con omissione di parole funzionali, copula, pronomi e articoli per adattare l'*input* a un basso livello di competenza (Kliefgen, 1985), ma non esistono studi sistematici al riguardo.

In termini di caratteristiche lessicali, esistono un buon numero di studi che riportano statistiche descrittive sulle differenze del parlato dell'insegnante L2 rivolto a parlanti nativi o non nativi ma purtroppo – anche in questo caso – non analizzando i risultati in termini di significatività statistica è difficile trarre delle conclusioni generali. Si ritiene che il parlato degli insegnanti L2 contenga elementi lessicali più frequenti o più basilari e per contro poche espressioni idiomatiche, pronomi indefiniti o lemmi troppo colloquiali (Chaudron, 1982; Henzl, 1979; Kliefgen, 1985; Mizon, 1981). Per riscontrare questo tipo di evidenze viene calcolata la *type-token ratio*<sup>4</sup> o la distribuzione di classi di parole, distinguendo tra parole contenuto e parole funzione. Tali analisi sono state effettuate sui dati trascritti di insegnanti di italiano L2, nei paragrafi 4.2 e 4.3.

### 3. CORPORA DI LEZIONI TRASCRITTE

La linguistica dei *corpora* richiede *corpora* rappresentativi rispetto al fenomeno che si andrà ad investigare, caratterizzando il concetto di rappresentatività in termini di dimensioni sufficienti e di bilanciamento delle classi analizzate. È purtroppo difficile avere a disposizione un *corpus* di dimensioni consistenti quando si analizza il parlato trascritto, perché le metodologie di trascrizione manuale sono particolarmente onerose. Tuttavia, le dimensioni dei due *corpora* di seguito presentati sono comparabili con quelle di *corpora* analoghi di *teacher talk* per la lingua inglese.

#### 3.1. *Corpus ParInIt*

Il *corpus* ParInIt (*Parlato di Insegnanti di Italiano*) è un *corpus* di lezioni trascritte<sup>5</sup> creato a partire da registrazioni effettuate in presenza in classi di italiano L2 appartenenti a corsi di due differenti livelli QCER, A1 e B1.

Le registrazioni sono state raccolte tra il 2018 e il 2019 presso classi di apprendenti adulti, mediamente composte da 20 allievi di diverse nazionalità. Al fine di limitare la

<sup>4</sup> Un rapporto *type-token ratio* (TTR) è il numero totale di parole diverse (tipi) diviso per il numero totale di parole (*token*) in un dato segmento di lingua o di testo. Più alto è il valore percentuale maggiore è la varietà del vocabolario usato in un testo.

<sup>5</sup> Salvati (2021), *Il parlato degli insegnanti di italiano L2: un'indagine in classi di livello elementare e intermedio* (tesi di dottorato non pubblicata).

variabilità nel parlato degli insegnanti eventualmente legata al livello di esperienza degli stessi o a caratteristiche intrinseche degli apprendenti, si è scelto di considerare soltanto classi di apprendenti adulti, con un'età media di 20 anni. Gli insegnanti registrati, tutti di madrelingua italiana, hanno invece soddisfatto il requisito del possesso di formazione specifica per la didattica dell'italiano a stranieri, unito ad almeno cinque anni di esperienza.

In totale sono state raccolte 20 ore di registrazione; le lezioni sono state successivamente trascritte adottando come metodologia la trascrizione ortografica.

Alla trascrizione è seguita l'anonimizzazione dei dati, che ha comportato la cancellazione di eventuali informazioni personali, quando presenti.

Il *corpus* finale è composto da 27.500 *tokens*, suddivisi in maniera non bilanciata nei due livelli A1 e B1 (Tabella 1).

Tabella 1. *Composizione del corpus ParInIt*

	<b>ore di registrazione</b>	<b>numero lezioni</b>	<b>tokens trascrizioni</b>
A1	12	14	16000
B1	7	8	11500
totale	19	22	27500

### 3.2. *Corpus Oneworlditaliano*

Si presume che molti degli adeguamenti operati dall'insegnante di L2 dipendano dalla compresenza, nel medesimo contesto comunicativo di una lezione, di insegnante e apprendenti. Tuttavia possiamo ipotizzare che entro certi limiti tali adeguamenti siano presenti anche in lezioni fruite dall'apprendente in maniera asincrona, senza possibilità di uno scambio comunicativo, come avviene nell'insegnamento delle lingue veicolato attraverso piattaforme digitali. Nello specifico, concentreremo la nostra analisi su un caso di studio relativo ad un canale YouTube per l'apprendimento dell'italiano con un buon seguito, Oneworlditaliano<sup>6</sup>. Una parte delle lezioni sono suddivise in livelli QCER. Le trascrizioni delle 12 lezioni di livello A1 e delle 12 lezioni di livello B1 vanno a costituire un *corpus* bilanciato di produzioni linguistiche comparabile con quanto raccolto e trascritto durante le lezioni in presenza (Tabella 2).

Tabella 2. *Composizione del corpus Oneworlditaliano*

	<b>Ore di videoregistrazione</b>	<b>numero lezioni</b>	<b>tokens trascrizioni</b>
A1	2 h 15 min	12	9940
B1	2 h 45 min	12	19678
totale	5 h	24	29618

<sup>6</sup> <https://www.youtube.com/user/Oneworlditaliano>.

#### 4. ANALISI DEGLI INDICATORI DI COMPLESSITÀ

Esiste una vasta letteratura sull'analisi della complessità della produzione linguistica, specialmente in forma testuale (Amancio *et al.*, 2012; Ristad, 1993; Yasseri *et al.*, 2012), che spesso rielabora sotto forma di indicatori statistici alcune delle ipotesi di lavoro avanzate per la complessità di una lingua in generale. Considerare la complessità strutturale di una lingua significa analizzare una sua realizzazione (sotto forma di testo o di trascrizione) a prescindere dalle caratteristiche soggettive di chi lo decodifica, ad esempio tenendo conto del rapporto tra parole contenuto e parole funzione o della frequenza di vocaboli più o meno specialistici.

Nel presente lavoro adottiamo una concezione di complessità meramente strutturale, intesa come complessità che deriva dai singoli elementi linguistici e dalla relazione tra di essi, esulando da questioni di costo cognitivo nel processamento e dalle dinamiche acquisizionali. In maniera complementare rispetto a quanto proposto da Pallotti (2015), ci focalizziamo su una serie di misure che vanno ad analizzare la complessità strutturale del parlato dell'insegnante di L2 (invece che, come nel suo caso, dell'interlingua).

Per misurare la complessità nel parlato del docente, si prendono in esame la morfologia (nominale e verbale), la diversità lessicale – ovvero la ricchezza del vocabolario impiegato, la densità lessicale (derivata dal rapporto tra elementi lessicali ed elementi grammaticali) e infine l'impiego delle parole appartenenti al Nuovo Vocabolario di Base (De Mauro, 2016).

##### 4.1. *Complessità morfologica*

Brezina e Pallotti (2016) propongono, come indicatore di complessità morfologica, il *Morphological Complexity Index* (MCI), una misura per calcolare la diversità flessionale in campioni standardizzati di una classe di parole in un testo, attualmente implementato per verbi e sostantivi. La morfologia flessionale è intesa come una serie di operazioni applicate sulle basi lessicali (ad esempio, aggiungere un affisso, lasciare la base invariata o cambiarla). La base lessicale è identificata operativamente come quella comune a più celle nel paradigma verbale. I processi flessionali sono tutti i cambiamenti possibili a partire da questa base.

Nel loro lavoro Brezina e Pallotti (2016) testano l'indicatore MCI su testi brevi scritti da parlanti nativi e non nativi in due lingue differenti da un punto di vista morfologico (italiano e inglese). La complessità morfologica dei testi prodotti da apprendenti di italiano di livello basso è minore rispetto a testi di livello più avanzato ed è inoltre positivamente correlata con altre misure di competenza, come la diversità lessicale e la lunghezza delle frasi. Per quanto riguarda l'inglese, invece, la complessità morfologica rimane costante indipendente dal livello dei parlanti, in quanto si tratta di una lingua con un sistema inflessionale relativamente semplice. L'italiano, al contrario, ha tre classi di coniugazione e per tutti i verbi è necessario specificare tempo, aspetto e modo; è quindi prevedibile che i parlanti dei livelli più bassi non padroneggino a pieno tutto il repertorio di processi morfologici disponibili.

Si può ipotizzare che anche nell'*input* fornito dall'insegnante agli apprendenti vi sia in maniera più o meno consapevole un controllo sulla complessità morfologica delle parole scelte. Per il calcolo di tale indicatore, applicato per la prima volta al parlato degli insegnanti, abbiamo utilizzato l'output di un'interfaccia web<sup>7</sup> che implementa l'indicatore MCI.

<sup>7</sup> [http://corpora.lancs.ac.uk/vocab/analyse\\_morph.php](http://corpora.lancs.ac.uk/vocab/analyse_morph.php).

Per quanto riguarda la complessità morfologica delle trascrizioni del livello B1 nel *corpus* ParInIt, essa risulta maggiore di quella del livello A1. La differenza tra i due livelli del *corpus* ParInIt (Tabella 3) è significativa per i verbi (Welch t-test:  $t = -3.50$ , DoF = 18.43,  $p = 0.0025$ ) ma non per i nomi (Welch t-test:  $t = -0.79$ , DoF = 13.45,  $p = 0.45$ ).

Lo stesso si riscontra per il *corpus* Oneworlditaliano (Tabella 4), differenza significativa per i verbi (Welch t-test:  $t = -5.91$ , DoF = 16.07,  $p = 0.000022$ ) ma non per i nomi (Welch t-test:  $t = -0.56$ , DoF = 20.55,  $p = 0.58$ ).

Tabella 3. *Complessità morfologica dei nomi. Test di significatività della differenza tra valori medi per i due livelli*

	<b>media A1 (deviazione standard)</b>	<b>media B1 (deviazione standard)</b>
<i>Corpus</i> ParInIt	2.59 (0.77)	2.88 (0.85)
<i>Corpus</i> Oneworlditaliano	3.82 (1.02)	4.1 (1.34)

Tabella 4. *Complessità morfologica dei verbi. Test di significatività della differenza tra valori medi per i due livelli*

	<b>media A1 (deviazione standard)</b>	<b>media B1 (deviazione standard)</b>
<i>Corpus</i> ParInIt	11.98 (0.83)	13.06 (0.62)
<i>Corpus</i> Oneworlditaliano	9.65 (2.07)	13.57 (1.02)

#### 4.2. *Diversità lessicale*

La conoscenza del vocabolario è un aspetto fondamentale nella comprensione di una lingua seconda, sia letta che ascoltata. Pertanto, lezioni che si indirizzano ad apprendenti di livelli differenti dovrebbero presentare delle caratteristiche, sul piano lessicale, che permettono – sulla base della complessità del vocabolario usato dall'insegnante – di classificarle come appartenenti ad un determinato livello.

Una variabile spesso presa in considerazione è la diversità lessicale ovvero la varietà del vocabolario, caratterizzata sulla base del numero di *types* usati in un testo o in una trascrizione di una comunicazione orale.

Più nello specifico, la misura frequentemente usata per la diversità lessicale è la *type-token ratio* (TTR) che indica il rapporto tra il numero di *tokens* unici (detti *types*) e il numero totale di *tokens* in un testo, ovvero tra la grandezza del vocabolario e la lunghezza del testo.

Il rapporto tra *types* e *tokens* decresce in maniera sistematica con l'aumentare della lunghezza del testo perché il parlante, o lo scrivente, tende a ripetersi. Ciò rende difficile la comparazione di testi differenti.

Essendo tale misura influenzata dalla lunghezza del testo che si va ad analizzare, si preferisce usare la *root* TTR (anche detta indice di Guiraud) e la TTR corretta che hanno al denominatore rispettivamente il logaritmo e la radice quadrata della lunghezza del testo e producono una misura meno condizionata dall'estensione del testo.

La TTR applicata con fini comparativi a classi di testi di lunghezza diversa non aiuta a capire quale delle due possa essere caratterizzata, in termini di valore medio della TTR,

come meno variabile a livello lessicale. Come si evince infatti dalla tabella 5 (prima colonna), la TTR media per le trascrizioni di livello A1 e B1 non aiuta a distinguerli. Calcolando invece la TTR sia con l'indice di Guiraud sia considerando solo le prime  $n$  parole delle trascrizioni (dove  $n$  sta per il numero di parole della trascrizione più breve, nel nostro caso rispettivamente 134 e 426 per i due *corpora*), la differenza tra i due livelli emerge con maggiore chiarezza (Tabella 5, Tabella 6)<sup>8</sup>.

Per quanto riguarda la differenza tra livelli, i valori medi per l'indice di Guiraud sono differenti in maniera significativa sia per il *corpus* ParInIt (Welch t-test:  $t = -5.05$ , DoF = 18.82,  $p = 0.000075$ ) che per il *corpus* Oneworlditalia (Welch t-test:  $t = -4.46$ , DoF = 16.65,  $p = 0.0003$ ).

Tabella 5. *Valori medi per diverse tipologie di TTR relativamente al corpus ParInIt. Tra parentesi il valore della deviazione standard*

	<b>TTR</b>	<b>root TTR (Guiraud)</b>	<b>TTR corretta</b>	<b>TTR con soglia 134</b>
A1	0.3012 (0.1104)	8.6760 (1.1880)	6.1348 (0.84)	0.5381 (0.0784)
B1	0.3075 (0.0667)	10.8732 (0.8449)	7.6885 (0.5974)	0.5832 (0.07)

Tabella 6. *Valori medi per diverse tipologie di TTR relativamente al corpus Oneworlditalia. Tra parentesi il valore della deviazione standard*

	<b>TTR</b>	<b>root TTR (Guiraud)</b>	<b>TTR corretta</b>	<b>TTR con soglia 426</b>
A1	0.3197 (0.061)	8.8570 (1.5766)	6.2628 (1.1148)	0.3926 (0.641)
B1	0.2838 (0.0417)	11.1484 (0.8289)	7.8831 (0.5861)	0.4631 (0.0278)

### 4.3. *Densità lessicale*

Se la diversità lessicale è una misura di quante parole differenti sono usate in un testo, la densità lessicale è una misura della proporzione degli elementi appartenenti a determinate classi lessicali che sintetizza la complessità della comunicazione, applicabile sia alla produzione scritta che parlata. È implementabile in vari modi, ma il più diffuso è senza dubbio il rapporto tra la frequenza degli elementi lessicali (sostantivi, verbi, aggettivi e avverbi) e la frequenza degli elementi grammaticali (congiunzioni, preposizioni, pronomi, determinativi e alcune classi di verbi e avverbi).

La densità lessicale influenza la comprensibilità di un testo ed è maggiore nello scritto in quanto si tratta di una modalità di comunicazione pianificata e soggetta a revisioni. Al contrario, nel parlato vi sono maggiori ripetizioni e si tende ad integrare o addirittura anticipare il *feedback* dell'interlocutore, producendo così una comunicazione lessicalmente meno densa.

<sup>8</sup> Un risultato parziale relativo all'analisi della TTR è illustrato in Coppola, Moretti, Salvati (2021).

Tabella 7. *Densità lessicale media nei due corpora. Tra parentesi il valore della deviazione standard*

	<b>media A1 (deviazione standard)</b>	<b>media B1 (deviazione standard)</b>
<i>Corpus ParInIt</i>	0.46 (0.05)	0.45 (0.02)
<i>Corpus Oneworlditaliano</i>	0.47 (0.04)	0.43 (0.03)

Se nel caso del *corpus* Oneworlditaliano la differenza tra i due livelli sembra essere significativa (Welch t-test:  $t = -3.43$ , DoF = 21.66,  $p = 0.003$ ), così non è per il *corpus* ParInIt (Welch t-test:  $t = -0.58$ , DoF = 18.41,  $p = 0.58$ ).

#### 4.4. *Uso del vocabolario di base*

Tra gli adeguamenti del *teacher talk*, l'uso di parole generalmente più frequenti è spesso riportato come un fattore importante. L'insegnante di lingua deve necessariamente essere compreso e pertanto si giustificano scelte lessicali adeguate al livello; d'altra parte fornisce un modello linguistico, anche al livello lessicale, e una parte importante dell'attività didattica è dedicata proprio al lessico che deve essere padroneggiato dagli apprendenti per avanzare di livello di competenza.

Il *Vocabolario di base della lingua italiana* (VdB), elaborato da Tullio De Mauro alla fine degli anni Settanta e pubblicato per la prima volta in appendice a *Guida all'uso delle parole* (De Mauro, 1980), è stato in seguito aggiornato nelle successive edizioni. Nel presente lavoro si fa riferimento all'ultima versione disponibile, il Nuovo Vocabolario di Base, composto da circa 7000 elementi e pubblicato a dicembre del 2016<sup>9</sup>. Il *Nuovo Vocabolario di Base*, fondato su un *corpus* di 18 milioni di parole, contiene lemmi nuovi (legati prevalentemente alle nuove tecnologie) e cambiamenti di fascia (diverse parole astratte diventate più frequenti) (Chiari, 2017).

Si tratta di una lista strutturata di parole, caratterizzanti la competenza lessicale di un parlante nativo, suddivise in tre fasce:

1. il vocabolario fondamentale, contenente le parole usate con maggiore frequenza;
2. il vocabolario di alto uso, contenente parole potenzialmente note a tutti coloro che hanno alle spalle 7-8 anni di scuola;
3. il vocabolario di alta disponibilità, contenente le parole 'disponibili' anche se relativamente poco frequenti.

Il vocabolario di base è già stato usato nell'analisi dei materiali per l'insegnante di italiano come L2. Secondo Gallina (2017: 338) esso «fornisce a chi si occupa di acquisizione, insegnamento, valutazione e certificazione dell'italiano L2 uno strumento importante per elaborare sillabi e curricula, materiali didattici, test di valutazione e certificazione delle competenze linguistiche comunicative». Si può quindi ipotizzare che nel parlato degli insegnanti di italiano L2 il lessico includa elementi del vocabolario di base con maggiore frequenza nel livello A1 rispetto al livello B1.

<sup>9</sup> <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

Tabella 8. *Uso del vocabolario di base nel corpus ParInIt*

	<b>A1</b>	<b>B1</b>
Fondamentale	86%	85%
Alto uso	8%	9%
Alta disponibilità	6%	6%
Fondamentale sul totale del <i>corpus</i>	78%	78%

Tabella 9. *Uso del vocabolario di base nel corpus Oneworlditalia*

	<b>A1</b>	<b>B1</b>
Fondamentale	82.6%	84.6%
Alto uso	10.5%	9.6%
Alta disponibilità	7%	6%
Fondamentale sul totale del <i>corpus</i>	75.6%	75.6%

Considerando la stratificazione delle tre fasce del VdB nei due *corpora* esaminati si riscontrano valori molto simili (Tabella 8 e Tabella 9), rendendo pertanto superfluo il test di Welch. Si vede tuttavia che il lessico usato è, in entrambi i livelli, principalmente fondamentale, con un lieve incremento, nel B1, soltanto della fascia di “alto uso”, di un punto percentuale nel *corpus* ParInIt.

Ci si aspettava una differenza più netta tra il livello A1 e il livello B1, ipotizzando che nel livello intermedio gli insegnanti di L2 adottassero maggiormente parole del lessico comune, ma i dati analizzati mostrano un uso analogo del VdB: si poteva allora supporre una crescita delle fasce di AD e AU nel livello B1, che in effetti esiste per l'AU, ma riguarda soltanto il corpus ParInIt e rimane minima.

In entrambi i *corpora* non emergono quindi differenze nell'uso del VdB tra i due livelli.

Sembrerebbe allora che questo tipo di adeguamento non segua i livelli, o per lo meno non sia dipendente dai livelli qui esaminati. Si può comunque ritenere positivo un uso così ampio dei vocaboli di base – anche se non in progressione rispetto al livello – sia al fine di agevolare la comprensione, sia in quanto obiettivo didattico: in questo modo, gli insegnanti favoriranno l'apprendimento di parole effettivamente utili e immediatamente “spendibili”.

## 5. CONCLUSIONI

Questo articolo propone una analisi quantitativa del parlato degli insegnanti di italiano L2 volta a verificare se esista un *continuum* a vari livelli linguistici, in termini di complessità, nell'*input* fornito ad apprendenti di due livelli distinti del CEFR (A1 e B1). Si è preso in

considerazione, oltre al *corpus* ParInIt, relativo a lezioni raccolte in presenza, anche il *corpus* Oneworld contenente lezioni somministrate *online*: lo scopo del confronto tra i due *corpora* è stato verificare la presenza o l'assenza di adeguamenti nel parlato anche quando non c'è compresenza fisica tra l'insegnante e gli apprendenti.

L'analisi si inserisce in un quadro teorico delineato da qualche decennio per l'inglese, con lo studio dei tratti distintivi del *teacher talk* inteso come parlato degli insegnanti di inglese L2. In termini metodologici, l'applicazione di misure derivate dalla linguistica dei *corpora* rende possibile testare su dati trascritti le ipotesi teoriche.

Le misure presentate in questo contributo mostrano che sembra sussistere una differenza tra livelli per molti degli indicatori di complessità del parlato dell'insegnante. Dal confronto tra i due *corpora* emerge invece che gli adeguamenti si verificano spesso in maniera simile, sia nel parlato in presenza che nel parlato unidirezionale a distanza. È allora possibile inquadrare gli adattamenti come caratteristiche intrinseche del parlato dell'insegnante di lingua che pianifica, in maniera implicita, come rendere l'*input* più fruibile, indipendentemente dai *feedback* degli apprendenti.

Per quanto riguarda la complessità morfologica, modellata secondo il *Morphological Complexity Index*, i risultati per il *corpus* ParInIt e il *corpus* Oneworlditalia sono analoghi: i valori medi dei livelli A1 e B1 differiscono in maniera significativa per quanto riguarda i verbi mentre lo stesso non accade per i nomi. Tale risultato non è sorprendente, considerando la maggiore complessità morfologica dell'italiano per la classe lessicale dei verbi. Inoltre, il controllo nell'uso e nella varietà dei tempi verbali è un fattore intuitivamente determinante per l'insegnante.

La diversità lessicale, implementata con l'indice di Guiraud e ritenuta una buona approssimazione della varietà del vocabolario usato dall'insegnante, permette di rilevare che i valori medi sono differenti in maniera significativa tra il livello A1 e B1 sia per il *corpus* ParInIt che per il *corpus* Oneworlditalia.

A livello di densità lessicale, intesa come rapporto tra la frequenza degli elementi lessicali e la frequenza degli elementi grammaticali, la differenza è significativa per i valori medi dei livelli A1 e B1 relativamente al *corpus* Oneworlditalia ma non per il *corpus* ParInIt. Questo risultato è in linea con l'analisi qualitativa delle trascrizioni, che risultano piuttosto pianificate nel caso delle lezioni online (e quindi più simili allo scritto) mentre nel caso del *corpus* dei lezioni in presenza vi sono molte ripetizioni proprie del parlato e sicuramente rese necessarie dal *feedback* degli apprendenti che appiattiscono la differenza tra livelli.

L'ultimo livello di complessità analizzato, l'uso del vocabolario di base, non risulta dirimente per la distinzione tra livelli in nessuno dei due *corpora* considerati, nonostante da più parti si ritenga la stratificazione tra vocabolario fondamentale, di alto uso e di alta disponibilità, un criterio importante per la pianificazione dei materiali di insegnamento di italiano L2. Tale risultato non è di facile interpretazione e varie ipotesi possono essere avanzate, rendendo necessarie analisi future sullo stesso argomento.

## RIFERIMENTI BIBLIOGRAFICI

- Amancio D. R., Aluisio S. M., Oliveira O. N. Jr., Costa L. da F., (2012), "Complex networks analysis of language complexity", in *Europhys. Lett.*, 100, 5 (s.p.):  
[https://www.researchgate.net/publication/235645229\\_Complex\\_networks\\_analysis\\_of\\_language\\_complexity/link/5755d78508ae0405a5757e0f/download](https://www.researchgate.net/publication/235645229_Complex_networks_analysis_of_language_complexity/link/5755d78508ae0405a5757e0f/download).
- Brezina V., Pallotti G. (2016), "Morphological complexity in written L2 texts", in *Second Language Research*, 35, 1, pp. 99-119.

- Chaudron C. (1988), *Second language classrooms: Research on teaching and learning*, Cambridge, NY, Cambridge University Press.
- Chaudron C. (1982), "Vocabulary elaboration in teachers' speech to L2 learners", in *Studies in Second Language Acquisition*, 4, pp. 170-80.
- Chiari I. (2017), "Parole per farsi capire", in *La Vita Scolastica*, 10, pp. 16-18.
- Coppola D., Moretti R., Salvati L. (2021), "L'apprendimento linguistico cooperativo in contesti plurilingui. Strategie dell'apprendente e dell'insegnante", in Caruana S., Chircop K., Pace M., Vassallo Gauci P. (a cura di), *Politiche e pratiche per l'educazione linguistica, il multilinguismo e la comunicazione interculturale*, SAIL, 18, Università Ca' Foscari, Venezia, pp. 161-169.
- Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge University Press / Council of Europe: <https://rm.coe.int/16802fc1bf>. Trad. it. (2002), *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*, La Nuova Italia-Oxford-RCS, Firenze-Milano.
- De Mauro T. (1980), *Guida all'uso delle parole: come parlare e scrivere semplice e preciso. Uno stile italiano per capire e farsi capire*, Editori riuniti, Roma.
- Ellis R. (1980), "Classroom interaction and its relation to second language learning", in *RELC Journal*, 11, pp. 29-48.
- Ellis R. (1985), *Understanding second language acquisition*, Cambridge University Press, Cambridge.
- Farrell A. (2020), *Corpus Perspectives on the spoken models used by EFL teachers*, Routledge, London.
- Gallina F. (2017), "Il Nuovo Vocabolario di Base e il lessico degli stranieri: una prima ricognizione quantitativa", in *Studi Italiani di Linguistica Teorica e Applicata*, XLVI, 2, pp. 337-359.
- Hakansson G. (1986), "Quantitative aspects of teacher talk", in G. Kasper (ed.), *Learning, teaching and communication in the foreign language classroom*, Aarhus University Press, Aarhus, pp. 83-98.
- Henzl V. (1979), "Foreigner talk in the classroom", in *International review of applied linguistics*, 18, pp. 159-167.
- Kleifgen L. A. (1985), "Skilled variation in a kindergarten teacher's use of foreigner talk", in Gass S. M., Madden C. G. (eds.), *Input in second language acquisition*, Newbury House, Rowley (MA), pp. 59-68.
- Laufer B., Nation P. (1995), "Vocabulary size and use: lexical richness in L2 written production", in *Applied Linguistics*, 16, 3, pp. 307-22.
- Lyster R. (1998), "Recasts, repetition, and ambiguity in L2 classroom discourse", in *Studies in Second Language Acquisition*, 20, pp. 51-81.
- Long M. H. (1983), "Native speaker/non-native speaker conversation and the negotiation of comprehensible input", in *Applied Linguistics*, 4, pp. 126-141.
- Mizon S. (1981), *Teacher talk: a case study from the Bangalore/Madras communicational ELT project*, M.A. thesis, University of Lancaster (UK).
- Pallotti G. (2015), "A simple view of linguistic complexity", in *Second Language Research*, 31, 1, pp. 117-134.
- Ristad E. S. (1993), *The language complexity game*, MIT Press, Cambridge, MA.
- Wesche M. B., Ready D. (1985), "Foreign talk in the university classroom", in Gass S. M. Madden C. G. (eds.), *Input in second language acquisition*, Newbury House, Rowley (MA), pp. 89-114.
- Yasseri T., Kornai A., Kertész J. (2012), "A Practical Approach to Language Complexity: A Wikipedia Case Study", in *PLoS ONE*, 7. (s. p.): <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0048386>.