UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

La ricorrenza del lessico di base in produzioni scritte in italiano L2 e L1

Vedder, I.; Benigno, V.

Publication date
2013
Document Version
Final published version
Published in
Linguistica e Filologia

Link to publication

Citation for published version (APA):
Vedder, I., & Benigno, V. (2013). La ricorrenza del lessico di base in produzioni scritte in italiano L2 e L1. *Linguistica e Filologia*, *33*, 59-84.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:07 Oct 2026

Ineke Vedder
(Università di Amsterdam)

Veronica Benigno
(Pearson Language Testing, London)

La ricorrenza del lessico di base in produzioni scritte di italiano L2 e L1

This paper discusses the results of an experiment designed to analyze lexical richness, operationalized in terms of lexical frequency, in the written production of a group of L2 learners and native speakers of Italian. The data are derived from the CALC study (Communicative Adequacy and Linguistic Complexity in L2 writing; *cfr. Kuiken / Vedder / Gilabert 2010; Vedder 2012), set up to investigate the relationship between communicative adequacy and linguistic complexity in L2 and L1 writing (Italian, Dutch, Spanish), in relation to the* Common European Framework of References *(CEFR; Council of Europe, 2001). The present study focuses on the data of the L2 learners and native speakers of Italian. The first research question discussed in the study concerns the relationship between lexical richness and the general proficiency level of L2 learners. The second question addresses lexical richness in the written texts of Dutch low-intermediate and intermediate learners of Italian, compared to that of native speakers of Italian and assessed by means of a lexical frequency profiling method (Laufer / Nation 1995).*
The participants are 39 Dutch university students of Italian, with a proficiency level ranging from A2 to B1, and 18 native speakers of Italian, enrolled in an Italian university. In the study the participants had to carry out two writing tasks, consisting of a short essay in which an opinion had to be reported about factual information. To establish the general language proficiency in L2 a C-test and the DIALANG placement test were administered. Results show that although a relationship between lexical richness and language proficiency in L2 could not be demonstrated, there appeared to be significant differences in lexical richness between the L2 and L1 writers. The paper concludes with a discussion of the development of lexical richness in Italian L2.

1. *Introduzione*

Oggetto del presente articolo è l'analisi della ricchezza lessicale di produzioni scritte di un gruppo di apprendenti olandesi di italiano L2 a confronto con un gruppo di controllo di parlanti nativi. Il contributo si inserisce nell'ambito dello studio denominato CALC (*Communicative*

Linguistica e Filologia 33 (2013): pp. 59-84.

Adequacy and Linguistic Complexity in L2 writing; cfr. Kuiken / Vedder / Gilabert 2010; Vedder 2012), il cui obiettivo principale è indagare la relazione esistente tra adeguatezza comunicativa e complessità linguistica in produzioni scritte di apprendenti di lingue seconde di livello A2-B1 (italiano L2, olandese L2, spagnolo L2), sulla base del Quadro Comune Europeo di Riferimento (QCER; Consiglio d'Europa 2002). Il secondo obiettivo del CALC è descrivere l'interlingua e il ruolo della competenza linguistica nella L2, tramite l'analisi di particolari strutture linguistiche, ad esempio l'impiego da parte degli apprendenti di frasi subordinate e coordinate, la complessità del sintagma nominale, e l'uso delle collocazioni e di particolari strutture lessicali. Infine, il CALC si propone di esplorare la dimensione pedagogica della scrittura accademica nella L2, in relazione alle scale e ai descrittori del QCER. Per un approfondimento rimandiamo al paragrafo 2.

Il nostro studio si pone due obiettivi di ricerca fondamentali. In primo luogo, misurare la ricchezza lessicale delle produzioni scritte di un gruppo di apprendenti di italiano L2, stabilendo la correlazione esistente tra utilizzo del lessico di base in termini di frequenza e competenza linguistica degli apprendenti. In secondo luogo, mettere a confronto la ricchezza lessicale delle produzioni degli apprendenti e delle produzioni del gruppo di controllo dei parlanti nativi. L'analisi si concentrerà sulla distribuzione del lessico in tre fasce di frequenza contenenti i 3000 vocaboli più usati in italiano e create sulla base di un corpus di dati orali (Bardel / Lindqvist 2011; Bardel et al. 2012; Lindqvist 2010; Lindqvist et al. 2011). L'indagine analizzerà un campione di 78 saggi argomentativi prodotti da 39 studenti di italiano L2 di madrelingua olandese. Al fine di operare il confronto tra le produzioni scritte in italiano L1 e L2, i risultati dei non nativi saranno confrontati con quelli ottenuti dall'esame di 36 testi scritti da un gruppo di controllo composto da 18 parlanti nativi.

Nella prima parte dell'articolo (paragrafi 2-6), sarà fornito un breve quadro teorico dei meccanismi in gioco nell'acquisizione lessicale di una lingua seconda. Offriremo una definizione di due concetti che sono centrali nel nostro studio, quello di ricchezza lessicale e di lessico di base, inteso come nucleo essenziale del lessico di una lingua, il cui apprendimento è di primaria importanza tanto nella lingua straniera quanto nella lingua materna. Nel paragrafo 6 saranno chiariti gli obiettivi dell'analisi e la metodologia adottata. Descriveremo altresì il corpus dei

dati su cui si basa lo studio: la ricerca CALC, il profilo degli apprendenti di lingua seconda e del gruppo di controllo dei nativi, i *task* di scrittura utilizzati per la raccolta dei dati e i test adottati per la valutazione della competenza linguistica nella L2 (C-test e DIALANG). Nella sezione finale (paragrafi 7-8), saranno riportati e discussi i risultati dell'analisi condotta sui testi prodotti dai nativi e dai non nativi e saranno ricavate delle implicazioni teoriche inerenti lo sviluppo della competenza lessicale in L2[1].

2. *Il progetto CALC*

Le produzioni degli apprendenti analizzate nel nostro studio provengono da un corpus più ampio, il corpus CALC *(Communicative Adequacy and Linguistic Complexity in L2 writing*) raccolto con l'obiettivo di indagare, attraverso un'analisi cross-linguistica, lo sviluppo della competenza linguistica scritta di tre gruppi di studenti universitari con un livello di padronanza linguistica (rispettivamente in italiano L2, olandese L2, e spagnolo L2) compreso tra i livelli A2 e B1 del *Quadro Comune Europeo di Riferimento per le lingue* (QCER; Consiglio d'Europa 2001)[2]. Come spiegato nell'introduzione, l'obiettivo della ricerca CALC è indagare la relazione esistente tra adeguatezza comunicativa e complessità linguistica in produzioni scritte di apprendenti di lingue seconde. Una seconda finalità del CALC è comprendere il ruolo del livello di competenza linguistica in L2 tramite l'indagine di alcuni tratti sintattici specifici: ad esempio, lo sviluppo della complessità sintattica, in termini di occorrenza di strutture subordinate e coordinate, e la complessità del sintagma nominale. Infine, il CALC si propone di esplorare la dimensione pedagogica della scrittura accademica in L2, in relazione alle scale e ai descrittori del QCER (Kuiken / Vedder / Gilabert 2010; Vedder 2012).

[1] Desideriamo ringraziare, innanzitutto, le colleghe Camilla Bardel e Anna Gudmundson (Università di Stoccolma) per l'assistenza fornita per l'analisi elettronica dei dati. Inoltre ringraziamo Margriet van de Leur (Università di Leida) per i preziosi suggerimenti e l'aiuto che ci ha dato e Tanja Speek (Università di Wageningen) per l'analisi statistica.

[2] Il livello di competenza linguistica in italiano L2, olandese L2 e spagnolo L2 degli informanti del CALC è stato stabilito attraverso DIALANG (http://www.lancs.ac.uk/researchenterprise/dialang/about).

Nel presente contributo discutiamo i dati ricavati dall'analisi delle produzioni di 39 apprendenti di italiano L2 e di 18 informanti di italiano L1. Agli apprendenti di italiano L2, studenti universitari di madrelingua olandese, sono stati somministrati due test volti a misurare il livello di competenza linguistica generale di partenza: il C-test e il DIALANG.

Il C-test è una prova di livello simile al *cloze* consistente in 4-6 testi (semi-)autentici in cui vengono cancellate parti di parole: la metà (o la metà più una lettera) di ogni seconda parola del testo, per esempio *capi_ _ _ _* per *capitolo*. Compito degli apprendenti è riscostruire le parti mancanti delle parole (il numero di cancellazioni ammonta di solito a 100)[3]. Il DIALANG è un software che permette all'apprendente di verificare la propria competenza in L2. Nello studio originario del CALC, i risultati degli informanti non nativi nel DIALANG sono stati utilizzati per avere un'indicazione globale del livello di competenza linguistica di partenza degli apprendenti in riferimento ai descrittori del QCER[4]. Nel presente studio, invece, si è deciso di basare l'analisi soltanto sui risultati degli informanti nel C-test. Il motivo dell'esclusione dei risultati ottenuti nel DIALANG è che quest'ultimo è risultato molto difficile per la maggior parte degli informanti, dunque non utile per operare una differenziazione netta tra gli apprendenti. Sono stati esclusi dall'analisi anche i risultati ottenuti nel C-test dagli studenti madrelingua perché molto alti (94-99 risposte corrette su 100) rispetto a quelli deipartecipanti non nativi (i risultati variano da 46 a 93 risposte corrette)[5].

A tutti gli informanti sono stati somministrati due compiti di scrittura. Nel primo compito è stato chiesto loro di indicare quale organizzazione no-profit, tra una lista di possibili candidate dovesse essere finanziata dall'università con una somma di 55.000 euro. I partecipanti dovevano motivare la propria scelta in un testo di circa 150 parole[6], fornendo almeno tre argomentazioni valide. Nel secondo compito si trattava di

[3] Per il C-test somministrato agli informanti dello studio e i possibili vantaggi del C-test rispetto al *cloze* si vedano Berns / Vedder (2007), Dörnyei / Katona (1992), Kuiken / Vedder / Gilabert 2010, Vedder (2012).

[4] Per ulteriori informazioni rimandiamo il lettore al sito del DIALANG: http://www.lancs.ac.uk/researchenterprise/dialang/about.

[5] Il punteggio medio ottenuto nel C-test in italiano L2 è 69,74, con una mediana di 71 e una moda di 79.

[6] Anche se tutti i partecipanti avevano ricevuto la consegna di scrivere un testo di circa 150 parole, si è riscontrata una certa variabilità nella lunghezza dei testi.

decidere, tra tre articoli di giornale proposti, quale meritasse di apparire come articolo di fondo, e di motivarne la scelta. Gli informanti di madrelingua italiana, studenti universitari della stessa età e provenienti da un curriculum di studi simile, sono stati sottoposti agli stessi test[7]. I partecipanti avevano a disposizione 35 minuti di tempo per ciascuno dei due compiti; l'uso del dizionario non era ammesso.

3. *Apprendimento del lessico nella L2*

Negli ultimi anni si è registrato un crescente interesse per l'apprendimento del lessico nella L2, come testimoniano le numerose ricerche finora condotte (Bogaards / Laufer 2004; Daller et al. 2007; Horst / Collins 2006; McCarthy / Jarvis, 2007; Milton et al. 2008; Nation 2001, 2006). Nonostante la maggior parte delle ricerche sull'apprendimento del lessico si interessi allo studio dell'inglese L2, negli ultimi anni sono state svolte diverse ricerche anche sull'apprendimento del lessico in francese L2 (Cobb / Horst 2004; Lindqvist 2010; Ovtcharov et al. 2006; Tidball / Treffers-Daller 2007; Wilks / Meara 2007; Treffers-Daller et al. 2008) e in altre lingue seconde (cfr. Eyckmans et al. 2007 per l'olandese L2; Lengyel et al. 2007 per il tedesco L2; Lorenzo-Dus 2007 per lo spagnolo L2). Nell'ambito dell'italiano L2, ricordiamo, tra gli studi piu importanti, le ricerche di Bernini (2004, 2006), Bozzone Costa (2002), Carloni / Vedovelli (2005), Ferrreri (2005), Lo Duca (2007), Spreafico (2003, 2005, 2008). Tali studi concordano nel ritenere che le parole non siano apprese come unità isolate, ma in un complesso intreccio di relazioni semantiche, in linea con i numerosi studi sull'apprendimento del lessico nella lingua materna (Aitchinson 2003; Hoey 2005; Lo Cascio 2007; Nation 2001).

Come è stato dimostrato da questi studi, l'apprendimento della lingua seconda si caratterizza da regolarità intrinseche e segue diverse fasi di apprendimento, in cui l'apprendente combina progressivamente principi pragmatici, fonetici, semantici, morfologici e sintattici, man mano che la sua produzione linguistica diventa più complessa. In base al noto modello funzionalista della Varietà Basica (*Basic Variety*) di Klein /

[7] Per una descrizione più dettagliata dei risultati, la metodologia adottata e le misure utilizzate per la valutazione delle produzioni scritte, rimandiamo a Kuiken / Vedder / Gilabert (2010) e Vedder (2012).

Perdue (1997), l'apprendimento del lessico nella L2 costituisce un percorso evolutivo e graduale, che va dalla pragmatica alla sintassi. Nelle fasi iniziali dell'apprendimento l'*output* si caratterizza per la presenza di brevi enunciati, prodotti ancora senza una vera e propria riflessione grammaticale, e le parole seguono un ordine pragmatico e discorsivo, più che sintattico. Le prime produzioni degli apprendenti sono spesso costituite da formule fisse (*chunks*) non produttive, perché le parole che le compongono non vengono utilizzate in altre combinazioni. Soltanto nelle fasi successive, quando queste formule fisse vengono decomposte ed analizzate, l'apprendente riesce ad assegnare le parole ad una classe morfologica (Bernini 2006).

Gli studi finora condotti coprono numerosi aspetti dell'apprendimento del lessico, quali l'importanza per l'apprendimento della frequenza delle parole nell'*input*, il ruolo del transfer dalla L1 o da altre lingue conosciute dall'apprendente, la differenza tra apprendimento implicito e esplicito, la correlazione tra conoscenza lessicale e livello di competenza linguistica generale, e la quantità e la qualità ('profondità') della conoscenza lessicale degli apprendenti (Bardel / Lindqvist 2011; Bardel et al. 2012; Laufer / Nation 1995; Lindqvist 2010; Lindqvist et al. 2011; Read 2000).

Stabilire quando le parole di una L2 possano considerarsi acquisite dagli apprendenti a tutti gli effetti non è facile (Carloni / Vedovelli 2005; Ferreri 2005). Ogni parola è portatrice di molteplici proprietà fonologiche, grafiche, semantiche, morfologiche, sintattiche e pragmatiche, ed è improbabile che tali proprietà vengano acquisite tutte contemporaneamente. Piuttosto che interpretare il concetto di 'conoscenza' di una parola in termini assoluti, sarebbe più corretto parlare di molti livelli o gradi di conoscenza possibile e raffigurare il processo di ampliamento del vocabolario come un processo ricorsivo in cui gli apprendenti scoprono proprietà e tratti nuovi delle parole della L2 (Lo Duca 2007: 135-136). Per essere in grado di descrivere lo sviluppo lessicale nella L2 è inoltre necessario affidarsi alle produzioni orali o scritte degli apprendenti; tuttavia esse non rappresentano la globalità del vocabolario conosciuto dagli apprendenti, ma semmai solo la parte 'controllata' del lessico, sollecitata dal compito e impiegata dall'apprendente durante l'esecuzione del compito. In altri tipi di compiti, con modalità di elicitazione dei dati diversa (intervista, telefonata, conversazione tra amici, compito in classe) è possibile che si ottengano risultati piuttosto diversi (Spreafico 2003: 110).

4. *Il lessico di base dell'italiano*

Il lessico di base rappresenta la parte più interna del lessico e costituisce il nucleo di parole essenziali per lo svolgimento degli atti comunicativi di base. Come è stato descritto nei paragrafi precedenti, le parole ricorrenti più di frequente nell'*input* fornito dai nativi, o le parole dotate di una forte funzionalità comunicativa giocano un ruolo fondamentale nell'apprendimento della lingua seconda. Esiste infatti una correlazione positiva tra fattori che favoriscono l'apprendimento lessicale nella L2 (per esempio brevità della parola e facile pronunciabilità, concretezza e trasparenza del significato, facilità di memorizzazione) e caratteristiche del lessico di base (Benigno 2007).

Per individuare il lessico di base dell'italiano si rivelano di grande utilità i dizionari di frequenza quali il *Lessico Italiano di Frequenza* (LIF 1972) e il *Lessico dell'Italiano Parlato* (LIP 1993), nonché la lista di frequenza del *Vocabolario di Base* (VdB 1980). Il LIF si basa su un corpus di testi scritti (testi di teatro, narrativa, cinema, periodici sussidiari), mentre il LIP è basato su un corpus di testi parlati (conversazioni, lezioni, dibattiti, interviste), raccolti a Milano, Firenze, Roma e Napoli. Diversamente dal LIF e dal LIP, il VdB, elaborato da De Mauro nel 1980, combina il criterio statistico della frequenza delle parole con un criterio più soggettivo, basato sulla conoscenza di parole meno frequenti ma 'disponibili' da parte di parlanti nativi (alunni di terza media e adulti con licenza di scuola media superiore)[8]. Il VdB comprende circa 7000 parole, suddivise in tre fasce: il Vocabolario Fondamentale, il Vocabolario di Alto Uso e il Vocabolario di Alta Disponibilità. Le prime due fasce sono state estratte dai primi 5001 lemmi del LIF, mentre la terza fascia è stata ricavata da De Mauro e i suoi collaboratori tramite interviste ai parlanti nativi. La fascia del Vocabolario Fondamentale, che comprende circa 2750 parole, è costituito dalle parole più frequenti dell'italiano, comprese da circa il 79% della popolazione italiana ('abitudine', 'arrabbiarsi'). Il Vocabolario di Alto Uso contiene le parole spesso usate dai parlanti in ogni situazione quotidiana. Si tratta di circa 2750 parole ('accap-

[8] Nelle edizioni successive, il VdB, ora incluso nel GRADIT (*Grande Dizionario Italiano dell'Uso*; De Mauro 1999), ha subito un numero di variazioni, soprattutto nella fascia delle parole di alta disponibilità. Recenti ricerche sul lessico del VdB hanno comunque dimostrato la stabilità del VdB, in quanto più che nuovi lemmi, si registrano cambiamenti interni di fascia (Bisconti 2012: 12).

patoio', 'abbassare'). Il Vocabolario di Alta Disponibilità (AD) comprende invece circa 2.300 parole ('abbigliamento', 'abbottonare') che "...può accaderci di non dire né tanto meno di scrivere mai o quasi mai, ma che sono legate a oggetti, fatti, esperienze ben noti a tutte le persone adulte nella vita quotidiana..." (De Mauro 2003: 162).

Uno strumento elettronico che consenta l'analisi della ricorrenza del lessico di base in produzioni di parlanti di italiano L2 e L1 non esiste. Nel presente lavoro abbiamo adoperato come strumento d'analisi lessicale per l'esame del lessico di base in italiano lo strumento sviluppato da Bardel / Gudmundson / Lindqvist (Lindqvist 2010; Lindqvist / Bardel / Gudmundson 2011; Bardel / Lindqvist 2012). Il software creato da Bardel / Gudmundson / Lindqvist distingue tre fasce di frequenza, ciascuna consistente di 1000 parole. Inoltre vi è una quarta categoria di parole fuori lista e una categoria combinata, che unisce la terza fascia (contenente i vocaboli aventi un rango di frequenza che va da 2000 a 3000) e la categoria delle parole fuori lista. Il programma si basa su due database sviluppati alla Karl Franzens Universität di Graz, la versione lemmatizzata del LIP ed il corpus per l'italiano parlato C-Oral-rom (*Integrated reference corpora for spoken Romance languages*), e misura la proporzione di parole di alta frequenza rispetto a quelle di bassa frequenza. Per una descrizione più dettagliata del software e la metodologia adottata nello studio rimandiamo il lettore al paragrafo 6.

5. *Richezza lessicale in L2*

Nella letteratura si trovano diverse definizioni del concetto di ricchezza lessicale. Read (2000), ad esempio, distingue vari aspetti della ricchezza lessicale: varietà, sofisticatezza, densità e numero di errori lessicali. Altri definiscono invece la ricchezza lessicale in termini di frequenza delle parole nell'input (Cobb / Horst 2004; Laufer / Nation 1995). Nel nostro studio seguiamo un approccio simile e intendiamo il concetto di ricchezza lessicale in termini di frequenza delle parole, ovvero in termini della loro ricorrenza nelle produzioni degli informanti nativi e non nativi (Lindqvist / Bardel / Gudmundson 2011).

Molti studi hanno rilevato l'importanza della frequenza delle parole nell'*input* per l'apprendimento del lessico nella L2 (Cobb / Horst 2004;

Ellis, 1997, 2002; Lindqvist et al. 2011; Milton 2007; Nation / Waring 1997; Nation 2001, 2006). Tali studi ritengono che l'apprendimento del lessico in L2 ai livelli iniziali e intermedi si caratterizzi per l'espansione prevalentemente quantitativa del vocabolario dell'apprendente. Con l'innalzarsi del livello di competenza linguistica nella L2 l'incremento quantitativo del numero di vobaboli rallenta, invece. In questa seconda fase dell'apprendimento del lessico si verifica una crescita anche qualitativa del lessico, in quanto l'apprendente impara ad usare le parole che già conosce in altri contesti scoprendone delle proprietà e dei significati nuovi (Meara 1996). Aumenta anche la facilità e la velocità con cui l'apprendente riesce ad accedere alle parole immagazzinate nel lessico mentale (Nation 2006). In uno studio sullo sviluppo del lessico di base in italiano L2, Spreafico (2003, 2005), servendosi di indici quantitativi tra i quali l'Indice di Guiraud (1954)[9] per analizzare le produzioni orali di un apprendente non nativo dell'italiano, descrive l'andamento del numero di parole conosciute ed utilizzate dall'informante nei diversi stadi dello sviluppo linguistico. I risultati principali che emergono dallo studio indicano che nelle produzioni orali dell'informante si osserva una crescita non lineare del numero di lemmi nel corso delle registrazioni, e una riduzione dell'incidenza percentuale delle parole contenuto rispetto a quella delle parole funzione. In base ai risultati ottenuti nello studio Spreafico suggerisce di adoperare, accanto a degli indici lessicali quantitativi come l'Indice di Guiraud, altre misure qualitative come il *Lexical Frequency Profile* (LFP) di Laufer / Nation (1995)[10].

Il *Lexical Frequency Profile* misura la proporzione del lessico di alta frequenza rispetto a quello di bassa frequenza nelle produzioni L2, suddividendo le parole in tre categorie (*frequency bands*), in base alla loro ricorrenza nell'*input*. La prima fascia del LFP contiene le 1000 parole più frequenti della L2, la seconda fascia le successive 1000 parole. La terza fascia consiste di circa 570 parole appartenenti all'*Academic Wordlist,* una lista estratta da un corpus di testi accademici (Coxhead 2000). Infine, le parole che non figurano in nessuna di queste tre fasce

[9] L'Indice di Guiraud valuta la ricchezza lessicale di campioni di testi. Similmente all'indice D (Malvern et al 2004; Malvern / Richards 2009) ha il vantaggio rispetto alla TTR (*Type* / *Token Ratio*, ovvero il rapporto tra il numero di parole totale e il numero di parole identiche in un testo) di essere meno sensibile alla lunghezza del testo.

[10] Per il risultati dell'esame della ricchezza lessicale nei testi degli informanti del CALC misurata con l'Indice di Guiraud rimandiamo a Kuiken / Vedder / Gilabert (2010) e Vedder (2012).

appartengono alla categoria delle parole fuori lista (*off list*). Una versione francese del LFP è stata sviluppata da Cobb / Horst (2004; si veda anche Lindqvist 2010). Per quanto concerne la versione italiana del LFP, Bardel / Gudmundson / Lindqvist (Lindqvist 2010; Lindqvist / Bardel / Gudmundson 2011; Bardel / Lindqvist 2012) hanno sviluppato uno strumento di analisi lessicale ispirato al LFP di cui, come già spiegato nel paragrafo 4, ci siamo servite nello studio discusso in questo articolo. Per la descrizione del *software* e la metodologia che abbiamo seguito rimandiamo alla sezione 6.

6. *Focus dello studio e metodologia*

L'obiettivo della nostra ricerca è verificare se le parole più frequenti siano apprese prima delle altre e se quindi la presenza di parole poco diffuse nell'*input* degli informanti sia indice di una maggiore competenza linguistica (Cobb / Horst 2004; Lindqvist et al. 2011; Meara 1996; Ovtcharov et al. 2006; Schmitt / Mc Carthy 1997; Spreafico 2003, 2005). Se così fosse, le produzioni degli apprendenti più avanzati si distinguerebbero da quelle degli apprendenti con un livello di competenza linguistica più basso per la presenza di un maggior numero di parole meno frequenti. Riconosciamo che tale assunto non può comunque avere valore assoluto, visto che il bagaglio lessicale di ogni parlante, persino del parlante non nativo ad un livello di apprendimento iniziale, include anche termini meno frequenti ma 'disponibili' (De Mauro 1980) o persino appartenenti al linguaggio specialistico perché appresi in base alle proprie esperienze o ai propri interessi personali e professionali (Bardel / Lindqvist 2012). Da qui l'importanza, nelle ricerche volte a indagare il vocabolario essenziale di una lingua, di tenere conto, accanto al criterio della frequenza, anche del criterio della disponibilità (Benigno 2012)[11].

[11] Benigno (2012) estrae da un corpus di testi reperiti nel web di più di un miliardo di parole, un campione di collocazioni appartenenti all'area semantica delle relazioni sociali e verifica quali combinazioni sono ritenute essenziali da parte di 90 parlanti nativi francesi. I risultati mostrano che la correlazione tra carattere fondamentale delle associazioni (così come percepito dal gruppo dei parlanti nativi) e loro frequenza è lineare ma non sistematica. Esistono infatti delle eccezioni alla linearità della correlazione, per cui combinazioni molto frequenti sono ignorate dai nativi, e viceversa, combinazioni poco frequenti sono contrassegnate come essenziali dalla maggior parte del

Lo strumento adoperato per misurare l'uso del lessico di base è stato il *profiler* lessicale, sviluppato da Bardel / Gudmundson (Lindqvist 2010; Lindqvist / Bardel / Gudmundson 2011; Bardel / Gudmundson / Lindqvist 2012), ispirato al LFP (Laufer / Nation 1995). Il *profiler* classifica le parole di un corpus di testi in base a dei criteri prestabiliti, vale a dire le occorrenze delle parole appartenenti alle fasce di frequenza create da Bardel e Gudmundson con cui i nostri dati sono stati confrontati. Le tre fasce di frequenza (*Band* 1, *Band* 2, *Band* 3) si basano su una combinazione di due corpora: il corpus delle versioni lemmatizzate del LIP, creato dalla Karl Franzens Universität di Graz, il corpus C-Oral-Rom (*Integrated reference corpora for spoken Romance languages*). La prima fascia, come spiegato già nel paragrafo 4, contiene le 1000 parole più frequenti dei due corpora di riferimento, la seconda fascia comprende le 1000 parole successive e la terza fascia le parole con una frequenza da 2000 a 3000. Come nel LFP di Laufer / Nation (1995) è stata creata poi una quarta categoria delle parole fuori lista (*off-list)*, che include le parole non appartenenti alle fasce 1-3; infine si è aggiunto una quinta categoria combinata che comprende le parole della terza fascia e le parole fuori lista[12].

Poiché le tre fasce di frequenza si basano sui lemmi e non sulle famiglie di parole, è stato necessario lemmatizzare i nostri dati con il programma *TreeTagger*[13]. Siccome *TreeTagger* riconosce soltanto le parole con una grafia corretta, sono state escluse dall'analisi le parole ritenute incomprensibili, mentre le parole con 'piccoli' errori di ortografia ('organizazione' invece di 'organizzazione'; 'scelgliere' invece di 'scegliere') sono state corrette e reinserite nel database, seguendo il metodo adottato anche da Laufer / Nation (1995: 315). Sono state eliminate inoltre le parole 'inventate', come risultato di un processo di transfer dalla L1 o da altre lingue seconde conosciute dagli apprendenti. Esempi di queste creazioni lessicali che *TreeTagger* ovviamente non riconosce,

campione intervistato. A un'analisi attenta delle combinazioni in oggetto, si è poi osservato che la fissità degli elementi che compongo l'unità fraseologica determina la scelta dei parlanti nativi e contribuisce, insieme alla frequenza, a definirne il carattere essenziale.

12 Per le tre fasce di frequenza è stato creato un *Perl script.* Poi, le fasce di frequenza sono state memorizzate in un database SQL (*Structured Query Language*).

13 Per ulteriori informazioni, si veda il sito http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

sono ‘developpare’ (‘sviluppare’); ‘adressare’ (‘rivolgere’), ‘envirimento’ (‘ambiente’).

Dopo che i testi degli studenti nativi e non nativi sono stati lemmatizzati con *TreeTagger*, le frequenze di ogni lemma sono state calcolate manualmente. I dati di tutti gli informanti sono stati confrontati con il database, al fine di classificare i lemmi usati da ciascuno di loro nelle varie fasce di frequenza. Per ogni fascia di frequenza sono poi stati calcolati le percentuali e i numeri totali dei lemmi adoperati dai due gruppi degli apprendenti L2 e gli studenti di madrelingua italiana.

7. *Risultati*

In questa sezione riportiamo i risultati ottenuti dall’analisi. Allo scopo di verificare se esiste una correlazione tra ricorrenza del lessico di base frequente in italiano L2 e livello di competenza linguistica generale degli apprendenti di italiano L2 sono stati calcolati i coefficienti di correlazione di Pearson[14] tra le due variabili: le percentuali delle fasce di frequenza (*Band* 1, *Band* 2, *Band* 3*, Off-list*, e infine la categoria combinata del *Band* 3 e delle parole *Off-list*) e i risultati ottenuti dagli apprendenti nel test di livello, il C-test. Per confrontare gli indici di ricchezza lessicale e le produzioni degli studenti L2 e degli informanti nativi è stato invece applicato un Test T[15].

I paragrafi 7.1 e 7.2 riportano le percentuali con cui i lemmi appartenenti alle varie categorie di frequenza compaiono nelle produzioni del gruppo degli apprendenti di italiano L2 e nel gruppo dei parlanti nativi nei due compiti svolti (Task 1 e Task 2), insieme al numero totale dei lemmi adoperati. Nel paragrafo 7.3 presentiamo i risultati inerenti alla prima domanda di ricerca dello studio, ovvero la relazione tra frequenza delle parole ricorrenti nelle produzioni degli apprendenti di italiano L2 e

[14] Il coefficiente di Pearson rappresenta il grado di correlazione lineare tra due variabili X e Y. Assume valori compresi tra –1 e +1. Una correlazione di 0 indica l’assenza di una correlazione lineare tra X e Y; –1 indica una perfetta correlazione negativa tra X e Y; +1 indica l’esistenza di una perfetta correlazione positiva tra le due variabili. Il simbolo *r* è usato per indicare la misura in cui due variabili X e Y risultano correlate. Il simbolo *p* indica il grado di probabilità statistica che il risultato osservato si verifichi.

[15] Il Test t permette di stabilire se tra due variabili vi siano delle differenze significative.

livello di competenza linguistica generale, misurato attraverso il C-test. Il paragrafo 7.4 contiene i risultati concernenti la seconda domanda che ha guidato la nostra ricerca, il confronto tra ricorrenza del lessico di base più frequente in italiano L2 e italiano L1.

7.1 *La ricorrenza del lessico di base in italiano L2*

La Tabella 1 contiene le percentuali medie insieme alla deviazione standard[16] dei lemmi appartenenti alle varie fasce di frequenza (*Band* 1, *Band* 2, *Band* 3, *Off-list*, *Band* 3 + *Off-list*), che compaiono nelle produzioni degli apprendenti di italiano L2 nei due compiti (Task 1 e Task 2).

	Band 1		Band 2		Band 3		Off-list		Band 3 + off-list	
	media (%)	σ	media (%)	σ	media (%)	σ	media (%)	σ	media (%)	σ
Task 1	85,52	3,88	6,91	3,09	1,33	1,31	6,24	3,16	7,56	3,18
Task 2	89,10	3,38	3,50	1,75	3,40	2,41	3,99	2,36	7,39	2,57

σ = deviazione standard

Tabella 1. Ricorrenza del lessico di base in italiano L2 (n = 39)

Dalla tabella si evince che nel primo compito (Task 1) le parole rientranti nella prima fascia di frequenza (Band 1) sono utilizzate dal gruppo degli apprendenti L2 con una percentuale media dell'85,52% (deviazione standard 3,88). I vocaboli che appartengono alla seconda fascia (Band 2) ricorrono con una frequenza molto più bassa (6,91%; deviazione standard 3,09). Ancora più rari i lemmi che fanno parte della terza fascia, la Band 3 (media: 1,33%; deviazione standard 1,31). I lemmi rientranti nella categoria delle parole fuori lista (*Off-list*), infine, figurano nelle produzioni degli apprendenti L2 con una media del 6,24% (deviazione standard 3,16), mentre quelli della categoria combinata (Band 3 + *Off-list*) con una media del 7,56% (devazione standard 3,18).

I risultati ottenuti dagli informanti nel Task 2 sono piuttosto simili a

[16] La deviazione standard è un indice di dispersione volto a calcolare il grado di dispersione tra le percentuali dei singoli apprendenti rispetto alla media.

quelli del Task 1. Tuttavia, nel Task 2 la percentuale di lemmi appartenenti alla prima fascia di frequenza (Band 1) risulta leggermente più alta di quella del Task 1 (media 89,1%; deviazione standard 3,38) e le percentuali dei vocaboli rientranti nelle fasce 2 e 3 e nella categoria fuori lista appaiono più vicine (Band 2: media 3,5%, deviazione standard 1,75; Band 3: media 3,4%, deviazione standard 2,41; *Off-list*: media 3,99%, deviazione standard 2,36). I lemmi appartenenti alla categoria combinata (Band 3 + *Off-list*) ricorrono con una media del 7,39% (deviazione standard 2,57), una percentuale che non si differenzia molto da quella rilevata nel Task 1.

La Tabella 2 qui di seguito riporta il numero totale dei lemmi ricorrenti nelle produzioni del gruppo degli apprendenti L2 (Task 1 e Task 2).

	minimo	massimo	media	σ
Task 1	54	247	140.21	36,71
Task 2	87	201	135,10	29,50

σ = deviazione standard

Tabella 2. Numero totale di lemmi in L2 (n = 39)

Come si ricava dalla tabella, la media del numero totale di lemmi adoperati dagli apprendenti nel corpus dei testi del Task 1 è 140,21, con un minimo di 54 lemmi e un massimo di 247. Queste differenze rilevate tra i singoli apprendenti si riflettono in una deviazione standard piuttosto elevata (36,71). La media del numero totale di lemmi contenuti nei testi del Task 2 è 135,1, con un numero minimo di 87 lemmi e un massimo di 201. Benché più bassa rispetto al Task 1, anche nel Task 2, come dimostrato dalla tabella, la deviazione standard risulta piuttosto elevata (29,5).

7.2 *La ricorrenza del lessico di base in italiano L1*

Nella Tabella 3 sono rappresentate le percentuali dei lemmi, appartenenti alle varie fasce di frequenza (Band 1, Band 2, Band 3, *Off-list*, Band 3 + *Off-list*), utilizzati nei due compiti (Task 1 e Task 2) dal gruppo di controllo degli informanti L1.

	Band 1		Band 2		Band 3		Off-list		Band 3 + Off-list	
	media (%)	σ	media (%)	σ	media (%)	σ	media (%)	σ	media (%)	σ
Task 1	78,69	3,34	7,06	1,97	2,91	1,42	11,34	3	14,25	3,11
Task 2	82,47	3,97	4,68	2,01	4,47	2,66	8,37	3,38	12,85	3,66

σ = deviazione standard

Tabella 3. Ricorrenza del lessico di base in italiano L1 (n = 18)

Nel Task 1, come emerge dalla tabella, i lemmi che fanno parte della prima fascia di frequenza (Band 1) ricorrono nelle produzioni del gruppo dei parlanti nativi con una media del 78,69% (deviazione standard 3,34). La percentuale dei lemmi appartenenti alla seconda e alla terza fascia risulta molto più bassa, così come osservato nei testi degli studenti L2 (Band 2: media 7,06%; deviazione standard 1,97; Band 3: media 2,91%, deviazione standard 1,42). Più frequenti rispetto all'impiego da parte degli studenti L2 appaiono i lemmi della categoria dei fuori lista e della categoria combinata (Off-list: media 11,34%, deviazione standard 3; Band 3 + Off-list: media 14,25%, deviazione standard 3,11).

I risultati dei parlanti L1 nel Task 2 non si differenziano molto dai risultati ottenuti nel Task 1. I lemmi che appartengono alla prima fascia di frequenza (Band 1) vengono utilizzati con una media dell'82,47% (deviazione standard 3,97). La presenza di lemmi della seconda fascia (Band 2) appare di nuovo molto bassa (media 4,68%, deviazione standard 2,01), di poco superiore ai lemmi appartenenti alla terza fascia (media 4,47%; deviazione standard 2,66). I lemmi fuori lista ricorrono nelle produzioni degli studenti nativi con una percentuale dell'8,37% (deviazione standard 3,38), mentre i lemmi della categoria combinata (Band 3 + *Off-list*) sono presenti nei testi degli studenti L1 con una percentuale media del 12,85% (deviazione standard 3,66).

	minimo	massimo	media	σ
Task 1	84	270	147	48,19
Task 2	50	189	125,06	35,59

σ = deviazione standard

Tabella 4. Numero totale di lemmi in L1 (n = 18)

Anche i testi degli studenti madrelingua, come si evince dalla tabella, differiscono notevolmente tra di loro per quel che riguarda il numero totale di lemmi adoperati, similarmente a quanto osservato nei testi degli apprendenti L2. La media del numero totale di lemmi usati nei testi del Task 1 è 147, con un minimo di 84 lemmi e un massimo di 270. Tale differenza si riflette in una deviazione standard molto elevata, che supera quella degli apprendenti L2 (48,19). I testi del Task 2 contengono mediamente 125,06 lemmi, con un numero minimo di 50 lemmi e un massimo di 189. La deviazione standard nel Task 2 è di 35,59[17].

7.3 *Ricorrenza del lessico di base e competenza linguistica in L2*

Per esaminare la relazione tra lessico di base, suddiviso nelle varie fasce di frequenza da una parte, e livello di competenza linguistica globale degli apprendenti L2 in base al C-test dall'altra, sono stati calcolati i coefficienti di correlazione di Pearson tra punteggio ottenuto nel C-test e percentuali delle fasce di frequenza. Le correlazioni per i due compiti, Task 1 e Task 2, sono raffigurate nelle tabelle 5 e 6.

Lemmi	r (Correlation coefficient)	p-value
Band 1	0.0239	0.885
Band 2	–0.146	0.375
Band 3	0.0788	0.633
Off-list	0.0811	0.623
Band 3 + off-list	0.113	0.494

Livello di significatività: $p<0.05$

Tabella 5. Correlazioni tra C-test e fasce di frequenza in italiano L2 nel Task 1 (n = 39)

[17] La lunghezza media dei testi degli studenti nativi e non nativi, come si evince dalle tabelle, è molto simile, ad eccezione dei testi del Task 2 prodotti dai nativi, che sono più brevi. Questa differenza potrebbe essere dovuta alla diversa somministrazione dei task in Olanda e in Italia: dei 39 apprendenti L2, metà ha iniziato con il Task 1 e l'altra metà con il Task 2, al fine di evitare l'effetto del *task sequencing*; al gruppo dei 18 studenti nativi, invece, è stato assegnato prima il Task 1 e poi il Task 2.

Lemmi	r (Correlation coefficient)	p-value
Band 1	0.00666	0.968
Band 2	-0.124	0.451
Band 3	-0.0770	0.641
Off-list	0.161	0.328
Band 3 + off-list	0.0758	0.646

Livello di significatività: $p<0.05$

Tabella 6. Correlazioni tra C-test e fasce di frequenza in italiano L2 nel Task 2 (n = 39)

Come si evince dalle tabelle 5 e 6, la relazione tra ricorrenza del lessico di base suddivisa nelle cinque fasce di frequenza e livello di competenza linguistica nella L2 misurato in base al C-test non risulta statisticamente significativa in nessun caso, né nel Task 1 né nel Task 2. Diversamente da quanto ipotizzato, il livello di competenza linguistica degli apprendenti di italiano L2 non risulta dunque correlato con l'uso dei lemmi appartenenti alle varie fasce di frequenza. In termini di ricchezza lessicale nella L2, ciò implica che dal corpus della nostra analisi non emerge nessuna correlazione significativa tra ricchezza lessicale, intesa come uso di lemmi appartenenti ad una particolare fascia di frequenza, e livello di padronanza linguistica nella L2, e ciò vale sia per gli apprendenti più avanzati che per quelli principianti.

7.4 *Confronto tra Italiano L2 e Italiano L1*

Per il confronto tra risultati degli apprendenti di italiano L2 e risultati degli studenti nativi è stato applicato un Test T. I risultati del Test T per il Task 1 sono riportati nella Tabella 7.

Dal confronto tra i due gruppi emerge che nel Task 1 gli apprendenti di italiano L2 usano più lemmi appartenenti alla prima fascia di frequenza (Band 1) rispetto ai parlanti nativi, conformemente a quanto avevamo ipotizzato. Dall'analisi statistica risulta che tale differenza è significativa: $t(55)= –6.332$, $p<0.001$. Quanto all'utilizzo dei lemmi che rientrano nella seconda fascia di frequenza (Band 2), i due gruppi non si differenziano

Lemmi	d.f. (degrees of freedom)	t-value	p-value
Band 1	55	-6,332	<0.001***
Band 2	55	0,178	0.860
Band 3	55	4,055	<0.001***
Off-list	55	5,669	<0.001***
Band 3 + off-list	55	7,297	<0.001***

Livello di significatività: p<0.05

Tabella 7. Confronto italiano L2 - italiano L1 (Task 1)

(t(55)= 0.178, p = 0.860). Osserviamo invece delle differenze significative tra i due gruppi, come ipotizzato, nel numero di lemmi appartenenti alla terza fascia di frequenza (Band 3): nelle produzioni degli studenti di madrelingua tale numero risulta più alto (t(55)= 4.055, p <0.001). Appare significativa anche la differenza nell'uso dei lemmi fuori lista (*Off-list*) da parte dei due gruppi, poiché la percentuale di lemmi *Off-list* presenti nei testi del gruppo di controllo supera quella che si registra nei testi del gruppo degli apprendenti (t(55)= 5.669, p <0.001). Questo vale anche per le occorrenze dei lemmi della categoria combinata, più numerosi nei testi degli studenti nativi (Band 3 + *Offlist*: t(55)= 7.297, p <0.001).

I risultati emersi dal Test T, volti a misurare la differenza esistente nel Task 2 tra uso dei lemmi suddivisi nelle varie fasce di frequenza da parte degli informanti nativi e da parte di quelli non nativi, sono raffigurati nella Tabella 8.

I risultati ricavati dal confronto tra gruppo degli apprendenti L2 e quello dei parlanti nativi nel Task 2 non si discostano molto da quelli ottenuti nel Task 1. Similmente a quanto osservato nel Task 1 si riscontrano delle differenze tra i due gruppi nella ricorrenza dei lemmi appartenenti alla prima fascia di frequenza (Band 1). Ad un'analisi statistica, tali differenze appaiono significative (t(55)= –6.391, p<0.001). Gli studenti non nativi, come accade nel Task 1 e conformemente alle nostre ipotesi di partenza, utilizzano più lemmi appartenenti alla prima fascia di frequenza (Band 1) rispetto al gruppo di controllo degli studenti nativi. I due gruppi si distinguono anche nel Task 2, in quanto il numero di lemmi appartenenti alla seconda fascia di frequenza è maggiore nei testi del

Lemmi	d.f. (degrees of freedom)	t-value	p-value
Band 1	55	-6,391	<0.001***
Band 2	55	2,211	0.031**
Band 3	55	1,482	0.144
Offlist	55	5,544	<0.001***
Band 3 + offlist	55	6,357	<0.001***

Livello di significatività: p<0.05

Tabella 8. Confronto italiano L2 - italiano L1 (Task 2)

gruppo di controllo che nel gruppo degli apprendenti L2 (Band 2: t(55)= 2.211, p <0.031). Benché i lemmi appartenenti alla terza fascia (Band 3) compaiano più spesso nei testi dei parlanti nativi, tale differenza non risulta significativa all'analisi statistica (t(55)= 1.482, p =0.144). Per quanto concerne l'uso dei lemmi fuori lista (*Off-list*), come nel Task 1 e in linea con le nostre ipotesi, osserviamo anche per il Task 2 delle differenze significative tra i due gruppi (t(55)=5.544, p <0.001), in quanto sono di nuovo i parlanti nativi ad adoperare più lemmi rispetto al gruppo degli apprendenti. Appare significativamente più alto nei testi degli studenti di madrelingua italiana anche il numero di lemmi rientranti nella fascia combinata (Band 3 + *Off-list*: t(55)= 6.357, p<0.001), un dato che era già emerso anche dall'analisi condotta sui testi del Task 1.

Le nostre ipotesi sulla ricorrenza del lessico di base e sul ruolo della ricchezza lessicale in italiano L2 e L1, in termini di una maggiore o minore frequenza dei lemmi presenti nelle produzioni degli informanti dei due gruppi, risultano dunque in gran parte confermate. In entrambi i compiti gli studenti non nativi usano più spesso vocaboli appartenenti alla fascia delle 1000 parole più frequenti (Band 1), mentre nei testi dei parlanti nativi figurano più spesso parole con una frequenza più bassa o parole non rientranti nel lessico di base (*Off-list*; Band 3 + *Off-list*). Per quanto riguarda le parole appartenenti alla seconda fascia (Band 2; ranghi 1000-2000) e alla terza fascia (Band 3; ranghi 2000-3000), i risultati nei due compiti non sono univoci: nel Task 1 i due gruppi non si differenziano nella seconda fascia (Band 2) e risulta più alto il numero di lemmi utilizzato dagli studenti L1 appartenenti alla terza fascia (Band

3). Nel Task 2, invece, i risultati dei due gruppi nell'impiego di lemmi del Band 2 sono simili, mentre l'uso di lemmi appartenenti alla terza fascia è maggiore nei testi dei parlanti L1.

7. *Conclusioni*

Il presente lavoro ha misurato la ricorrenza del lessico di base, da un lato, in un corpus di 78 testi argomentativi scritti da un gruppo di 39 apprendenti di madrelingua olandese di italiano L2, la cui conoscenza di italiano L2 si situa tra i livelli A2-B1 del QCER); e dall'altro, in un corpus di 36 testi argomentativi scritti da un gruppo di controllo composto da 18 studenti di madrelingua italiana. L'obiettivo principale del nostro studio è stato verificare se esiste una correlazione tra ricchezza lessicale delle produzioni dei parlanti non nativi e livello di competenza linguistica generale, misurato attraverso un C-test. Inoltre, ci siamo proposti di sottolineare le differenze o analogie esistenti tra i due gruppi degli apprendenti e dei parlanti nativi.

La nostra ricerca, in primo luogo, si poneva come obiettivo di verificare se la competenza linguistica generale degli apprendenti influisse sull'uso del lessico di base. Dalla nostra analisi, non è emersa alcuna correlazione tra la ricorrenza dei lemmi appartenenti alle varie categorie di frequenza del lessico di base nelle produzioni degli apprendenti olandesi ed i risultati degli stessi nel C-test, volto a misurare il livello di competenza linguistica. Una possibile spiegazione va forse cercata nel fatto che i livelli degli apprendenti (tutti quanti studenti universitari di primo anno) erano molto simili. In studi futuri sarebbe interessante disporre di informanti di diverso livello di competenza, in particolare principianti e avanzati, per verificare se l'uso del lessico di base cambi in modo significativo ad uno stadio di apprendimento più avanzato.

I risultati ottenuti dal confronto delle produzioni degli apprendenti e dei parlanti nativi sono interessanti soprattutto per quanto riguarda i segmenti più estremi delle fasce di frequenza (Band 1; Band 3 e *Off-list*). Dall'analisi è emerso che gli apprendenti di italiano L2, come avevamo ipotizzato, usano maggiormente le parole più frequenti del lessico di base, ma sempre in misura minore rispetto ai parlanti nativi, che invece ricorrono più spesso a parole non appartenenti al lessico di base. In particolare, abbiamo rilevato che gli apprendenti olandesi usano signifi-

cativamente più spesso parole appartenenti alla prima fascia di frequenza (Band 1), mentre i parlanti nativi usano più spesso parole che appartengono alla categoria delle parole fuori lista o alla categoria combinata (Band 3 e *Off-list*). Ciò conferma l'assunto secondo cui le parole del lessico di base sono quelle più facili da apprendere, perché maggiormente presenti nell'input e perché aventi caratteristiche che ne favoriscono l'apprendimento, ad esempio brevità e morfologia semplice. Le parole meno frequenti nell'*input*, invece, vengono acquisite in un secondo momento, quando gli informanti sono stati esposti per un periodo maggiore all'*input* e hanno sviluppato abilità linguistiche che gli consentano di passare ad uno stadio più avanzato di competenza lessicale. Per quanto riguarda la seconda e la terza fascia di frequenza (Band 2 e Band 3), si rilevano percentuali d'uso molto più basse rispetto alle altre fasce, e ciò sia nei testi dei nativi che nei testi dei non nativi. Inoltre, i risultati sono meno chiari e variano nei due compiti di scrittura: i due gruppi si differenziano o nella Band 2 o nella Band 3, anche se sono sempre gli studenti nativi ad adoperare la percentuale più alta di lemmi.

La nostra ricerca, focalizzando sugli aspetti quantititativi dello sviluppo lessicale, esclude gli aspetti qualitativi della conoscenza del lessico, dei tratti semantici e dei significati che le parole possono assumere in contesti diversi. Non sono state prese in considerazione le relazioni sintagmatiche in cui le parole di base compaiono e le co-occorrenze delle combinazioni più frequenti. Lo stesso può dirsi delle collocazioni, ovvero le unità prefabbricate il cui status prioritario nella produzione linguistica nella L1 come nella L2 è ormai pienamente riconosciuto (tra i numerosissimi studi, si veda ad esempio quello di Wray, 2002). Un'analisi più approfondita delle combinazioni semantiche privilegiate dagli apprendenti, potrà probabilmente contribuire a gettare maggiore luce sulla relazione esistenza tra competenza linguistica e uso del vocabolario di base (inteso come insieme di parole il cui senso si attualizzano in contesto) e sulle possibili differenze tra nativi e non nativi.

Ineke Vedder
Universiteit van Amsterdam
Spuistraat 210,
1012 VT Amsterdam - NEDERLAND
i.vedder@uva.nl

Veronica Benigno
Pearson Language Testing
Research Manager
Strand, 80
WC2R 0RL London U.K.
veronicabenigno@hotmail.com

Bibliografia

Aitchinson, Jean, 2003, *Words in the mind. An introduction to the mental lexicon*, Oxford, Blackwell Publishing.

Bardel, Camilla / Lindqvist, Christina, 2011, "Developing a lexical profiler for spoken French and Italian L2. The role of frequency, cognates and thematic vocabulary". In: Roberts, Lea / Pallotti, Gabriele / Bettoni, Camilla (eds.), *EUROSLA Yearbook 11*, Amsterdam, John Benjamins: 75-93.

Bardel, Camilla / Gudmundson, Anna / Lindqvist, Christina, 2012, "Aspects of lexical sophistication in advanced learners'oral production: Vocabulary acquisition and use in L2 French and Italian". *Studies in Second Language Acquisition*, 34/2, Special Issue: 269-290.

Benigno,Veronica, 2007, "Il vocabolario di base: tratti costitutivi, rilevanza cognitiva e acquisizione in italiano L2". In: Lo Cascio, Vincenzo (a cura di), *Parole in rete: Apprendimento e teoria nell'era elettronica*, Novara, Utet-Università: 151-174.

Benigno, Veronica, 2012, *La notion de collocation fondamentale. Étude de corpus en vue d'une exploitation didactique*. Tesi di dottorato in cotutela tra l'Università degli Studi di Palermo e l'Université Stendhal III di Grenoble.

Bernini, Giuliano, 2004, "Come si imparano le parole. Osservazioni sull'acquisizione del lessico in L2". *ITALS* 2 / 1: 23-47.

Bernini, Giuliano, 2006, "Strategie di lessicalizzazione e input ambiguo nell'acquisizione di L2. I verbi di moto in italiano". In: Grandi, Nicola / Gabriele Iannàccaro (eds.), *Zhì. Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*, Cesena/Roma, Caissa Italia editore: 65-84.

Bernini, Giuliano / Spreafico, Lorenzo / Valentini, Ada (a cura di), 2008, *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*, Perugia, Guerra.

Berns, Renske / Vedder, Ineke, 2007, "In carcere per Lara. L'uso delle strategie di ricostruzione testuale e le scelte linguistiche in italiano L2". *Studi italiani di linguistica teorica e applicata* 36/2: 297-310.

Bisconti, Valentina, 2012, "La svolta lessicografica di Tullio De Mauro e i dizionari contemporanei". *Chroniques italiennes web* 23/2: 1-26.

Boogaards, Paul / Laufer, Batia (eds.), 2004, *Vocabulary in a second language*: *Selection, acquisition and testing*, Amsterdam, John Benjamins.

Bortolini, Umberto / Tagliavini, Carlo / Zampolli, Antonio, 1971, *Lessico di frequenza della lingua italiana contemporanea*, Milano, Garzanti.

Bozzone Costa, Rosella, 2002, "Rassegna degli errori lessicali in testi scritti da ap-

prendenti elementari, intermedi ed avanzati di italiano L2 (ed implicazioni didattiche)". *Linguistica e Filologia* 14: 37-67.

Carloni, Fiammetta / Vedovelli, Massimo, 2005, "Il vocabolario di base dell'italiano degli stranieri". In: De Mauro, Tullio / Chiari, Isabella (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne: 247-275.

Cobb, Tom / Horst, Marlise, 2004, "Is there room for an academic wordlist in French?" In: Boogaards, Paul / Laufer, Batia (eds.), *Vocabulary in a second language: Selection, acquisition, and testing*, Amsterdam, John Benjamins, 15-38.

Council of Europe, 2001, *Common European Framework of Reference for languages: Learning, teaching, assessment*, Cambridge, Cambridge University Press.

Consiglio d'Europa, 2002, *Quadro Comune Europeo di Riferimento per le lingue: Apprendimento, insegnamento, valutazione*, Firenze, La Nuova Italia/Oxford University Press.

Coxhead, Averil, 2000,"A new academic word list". *TESOL Quarterly*, 34/2: 213-238.

Daller, Helmut / Milton, James / Treffers-Daller, Jeanine (eds.), 2007, *Modelling and assessing vocabulary knowledge*. Cambridge, Cambridge University Press.

De Mauro, Tullio, 1980/2003, *Guida all'uso delle parole,* Roma: Editori Riuniti (12a edizione 2003).

De Mauro, Tullio / Mancini, Federico / Vedovelli, Massimo / Voghera, Miriam (a cura di), 1993, *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.

De Mauro, Tullio, 1999, *GRADIT*. *Grande Dizionario Italiano dell'uso,* Torino: Utet.

Dörnyei, Zolta / Katona, Lucy, 1992, "Validation of the C-test amongst Hungarian EFL learners". *Language Testing* 9/2: 187-206.

Ellis, Nick, 1997, "Vocabulary acquisition: Word structure, collocation, word-class, and meaning". In: Schmitt, Norbert / McCarthy, Michael (eds.), *Vocabulary: Description, acquisition and pedagogy*, Cambridge, Cambridge University Press: 122-139.

Ellis, Nick. 2002, "Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition". *Studies in Second Language Acquisition* 24/2: 143-188.

Eyckmans, June / Van de Velde, Hans / Van Hout, Roeland / Boers, Frank, 2007, "Learners' response behaviour in yes/no vocabulary tests". In: Daller, Helmut / Milton, James / Treffers-Daller, Jeanine (eds.), *Modelling and assessing vocabulary knowledge,* Cambridge, Cambridge University Press: 59-76.

Ferreri, Silvana, 2005, "L'estensione delle conoscenze lessicali individuali". In: De Mauro, Tullio / Chiari, Isabella (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne: 307-334.

Guiraud, Pierre, 1954, *Les caractères statistiques du vocabulaire. Essai de méthodologie*. Paris, Presses universitaires de France.

Hoey, Michael, 2005, *Lexical priming. A new theory of words and language*, London, Routledge.

Horst, Marlise / Collins, Laura, 2006, "From faible to strong: How does their vocabulary grow?" *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 63/1: 83-106.

Klein, Wolfgang / Perdue, Clive, 1997, "The Basic Variety (or: Couldn't natural languages be much simpler?)". *Second Language Research* 13/4: 301-347.

Kuiken, Folkert / Vedder, Ineke / Gilabert, Roger, 2010, "Communicative adequacy and linguistic complexity in L2 writing". In: Bartning, Inge / Martin, Maisa / Vedder, Ineke (eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, Eurosla Monographs 1: 81-100.

Laufer, Batia / Nation, Paul, 1995, "Vocabulary size and use: Lexical richness in L2 written production". *Applied Linguistics* 16/3: 307-322.

Lengyel, Zsolt / Navracsics, Judit / Szilágyi, Anikó, 2007, "Analysing L2 lexical processes via C-Test". In: Lengyel, Zsolt / Navracsics, Judit (eds.), *Second language lexical processes. Applied linguistic and psycholinguistic perspectives,* Clevedon: Multilingual Matters: 166-185.

Lindqvist, Christina, 2010, "La richesse lexicale dans la production orale de l'apprenant avancé de français". *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes* 66/3, 393-420.

Lindqvist, Christina / Bardel, Camilla / Gudmundson, Anna, 2011, "Lexical richness in the advanced learner's oral production of French and Italian L2". *IRAL* 49/3: 221-240.

Lo Cascio, Vincenzo (a cura di), 2007, *Parole in rete. Teorie e apprendimento nell'era digitale*, Novara, Utet Universitaria.

Lo Duca, Maria Giuseppina, 2007, "Quante e quali parole nell'insegnamento dell'italiano L2? Riflessioni in margine alla costruzione di un sillabo". In: Pistolesi, Elena (a cura di), *Lingua, scuola e società. I nuovi bisogni comunicativi nelle classi multiculturali*, Istituto Gramsci del Friuli Venezia Giulia, Trieste: 135-150.

Lorenzo-Dus, Nuria, 2007, "The best of both worlds? Combined methodological

approaches to assessment of vocabulary in oral proficiency interviews". In: Daller, Helmut / Milton, James / Treffers-Daller, Jeanine (eds.), *Modelling and assessing vocabulary knowledge,* Cambridge, Cambridge University Press: 220-233.

Malvern, David / Richards, Brian / Chipere, Ngoni / Durán, Pilar, 2004 (eds.), *Lexical diversity and language development. Quantification and assessment*, Basingstoke, New York, Palgrave Macmillan.

Malvern, David / Richards, Brian, 2009, "A new method of measuring rare word diversity: The example of L2 learners of French". In: Richards, Brian / Daller, Helmut / Malvern, David / Meara, Paul / Milton, James / Treffers-Daller, Jeanine (eds.), *Vocabulary Studies in first and second language acquisition*, Basingstoke, New York, Palgrave Macmillan:164-178.

McCarthy, Michael / Jarvis, Scott H., 2007, "Vocd: A theoretical and empirical evaluation". *Language Testing* 24/4: 459-488.

Meara, Paul, 1996, "The dimension of lexical competence". In: Brown, Gillian / Malmkjaer, Kirsten / Williams, John (eds.), *Performance and competence in second language acquisition*, Cambridge, Cambridge University Press: 33-53.

Milton, James / Daller, Helmut / Malvern, David / Meara, Paul / Richards, Brian / Treffers-Daller, Jeanine (eds.), 2008, "Vocabulary", Special Issue, *Language Learning Journal* 36/2.

Nation, Paul, 2001, *Learning vocabulary in another language*, Cambridge, Cambridge University Press.

Nation, Paul, 2006, "How large a vocabulary is needed for reading and listening?" *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* 63/1, 59-82.

Nation Paul / Waring, Robert, 1997, "Vocabulary size, text coverage and word lists". In: Schmitt, Norbert / McCarthy, Michael (eds.), *Vocabulary, description, acquisition and pedagogy*, Cambridge, Cambridge University Press: 6-19.

Ovtcharov, Valentin / Cobb, Tom / Halter, Randall, 2006, "La richesse lexicale des productions orales: mesure fiable du niveau de compétence langagière". In: *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* 63/1, 107-125.

Read, John, 2000, *Assessing vocabulary,* Cambridge, Cambridge University Press.

Schmitt, Norbert / McCarthy, Michael (eds.), 1997, *Vocabulary, description, acquisition and pedagogy*, Cambridge, Cambridge University Press.

Spreafico, Lorenzo, 2003, "Misurare le parole. Analisi lessicale quantitativa di un apprendente di italiano L2", *Linguistica e Filologia* 17: 93-125.

Spreafico, Lorenzo, 2005, "Lo sviluppo lessicale di un apprendente di italiano L2. Problemi e metodi di analisi quantitativa". In: Banti, Giorgio / Marra, Antonietta / Vineis, Edoardo (a cura di), *Atti del 4° congresso di studi dell'Associazione Italiana di Linguistica Applicata,* Modena, 19-20 febbraio 2004, Perugia, Guerra Edizioni: 241-257.

Spreafico, Lorenzo, 2008, "Lessicalizzazione di eventi di moto in italiano. Problemi di tipologia". In: Bernini, Giuliano / Spreafico, Lorenzo / Valentini, Lorenzo (a cura di), *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*, Perugia, Guerra: 139-159.

Tidball, Françoise / Treffers-Daller, Jeanine, 2007, "Exploring measures of vocabulary richness in semispontaneous French speech. A quest for the Holy Grail?" In: Daller, Helmut / Milton, James / Treffers-Daller, Jeanine (eds.), *Modelling and assessing vocabulary knowledge*, Cambridge, Cambridge University Press: 133-149.

Treffers-Daller, Jeanine / Daller, Helmut / Malvern, David / Richards, Brian /Meara, Paul / Milton, James (eds.), 2008, "Knowledge and use of the lexicon in French as a second language". Special Issue, *Journal of French Language Studies* 18/3.

Valentini, Ada, 2008, "Strategie di lessicalizzazione degli eventi di moto nell'italiano L2 di apprendenti (semi) guidati". In: Bernini, Giuliano / Spreafico, Lorenzo / Valentini, Ada (a cura di), *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*, Perugia, Guerra: 179-203.

Vedder, Ineke, 2012, "Adeguatezza comunicativa e complessità linguistica in italiano L2 e italiano L1: la valutazione delle produzioni scritte di nativi e non nativi". In: Bernini, Giuliano / Lavinio, Christina / Valentini, Ada / Voghera, Miriam (a cura di), *Competenze e formazione linguistiche. In memoria di Monica Berretta.* Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata, Bergamo, 9-11 giugno 2011, Perugia, Guerra Edizioni: 139-159.

Wilks, Clarissa / Meara, Paul, 2007, "Implementing graph theory approaches to the exploration of density and structure in L1 and L2 word association networks". In: Daller, Helmut / Milton, James / Treffers-Daller, Jeanine (eds.), *Modelling and assessing vocabulary knowledge*. Cambridge, Cambridge University Press: 167-181.

Wray, Alison, 2002, *Formulaic language and the lexicon*, Cambridge, Cambridge University Press.