

MASSIMO MONEGLIA (FIRENZE)

LABLITA-SUITE. RISORSE PER L'ACQUISIZIONE DELL'ITALIANO L2

ABSTRACT

LABLITA-Suite. Resources for the acquisition of Italian as a second language – LABLITA-suite provides technology-enhanced learning resources for the acquisition of Italian L2. IMAGACT allows for mastering the semantic properties of action verbs in the early phases of language acquisition. The LABLITA corpus of Spoken Italian can be used for training learners for face to face conversations. RIDIRE and CORDIC provide corpus linguistic tools for accessing Italian phraseology, which is useful for enhancing writing capabilities in the various domains of language usage.

KEYWORDS: Italian L2, Technology Enhanced Learning, Spontaneous speech resources, Web corpora, Action Ontology

1. INTRODUZIONE

Questo lavoro presenta sinteticamente le risorse prodotte da LABLITA nei vari ambiti dello studio empirico del linguaggio e introduce al loro utilizzo in italiano L2. Nel secondo paragrafo, utilizzando l'ontologia dei verbi d'azione IMAGACT, vedremo come, nelle prime fasi di acquisizione di L2, l'apprendimento del lessico verbale di base possa configurarsi, come acquisizione della semantica piuttosto che del lessico. Nel terzo paragrafo affronteremo la questione della comprensione della L2 nella varietà orale, tema che si pone appena l'apprendente interagisce con i parlanti madrelingua. Vedremo come la disponibilità di corpora orali, trattati e allineati al suono come il Corpus di Italiano Parlato LABLITA, permettono di sviluppare attività basate sull'accesso a tali corpora. Il quarto paragrafo è dedicato al consolidamento della conoscenza lessicale negli apprendenti di livello universitario. Vedremo come l'accesso al grande corpus RIDIRE e al piccolo corpus CORDIC, attraverso strumenti propri della *Corpus Linguistics*, risponde alla necessità di acquisire la fraseologia italiana nei suoi vari domini dell'uso.

2. ACQUISIZIONE DEL LESSICO DI BASE E ACQUISIZIONE DELLA SEMANTICA. IMAGACT E LA CAPACITÀ DI RIFERIRSI ALLE AZIONI IN L2

IMAGACT raccoglie 1010 concetti azionali e i verbi che li possono identificare in molte lingue. Italiano, inglese, cinese, spagnolo, portoghese, hindi-urdu, arabo, serbo, polacco, danese, greco, tedesco, e giapponese sono lingue già sviluppate, e altre sono in elaborazione.

Con IMAGACT è possibile osservare che verbi di lingue diverse in relazione di traduzione categorizzano in modo diverso le azioni della vita quotidiana. Le variazioni semantiche sono evidenziate mostrando i prototipi delle azioni che ogni verbo può identificare, che IMAGACT fornisce in brevi filmati o animazioni 3D.

Per esempio, la differenza semantica tra *take* e *prendere*, i verbi d'azione a più alta frequenza in italiano e inglese, può essere evidenziata in Figura 1. I verbi sono co-estensivi in molti casi, ma *prendere* non può essere usato quando *take* è equivalente a *bring* e a *lead*, mentre *take* non può essere applicato alle azioni in cui *prendere* equivale a *colpire* (*hit*) o *acchiappare* (*catch*). La differenza può essere compresa dall'apprendente in modo semplice, osservando gli ambiti di co-estensione (al centro) e i rispettivi ambiti di sovra-estensione (a destra e a sinistra).

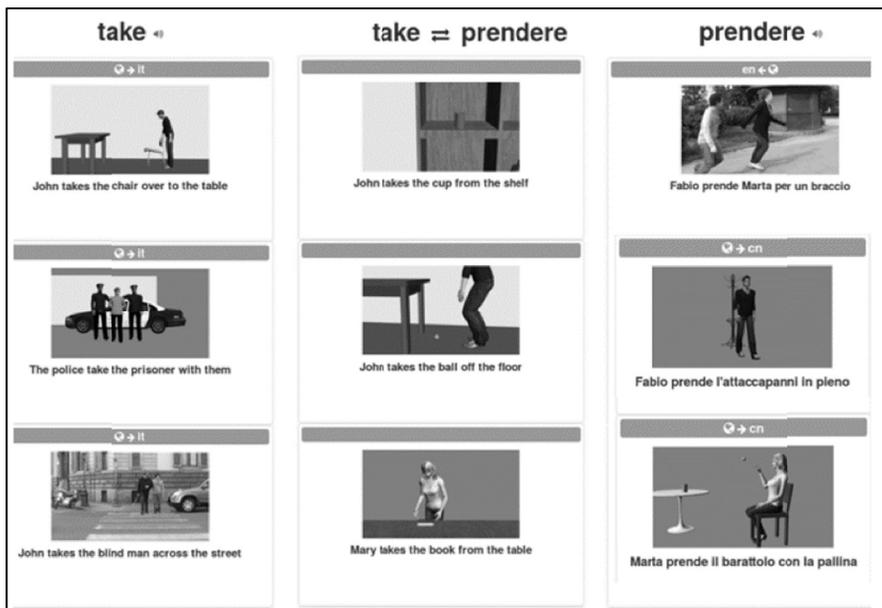


Figura 1. Comparazione *take* / *prendere*

Questa semplice informazione non può essere derivata dai dizionari bilingui, che danno notizie sulla variazione dei sensi di un verbo, ma non forniscono il loro differenziale. Al contrario l'informazione è cruciale per apprendere il significato

di un verbo nella L2 target, e si ottiene mappando i verbi italiani e inglesi sulla stessa ontologia.

Questa possibilità è implementata per ogni lingua presente in IMAGACT. Di conseguenza, il lessico verbale di azione di lingue come l'italiano e il polacco potrà essere confrontato. Per esempio, il verbo italiano *prendere* potrà essere comparato al traduce polacco tipico *brać* (Figura 2) e l'utente italiano potrà apprezzare che *prendere* è più generale di *brać*, che non è utilizzabile in molte delle estensioni di *prendere* (sulla sinistra), ovvero quando *prendere* è equivalente a *estrarre*, *colpire*, *togliere a qualcuno*, *dare* e *acchiappare*. Allo stesso modo l'utente polacco potrà conoscere la variazione in sovra-estensione del verbo italiano.

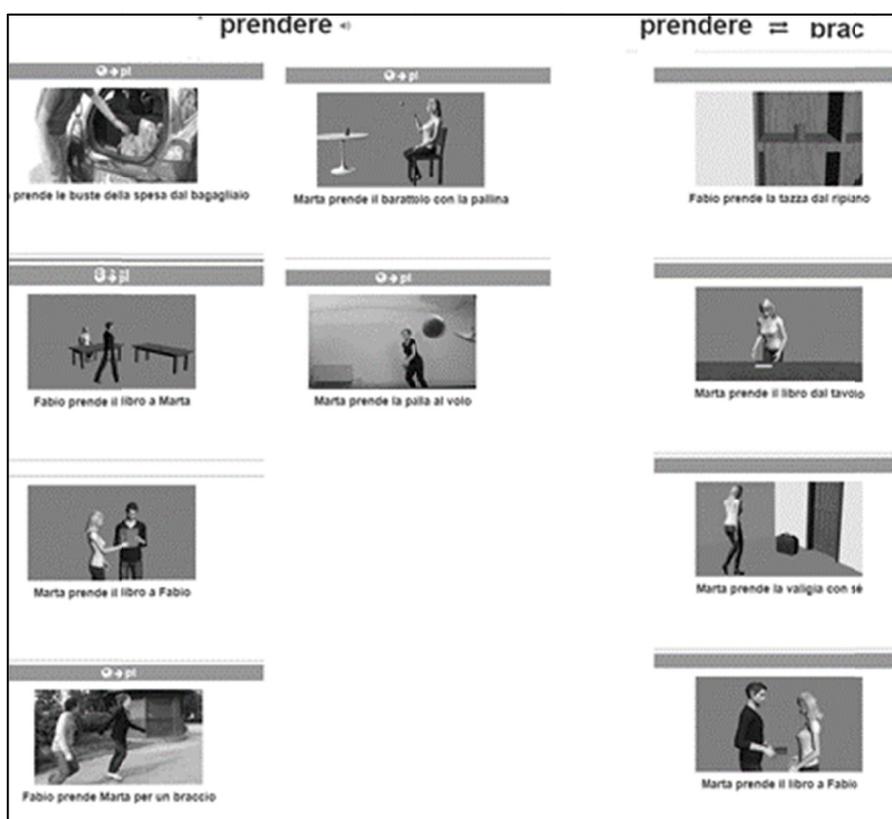


Figura 2. Comparazione *prendere* / *brać*

Un apprendente può cercare in IMAGACT in due modi principali (Moneglia et al. 2014; Moneglia et al. 2013):

1. in modalità *Dictionary*, come in un dizionario bilingue per immagini;
2. in modalità *Compare*, attraverso una comparazione esplicita delle azioni identificate da due verbi (della stessa lingua o di lingue diverse).

DICTIONARY

Un utente ha di solito interesse a sapere come un verbo della sua lingua si traduce in un'altra lingua, non solo in generale, ma per risolvere casi di traduzione. Facciamo l'esempio di un apprendente italiano che studia il polacco e immaginiamo il problema posto dalla scelta di un verbo polacco per tradurre una frase con il verbo *girare* in due istruzioni culinarie:

- (1) gira la scatola del prodotto per vedere gli ingredienti utilizzati
- (2) gira lo spezzatino frequentemente.

Cercando il verbo *girare*, IMAGACT funziona come un dizionario e risponde proponendo una o più animazioni che illustrano la variazione di questo verbo sui diversi tipi di azione a cui questo si può riferire (Figura 3):

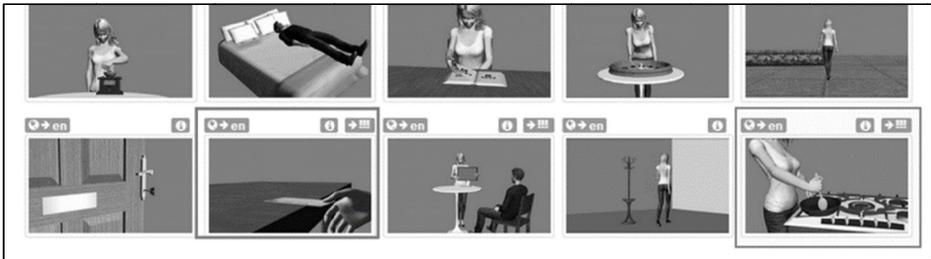


Figura 3. La variazione di *girare* per tipi di azione

Guardando i prototipi, un utente intelligente, che conosce il significato delle frasi italiane, può riconoscere quali sono le azioni che corrispondono alle frasi da tradurre. Per esempio, l'azione della frase (1) è del tipo cerchiata sulla sinistra e il secondo caso corrisponde all'azione cerchiata sulla destra. Su questa base, selezionando il polacco come lingua *target*, l'utente vedrà come ci si riferisce a quelle specifiche azioni in questa lingua.

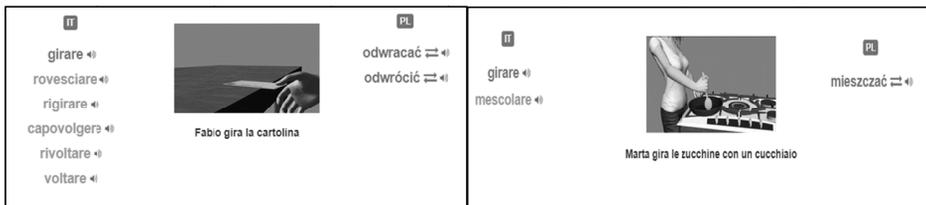


Figura 4. La traduzione di *girare* in polacco in due tipi di azione della variazione del verbo

Figura 4 mostra che in polacco, diversamente dall'italiano, non si può utilizzare lo stesso verbo per le due azioni, e si utilizzano rispettivamente *odwracać* e *mieszczać*.

Accanto ai prototipi, il sistema restituisce, oltre al verbo cercato nella lingua sorgente e il suo traduce nella lingua target in quel prototipo, anche una serie di altri verbi che possono indicare equivalentemente le azioni nella variazione del verbo cercato. Figura 4 mostra che, in italiano, ciascuna delle azioni individuate può essere espressa equivalentemente, oltre che da *girare*, anche da altri verbi, ovviamente non applicabili a tutte le azioni a cui *girare* si estende. In particolare, *girare* equivale a *rovesciare*, *capovolgere*, *voltare*, *rivoltare* nel primo caso e a *mescolare* nel secondo caso. L'apprendente potrà quindi indurre delle regolarità nelle modalità di traduzione, ovvero che *odwracać* si usa quando *girare* equivale a *voltare* e *mieszczać* quando *girare* equivale a *mescolare*.

COMPARE

In IMAGACT, come abbiamo visto per *prendere* e *take*, è possibile comparare direttamente verbi di lingue diverse che si propongono come la traduzione dell'uno nell'altro, evidenziando le differenze. Figura 5 mostra la comparazione tra *girare* e il traduce *kręcić* e rende esplicita la loro differenza.

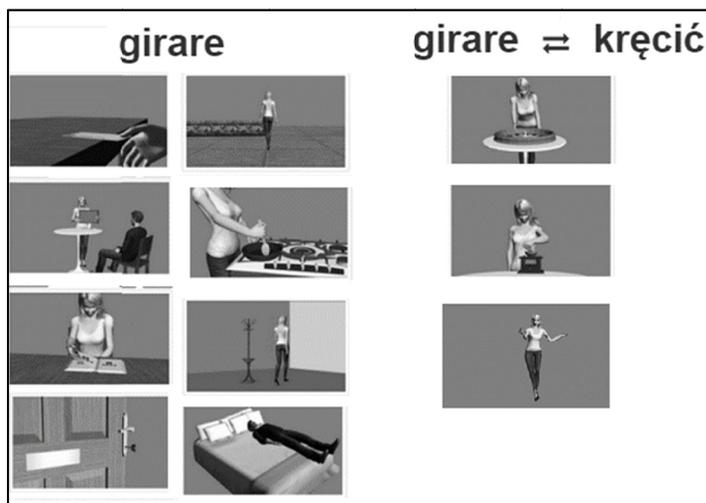


Figura 5. *Girare* vs *kręcić*: la sotto-estensione del verbo polacco

Kręcić si applica specificamente a “rotazioni continue di un oggetto sul suo asse verticale”, mentre *girare* è un verbo che si estende a molti altri tipi di rotazioni.

L'utente italiano potrà domandarsi quale verbo polacco si usa nei vari campi che non sono coperti da *kręcić* e potrà ottenere un panorama via via sempre più fine delle differenze tra il lessico della sua lingua e quello della lingua target. Per esempio, potrà domandarsi quale verbo si usa per *girare* nel senso di *svoltare*.

Cliccando sull'icona corrispondente a questa interpretazione di *girare*, otterrà il verbo polacco *skrećać*.

Complessivamente quindi, data la variazione del verbo italiano ottenuta attraverso la funzione Dictionary, l'apprendente potrà interrogarsi su quali verbi polacchi servono nei vari casi e ottenere poi, attraverso la funzione Compare, vedere i verbi che esprimono tutte le varie azioni che in italiano sono *girare*. In particolare, potrà indurre la generalizzazione che il polacco non ha verbi altrettanto generali per esprimere il concetto italiano, ma che distingue tre grandi tipi di movimenti rotatori, *skrećać*, per i movimenti nello spazio, quando girare equivale a *svoltare*, *kręcić* per i movimenti rotatori continui degli oggetti su sé stessi, *odwracać* per i movimenti rotatori discontinui, che portano a evidenziare parti non visibili dell'oggetto, ovvero quando *girare* equivale a *voltare*.

3. COMPRENDERE L'ITALIANO NELLE INTERAZIONI FACCIA A FACCIA: IL CORPUS DI ITALIANO PARLATO LABLITA

3.1 IL CORPUS LABLITA E L'ALLINEAMENTO TESTO SUONO

Le principali difficoltà nella comprensione della varietà scritta si situano a livello della maggiore o minore ricchezza del lessico e nella complessità morfo-sintattica del testo. Il parlato pone invece primariamente un problema di *ricoscimento*. Fatta salva la conoscenza delle parole e della loro pronuncia, il parlato spontaneo è oggettivamente difficile da comprendere per gli apprendenti, che necessitano di un lungo processo di pratica per poter affrontare le interazioni orali con parlanti madrelingua. Le unità di parola sono infatti conosciute dall'apprendente in isolamento e non nelle condizioni olistiche di enunciazione in cui si verificano. Ne deriva in primo luogo una difficoltà nell'assegnazione di porzioni di *speech* a unità informative minime (parole) e complessivamente una difficoltà di segmentazione del parlato in unità pragmaticamente significanti. In termini generici, se anche qualche parola la coglie, l'apprendente "perde il filo" a causa di fenomeni come i seguenti:

- alta velocità di locuzione;
- frequente co-articolazione delle parole;
- aferesi, elisione, troncamento;
- disfluenze;
- scarsa prominenza fonetica degli elementi morfologici;
- strutture ritmico-accidentali diverse rispetto alla lingua madre;
- rilevanza dell'informazione prosodica.

L'accesso all'informazione acustica è ovviamente il requisito minimo per lo sfruttamento dei corpora che riportano l'esecuzione linguistica orale (Sinclair 1994). LABLITA ha raccolto un grande corpus multimediale di parlato spontaneo italiano,

pubblicato in varie forme negli ultimi venti anni,¹ che può costituire la base per attività didattiche dedicate allo sviluppo della capacità di comprendere l'italiano nella sua varietà orale e conseguentemente a facilitare l'interazione con parlanti madrelingua nei contesti naturali faccia a faccia (Tyne 2009; Nicolas 2012).

I corpus LABLITA, seguendo la Teoria della Lingua in Atto (L-AcT) di E. Cresti, fornisce la chiave per rendere simultaneo l'apprezzamento di informazione acustica e testuale e quindi per rendere sfruttabili i corpora di parlato a fini acquisizionali. L-AcT definisce l'unità naturale di interpretazione del parlato (unità di riferimento), che non è la "frase" come nello scritto, ma l'"enunciato" (Cresti 2000). Costituisce *enunciato* ogni sequenza minima di parole interpretabile pragmaticamente. L'enunciato è segnalato dalla prosodia ed è identificato nel flusso del parlato da *breaks* prosodici percepiti come terminali (Cresti, Moneglia 2005).

I corpora orali trattati secondo L-AcT, non si limitano quindi a fornire trascrizioni, ma comprendono un allineamento testo / suono che segmenta il parlato nelle unità rilevanti per l'interpretazione, gli enunciati. Gli enunciati sono brevi e l'accesso simultaneo all'informazione acustica e testuale consente all'apprendente di apprezzare separatamente, una dietro l'altra, le sequenze di parole dotate di senso autonomo e compiuto.

Dal punto di vista tecnico, l'allineamento è gestito da un software (WinPitch), usato per il trattamento dei corpora di LABLITA. WinPitch consente di affrontare e ridurre progressivamente il problema della segmentazione del parlato spontaneo garantendo varie modalità di accesso alla sorgente, dalle più semplici alle più complesse:

- accesso a velocità ridotta;
- accesso con segmentazione enunciato per enunciato (di *default*);
- accesso per selezione di più enunciati;
- accesso per segmento di un enunciato selezionato dall'utente.

In particolare, attraverso l'allineamento per enunciati l'utente può selezionare e ascoltare porzioni di testo significative dal punto di vista semantico, diminuendo via via la difficoltà di riconoscimento dell'informazione acustica.

Figura 6 riporta l'allineamento di un brano in cui una parlante racconta di quando è andata ad un concerto. L'utente può ascoltare porzioni di testo più o meno lunghe, selezionandole nella finestra testuale (in alto). In questa modalità può scegliere di accedere separatamente ai turni dialogici di ciascun parlante (come nella selezione di enunciati evidenziata in figura). Può altrimenti focalizzare la sua attenzione separatamente su ciascun enunciato di un turno, ottenendo così l'ascolto di una porzione breve ma compiuta del testo, la cui segmentazione in parole nella trascrizione diviene più facilmente riconoscibile. Può fare questo sia cliccando su ciascun enunciato nel testo, sia cliccando sui segmenti di trascrizione allineati all'onda e all'analisi dell'F0 (in basso nella figura).

¹ Cresti 2000 (60.000 parole), la parte italiana del corpus C-ORAL-ROM (Cresti, Moneglia 2005) 300.000 parole, il Corpus Stammerjohan e il corpus di Confronto di 100.000 parole ciascuno (Scarano, Moneglia 2008).

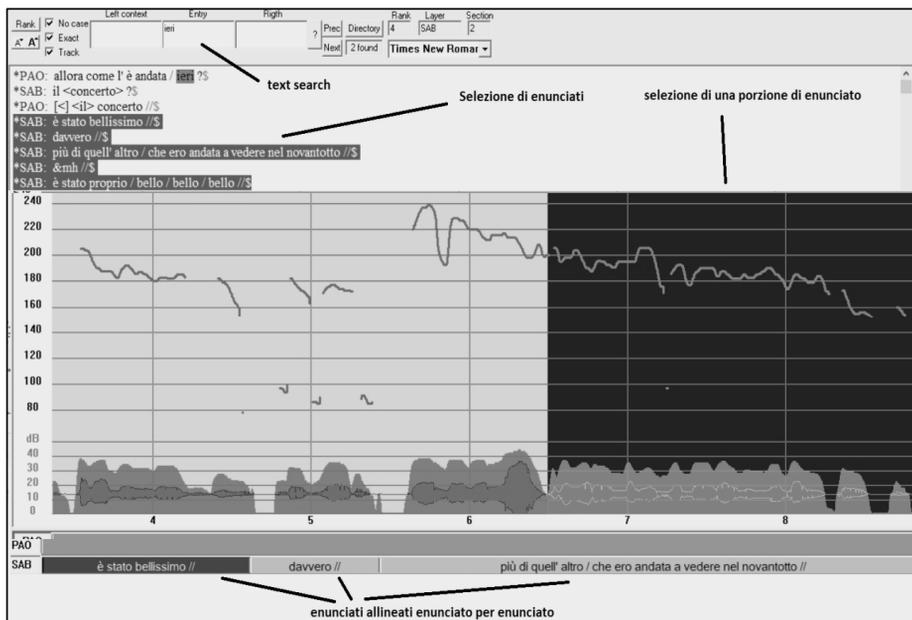


Figura 6. L'accesso all'informazione acustica attraverso l'allineamento testo/ suono in WinPitch

Come si può vedere dall'esempio, gli enunciati nel parlato spontaneo sono sequenze brevi e spesso nominali (*il concerto; davvero*), ovvero non corrispondono a frasi sintattiche, ma nondimeno costituiscono unità di senso autonome, pragmaticamente interpretabili nel momento in cui se ne sente il suono e se ne può apprezzare l'intonazione. Gli enunciati, come mostra l'onda riportata alla base della figura, non sono divisi da pause, ma sono comunque percepiti separati prosodicamente. L'andamento prosodico è riportato nel formato di trascrizione attraverso sbarre doppie (//), che indicano le interruzioni prosodiche terminali. Le sbarre semplici (/) indicano interruzioni prosodiche all'interno dell'enunciato, che possono anche avere un rilievo anche per l'apprendente (Scarano 2005). L'apprendente potrà così interiorizzare i modelli di co-articolazione di parole così come si realizzano nell'uso, potrà riconoscerli in ricezione e metterli in opera a livello produttivo.

3.2. LO SFRUTTAMENTO DELLA VARIAZIONE DIAFASICA PER LA GESTIONE DELLE DIFFICOLTÀ

Il Corpus LABLITA, in tutte le sue pubblicazioni, ha la forma di un corpus di riferimento per la rappresentazione dell'italiano² e a questo fine presenta una notevole variazione diafasica, che può essere molto utile per lo sfruttamento per

² Il corpus non presenta variazione diatopica, ma rappresenta un taglio delle varietà che possono co-occorrere a Firenze e in Toscana.

l'italiano L2. Tabella 1 mostra la variazione del corpus che riguarda in primo luogo la distinzione tra testi di parlato informale (non regolato) e parlato formale (regolato) sia in contesto naturale che attraverso i media.

Tabella 1. La variazione diafasica del corpus di parlato italiano LABLITA e la sua consistenza

Diaphasic Variation of LABLITA CORPUS			Samples	Words	Ref.Units
Free	Family/Private	Monologues	28	55,141	5,795
turn	Context	Dialogues/Multi-dialogues	140	254,085	39,593
taking	Public Context	Monologues	1	1,723	198
		Dialogues/Multi-dialogues	39	67,434	5,842
<i>total</i>			208	378,383	51,428
Regulated	Family/Private	Monologues	1	3,023	193
turn	Context	Dialogues/Multi-dialogues	28	52,915	5,713
taking	Public Context	Monologues	38	105	4,678
		Dialogues/Multi-dialogues	40	186,103	12,944
<i>total</i>			107	242,308	23,528
Broadcasting			49	131,539	12,593
Telephone			74	32,541	4,445
Total Adult Corpus			438	784,771	91,994
Cinema (written to be spoken)			189	70,414	N.C.
Talking to Children			276	260,595	N.C.
<i>Total</i>			903	1,115,780	N.C.

I vari contesti diafasici manifestano gradi di difficoltà diversi in relazione alla velocità di eloquio, all'identificabilità di un argomento, al grado di testualità realizzata nel prodotto orale e soprattutto in relazione al grado di interattività del contesto. Il parlato regolato e dei media ha una velocità minore e una lunghezza maggiore degli enunciati, necessaria ad articolare un discorso, e in genere produce testi nei quali è sviluppato un argomento. Il parlato in contesto interattivo ha invece enunciati brevi e molto più veloci, e produce sequenze nelle quali l'argomento cambia di frequente. Il parlato interattivo non è finalizzato alla realizzazione di un testo ma ha scopi pratici. In sintesi, il monologo, oltre ad essere più lento del dialogo, ha natura testuale, mentre il dialogo interattivo ha natura eminentemente pragmatica e corrisponde ad un cambiamento continuo di argomenti e di prospettive, dovuto all'interazione tra i parlanti e alle sue finalità, non testuali. La conversazione a più voci è anche di solito più frammentata e costituisce sicuramente il contesto più difficile.

Nel processo di apprendimento è quindi possibile modulare la tipologia dei testi orali da proporre ad un apprendente, consentendo un suo progressivo avvicinamento alla lingua italiana parlata, dai contesti più semplici ai più complessi:

- monologhi in ambito formale;
- emissioni media;
- monologhi informali;
- parlato dialogico interattivo in contesto naturale;
- conversazioni a più voci private e familiari.

4. CORPORA IN RETE PER L'ACQUISIZIONE DEL LESSICO E DELLA FRASEOLOGIA: RIDIRE E CORDIC

L'utilizzo dei corpora per l'acquisizione delle lingue seconde non è immediato per gli apprendenti come lo sono le grammatiche e i dizionari, e neppure per i docenti il loro uso è comune³. La disponibilità di strumenti computazionali per l'estrazione di informazione viva da grandi corpora rappresentativi dell'uso è però un'opportunità importante per l'acquisizione della capacità di utilizzare le lingue seconde e costituisce il portato naturale dell'evoluzione tecnologica.

LABLITA-*suite* propone due risorse di rete finalizzate all'utilizzo dei corpora per l'apprendimento dell'italiano: CORDIC (Corpora Didattici Italiani di Confronto) e RIDIRE (Risorsa Italiana di Rete).

CorDIC è uno strumento che introduce alle tecniche di accesso ai corpora e permette in particolare il confronto tra le varietà scritta e orale dell'italiano. La risorsa è stata realizzata a supporto di Cresti, Panunzi (2013) e costituisce un ambiente adatto alla formazione linguistica degli apprendenti. CorDIC è basato su due piccoli corpora (scritto e parlato) costruiti per essere strettamente comparabili. Attraverso CORDIC, che ha una interfaccia di ricerca in rete molto semplice e efficiente, l'apprendente può familiarizzare con le nozioni di base necessarie per l'accesso ai corpora e fare ricerche che producono le seguenti informazioni:

- liste di frequenza;
- concordanze di parola (per forma o per lemma);
- ordinamento delle concordanze;
- *patterns* di parole e parti del discorso;
- collocazioni.

RIDIRE⁴ è un grande archivio pubblico di testi derivati dal web italiano (Cresti, Moneglia 2012), che da un lato rappresenta la lingua italiana nel suo insieme e dall'altro rende possibile selezionare il dominio specifico di interesse dell'apprendente, in un'ampia scelta significativa per la vita e la cultura italiane. La base dati interrogabile ha una dimensione di circa 1.3 miliardi di parole e struttura i domini in due tipologie (Figura 7):

Domini Semantici, che identificano i campi dell'eccellenza italiana nel mondo:

- Letteratura;
- Moda;
- Design-architettura;
- Cucina;

³ Si vedano Sinclair (2004); Tognini Bonelli (2000); Conrad (2006) e la principale conferenza di settore TALC come esempi delle proposte di utilizzo.

⁴ Il progetto RIDIRE.it (Risorsa Dinamica Italiana di Rete), nasce su suggerimento di John Sinclair, è stato promosso dalla Società Internazionale di Linguistica e Filologia Italiana (SILFI) in collaborazione con un consorzio di Università italiane e finanziato sul Fondo Italiano per la Ricerca di Base (FIRB).

- Sport;
- Religione;
- Arti figurative;
- Cinema;
- Musica.

Domini Funzionali, nei quali la lingua ha convenzioni legate alla funzione, indipendentemente dal contenuto semantico:

- Informazione;
- Economia e affari;
- Amministrazione e legislazione.

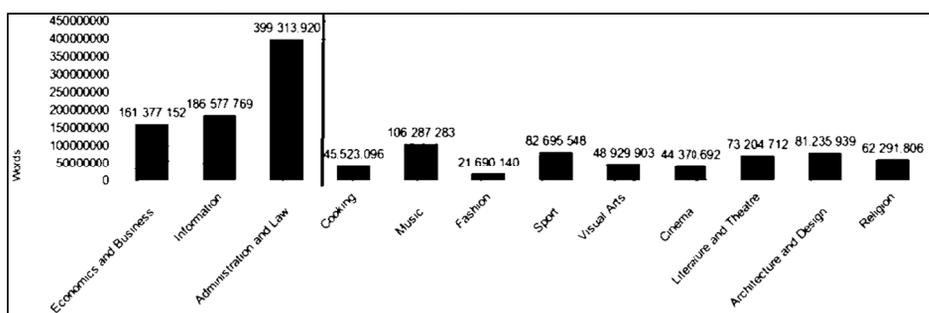


Figura 7. Il Corpus Design di RIDIRE

RIDIRE è orientato alla formazione universitaria, che consente l'applicazione di strumenti con un certo grado di complessità e ha come obiettivo il potenziamento dell'uso italiano nella sua varietà scritta, nella quale le specificità dell'uso linguistico settoriale e la necessità di adeguarsi alle forme convenzionali dell'uso si manifesta.

Le funzioni di ricerca generano le seguenti informazioni, rispettivamente sull'intero corpus o in ciascun sotto-corpus:

- liste di frequenza;
- concordanze ordinabili e raggruppabili (per forma e per lemma);
- PoS patterns;
- collocazioni;
- *colligations* e *Sketches*⁵.

Facciamo qui alcuni esempi salienti delle funzioni di RIDIRE. Gli apprendenti si domandano spesso quale preposizione si deve utilizzare in relazione ad un certo verbo in L2, perché le lingue si comportano in modo idiosincratico e l'interferenza causa errori molto frequenti in L2. La ricerca di *patterns* all'interno delle concordanze dell'intero corpus RIDIRE può fornire risposte molto significative a questo proposito.

⁵ Si veda in particolare Sketch Engine (Kilgarriff *et al.* 2015) e Colombo *et al.* 2014 per una descrizione dettagliata delle funzioni di RIDIRE.

collocano con i due verbi in questo dominio in posizione di oggetto (*postV_N*) ed è chiarissima. Nei toni di grigio in basso vediamo le collocate post verbali di *richiedere*, in grigio in alto le poche collocate di *chiedere* e in bianco le espressioni che collocano con entrambi i verbi.

postV_N				
trascrizione	102	0	7,41	0
ristoro	71	423	7,78	7,68
esibizione	18	249	5,75	6,91
risoluzione	106	663	6,51	7,84
chiarimento	31	474	5,67	7,66
restituzione	41	721	5,63	8,14
risarcimento	318	3327	6,21	9,11
annullamento	186	3610	6,94	10,12
condanna	109	2580	5,62	9,35
rigetto	49	3746	5,92	10,53
cosa	0	544	0	6,91
rilascio	0	409	0	6,93
autorizzazione	0	734	0	7,15
rinnovo	0	378	0	7,25
rinvio	0	489	0	7,25
iscrizione	0	607	0	7,28
sospensione	0	617	0	7,28
riconoscimento	0	713	0	7,61
ammissione	0	582	0	7,68
ricongiungimento	0	429	0	7,76
trasferimento	0	1003	0	8,04
accoglimento	0	777	0	8,08
rimborso	0	847	0	8,08
parola	0	2017	0	8,19
conferma	0	888	0	8,51

Figura 9. Sketch difference: *Chiedere* vs *Richiedere* – Nomi oggetto in Amministrazione

preN_V				
accusare	78	0	8,88	0
lamentare	38	0	8,47	0
camuffare	4	0	6,76	0
riacutizzare	4	0	6,59	0
smaltire	6	0	6,51	0
anestetizzare	3	0	6,38	0
avvertire	102	13	9,85	6,01
sentire	230	83	8,03	6,1
provocare	40	28	8,52	7,41
procurare	18	25	7,72	7,38
atternuare	4	6	6,56	6,45
sopportare	17	48	7,92	8,29
causare	8	17	6,74	7,37
lenire	5	23	7,11	8,57
alleviare	3	29	6,32	8,58
arrecare	0	5	0	6,2
calmare	0	5	0	6,24
eliminare	0	10	0	6,34
seminare	0	11	0	6,36
superare	0	31	0	6,54
sperimentare	0	29	0	6,79
patire	0	12	0	6,94
condividere	0	55	0	7,13
mitigare	0	10	0	7,37
provare	0	101	0	8,45

Figura 10. Sketch difference Domain: *dolore* – verbi di cui è oggetto in Sport e Religione

Se la funzione *Sketch Difference* evidenzia il contrasto tra le associazioni di due lemmi, la funzione *Sketch Difference Domains* mostra la variazione di associazioni lessicali preferenziali di un particolare lemma in due domini diversi. Figura 10 mostra i verbi con cui si associa il nome *dolore* rispettivamente nello Sport (in grigio nella parte alta) e in Religione (in grigio nella parte bassa). La differenza nei due domini risulta evidente dai verbi dei quali il nome *dolore* è oggetto (*PreN_V*): per lo Sport il dolore è “fisico” e in Religione è “psicologico”.

5. CONCLUSIONI

Nelle diverse fasi dell'acquisizione dell'italiano L2 LABLITA-suite risponde a esigenze concrete degli apprendenti attraverso risorse e sistemi di *Technology Enhanced Learning*.

IMAGACT fornisce uno strumento di acquisizione delle peculiarità semantiche del lessico verbale di azione italiano, essenziale per comunicare con sicurezza in contesti ordinari, e può essere utilizzato nelle prime fasi dell'acquisizione, quando è necessario apprendere il lessico di base della L2 e distinguere le sue specificità dal lessico della lingua materna. A livello metodologico IMAGACT propone il passaggio dall'acquisizione del lessico all'acquisizione della semantica.

Il Corpus di Parlato LABLITA, con le tecnologie per lo sfruttamento dell'informazione acustica che lo accompagnano (WinPitch) permette di preparare gli apprendenti all'uso dell'italiano nell'interazione faccia a faccia e configura attività significative quando la conoscenza del lessico e della sintassi inizia a consolidarsi. A livello metodologico, l'approccio riflesso all'italiano orale fronteggia problemi sensibili non affrontati nelle pratiche didattiche correnti.

RIDIRE e CORDIC propongono agli apprendenti di livello avanzato l'utilizzo dei concetti della linguistica dei corpora per accedere alle convenzioni dell'italiano scritto e lo fanno in modo da corrispondere in modo semplice e diretto a esigenze pratiche. A livello metodologico, RIDIRE e CORDIC forniscono gli strumenti per affrontare la linguistica della *parole*, affiancandola allo studio tradizionale della *langue*.

BIBLIOGRAFIA

- COLOMBO, S., MONEGLIA, M., PANUNZI, A. (2014): "Una Guida alle funzioni di ricerca linguistica nel web corpus RIDIRE", in GARAVELLI, E., SUOMELA, H. (eds.): *Dal Manoscritto al Web, Canali, Modalità di Trasmissione dell'italiano* (Atti del XII congresso SILFI, Helsinki, 18–20 Giugno 2012), Cesati, Firenze, 481–502.
- CONRAD, S. (2006): "Challenges for English corpus linguistics in second language acquisition research", in KAWAGUCHI, Y., ZAIMA, S., TAKAGAKI, T. (eds.): *Linguistic informatics and spoken language corpora: Contributions of linguistics, applied linguistics and computer science*, John Benjamins, Amsterdam, 67–88.
- CRESTI, E. (2000): *Corpus di italiano parlato*, Accademia della Crusca, Firenze.
- CRESTI, E., MONEGLIA, M. (eds.) (2005): *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*, John Benjamins Publishing Company, Amsterdam / Philadelphia.
- CRESTI, E., MONEGLIA, M. (2012): "Risorse di rete per l'insegnamento dell'italiano", in BIANCHI, P., DE BLASI, N., DE CAPRIO, C., MONTUORI, F. (eds.): *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali* (Atti dell'XI Congresso SILFI Società Internazionale di Linguistica e Filologia Italiana), Cesati, Firenze, 597–613.
- CRESTI, E., PANUNZI, A. (2013): *Introduzione ai corpora dell'italiano*, Il Mulino, Bologna.

- KILGARRIFF, A., MARCOWITZ, F., SMITH, S., THOMAS, J. (2015): *Corpora and Language Learning with the Sketch Engine and SKELL*, disponibile in <https://www.sketchengine.eu/wp-content/uploads/2015/04/2015-corpora-submitted.pdf>.
- MARTÍNEZ, N.C. (2012): *C-Or-DiAL (Corpus Oral Didáctico) Anotado Lingüísticamente*, Liceus, Madrid.
- MONEGLIA, M., BROWN, S., FRONTINI, F., GAGLIARDI, G., KHAN, F., MONACHINI, M., PANUNZI, A. (2014): “The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action”, in CALZOLARI, N. et al. (ed.): *Proceedings of LREC’14*, ELRA – European Language Resources Association, Reykjavik, 3425–3432.
- MONEGLIA, M., PANUNZI, A., GAGLIARDI, G., MONACHINI, M., RUSSO, I., DE FELICE, I., KHAN, F., FRONTINI, F. (2013): “IMAGACT E-learning Platform for Basic Action Types”, in *Proceeding of the international conference ICT for Language Learning* (Florence 14–15 November 2013), Libreria Universitaria, Firenze, 85–90.
- SCARANO, A. (2005): “L’accesso all’informazione prosodica in C-ORAL-ROM. Pragmatica e insegnamento delle lingue seconde”, in MARTINEZ, C.N., SCOTT, S. (eds.): *Studi per l’insegnamento delle lingue europee*, Firenze University Press, Firenze, 81–95.
- SCARANO, A., MONEGLIA, M. (2008): “Il Corpus Stammerjohann. Il primo corpus di italiano parlato, in rete nella base dati di LABLITA”, in PETTORINO, M. (ed.): *La comunicazione parlata* (Atti del Congresso Internazionale, Napoli 23–25 febbraio 2006), Liguori, Napoli.
- SINCLAIR, J.M. (1991): *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- SINCLAIR, J.M. (1994): *Spoken Language [3B], Phonetic /Phonemic and Prosodic Annotation [5.2]. NERC-1 Network of European Reference Corpora. Final Report*, NERC 1994, Istituto di Linguistica Computazionale – CNR, Pisa.
- SINCLAIR, J.M. (ed.) (2004): *How to Use Corpora in Language Teaching*, John Benjamins, Amsterdam-Philadelphia.
- TOGNINI-BONELLI, E. (2000): “Il corpus in classe: da una nuova concezione della lingua a una nuova concezione della didattica”, in ROSSINI FAVRETTI, R. (ed.): *Linguistica e informatica: multimedialità, corpora e percorsi di apprendimento*, Bulzoni, Roma, 93–108.
- TYNE, H. (2009): “Corpus oraux par et pour l’apprenant”, in BOULTON, A. (ed.): *Des documents authentiques oraux aux corpus: Questions d’apprentissage en didactique des langues*, Mélanges CRAPEL, Nancy, 91–111.

SITOGRAFIA

- C-ORAL-ROM: <https://benjamins.com/series/scl/15/scl_15.iso>
 CORDIC: <<http://corporadidattici.lablita.it/>>
 IMAGACT: < <http://imagactpp.imagact.it/imagact/query/dictionary.seam>>
 RIDIRE: < <http://www.ridire.it/>>
 Sketch Engine <http://www.sketchengine.co.uk/>
 WinPitch: <https://www.winpitch.com/>