

*Studi Italiani di Linguistica Teorica e Applicata*  
*anno XLVI, 2017, numero 2*

**BOLLETTINO A CURA DEL CENTRO DI ECCELLENZA  
UNIVERSITÀ PER STRANIERI DI SIENA**



FRANCESCA GALLINA

*Università di Pisa*

## IL NUOVO VOCABOLARIO DI BASE E IL LESSICO DEGLI STRANIERI: UNA PRIMA RICOGNIZIONE QUANTITATIVA

### ABSTRACT

The *New Basic Vocabulary of Italian* (NVDB) published in 2016 by Tullio de Mauro represents an important resource for those, among others, interested in varieties of Italian as L2, and especially in learning processes of vocabulary of Italian as L2. The aim of this paper is to compare the most frequently used vocabulary by native speakers with the vocabulary used by non-native speakers. The NVDB will be compared with the frequency and usage lists driven by the *LIPS – Lexicon of Italian Spoken by Foreigners* (Gallina, 2015) in order to see, first of all, how many lexical units belonging to the NVDB are present also in the LIPS corpus. The research hypothesis is that lexical uses of non-native speakers are very similar to those of native speakers, which represents the main learning input for Italian L2 learners, and that the vocabulary used by non-native speakers is even more similar to that of native speakers in the first two layers of NVDB, namely the Fundamental vocabulary and the High Usage vocabulary, whereas the High Availability vocabulary is not widely used among foreigners. We will present also a first overview of the results of an analysis of NVDB words used also by non-native speakers and their frequency and usage range in the LIPS.

### 1. INTRODUZIONE

La pubblicazione del Nuovo Vocabolario di Base (NVDB) nel dicembre 2016, a cura di Tullio De Mauro e Isabella Chiari, ha segnato un momento importante per quanti si occupano di temi legati alla struttura del lessico dell'italiano e in particolare agli usi lessicali degli italiani, aggiornando uno strumento come il Vocabolario di Base (VDB), sempre di Tullio De Mauro, pubblicato nel 1980 in appendice a "Guida all'uso delle parole" presso gli Editori Riuniti. Il VDB ha avuto ampia risonanza non solo per il valore descrittivo degli usi lessicali dei parlanti italiani, ma anche per le sue appli-

cazioni in ambito educativo e in particolar modo nei contesti dove l'apprendimento della lingua italiana, il suo insegnamento e la sua valutazione hanno fatto riferimento appunto alla raccolta dei vocaboli più frequenti del lessico italiano che è il vocabolario di base<sup>1</sup>.

L'interesse e l'utilità di un vocabolario di base dell'italiano non riguardano naturalmente solo chi si occupa della lingua italiana scritta e parlata da nativi, ma toccano altresì chi si occupa di varietà di apprendimento della lingua italiana e osserva con attenzione il repertorio di vocaboli di base dell'italiano in quanto termine di confronto anche per studiare il lessico utilizzato dagli stranieri che, in Italia o all'estero, apprendono l'italiano come L2. Non solo, la presenza di un vocabolario di base per l'italiano 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 linguisticocomunicative.

La pubblicazione del NVDB investe dunque il campo della linguistica educativa e dell'educazione linguistica, ma anche della ricerca in ambito lessicologico e sociolinguistico, oltre che più specificamente della didattica dell'italiano L2, e rende pertanto necessario e possibile porsi delle domande su quale sia il rapporto tra il vocabolario di base degli italiani e il lessico degli stranieri, sia con intenti di studio, sia per i risvolti applicativi che ne derivano. Il confronto tra il lessico dei nativi e dei non-nativi, che ci accingiamo a svolgere in questa sede, consente, infatti, di scattare una fotografia dello *status* delle varietà di apprendimento dell'italiano in rapporto all'italiano d'uso dei nativi sotto il profilo lessicale, aprendo anche delle prospettive in chiave acquisizionale e conseguentemente in chiave didattica.

## 2. IL VOCABOLARIO DI BASE

Prima di procedere è opportuno richiamare brevemente alcuni concetti fondamentali, chiedendosi innanzitutto in che cosa consiste un vocabolario di base. Come afferma Ferreri (2005: 57), il vocabolario di base di una lingua è "l'insieme di vocaboli noto a tutte le persone di una comunità linguistica, indipendentemente dalla loro professione e con grado di istruzione corrispondente alla scuola di base". Il vocabolario di base include almeno per una sua componente le parole che vengono maggiormente utilizzate dai parlanti e rappresenta pertanto il nucleo centrale del lessico di una lingua. Si caratterizza per l'alta copertura lessicale delle unità lessicali dotate di maggiore frequenza che ne fanno parte, una forte polisemia e la continuità con le parole di più vecchia datazione di una lingua. Basti pensare che il lessico

della Divina Commedia ha giocato un ruolo rilevante nella costruzione del vocabolario di base dell'italiano (Chiari, De Mauro, 2012: 23). Il vocabolario di base di una lingua tende infatti a mantenersi stabile lungo l'asse diatopico, diafasico e diacronico.

Il vocabolario di base dell'italiano nasce sulle orme degli studi lessicologici francesi che avevano fatto ampio ricorso a strumenti statistici basati sulla frequenza per descrivere e analizzare il *Français Élémentaire* e d'altro canto che avevano individuato il lessico di alta disponibilità, caratterizzato da quella che è stata definita 'disponibilità lessicale', ovvero il tratto per cui una unità lessicale non è frequente dal punto di vista statistico, ma è molto legata a contesti e situazioni concreti e pertanto fortemente disponibile per i parlanti, oltre che maggiormente legata alla dimensione del parlato (Gougenheim, 1955)<sup>2</sup>.

Due sono dunque i principi che guidano la costruzione di un vocabolario di base, la frequenza e la disponibilità (Gougenheim, 1955: 409). La selezione delle unità lessicali che vengono incluse in un vocabolario di base avviene infatti innanzitutto su base statistica guardando alla frequenza e all'uso delle unità lessicali, sia con funzione grammaticale che lessicale. Nel caso del VDB la scelta dei vocaboli è stata fondata sul *Lessico di frequenza della lingua italiana contemporanea* (LIF) (Bortolini *et al.*, 1972) da cui sono stati estratti i primi 5.001 lemmi in ordine di rango d'uso, e successivamente sul *Lessico di frequenza dell'italiano parlato* (LIP) (De Mauro *et al.*, 1993)<sup>3</sup>. Non solo, nel VDB rientrano anche unità lessicali che risultano di forte disponibilità, ovvero quegli elementi che, come abbiamo già ricordato, non trovano grande uso nei testi di una lingua dal punto di vista statistico, ma che per i parlanti sono ben noti e disponibili, spesso percepiti addirittura come dotati di una maggiore disponibilità rispetto alle unità lessicali che hanno maggior frequenza. La selezione di tali vocaboli è fatta tramite una rilevazione presso un campione di parlanti di una lingua con approcci di tipo psicolinguistico. L'ultima versione del VDB fu pubblicata nel 2007 in appendice al Grande Dizionario Italiano dell'Uso, GRADIT (De Mauro, 1999 I ed.) con un insieme di circa 6.700 lemmi.

Ci preme infine sottolineare come nella prassi didattica dell'insegnamento delle lingue il vocabolario di base è stato preso in considerazione come termine di riferimento nella selezione del lessico da proporre agli apprendenti in contesto formativo. Ne sono testimonianza, ad esempio, le prime liste di frequenza per l'apprendimento dell'italiano come Thompson (1927), Skinner (1935), Russo (1947), Juilland, Traversa (1973), Sciarone (1977)<sup>4</sup>.

### 3. IL PERCHÉ DI UN CONFRONTO TRA IL NVDB E IL LESSICO DEGLI STRANIERI

Nel presente contributo ci proponiamo di intraprendere un primo confronto tra il NVDB e il lessico utilizzato da parlanti per i quali l'italiano è una L2, facendo riferimento al primo lessico di italiano parlato da stranieri. L'elaborazione del *LIPS – Lessico di Italiano Parlato da Stranieri* (Gallina, 2015) si è basata su un *corpus* di italiano parlato da non nativi da cui è stata estratta una lista d'uso che ha messo in rilievo le scelte lessicali degli stranieri, prestando attenzione al rapporto tra lessico prodotto e contesto di apprendimento da un lato e tra lessico prodotto e livello di competenza linguistico-comunicativa dall'altro lato. Una lista d'uso si caratterizza per il fatto di non prendere in considerazione solo la frequenza di un'unità lessicale, ma di considerarne l'uso, ovvero il prodotto della sua frequenza e della sua dispersione, che considera anche le frequenze parziali normalizzate, ovvero indica se e dove un lemma si concentra in un *corpus* o in una tipologia testuale<sup>5</sup>. Una lista d'uso consente dunque di tenere nella debita considerazione anche quanto un'unità lessicale è dispersa all'interno di un *corpus*, garantendo una maggiore affidabilità nell'interpretazione anche della sua frequenza.

La prassi di mettere a confronto un testo, o un *corpus*, con un termine di paragone esterno al testo rientra in quelle misure di valutazione del lessico definite *Extrinsic Measures of Lexical Variety* da Meara e Bell (2001: 6), che di solito è individuato negli usi dei nativi. Tali misure hanno il pregio di non valutare semplicemente la quantità di parole usate dai non-nativi, ma di considerare anche il tipo di parole usate, la loro frequenza negli usi dei nativi e nell'input e quindi il loro rilievo per l'apprendente. L'elaborazione di misure simili nasce dall'esigenza di valutare anche sotto il profilo qualitativo il lessico degli stranieri, partendo dall'ipotesi che le parole più ricorrenti negli usi nativi siano quelle che vengono apprese prima dai non-nativi e che quindi siano altrettanto frequenti nei loro usi. Inoltre, tali misure muovono dal presupposto che un apprendente con un lessico ristretto farà maggiore uso di parole con frequenza alta e minore uso di parole con frequenza bassa, che ricorrono di rado nell'input a disposizione degli apprendenti, e viceversa per apprendenti con un lessico ampio. Il confronto con il NVDB può dunque a nostro parere consentire un'indagine più in profondità degli usi lessicali dei non-nativi, evidenziando la relazione tra tali usi e le fasce di frequenza del lessico italiano, e può essere inteso come una “griglia cui riportare la struttura dei processi di sviluppo lessicale nelle interlingue individuali e nelle loro modellizzazioni sistematiche” (Vedovelli, Carloni, 2005: 255). Il NVDB rappresenta, infatti, la norma dei nativi, e forma pertanto l'input di apprendimento per i non-nativi esposti nel contatto con l'italiano alla presenza ricorrente dei vocaboli che compongono il NVDB.

Ciò che ci proponiamo in questa sede è dunque di mettere a confronto la lista d'uso del LIPS per osservare e misurare quante unità lessicali appartenenti al NVDB siano presenti anche nel lessico utilizzato dagli stranieri e in quali fasce di frequenza del NVDB si collochino tali elementi. Andremo a individuare i punti di contatto tra il LIPS e il NVDB per osservare se e in che modo gli stranieri utilizzino unità lessicali che sono parte del NVDB e che non erano invece incluse nella versione del VDB del 1980 o che viceversa erano parte della lista del VDB e sono fuoriuscite da quella più recente del NVDB.

L'ipotesi che andremo a verificare è che gli usi dei non-nativi rispettino quelli dei nativi, che ovviamente costituiscono l'input di apprendimento per chiunque intraprenda un percorso di acquisizione o di apprendimento dell'italiano come L2, e che tale vicinanza negli usi sia maggiore per la fascia del NVDB che ha maggiore frequenza, ovvero il vocabolario fondamentale. Verificheremo anche se gli usi meno frequenti del NVDB come i vocaboli di alto uso e in particolar modo i vocaboli della fascia di alta disponibilità trovino uno scarso impiego, o per lo meno un uso minore, nel lessico prodotto dagli stranieri.

Come vedremo nel par. 5, il NVDB si basa su un *corpus* di italiano scritto, e in parte parlato, da cui sono estratte le unità lessicali più frequenti. Allo stesso modo il LIPS si fonda su un *corpus* di italiano che rappresenta però solo la dimensione orale e che è parlato da stranieri. Il fatto che il primo si basi prevalentemente sullo scritto, mentre il secondo si fondi esclusivamente sul parlato potrebbe a prima vista compromettere le possibilità di confronto tra i due. Tuttavia, alcune considerazioni possono essere fatte in merito a tale questione. Innanzitutto, come hanno dimostrato altri confronti tra l'italiano scritto e l'italiano parlato a livello lessicale le differenze nelle fasce di frequenza d'uso sono maggiori nella testa delle liste d'uso e vanno riducendosi al diminuire della frequenza, segnando un progressivo avvicinamento tra il lessico orale e quello scritto per cui, ad eccezione ad esempio degli elementi che caratterizzano prevalentemente o esclusivamente l'interazione orale, una lista di italiano scritto può ben rappresentare almeno gli elementi più frequenti di una lista di italiano parlato<sup>6</sup>. Ne deriva che il confronto tra il NVDB e il LIPS, ovvero tra le parole più frequenti dell'italiano parlato e scritto da nativi e tra le parole più frequenti del parlato italiano di stranieri diventa una operazione possibile nonostante la diversa dimensione diamesica che le contraddistingue.

Allo stato attuale non sono noti, se non per linee generali, i dettagli della composizione del NVDB, per la quale rimandiamo al par. 5. Tuttavia, pur tenendo nella debita considerazione il fatto che sarebbe opportuno approfondire in dettaglio quanti e quali testi compongono il *corpus* da cui

deriva il NVDB, crediamo di poter ipotizzare che ai testi da cui è stato tratto il *corpus* possa essere attribuito un certo grado di rappresentatività dell'input, sia esso scritto o orale, cui sono esposti i non-nativi nel percorso di sviluppo della competenza linguistico-comunicativa, per cui viene riproposto a loro volta dagli stranieri nei loro usi lessicali. A tal proposito, prima di prendere in esame nel dettaglio il NVDB e l'utilizzo delle unità lessicali che vi rientrano da parte degli stranieri, è opportuno richiamare un aspetto fondamentale delle lingue e dei linguaggi ben illustrato anche da De Mauro, che rimarca il fatto che non sono le forme linguistiche a significare, esse non possiedono infatti "alcuna intrinseca capacità semantica: esse sono strumenti, espedienti, più o meno ingegnosi, senza vita e senza valore fuori delle mani dell'uomo, delle comunità storiche che ne facciano uso", ma sono gli uomini a significare mediante le forme linguistiche, ponendo così le condizioni necessarie agli individui per poter comunicare (De Mauro, 1965: 31). Non solo, ma è la vaghezza del significato che fa sì che il significato debba essere descritto in rapporto agli utenti in un tempo dato per descriverne soddisfacentemente le possibilità di estensione (De Mauro, 1982: 102).

#### 4. GLI USI LESSICALI, LE ANALISI QUANTITATIVE DEL LESSICO

Le ragioni che ci spingono a valorizzare l'analisi degli effettivi usi lessicali dei parlanti e l'applicazione di misure quantitative allo studio del lessico sono riconducibili alle elaborazioni teoriche e alle considerazioni di Hjelmslev e soprattutto di Coseriu, che sulla scia di Saussure e ricostruendo il senso autentico del pensiero saussuriano hanno interpretato la lingua non solo come puro congegno di relazioni formali opposto al volatile uso del piano della *parole*, ma hanno distinto nella lingua una dimensione formale, una dimensione normativa e infine quella relativa all'uso linguistico. Hjelmslev ha definito la lingua contemporaneamente come forma pura, come forma materiale e come insieme di abitudini cioè come uso (Hjelmslev, 1981, ed. orig. 1959). L'uso rappresenta un punto di svolta per lo studio del lessico, poiché su di esso è possibile basare lo studio dei significati, proprio perché esso costituisce il veicolo dei significati di una lingua. In questo senso ricordiamo quanto dice Coseriu non solo a proposito dell'opportunità di distinguere sistema, norma e *parole* per lo studio del linguaggio, ma anche a proposito delle difficoltà nello studio del lessico dovute alla grande varietà delle opposizioni lessicali che ne rendono difficile lo studio sistematico (Coseriu, 1971), per cui l'unica via di uscita rimane quella di impostare lo studio del lessico sugli usi lessicali, attraverso analisi di tipo quantitativo, capaci di ridurre, almeno in parte, la grande variabilità lessicale delle lingue

e di cogliere in qualche modo la plurideterminabilità degli usi linguistici (De Mauro, 1982). Per quanto concerne il ruolo che l'uso ha nella ricerca sul linguaggio, è senz'altro doveroso fare riferimento anche a Wittgenstein, che nelle *Ricerche filosofiche* ha offerto un contributo fondamentale per lo studio del lessico a partire dall'uso linguistico: per Wittgenstein "il significato di una parola è il suo uso nel linguaggio" (Wittgenstein, 1953: § 43)<sup>7</sup>. Wittgenstein ha fondato il significato di una parola sul suo uso nella lingua, così che il significato diventa funzione dell'uso socialmente regolato e coordinato. Le forme linguistiche acquistano significato solo nella misura in cui sono usate dall'uomo, è il legame con la società che permette al significato di essere compreso, non il legame con altre forme appartenenti al sistema linguistico, con oggetti extralinguistici o con immagini mentali. È il comportamento linguistico degli individui, cioè il loro uso della lingua e dei suoi significati in riferimento al comune repertorio della comunità linguistica di appartenenza, che garantisce il collegamento tra forma e significato e che consente all'uomo di significare, di trasmettere dei significati comprensibili nonostante la comunicazione non sia sempre perfetta e nonostante il fatto che non esistono due individui con lo stesso identico patrimonio linguistico e lessicale. "Il vantaggio maggiore della definizione del significato come uso proposta da Wittgenstein è quello, per dir così, di riconciliarci con la realtà delle cose, ossia con la grande complessità del valore semantico di una forma" (De Mauro, 1965: 214), poiché ciò permette di cogliere la pluralità dei sensi che assume un significato nei differenti contesti di enunciazione e negli usi alternativi che se ne possono fare.

Per analizzare gli usi lessicali dei parlanti è possibile sfruttare l'apporto che la statistica può fornire alle scienze linguistiche<sup>8</sup>. Secondo quanto riporta De Mauro anche Saussure riconobbe l'importanza della relazione tra fatti di lingua e calcoli: "Il arrivera un jour (...) où on reconnaîtra que les quantités du langage et leurs rapports sont régulièrement exprimables, dans leur nature fondamentale par des formules mathématiques" (De Mauro, 1982: 85). Per studiare la variabilità e l'imprevedibilità delle lingue gli strumenti statistici possono rivelarsi particolarmente utili, soprattutto se consideriamo le dimensioni che definiscono una lingua, ovvero la suddivisione in sistema, norma e *parole*. A causa delle pressioni del tempo e della massa parlante, come aveva riconosciuto Saussure, gli usi linguistici modificano continuamente il sistema, rendendo necessario l'impiego di strumenti statistici proprio per accertare i cambiamenti che toccano il sistema del linguaggio, poiché solo questa "è la via che ci permette di intendere meglio il gioco dialettico tra sistema astratto e spinte concrete, storiche, sociali, intellettuali, stilistiche della massa parlante" (De Mauro, 1994: 104). La statistica può essere applicata alla linguistica non perché la lingua sia un calcolo, ma pro-

prio perché “gli accertamenti linguistici così gettano luce da un lato sulla lingua, sulla consistenza delle sue unità e strutture, dall’altra su chi la lingua utilizza, sui locutori e sulle diverse modalità con cui essi di volta in volta, da un tipo all’altro di situazione, mettono a frutto le risorse loro offerte dalla lingua” (De Mauro, 2005a: 17). Inoltre un approccio quantitativo consente di confermare o smentire le percezioni e le intuizioni che possono sorgere nel corso delle analisi qualitative, integrando dunque i due approcci. Il lessico è uno dei settori maggiormente trattati tramite analisi statistiche sugli usi, in virtù del fatto che la sua grande instabilità costituisce un ottimo terreno di applicazione delle analisi quantitative.

Non possiamo che condividere quanto dice De Mauro sul fatto che “dal caos delle occorrenze delle unità linguistiche l’analisi statistica consente di vedere emergere regolarità, la quali sono cariche, esse stesse, di informatività per i parlanti e per l’osservatore che voglia intendere il concreto funzionamento di una lingua e gli aspetti centrali del funzionamento del linguaggio umano” (De Mauro, 1994: 106).

Di fronte alla natura mutevole del lessico, che Hjelmslev ha giustamente definito “capriccioso” (Hjelmslev, 1981: 129), ci sembra che un approccio statistico sia l’unico in grado di rilevare le regolarità e le linee di tendenza che ci consentono di fornire una descrizione adeguata del lessico stesso e ci sembra altresì che a partire dagli usi lessicali e dalle analisi di tipo quantitativo sia possibile individuare l’unica via percorribile per fondare anche uno studio acquisizionale dei processi di apprendimento lessicale.

Sulla scorta delle considerazioni che precedono, si rende evidente l’utilità dell’osservazione degli usi lessicali più frequenti dei nativi e dei non nativi quali emergono dal NVDB e dal LIPS rispettivamente, così come il confronto tra tali usi, ricorrendo anche ad analisi quantitative come andremo a fare nei paragrafi seguenti.

## 5. IL NUOVO VOCABOLARIO DI BASE

Ci sembra innanzitutto rilevante ricordare come gli scopi che erano propri del VDB e che sono rinnovati anche per il NVDB sono di tre tipi: linguistico, educativo e regolativo. L’obiettivo linguistico è sia teorico che descrittivo e mira a rappresentare in modo sistematico e trasparente il cuore del lessico dell’italiano sulla scia dei tanti studi linguistici che dagli anni Trenta hanno focalizzato la propria attenzione sulla statistica applicata ai fatti di lingua e sulla frequenza dei vocaboli come Zipf, Mandelbrot, Guiraud, Herdan, oltre naturalmente a coloro che abbiamo già ricordato nel par. 2 e nella nota 3 del presente contributo. L’obiettivo educativo consiste nel for-

nire un mezzo utile a chi apprende o insegna l'italiano come L2, ma anche a chi ha necessità di valutare le conoscenze lessicali degli alunni della scuola primaria, oltre che di elaborare un piano di sviluppo della competenza lessicale in L1 e L2 (Ferrerri, 2005)<sup>9</sup>. L'obiettivo regolativo si richiama invece a quella tradizione di scrittura capace di rendere un testo comprensibile per il maggior numero possibile di individui e che ha cercato di regolare la scrittura soprattutto dei testi amministrativi o di comunicazione pubblica in generale, fornendo delle specifiche linee guida (Piemontese, 1996; Fioritto, 1997).

Il NVDB nasce a fronte di una serie di cambiamenti che come ricorda De Mauro (2016) hanno trasformato la società italiana a partire dagli anni Settanta del Novecento ad oggi, anche in ambito linguistico: non solo è cresciuto l'indice di scolarità della popolazione italiana, ma una quota sempre maggiore di tale popolazione è diventata via via capace di utilizzare sempre più l'italiano, in esclusiva o in alternativa al dialetto (si vedano in proposito i dati sull'uso della lingua italiana, dei dialetti e di altre lingue in Italia dell'I-STAT, 2014). A fronte di tali cambiamenti e sulla base dell'ipotesi di ricerca che guida la costruzione del NVDB, ovvero che i cambiamenti nella società abbiano lasciato un segno anche nell'insieme dei vocaboli maggiormente usati e più disponibili ai parlanti, è nato il NVDB che dà conto di come parecchie centinaia di vocaboli sono usciti dal vocabolario di base perché oggi meno utilizzati e d'altro canto sono centinaia i vocaboli che prima non vi rientravano e ora sono invece parte del NVDB o erano collocati in fasce di minore frequenza.

Il NVDB è frutto della selezione su un *corpus* di circa 18 milioni di occorrenze, di cui 15 milioni di testi scritti e 3 milioni di testi orali, che sono state estratte da testi raggruppabili in sei categorie: stampa (quotidiani e settimanali), saggistica (saggi divulgativi, testi e manuali scolastici e universitari), testi letterari (narrativa, poesia), spettacolo (copioni cinematografici, teatro), comunicazione mediata dal computer (chat, ecc.), registrazioni di parlato<sup>10</sup>. Per ulteriori dettagli sulla composizione del *corpus* e dei sottocorpora, sul trattamento del *corpus*, sulle procedure di *tagging* grammaticale e di lemmatizzazione, oltre che di calcolo della frequenza e uso, rimandiamo a Chiari, De Mauro (2012). Facciamo un breve richiamo solo al fatto che le procedure di *tagging* e di lemmatizzazione del NDVB sono in parte differenti rispetto a quelle impiegate nel VDB, per cui alcune unità lessicali ora inserite nel lessico di base possono essere frutto anche di una diversa lemmatizzazione.

Il NVDB contiene circa 2.000 lemmi del Lessico Fondamentale (FO), 3.000 lemmi di Alto Uso (AU) e 2.500 lemmi di Alta Disponibilità (AD). Secondo le prime rilevazioni presentate da Chiari e De Mauro (2014: 116), il 73,3% del vecchio lessico FO è tuttora appartenente a tale categoria, così

come è confermato il 47% del lessico di AU. Il 15% dei nuovi vocaboli del lessico FO appartenevano nel VDB al lessico di AU, come ad esempio *adulto, anziano, individuo, insegnante, lavoratore, maggioranza, procedura, reagire, telefonata*. Mentre fuoriescono dal lessico FO vocaboli come *aggiustare, agricoltura, carità, ebbene, fidanzato, mela, patria, pietà, volgare*. Chiari e De Mauro (2012) evidenziano anche come molti vocaboli che prima appartenevano al lessico comune dell'italiano sono ora entrati nel lessico FO, come ad esempio *auto, foto, chiuso, interessante, nascosto, evento, villaggio*.

## 6. IL NVDB E IL LESSICO DEGLI STRANIERI: LINEE DI ANALISI

Nel corso del presente lavoro di ricerca le linee di analisi che sono state esplorate hanno visto l'avvicinarsi di tre fasi.

In una prima fase abbiamo calcolato in termini quantitativi quanta parte del NVDB e quanta parte del lessico FO, AU e AD sia presente nel LIPS, sulla base di quanto già fatto con il LIPS e il VDB in Gallina (2015). Prima di procedere al calcolo, abbiamo sottratto al LIPS tutte le unità lessicali che non era possibile confrontare con il NVDB perché si tratta di unità non rilevate dal NVDB stesso, come ad esempio le forme che sono espressione dell'interlingua o unità appartenenti ad altre lingue parlate dai non nativi del LIPS e che pure compaiono in questo *corpus*. Abbiamo quindi effettuato il conteggio di quanta parte del NVDB e delle tre fasce che lo compongono vi sia nel LIPS sia nel suo complesso, sia selezionando solo alcuni raggruppamenti di testi prodotti da non-nativi secondo le due linee di analisi sulla base delle quali è interrogabile il LIPS, ovvero il livello di competenza linguistico-comunicativa dei parlanti e il contesto di apprendimento. Questa prima fase di analisi ha permesso di scattare un'istantanea capace di dare una misura globale dei vocaboli del LIPS che appartengono al NVDB, ovvero di avere un quadro d'insieme di quanto gli stranieri utilizzino vocaboli del NVDB anche a seconda del loro livello di competenza e del contesto di apprendimento dell'italiano L2.

In una seconda fase sono state messe a confronto la lista del VDB e la lista del NVDB, per individuare i vocaboli che prima erano parte del lessico di base e che ora ne sono fuoriusciti o, viceversa, per evidenziare quelli che prima non erano parte del VDB e ora sono invece parte del NVDB. I dati ad oggi resi disponibili dagli autori del NVDB sono infatti ancora preliminari e pertanto abbiamo sentito la necessità di un confronto puntuale tra le due liste che ci consentisse di procedere alla terza fase con maggior conoscenza della composizione attuale del vocabolario italiano di base. Tuttavia, in questa sede non presenteremo in dettaglio i dati di questa fase di lavoro, per-

ché nell'economia della nostra ricerca confrontare la vecchia con la nuova versione del vocabolario di base assume la funzione di darci un quadro più preciso degli usi lessicali degli stranieri rispetto a quelli dei nativi. Lo scopo della seconda fase della nostra indagine, infatti, non era quello di puntualizzare le differenze tra le due versioni, sebbene tale linea di analisi sia di notevole rilievo e possa avere ricadute sia sul piano teorico di descrizione e analisi dell'italiano che sul piano applicativo della didattica dell'italiano come L1 e come L2, ma piuttosto quello di consentirci di fare un confronto puntuale tra il LIPS e il vocabolario di base nella terza fase della ricerca.

Nella terza fase di analisi abbiamo confrontato puntualmente la lista d'uso del LIPS con i nuovi vocaboli che sono entrati nel NVDB e con i vocaboli che si sono ricollocati diversamente all'interno del NDVB rispetto alla collocazione precedente nel VDB. Lo scopo di tale confronto è quello di verificare innanzitutto se i nuovi vocaboli del NDVB fossero già presenti nel LIPS e quindi fossero già stati appresi e fossero già entrati nella competenza d'uso dei non-nativi, oppure se fossero assenti anche negli usi degli stranieri. Nel caso in cui le nuove unità lessicali siano già attestate negli usi dei non-nativi l'ipotesi più plausibile è che il *corpus* LIPS, la cui costruzione ha una datazione di gran lunga posteriore rispetto al *corpus* da cui è stato ricavato il VDB, rispecchi fedelmente il lessico parlato dai nativi secondo la sua evoluzione più recente e ciò rende ancora più evidente la necessità di una revisione del VDB a favore del NVDB. Tuttavia, la sola presenza/assenza di un vocabolo nuovo del NDVB non è un indicatore sufficiente per capire in che modo tale vocabolo si situi all'interno delle scelte lessicali e degli usi lessicali degli stranieri rispetto al NVDB. Abbiamo pertanto deciso di considerare anche la frequenza e il rango attestati nel LIPS dei nuovi vocaboli del NVDB per capire quanto essi siano utilizzati dai non-nativi, ovvero quale rilievo abbiano negli usi lessicali e quindi nei processi di sviluppo della competenza lessicale, processi che sono necessariamente fondati sull'input offerto dai nativi. Naturalmente, l'assenza dal LIPS dei nuovi vocaboli del NVDB rende tali vocaboli candidati a diventare obiettivo di apprendimento e di insegnamento dell'italiano L2, dal momento che la loro nuova collocazione nel lessico di base dell'italiano li rende non solo maggiormente frequenti e disponibili, ma anche necessari per poter sostenere più del 90% dei discorsi quotidiani in italiano. Infine, abbiamo confrontato la lista dei vocaboli usciti dal NVDB e che erano parte del VDB per verificare ugualmente se fossero presenti nel LIPS e con quale frequenza e uso. Tale confronto consente, infatti, di evidenziare se e come gli stranieri facciano uso di parole che prima erano parte del lessico di base o ora invece non lo sono più, così come permette di mettere in luce l'assenza nel LIPS di vocaboli che sono poi fuoriusciti anche dal NVDB, sempre nel quadro dell'ipotesi per cui il processo

di sviluppo della competenza e quindi gli usi lessicali, e più genericamente linguistici, dei non nativi si fondano sull'input offerto dall'italiano parlato o scritto dai nativi. I risultati di tale fase di lavoro non sono ancora definitivi, pertanto in questa sede ci limiteremo a una presentazione parziale di alcune tendenze generali rilevate ad oggi.

Nel prossimo paragrafo andremo dunque ad analizzare i risultati delle linee di analisi qui delineate, cercando di sottolineare eventuali punti di convergenza o di divergenza tra il lessico più usato dagli italiani e quello più usato dagli stranieri.

## 7. ANALISI DEI RISULTATI

Dall'analisi preliminare che abbiamo condotto confrontando il VDB e il NVDB abbiamo rilevato come di tutti i vocaboli che sono entrati nel NVDB circa l'8% è entrato nel lessico FO, il 62% nel lessico AU e il 29% nel lessico AD. Le maggiori variazioni si hanno dunque nel lessico di alto uso e nel lessico che dà conto della rilevanza attribuita dai parlanti ad alcuni vocaboli, mentre le unità lessicali dotate della massima frequenza del lessico italiano subiscono complessivamente pochi cambiamenti, dimostrando quindi una certa stabilità diacronica. D'altro canto, se osserviamo i vocaboli che erano compresi nel VDB e ora non sono stati inclusi nel NVDB, vediamo come il 3% circa di essi sia del lessico FO, il 69% del lessico AU e il 26% del lessico di AD. Andiamo dunque a osservare nel prossimo paragrafo in che rapporto si pone in LIPS con il NVDB innanzitutto sotto il profilo quantitativo.

### 7.1 Quanto NVDB c'è nel LIPS?

Lo svolgimento di un'analisi quantitativa che si fondi sul calcolo di quanti vocaboli del NVDB sono presenti nel LIPS consente di restituire un'immagine immediata della misura in cui gli stranieri conoscono e utilizzano i vocaboli del NVDB. La tabella 1 illustra i valori percentuali del numero di vocaboli del NVDB e del VDB che sono registrati nel LIPS, con in dettaglio anche i valori relativi alle tre fasce di frequenza del lessico FO, AU, AD.

**Tabella 1.** Valori percentuali di vocaboli NVDB e VDB presenti nel LIPS

NVDB	FO	AU	AD
81,01	53,26	34,40	8,32
VDB	FO	AU	AD
47,64	42,19	39,86	17,93

Rispetto al confronto con il VDB discusso in Gallina (2015) si ha quindi un notevole aumento dei vocaboli comuni al LIPS e al vocabolario italiano di base, consentendoci di ipotizzare che il lessico parlato da stranieri è molto più vicino al NVDB che al VDB e mostrando come l'input offerto dai nativi, soprattutto quello più frequente, influenzi notevolmente lo sviluppo del bagaglio lessicale di chi apprende l'italiano come L2.

Essendo il LIPS interrogabile anche sulla base di una selezione per livello di competenza linguistico-comunicativa del campione che ne costituisce la base, abbiamo conteggiato per ogni livello dall'A1 al C2 (Council of Europe, 2001) la percentuale di NVDB, ottenendo i risultati esposti in tabella 2.

**Tabella 2.** Valori percentuali di vocaboli NVDB presenti nel LIPS per livello di competenza

A1	A2	B1	B2	C1	C2
91,24	89,87	87,54	88,68	88,6	83,61

In termini generali, al crescere del livello di competenza si evidenzia un calo progressivo della quantità relativa di vocaboli NVDB, ovvero il livello A1 è il livello che maggiormente utilizza unità lessicali del NVDB, mentre il livello C2 è quello che ne utilizza meno, come ci si poteva aspettare. La decrescita di vocaboli NVDB è piuttosto lenta almeno fino al livello C1, mentre il calo si fa più marcato nel livello C2. Il NVDB è dunque una parte importante del lessico utilizzato dagli stranieri in tutti i livelli, ma risulta maggiormente utilizzato in quei livelli in cui i domini, i contesti, le situazioni comunicative sono più legati alla quotidianità, mentre risulta meno utilizzato nel livello che si caratterizza, secondo la definizione stessa del Quadro Comune Europeo, come la competenza di un utente capace di gestire e dominare anche situazioni comunicative professionali o comunque svincolate dalla comunicazione quotidiana per la quale invece è necessario fare ricorso al NVDB.

I dati raccolti nella tabella 3 illustrano la distribuzione delle fasce che compongono il NVDB per livello di competenza.

**Tabella 3.** Valori percentuali di vocaboli delle tre fasce del NVDB presenti nel LIPS per livello

	A1	A2	B1	B2	C1	C2
FO	88,8	88,25	74,08	74,56	76,76	76,31
AU	9,60	9,93	21,33	21,20	20,29	22,06
AD	1,60	1,80	4,57	4,22	2,94	1,62

Ciò che emerge chiaramente dall'osservazione dei dati è che si ha una prima consistente differenza tra i livelli A1 e A2 e i livelli successivi: nei livelli A, infatti, la percentuale di vocaboli FO è altamente superiore rispetto agli altri livelli, mettendo in luce come la competenza degli apprendenti con

una varietà di apprendimento basica si fonda soprattutto sulle unità lessicali più frequenti in assoluto del lessico italiano e meno su quelle di alto uso o di alta disponibilità. I limiti delle conoscenze lessicali, almeno sotto il profilo quantitativo, dei livelli A1 e A2 sono quindi ben evidenziati dai dati discussi. Per ciò che concerne il lessico di AU vediamo come esso trovi un uso ampiamente maggiore a partire dal livello B1, segnando una differenza marcata tra livelli iniziali di apprendimento e livelli successivi, a conferma dell'ipotesi di Carloni e Vedovelli (2005) che individuavano proprio nel lessico AU l'elemento discriminante tra apprendenti con livelli di competenza differenti, sostenendo che il lessico di AU sia meno presente nei livelli iniziali rispetto a quelli più avanzati. Osservando invece i livelli di competenza più avanzati, risulta interessante il fatto che la percentuale di lessico AD è maggiore nei livelli B1 e B2, con un calo progressivo che si fa più accentuato nel livello C1 e soprattutto nel livello C2, come era debito aspettarsi. Il lessico di AD è, infatti, per definizione un lessico che fa riferimento alla quotidianità, a momenti, fatti, eventi usuali e ripetuti, per cui è più facile riscontrarne la presenza nei livelli B1 e B2 piuttosto che nei livelli C1 e C2. Possiamo quindi ricavare da tali dati alcune considerazioni generali: i livelli A si caratterizzano per un uso maggiore del lessico FO rispetto agli altri livelli; inoltre, il lessico AU segna un'ulteriore differenza tra i livelli di varietà basica e i livelli più avanzati, facendo emergere quelli che sono i limiti della competenza lessicale dei livelli iniziali; ciò che distingue invece i livelli B e C è il lessico AD, che è ben presente negli usi, e quindi nelle situazioni comunicative, dei livelli B1 e B2 e viene invece a ridursi notevolmente nei livelli C1 e C2, dove il riferimento a ciò che è abituale e quotidiano è meno comune.

Volendo interrogare il LIPS sulla base dei contesti di apprendimento in cui gli stranieri che compongono il campione del *corpus* LIPS hanno appreso l'italiano e raggruppando tali contesti in tre macro gruppi che sono a. Italia; b. estero; c. misto – ovvero apprendenti che hanno imparato l'italiano rispettivamente in Italia, in altri Paesi e sia in Italia che all'estero – è possibile procedere calcolando anche per questi raggruppamenti la percentuale di vocaboli NVDB per verificare se vi siano delle differenze tra apprendenti che vivono in Italia, all'estero o hanno vissuto sia in Italia che all'estero, e che sono dunque esposti a input e processi di apprendimento misti. I dati del gruppo Italia mettono in luce come appartengono al NVDB l'85,81% di unità lessicali utilizzate da questo gruppo, con un risultato molto simile al gruppo estero, per il quale l'85,24% rientra nel NVDB. Il gruppo misto è invece quello che si differenzia maggiormente con un 90,02% di unità lessicali che sono parte del NDVB. Tali dati sono in linea con quanto già discusso in Gallina (2015). Vediamo dunque la distribuzione nelle fasce di frequenza del NVDB all'interno dei tre gruppi nella tabella 4.

**Tabella 4.** Valori percentuali dei vocaboli delle tre fasce del NVDB presenti nel LIPS per contesto di apprendimento

	estero	Italia	misto
FO	69,76	72,8	78,99
AU	25,95	23,59	18,59
AD	4,29	3,59	2,40

Come in Gallina (2015) il gruppo estero è il gruppo che usa il maggior numero di vocaboli AU e AD, mentre il gruppo Italia si colloca in una posizione intermedia tra gli altri due gruppi. Rimandiamo a Gallina (*ibid.*) per una discussione approfondita di tali dati, tuttavia ricordiamo qui come la presenza di unità lessicali anche di AU e AD in tutti i gruppi è indizio della capacità degli apprendenti, all'interno del percorso di sviluppo della competenza in italiano L2, di rendere oggetto di apprendimento e poi di uso i vocaboli appartenenti anche a queste fasce del vocabolario di base, avvicinandosi così agli usi dell'italiano da parte dei nativi. Il contatto con tali unità lessicali può avvenire tramite l'input offerto in contesto didattico e i materiali didattici, soprattutto per chi apprende l'italiano all'estero, ma anche tramite la negoziazione del senso che si ha nel contatto con i nativi per chi vive in Italia e l'elaborazione dell'input nel processo acquisizionale che diventa quindi un processo innanzitutto sociale<sup>11</sup>.

### *7.2 Frequenza e uso del NVDB da parte dei non-nativi: alcune rilevazioni preliminari*

Per confrontare il NVDB con il LIPS in maniera puntuale abbiamo innanzitutto preso in esame il lessico NVDB suddiviso per fasce e verificato non solo quali vocaboli fossero presenti o assenti nel LIPS, ma anche con quale frequenza e uso comparissero nel lessico parlato da stranieri. In questa sede ci limiteremo solo a presentare alcune tendenze emerse nel corso di questa fase di analisi.

In tutte e tre le fasce del NVDB abbiamo innanzitutto rilevato come parte dei nuovi vocaboli del NVDB presenti anche nel LIPS siano entrati nel NVDB per motivi legati a una diversa procedura di *tagging* grammaticale e di lemmatizzazione rispetto al VDB. Ne sono un esempio numerosissimi avverbi che terminano in *-mente* o molti vocaboli come *determinato*, *divertente*, *legato*, *nascosto*, *sbagliato*, *seguito*, *scorso*, che hanno nel LIPS frequenze diverse tra loro, alcuni caratterizzandosi infatti per una frequenza e un uso molto elevati, come ad esempio *interessante*, e altri invece essendo poco frequenti e usati, seppur presenti nel LIPS, come ad esempio *compreso*.

I cambiamenti registrati nel NVDB e attestati anche nel LIPS che risultano più interessanti sono però quelli legati all'evoluzione della società italiana e della lingua italiana. Osservando ad esempio alcune categorie grammaticali come quella del nome, è evidente che i nuovi vocaboli del NVDB già presenti nel LIPS sono, in larga parte, testimonianza concreta dei cambiamenti in seno alla società. Ad esempio tra i nomi che erano già attestati nel LIPS e ora sono parte del lessico FO ci sono numerosi vocaboli legati alla tecnologia come *web* e *file*, che hanno però pochissime occorrenze nel LIPS, e altri invece che ricorrono più spesso nel parlato degli stranieri come *internet*, *cellulare*, *e-mail*, *televisivo*, *video*. Tra i vocaboli che maggiormente risentono dell'uso del parlato erano già comprese nel LIPS elementi come *foto* e *tivù*, con una frequenza molto elevata, e *auto*. Pochissimi sono i vocaboli FO che sono assenti nel LIPS e tra questi vi sono *blog*, *capitolo*, *editore*, *essi*, *generare*, *individuare*, *sentenza*. Si tratta prevalentemente di vocaboli che sono probabilmente riconducibili più alla dimensione scritta della lingua che a quella orale e pertanto non ci stupisce la loro assenza nel LIPS. Inoltre non sono presenti nel LIPS le parole di accentuata volgarità presenti nel NVDB e alcune unità lessicali appartenenti a categorie grammaticali come articoli e pronomi che sono state trattate diversamente rispetto al NVDB, come ad esempio *le*, *una*, *li*, ecc.

Tra i pochi vocaboli del lessico FO del VDB che non sono parte del NVDB vi sono diverse interiezioni che sono presenti, alcune con frequenze molto elevate, nel LIPS, come ad esempio *ah*, *basta*, *eh*, *oh*, il cui utilizzo caratterizza il parlato più che lo scritto e che pertanto non ci sorprende essere assenti nel *corpus* che costituisce il NVDB. Vi sono inoltre alcuni vocaboli con frequenze bassissime nel LIPS come *allevamento*, *biancheria*, *carrozza*, *motocicletta* o addirittura assenti dal LIPS come *appetito*, *benedire*, *giovanotto*, *vigliacco*. Tutto ciò mette in luce come soprattutto nelle fasce di frequenza più elevate il lessico dei non-nativi sia molto simile negli usi al lessico dei nativi e inoltre fa emergere anche negli usi dei non-nativi da un lato la stabilità del lessico fondamentale sulla dimensione diacronica e dall'altro lato i riflessi dei cambiamenti della società italiana e i processi di evoluzione degli usi e delle scelte lessicali anche presso gli stranieri.

Nel lessico di AU tra i nuovi vocaboli presenti nel LIPS ci sono ad esempio *telefonino*, *supermercato*, *stress* a segnalare i cambiamenti avvenuti nella società italiana, ma anche molti esotismi legati alla tecnologia come *community*, *link*, *network*, o più generici come *gay*, *global*, *party*, *style*, *trend*, ecc. Tra i vocaboli AU che non sono più parte del vocabolario

di base ve ne sono moltissimi che non erano attestati nemmeno nel parlato degli stranieri come ad esempio *castigo, disattenzione, gabbiano, invidia, pagliaccio, raggruppare*, ma anche molti vocaboli che nel LIPS avevano frequenza molto bassa come *battesimo, esportazione, inseguire, mediocre, ombrello, squillo*. Anche in questo caso dunque il lessico degli stranieri aveva già fatto affiorare lo scarso uso di tali vocaboli e mostra una forte somiglianza con gli usi lessicali dei nativi.

Se diamo un ultimo sguardo al lessico AD a fronte del LIPS una prima considerazione riguarda il fatto che la frequenza e l'uso dei vocaboli di AD che sono presenti nel LIPS sono estremamente bassi, marcando così una differenza netta rispetto al lessico AU e soprattutto FO già attestato nel LIPS. Osservando i vocaboli di AD del VDB che ora sono stati esclusi dal NVDB, ciò che emerge nettamente è che sono solo alcune unità a essere attestate nel LIPS, peraltro con frequenze bassissime, tra cui: *barbiere, corsa, distaccare, maglieria, saldare, vanitoso*. Per il resto, i vocaboli di AD usciti dal NVDB non sono registrati nella lista d'uso del LIPS principalmente perché nel LIPS, così come nel NVDB, aggettivi e nomi sono lemmatizzati alla forma maschile, diversamente da quanto era stato fatto nel VDB, ma anche perché alcune unità lessicali che prima erano sentite come disponibili dai parlanti di fatto non rientrano tra gli usi reali degli stranieri, come ad esempio *asciugacapelli, mezzadria, mittente, oppressione*.

## 8. CONCLUSIONI

Quali sono dunque le considerazioni che possiamo trarre da questa lettura preliminare dei dati del NVDB a fronte del LIPS? A nostro avviso pare possibile individuare tre chiavi di lettura.

Una prima chiave di lettura consente di vedere nel lessico dei non-nativi un riflesso abbastanza fedele del lessico dei nativi: gli stranieri utilizzano il NVDB in misura consistente quando parlano, mettendo in rilievo di conseguenza il ruolo che l'italiano prodotto da nativi gioca nei processi di apprendimento dell'italiano L2 e di sviluppo della competenza lessicale, processi nei quali è proprio l'italiano d'uso a costituire un input di apprendimento di fondamentale importanza. Come i dati del NVDB hanno dimostrato, anche il lessico dei non-nativi mostra, infatti, come il cambiamento della società italiana nelle sue diverse dimensioni, linguistiche, comunicative, tecnologiche, produttive, ecc. sia marcato e soprattutto sia ben rap-

presentato anche nei vocaboli che i parlanti stranieri scelgono di utilizzare nello scrivere e nel parlare in italiano. Sotto questo profilo quindi il LIPS ha anticipato alcune tendenze che sono poi state rese manifeste nella nuova edizione del vocabolario di base, per lo meno nella misura in cui esso è il prodotto degli usi lessicali dei nativi riflessi negli usi degli stranieri.

Analizzando con maggior attenzione le fasce di cui si compone il NVDB affiora un altro elemento capitale: il maggior grado di somiglianza tra il lessico dei nativi e dei non-nativi si ha nel lessico più frequente o di alto uso, con un progressivo allontanamento nel lessico di alta disponibilità. Questo ci conduce a proporre una seconda chiave di lettura, ovvero ciò che è più frequente negli usi dei nativi è l'input che meglio si radica nella competenza dei non-nativi, come i dati del confronto tra il lessico FO e il LIPS fanno risaltare: è infatti tra questi raggruppamenti di vocaboli che si trova il maggior numero di elementi comuni, a sottolineare ancora una volta il rilievo della frequenza e dell'input nei processi di sviluppo dell'interlingua. Nella fascia di alto uso decresce il numero di vocaboli che sono attestati anche nel LIPS, a causa, a nostro avviso, della loro minore frequenza nell'input di apprendimento rispetto al lessico fondamentale. Diminuisce ulteriormente il numero di vocaboli di alta disponibilità rilevato nel LIPS, mostrando come la percezione dei nativi del rilievo di tali elementi non sia sempre sufficiente per la loro acquisizione, soprattutto laddove il loro uso non sia frequente. L'entrata nella competenza lessicale del lessico di alta disponibilità può infatti dipendere non tanto dagli usi reali, ma dalla loro presenza nei materiali didattici, per chi studia l'italiano L2 in contesto formale, oppure dall'interazione con i nativi, attraverso dei processi di negoziazione del senso.

Vogliamo inoltre sottolineare come la convergenza tra gli usi dei non-nativi e il NVDB, soprattutto nelle fasce di lessico FO e di AU, evidenzia come gli autori dei materiali didattici utilizzati nei percorsi formali di apprendimento dell'italiano L2 siano capaci di valide scelte del lessico da proporre agli apprendenti, selezionando quelle unità lessicali che sono le più ricorrenti negli usi dei nativi e pertanto le più utili per gli apprendenti.

Una terza chiave di lettura si lega a una delle caratteristiche costitutive del lessico di una lingua ovvero la grande variabilità che qualifica la dimensione lessicale e che fa sì che anche i processi di sviluppo della competenza lessicale siano contrassegnati dalla medesima variabilità, influenzando quindi fortemente sul bagaglio lessicale di ogni singolo apprendente. Per quanto il supporto delle nuove tecnologie che consentono di rac-

cogliere quantità di dati linguistici consistenti e gli strumenti offerti dalla statistica linguistica siano di grande aiuto per individuare delle tendenze generalizzanti nell'ambito del lessico, che altrimenti per natura rimarrebbe difficilmente analizzabile, resta pur sempre aperto uno spazio di variabilità e irregolarità, come dimostrano alcuni cambiamenti nel vocabolario di base o la presenza/assenza di alcuni vocaboli del NVDB dal LIPS. La complessità del lessico e la sua variabilità possono essere almeno in parte controllate tramite analisi quantitative dei fatti linguistici, ma è altresì necessario riconoscere che è soprattutto nei processi di apprendimento, e nello specifico di apprendimento dell'italiano L2, che la dialettica tra regolarità e deviazione emerge con tutta la propria forza, mettendo in rilievo come la varietà dei contesti di apprendimento e le diverse capacità cognitive e sociali dei singoli individui possono anch'esse essere determinanti nello sviluppo delle varietà di apprendimento. I processi di apprendimento possono dunque diventare il luogo a cui riportare le analisi degli usi lessicali di nativi e non-nativi sia per riflettere sullo sviluppo della competenza lessicale in prospettiva acquisizionale sia per sviluppare una proposta didattica realmente capace di includere ciò è più frequente e disponibile nell'input e pertanto più utile per l'apprendente.

#### NOTE

<sup>1</sup> Si pensi a solo titolo di esempio alla collana dei Libri di Base degli Editori Riuniti, al giornale di facile lettura "Due parole" e ai materiali didattici realizzati da case editrici quali Principato, Giunti, Minerva Italiana.

<sup>2</sup> Si veda anche Gougenheim *et al.* (1956). La nozione di disponibilità è stata introdotta a partire da uno spoglio lessico-statistico di campioni di lingua parlata, che aveva evidenziato come tra le unità lessicali più frequenti mancassero le unità legate ai contesti quotidiani e concreti di vita. A fronte di tale assenza, i vocaboli di alta disponibilità sono stati individuati tramite un'inchiesta che chiedeva a un campione di studenti delle scuole francesi di indicare le 20 parole che ritenevano maggiormente utili a proposito di 16 centri di interesse (tra cui il corpo umano, la casa, i pasti, ecc.) e rivedendo e integrando in una seconda fase questo elenco in seno a una commissione di esperti.

<sup>3</sup> Per un approfondimento della nozione di vocabolario di base dell'italiano anche alla luce della stratificazione diacronica del lessico italiano si veda Giuliani *et al.* (2005). Giuliani *et al.* ricordano come dei primi 5.001 lemmi del LIF ne siano stati esclusi 261 perché risultati poco compresi da persone con un titolo di studio non superiore alla licenza media. Riportano inoltre come i primi 1.991 lemmi del LIF siano stati definiti come *vocabolario fondamentale* e i restanti 2.749 lemmi siano stati definiti come *vocabolario di alto uso*. A questi si aggiungono nella versione del VDB del 1980 2.336 lemmi di alta disponibilità. Sui cambiamenti nella composizione del lessico di alta disponibilità del VDB si veda De Renzo (2005).

<sup>4</sup> Ricordiamo anche Knease (1933) che ha lavorato su un *corpus* di testi letterari e Migliorini (1943) che si è basato invece sulle proprie intuizioni per la selezione delle unità lessicali. Come riporta Bogaards (1994) una delle prime liste costruite nell'ottica dei bisogni specifici dell'insegnamento e dell'apprendimento è quella di Keniston (1920) per lo spagnolo, preceduto però da

Kaeding (1898) per il tedesco e a cui hanno fatto seguito, a partire dagli anni Venti, numerose liste per varie lingue per le quali rimandiamo alla sintesi in De Mauro, Chiari (2005: 188-9). Rimandiamo inoltre a Katerinov et al. (1991) sia per un'analisi e un confronto delle liste di frequenza di italiano parlato e scritto sia per il loro uso in ambito didattico.

<sup>5</sup> Sul concetto di dispersione e sulle formule per calcolarla si vedano Bolasco (1999) e Chiari (2005). Nel caso del LIPS per la redazione della lista d'uso si è calcolato la dispersione delle unità lessicali adottando la stessa formula già impiegata nel LIF e nel LIP cui rimandiamo.

<sup>6</sup> Si veda a tal proposito Vedovelli (1993) che, confrontando il LIP (De Mauro *et al.*, 1993) con LIF (Bortolini, Tagliavini, Zampolli, 1971), ha rilevato come i dati sulla copertura delle liste di frequenza del parlato e dello scritto si differenziano maggiormente fino alle prime 2000 parole più frequenti, corrispondenti a quelle parole che nel Vocabolario di Base appartengono al Lessico Fondamentale, mentre le liste vanno progressivamente avvicinandosi scorrendo le liste stesse fino a trovare il maggior punto di prossimità intorno a 2500 lemmi. Le differenze tra scritto e parlato sono dunque significative se si considerano le teste delle liste d'uso, in cui il parlato ricorre a un nucleo di parole più ristretto rispetto allo scritto per realizzare uno spettro di testi più ampio, e vanno invece diminuendo via via che si considerano le parole con frequenza decrescente.

<sup>7</sup> Ricordiamo come De Mauro (1965) abbia per primo dato una lettura delle *Ricerche filologiche* in tal senso. Le considerazioni che seguono sono tratte dallo stesso De Mauro (1965), che è ritornato più volte nel corso del tempo su temi simili.

<sup>8</sup> Si veda in proposito il volume curato da Tullio De Mauro e Isabella Chiari sull'applicazione delle analisi quantitative ai fatti di lingua (De Mauro, Chiari, 2005). De Mauro è da sempre uno dei più convinti sostenitori nel panorama italiano dell'utilità dell'applicazione della statistica allo studio delle lingue storico-naturali, come dimostrano, tra gli altri, lavori come il contributo sulla statistica linguistica in *Enciclopedia Italiana* (1961).

<sup>9</sup> Come afferma Gougenheim (1955: 404) anche il *Français élémentaire* era stato creato come supporto alla diffusione della lingua francese come L2.

<sup>10</sup> Nel NVDB entrano dunque anche testi della comunicazione mediata dal computer, che pone non poche questioni nell'analisi della lingua italiana soprattutto in relazione al rapporto tra parlato e scritto. Per un approfondimento rimandiamo a D'Achille (2003), Lorenzetti, Schirru (2006), Tavosanis (2011).

<sup>11</sup> Rimandiamo al contributo di Villarini (2008) per un'analisi della presenza del VDB nei materiali didattici destinati all'apprendimento dell'italiano L2.

## BIBLIOGRAFIA

- Barni M., Bagna C., Troncarelli D. (a cura di), (2008), *Lessico e apprendimenti. Il ruolo del lessico nella linguistica educativa*, Milano, Franco Angeli.
- Bolasco S. (1999), *Analisi multidimensionale dei dati. Metodi, strategie e criteri di interpretazione*, Roma, Carocci.
- Bortolini U. Tagliavini C., Zampolli A. (1971), *Lessico di frequenza della lingua italiana contemporanea*, Milano, IBM.
- Carloni F., Vedovelli M. (2005), "Il vocabolario di base dell'italiano degli stranieri", in De Mauro T., Chiari I., (a cura di), *Parole e numeri, analisi quantitative dei fatti di lingua*, Roma, Aracne pp. 247-275.
- Chiari I. (2005), *Informatica e lingue naturali. Teorie e applicazioni computazionali per la ricerca sulle lingue*, Roma, Aracne.

- Chiari I., De Mauro T. (2012), "The new basic vocabulary of Italian: problems and methods", *Statistica applicata*, 22(1), pp. 21-35.
- Chiari I., De Mauro T. (2014), "The New Basic Vocabulary of Italian as a Linguistic Resource", Basili R., Lenci A., Magnini B., *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014*, Pisa, Pisa University Press, pp. 113-6.
- Coseriu E. (1971), *Teoria del linguaggio e linguistica generale*, Roma-Bari, Laterza.
- D'Achille P. (2003), *L'italiano contemporaneo*, Bologna, il Mulino.
- De Mauro T. (1961), "Statistica linguistica", *Enciclopedia italiana*, Appendice III, 2: 820-821.
- De Mauro T. (1965), *Introduzione alla semantica*, Roma-Bari, Laterza.
- De Mauro T. (1980), *Guida all'uso delle parole*, Roma, Editori Riuniti.
- De Mauro T. (1982), *Minisemantica dei linguaggi non verbali e delle lingue*, Roma-Bari, Laterza.
- De Mauro T. (1994), *Capire le parole*, Roma-Bari, Laterza.
- De Mauro T. (1999), *Grande dizionario Italiano dell'Uso*, Torino, UTET.
- De Mauro T. (2005a), "Introduzione", in De Mauro T., Chiari I., (a cura di), *Parole e numeri, analisi quantitative dei fatti di lingua*, Roma, Aracne, pp. 9-17.
- De Mauro T. (2016), "Il Nuovo vocabolario di base della lingua italiana", *Internazionale*, <http://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana> (ultimo accesso 24.4.2017).
- De Mauro T., Mancini F., Vedovelli M., Voghera M. (1993), *Lessico di frequenza dell'italiano parlato*, Milano, ETASLIBRI.
- De Mauro T., Chiari I. (a cura di), (2005), *Parole e numeri, analisi quantitative dei fatti di lingua*, Roma, Aracne.
- De Renzo F. (2005), "Nuove rilevazioni sul vocabolario di base e di alta disponibilità", in De Mauro T., Chiari I., (a cura di), *Parole e numeri, analisi quantitative dei fatti di lingua*, Roma, Aracne, pp. 215-32.
- Ferreri S. (2005), *L'alfabetizzazione lessicale: studi di linguistica educativa*, Roma, Aracne.
- Fioritto A. (1997), *Manuale di stile: strumenti per semplificare il linguaggio delle amministrazioni pubbliche*, Presidenza del Consiglio dei Ministri, Milano, Il Mulino.
- Gallina F. (2015), *Le parole degli stranieri. Il Lessico dell'Italiano Parlato da Stranieri*, Perugia, Guerra Edizioni.
- Gensini S. (a cura di) (2006), *Fare comunicazione. Teoria ed esercizi*, Roma, Carocci.
- Giuliani A., Iacobini C., Thornton A. M. (2005), "La nozione di vocabolario di base alla luce della stratificazione diacronica del lessico dell'italiano", in De

- Mauro T., Chiari I., (a cura di), *Parole e numeri, analisi quantitative dei fatti di lingua*, Roma, Aracne, pp.193-213.
- Gougenheim G. (1955), “Le français élémentaire”, *International Review of Education/ Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'éducation*, 1: 401-412.
- Gougenheim G., Michéa R., Rivenc P., Sauvageot A. (1956), *L'élaboration du français élémentaire: Étude sur l'établissement d'un vocabulaire et d'une grammaire du base*, Paris, Didier.
- Hjelsmlev L. (1981), *Saggi di linguistica generale*, Pratiche Editrice, Parma (Ed. orig. 1959, *Essais linguistique*, Nordisk Sprog- og Kulturforlag, Copenhagen).
- ISTAT (2014), *L'uso della lingua italiana, dei dialetti e di altre lingue in Italia*, [http://www.istat.it/it/files/2014/10/Lingua-italiana-e-dialetti\\_PC.pdf](http://www.istat.it/it/files/2014/10/Lingua-italiana-e-dialetti_PC.pdf) (ultimo accesso 4.5.2017).
- Juilland A.G., Traversa V. (1973), *Frequency Dictionary of Italian Words*, The Hague, Mouton.
- Katerinov K., Boriosi Keterinov M.C., Sciarone A.G. (1991), *Calcolo di rendimento di liste di base: italiano parlato, italiano scritto, italiano soglia (Ricerche di linguistica computazionale)*, Perugia, Guerra.
- Keniston H. (1920), “Common words in Spanish”, *Hispania* III: 85-96.
- Knease T.M. (1933), *An Italian Word List from Literary Sources*, Toronto, The University of Toronto Press.
- Lorenzetti L., Schirru G. (2006), “La lingua italiana nei nuovi mezzi di comunicazione: SMS, posta elettronica e Internet”, in Gensini S. (a cura di), *Fare comunicazione. Teoria ed esercizi*, Roma, Carocci: 71-98.
- Meara P., Bell H. (2001), *P\_Lex: A simple and effective way of describing the lexical characteristics of short texts*, *Prospects*, 16, 3, pp. 5-19.
- Migliorini B. (1943), *Der grundlegende Wortschatz des Italienischen*, Marburg, Elwert.
- Piemontese M.E. (1996), *Capire e farsi capire: teorie e tecniche della scrittura controllata*, Bari, Tecnodid.
- Russo G.A. (1947), “A combined Italian word list”, *The Modern Language Journal*, 31, pp. 218-40.
- Sciarone A.G. (1977), *Vocabolario fondamentale della lingua italiana*, Perugia, Guerra.
- Skinner L.H. (1935), “A comparative study of the vocabularies of forty-five Italian text-book”, *The Modern Language Journal*, 20, pp. 67-84.
- Tavosanis M. (2011), *L'italiano del web*, Roma, Carocci.
- Thompson M.E. (1927), *A Study in Italian Vocabulary Frequency*, tesi di laurea non pubblicata, University of Iowa.
- Vedovelli M. (1993), “Confronti fra il LIP e le altre liste di frequenza dell'italiano”, in De Mauro T., Mancini F., Vedovelli M., Voghera M., *Lessico di frequenza dell'italiano parlato*, Milano, ETASLIBRI.

Villarini A. (2008), “Il lessico dei materiali didattici usati nei corsi di italiano per migranti”, in Barni, Bagna C., Troncarelli D., (a cura di), *Lessico e apprendimenti. Il ruolo del lessico nella linguistica educativa*, Milano, Franco Angeli, pp. 165-177.

Wittgenstein L. (1953), *Ricerche filosofiche*, Torino, Einaudi.



SILTA XLVI, 2/2017

Indice degli Autori

Barbero Bernal Juan Carlos, *Università di L'Aquila*, juancarlos.barberobernal@univaq.it

Capra Daniela, *Università di Modena e Reggio*, daniela.capra@unimore.it

Celani Simone, *Università La Sapienza di Roma*, simone.celani@uniroma1.it

Mattarucco Giada, *Università per Stranieri di Siena*, mattarucco@unistrasi.it

Polo Anna, *Università di Padova*, anna.polo@unipd.it

Romanelli Norma, *Université de Paris VII Diderot e Università di Siena*, normaromanelli@gmail.com

San Vicente Felix, *Università di Bologna*, felix.sanvicente@unibo.it

Silvestri Paolo, *Università di Siviglia*, psilvestri@us.es

Swiggers Pierre, *Università di Lovanio*, pierre.swiggers@kuleuven.be

Szoc Sara, *Università di Lovanio*, sara.szoc@kuleuven.be

## INDICE DEI NUMERI TEMATICI DI *SILTA*

1. Enrico Arcaini, Michèle Fourment-Berni Canani, Danielle Lévy-Mongelli, (a cura di), 1989, *Analisi comparativa: Francese/Italiano. Ricerca linguistica - Insegnamento delle lingue*, *SILTA*, XVIII, 1-2.
2. Franca Orletti (a cura di), 1991, *Problema sociolinguistico dell'immigrazione*, *SILTA*, XX, 2.
3. Franco Crevatin (a cura di), 1993, *Incontri con l'Africa*, *SILTA*, XXII, 1.
4. Carlo Serra Borneto (a cura di), 1993, *Aspects of Cognitive Grammar*, *SILTA*, XXII, 3.
5. Massimo Vedovelli (a cura di), 1994, *Aspetti dell'apprendimento spontaneo e guidato dell'italiano in contesto migratorio*, *SILTA*, XXIII, 2.
6. Enrico Arcaini, Michèle Fourment-Berni Canani, Danielle Lévy Mongelli, (a cura di), 1994, *Lingue e culture a confronto Ricerca linguistica - Insegnamento delle lingue*, *SILTA*, XXIII, 3.
7. Arianna Uguzzoni (a cura di), 1995, *L'aspetto fonico del linguaggio. Studi e ricerche*, *SILTA*, XXIV, 2.
8. Michele Prandi (a cura di), 1996, *La subordinazione non completiva. Un frammento di grammatica filosofica*, *SILTA*, XXV, 1.
9. Gaston Gross, Pierre Lerat, Christian Molinier (a cura di), 1996, *L'adjectif: une catégorie hétérogène*, *SILTA*, XXV, 3.
10. Gilbert Lazard (a cura di), 1997, *La typologie actancielle*, *SILTA*, XXVI, 2, 1997.
11. Marina Chini e Anna Giacalone Ramat (a cura di), 1998, *Strutture testuali e principi di organizzazione dell'informazione nell'apprendimento linguistico*, *SILTA*, XXVII, 1.
12. Cristoph Schwarze (a cura di), 1999, *Semantica lessicale*, *SILTA*, XXVIII, 2.
13. Sabine E. Koesters Gensini, (a cura di), 2000, *Dieci anni di linguistica sincronica tedesca (1989-1999)*, *SILTA*, XXIX, 1.
14. Raffaele Simone (a cura di), con la collaborazione di Edoardo Lombardi Vallauri e Silvia Pieroni, 2000, *Classi di parole e conoscenza lessicale*, *SILTA*, XXIX, 2.
15. Stefania Nuccorini (a cura di), 2001, *When a Torch becomes a Candle: Variation in Phraseology*, *SILTA*, XXX, 2.
16. Irina Čelyševa e Anna Jampol'skaia (a cura di), 2002, *Studi attuali di italianistica in Russia*, *SILTA*, XXXI, 2.

17. Claudia Lasorsa Siedina e Anna Jampol'skaia (a cura di), 2004, *Studi attuali di russistica in Italia*, *SILTA*, XXXIII, 2.
18. Harro Stammerjohann (a cura di), 2005, *La linguistica italiana nei paesi germanofoni (Germania, Austria, Svizzera tedesca)*, *SILTA*, XXXIV, 1.
19. Vincenzo Orioles e Fiorenzo Toso, (a cura di), 2005, *Le eteroglossie interne. Aspetti e problemi*, *SILTA*, XXXIV, 3.
20. Giovanni Rovere, (a cura di), 2007, *Per una teoria generale della lessicografia: l'impostazione di Herbert Ernst Wiegand*, *SILTA*, XXXVI, 1.
21. Antonia Rubino, (a cura di), 2007, *Lingua, identità e comunicazione in contesti anglofoni e italiani*, *SILTA*, XXXVI, 3.
22. Giovanni Iamartino, (a cura di), 2008, *Linguistica inglese: studi recenti e prospettive di ricerca*, *SILTA*, XXXVII, 1.
23. Marina Chini, (a cura di), 2009, *Plurilinguismo e immigrazione nella società italiana. Repertori, usi linguistici e fenomeni di contatto*, *SILTA*, XXXVIII, 1.
24. Lucyna Gebert e Maria Załęska (a cura di), 2009, *Linguistica polacca in Italia. Linguistica italiana in Polonia*, *SILTA*, XXXVIII, 3.
25. Vincenzo Orioles (a cura di), 2010, *Il metalinguaggio. Temi e costrutti*, *SILTA*, XXXIX, 1.
26. Gaetano Berruto (a cura di), 2011, *La variazione: un terreno d'incontro fra sociolinguistica e teoria linguistica*, *SILTA*, XL, 2.
27. Margarita Borreguero Zuloaga e Luis Luque Toro (a cura di), 2011, *Linguistica italiana in Spagna, linguistica spagnola in Italia*, *SILTA*, XL, 3.
28. Marina Benedetti (a cura di), 2012, *Lingue e Grammatiche. Contatti, divergenze, confronti*, *SILTA*, XLI, 3.
29. Laura Mariottini, Franca Orletti (a cura di), 2013, *Lingua, interazione, mediazione. La migrazione latinoamericana in Italia*, *SILTA*, XLII, 3.
30. Carla Bagna (a cura di), 2014, *L'industria delle lingue: prospettive dal mondo delle imprese e dalla ricerca linguistica*, *SILTA*, XLIII, 2.
31. Vincenzo Orioles (a cura di), 2015, *Il modello di Žarko Muljačić*, *SILTA*, XLIV, 1.
32. Franca Orletti, Anna Cardinaletti, Francesca M. Dovetto (a cura di), 2015, *Il ruolo del linguista tra linguistica clinica e linguistica medica*, *SILTA*, XLIV, 3.
33. Bruno Moretti, Elena Maria Pandolfi, Matteo Casoni, Sabine Christopher (a cura di), 2016, *L'italiano in Svizzera*, *SILTA*, XLV, 3.

Finito di stampare nel mese di Novembre 2017  
presso le Industrie Grafiche della Pacini Editore s.r.l.  
Via A. Gherardesca • 56121 Ospedaletto • Pisa  
Telefono 050 313011 • Telefax 050 3130300  
[www.pacineditore.it](http://www.pacineditore.it)

