



MERLIN

**Piattaforma multilingue per i livelli di riferimento
europei - contestualizzare il linguaggio degli
apprendenti**

Manuale per l'utente

Versione 1



Questo progetto è stato finanziato con il sostegno della Commissione europea. Questo sito internet riflette solo le opinioni dell'autore e la Commissione non può essere ritenuta responsabile per qualsiasi uso che possa essere fatto delle informazioni ivi contenute.

Citare come: Progetto MERLIN (2014): *Manuale per l'utente*. www.merlin-platform.eu.

Indice

Tabelle riassuntive/grafici

PARTE I: Informazioni di base

1	Il progetto MERLIN: motivazione e obiettivi	5
2	Metodologia.....	7
2.1	Raccolta dei dati	7
2.2	Trascrizione.....	7
2.3	Rivalutazioni	8
2.4	Annotazione manuale	9
2.4.1	Target Hypotheses (TH).....	9
2.4.2	Annotazione delle caratteristiche del linguaggio degli apprendenti	11
2.5	Annotazioni automatizzate	16
2.6	Controllo qualità.....	17

PARTE II: Guida per l'utente

3.	La sezione di documentazione MERLIN.....	20
3.1	<<Usare MERLIN>>.....	21
3.2	<<MERLIN per la ricerca>>	22
3.3	<<Corpus MERLIN>>.....	23
3.4	<<Annotazioni MERLIN>>.....	27
3.5	<<Documentazione>>	29
3.6	<<Scarica l'intero corpus>>	30
4.	Le funzionalità di ricerca MERLIN	31
4.1	<<Crea un sottocorpus>>	32
4.2	<<Ricerca semplice>>.....	36
4.3	<<Ricerca avanzata>>.....	38
4.4	<<Statistiche>>	44
5.	Aiuto	48
5.1	Manuale per l'utente	48
5.2	Familiarizzare con MERLIN: presentazione video	48
5.3	Navigare nell'interfaccia MERLIN: il tutorial screencast	49
5.4	Come adoperare MERLIN: scenari d'uso.....	49
5.5	Comprendere l'output della <<Ricerca avanzata>>.....	49
5.6	Glossario.....	51
5.7	Domande frequenti.....	51

Bibliografia

Tabelle riassuntive/grafici

Parte 1: Informazioni di base	
Tabella 1	Scala esemplificativa QCER, “Repertorio linguistico generale” (CoE 2001: 110)
Grafico 1	Struttura delle annotazioni MERLIN
Tabella 2	Target hypothesis 1 (TH1), esempio
Tabella 3	Fenomeni non catturati dalla TH1, esempio
Tabella 4	Target hypothesis 2 (TH2), esempio
Tabella 5	Caratteristiche del linguaggio degli apprendenti annotate (tag e definizioni)
Part 2: User guide	
Grafico 2	<<documentazione>>
Grafico 3	<<usare MERLIN>>
Grafico 4	<<usare MERLIN>>, <<Usare MERLIN per l’insegnamento delle lingue>>
Grafico 5	<<MERLIN per la ricerca>>
Grafico 6	<< corpus MERLIN>>
Grafico 7	Compiti d’esame MERLIN per l’italiano (<<corpus MERLIN>>)
Grafico 8	Dettaglio della descrizione di un compito (<<MERLIN corpus>>)
Grafico 9	Numero dei testi per ogni livello QCER & numero delle valutazioni per ogni livello QCER (<<corpus MERLIN>>)
Grafico 10	Il corpus MERLIN in cifre: numero di testi con target hypotheses 1 & 2 e error annotation 1 & 2 (<<corpus MERLIN>>)
Grafico 11	<<annotazioni MERLIN>>
Grafico 12	Struttura delle annotazioni MERLIN, corpus completo (<<Annotazioni MERLIN>>)
Grafico 13	Lista di tag di annotazione con esempi (<<Annotazioni MERLIN>>)
Grafico 14	<<documentazione>>
Grafico 15	<<scarica il corpus>>
Grafico 16	Area di ricerca
Grafico 17	Interfaccia di ricerca <<Crea un sottocorpus>>
Grafico 18	Specificazione delle caratteristiche di un sottocorpus, caratteristiche del linguaggio degli apprendenti (<<Crea un sottocorpus>>)
Grafico 19	Esempio di output, <<crea un sottocorpus>>
Grafico 20	Esempio di output per “Visualizza testo dell’apprendente e TH” in <<Crea un sottocorpus>>
Grafico 21	Esempio di output per “Visualizza informazioni sull’apprendente e valutazioni”, nell’output di <<crea un sottocorpus>>
Grafico 22	Interfaccia di ricerca <<ricerca semplice>>
Grafico 23	Esempio di output, <<ricerca semplice>>
Grafico 24	Esempio di output per <<visualizza informazioni sugli apprendenti e valutazioni>> nell’output di <<ricerca semplice>>
Grafico 25	Interfaccia di ricerca <<Ricerca avanzata>>
Grafico 26	Definizione dei tag delle parti del discorso (POS), <<ricerca avanzata>> (dettaglio, italiano)
Grafico 27	Definizione di <<caratteristica - dettaglio>>, <<Ricerca avanzata>> (dettaglio, grammatica)
Grafico 28	Esempio di ricerca di un lemma, <<Ricerca avanzata>>
Grafico 29	Esempio di output, ricerca di un lemma, <<Ricerca avanzata>>
Grafico 30	Esempio di output, ricerca di un lemma, meta informazioni nascoste, <<Ricerca avanzata>>
Grafico 31	Modifica dell’estensione del contesto di un lemma nell’output di <<Ricerca avanzata>>
Grafico 32	Visualizzazione delle annotazioni automatiche nell’output di <<Ricerca avanzata>>
Grafico 33	Interfaccia di ricerca <<Statistiche>>
Grafico 34	Esempio di calcolo di frequenza assoluta (tutti gli errori grammaticali, italiano), <<Statistiche>>
Grafico 35	Esempio di output, calcolo di frequenza assoluta (tutti gli errori grammaticali, italiano), <<Statistiche>>
Grafico 36	Tutti i risultati di un calcolo di frequenza (grafici 34-35), nell’output di <<Ricerca avanzata>>
Grafico 37	Aiuto
Grafico 38	Funzioni di guida nell’output di <<Ricerca avanzata>>
Grafico 39	Funzione di guida <<Sul risultato di questa ricerca>> nell’output di <<ricerca avanzata>>
Grafico 40	Lista di abbreviazioni usate nell’output di <<ricerca avanzata>>

Parte I: Informazioni di base

1 Il progetto MERLIN: motivazione e obiettivi

Introduzione

Il Quadro Comune Europeo di Riferimento per le Lingue (QCER) è il principale strumento per l'insegnamento e la certificazione delle lingue in Europa. Il suo nucleo essenziale, sebbene non sia l'unico contributo del Quadro alla standardizzazione e agli importanti progressi nell'insegnamento e nella valutazione delle lingue, è costituito dal noto sistema di livelli di competenza QCER, descritti attraverso scale esemplificative. Nonostante ne venga fatto ormai ampio uso (è difficile trovare un esame di lingua, un programma scolastico o un manuale che non faccia riferimento ai livelli delle scale), le scale QCER spesso non sono sufficientemente illustrate in termini di dati autentici del linguaggio degli apprendenti. Questa preoccupazione è maggiore se si considerano lingue diverse dall'inglese (cfr. ad es. Fulcher 2004, Hulstijn 2007, North 2000, Wisniewski 2014).

Il progetto MERLIN: "Piattaforma multilingue per i livelli di riferimento europei: contestualizzare il linguaggio degli apprendenti" ha il fine di migliorare questa situazione offrendo un contributo all'illustrazione e alla validazione del sistema di livelli QCER. MERLIN (2012-2014) è stato co-finanziato dall'Unione Europea (Programma di apprendimento permanente, 518989-LLP-1-2011-1-DE-KA2-KA2MP). MERLIN mira a esaminare e ampliare le basi empiriche delle scale QCER mediante l'allestimento di un corpus scritto di apprendenti per il ceco, il tedesco e l'italiano L2 (cfr. Wisniewski et al. 2013, Abel et al. 2014).

Contesto: scale QCER

Il QCER afferma di essere applicabile per tutte le lingue europee. Perciò, le descrizioni dei livelli dovevano essere generali, come nell'esempio seguente:

B1	Dispone di un repertorio linguistico sufficiente per descrivere situazioni non prevedibili, spiegare con ragionevole precisione i punti salienti di un concetto o di un problema ed esprimere pensieri su argomenti astratti o di cultura, quali la musica e i film.
	Dispone di strumenti linguistici e di lessico sufficienti per riuscire ad esprimersi, con qualche esitazione e parafrasi, su argomenti quali la famiglia, gli hobby e gli interessi, il lavoro, i viaggi e l'attualità, ma i limiti lessicali lo/la portano a ripetere e ad avere a volte anche qualche difficoltà di formulazione.
A2	Dispone di un repertorio linguistico elementare che gli/le permette di cavarsela in situazioni correnti di contenuto prevedibile, anche se generalmente deve cercare le parole e semplificare il messaggio.
	È in grado di formulare brevi espressioni di uso corrente per soddisfare semplici bisogni di tipo concreto: dati personali, <i>routine</i> quotidiane, desideri e bisogni, richieste di informazione. È in grado di usare strutture di base ed espressioni memorizzate, gruppi di poche parole e frasi fatte per parlare di se stesso/a e di altre persone, di ciò che si fa, di luoghi e di cose che si possiedono. Dispone di un repertorio limitato di brevi espressioni memorizzate che rispondono a situazioni prevedibili "di sopravvivenza"; nelle situazioni poco usuali si verificano frequenti interruzioni e fraintendimenti.
A1	Dispone di un repertorio molto elementare formato da espressioni semplici relative a dati personali e bisogni di tipo concreto.

Tabella 1: Scala esemplificativa QCER, "Repertorio linguistico generale" (CoE 2001: 110)

Per MERLIN sono state utilizzate le scale del capitolo 5 ('Competenze linguistico-comunicative'): repertorio linguistico generale | correttezza grammaticale | ampiezza del lessico | padronanza lessicale | ortografia | coerenza & coesione | appropriatezza sociolinguistica. Il QCER può essere scaricato dal sito internet del Consiglio d'Europa.¹

Illustrazione dei livelli QCER

D'altra parte, ci si è resi conto della necessità di arricchire i descrittori con illustrazioni aggiuntive e specifiche per ogni lingua. In considerazione di questa esigenza di integrare il QCER, fin dal 2001 lo stesso Consiglio d'Europa ha incoraggiato lo sviluppo di strumenti di supporto che potessero meglio illustrare le caratteristiche delle singole lingue. Un passo in questa direzione è stato quello di promuovere la pubblicazione delle *Descrizioni dei Livelli di Riferimento* (DLR) per le lingue nazionali e regionali.

Sempre più DLR tendono a essere basati su corpora di apprendenti, come nel caso dell'inglese (www.englishprofile.org), ma anche dell'italiano (Spinelli/Parizzi 2010) e dei Profili norvegesi (Carlsen 2013).

Anche MERLIN mira a illustrare i livelli QCER per determinate lingue, ma allo stesso tempo si distingue in quanto segue, per la prima volta, un approccio multilingue. Si rivolge quindi a tre lingue appartenenti a famiglie diverse (lingue slave, germaniche e romanze) e incoraggia il confronto interlinguistico. Inoltre, MERLIN si differenzia da altre iniziative di questo genere poiché offre libero accesso ai testi completi, ai compiti d'esame e a un'ampia gamma di annotazioni linguistiche e degli errori, mediante una piattaforma online orientata alla didattica. MERLIN inoltre contribuisce anche alla validazione delle scale QCER.

Validazione delle descrizioni dei livelli QCER

Lo sforzo del Consiglio d'Europa di classificare i descrittori del QCER (CoE 2001; North 2000; Schneider/North 2000) ha portato a enormi miglioramenti in termini di standardizzazione e trasparenza nell'apprendimento, nell'insegnamento e nella valutazione delle lingue straniere. Decisioni importanti per gli apprendenti di lingue sono prese facendo riferimento ai livelli del QCER. Un aspetto cruciale che è ancora troppo poco compreso è la validità empirica delle scale del QCER (Fulcher 2004; Hulstijn 2007): se le scale sono usate per descrivere o valutare il linguaggio degli apprendenti, esse devono riflettere ciò che gli apprendenti realmente fanno (Alderson 1991).

Dal momento che i livelli QCER sono sempre più usati in contesti ad alto rischio, nei quali decisioni importanti per le vite delle persone dipendono dall'interpretazione delle scale del QCER (ad es. ammissione all'università, acquisizione di cittadinanza), è particolarmente importante essere certi che le scale rispecchino realmente il linguaggio degli apprendenti. Da questo punto di vista, pochissima ricerca è stata condotta (cfr. ad es. Alderson et al. 2006;

¹ http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Alderson 2007; Fulcher 2004; Hulstijn 2007; Hulstijn et al. 2010; Little 2007; Wisniewski 2013, 2014).

La taratura delle scale del QCER si basa sulle opinioni espresse da professionisti del settore, attraverso valutazioni, riguardo la competenza della lingua seconda. Tuttavia, non è chiaro fino a che grado le valutazioni riflettano realmente i contenuti delle scale (Arras 2010; Eckes 2008; Pollitt/Murray 1996; Vaughan 1991). Non è stata effettuata alcuna analisi del linguaggio degli apprendenti durante il processo di classificazione del QCER per supportarne la validità empirica. MERLIN ha il fine di contribuire alla ricerca in merito alla validità delle scale QCER. I correlati linguistici ai contenuti delle scale principali del capitolo 5 sono stati operazionalizzati e sono ricercabili nell'interfaccia.

La ricerca sull'elaborazione del linguaggio naturale (NLP)

Il corpus MERLIN fornisce dati preziosi per lo sviluppo e la valutazione di strumenti per l'elaborazione del linguaggio naturale applicati al linguaggio degli apprendenti (Meurers 2012). Il corpus e le meta-informazioni sugli apprendenti e sulle valutazioni in esso contenute sostengono in modo immediato la ricerca sull'identificazione automatica della lingua madre, consentendole di superare l'attuale focus sull'apprendente di madrelingua inglese. Analogamente, il corpus è stato già utilizzato per la ricerca sulla classificazione automatica delle competenze per il tedesco (Hancke 2013). Il corpus MERLIN fornisce dati riccamente annotati anche per lo sviluppo e l'adattamento di strumenti di NLP e di applicazioni che aiutino gli apprendenti di lingue a migliorare l'uso del lessico, la coerenza, l'ortografia e la correttezza grammaticale.

2 Metodologia

2.1 Raccolta dei dati

I testi MERLIN sono tratti dalle sezioni di produzione scritta di esami standardizzati di alta qualità collegati al QCER e forniti da telc (Francoforte sul Meno, esami di italiano e di tedesco, www.telc.net) e ÚJOP (Praga, esami di ceco, www.ujop.cuni.cz). Queste istituzioni sono membri dell'ALTE (www.alte.org). I compiti sono stati utilizzati fino al 2013 e sono ora disponibili gratuitamente sulla piattaforma.

2.2 Trascrizione

I testi originali, scritti a mano dagli apprendenti, sono stati trascritti utilizzando un editor xml (xml mind©) dagli enti certificatori (telc and ÚJOP). I trascrittori hanno seguito delle linee guida per la trascrizione (disponibili nell'interfaccia) e l'affidabilità delle trascrizioni è stata verificata, inizialmente per un campione del 5% dei testi per ogni livello QCER. Dal momento

che sono stati rilevati molti errori di trascrizione, alla fine quasi tutti i testi sono stati sottoposti a una fase di revisione.

Nelle linee guida per la trascrizione sono stati inclusi tag (annotazione in linea) per caratteristiche testuali di base come tratti linguistici illeggibili o ambigui, parole in lingua straniera, emoticon, immagini, capoversi, parole copiate dalle tracce o formule di saluto. L'anonimizzazione (nomi, luoghi) ha fatto parte del processo di trascrizione ed è stata portata a termine in base alle linee guida. Le trascrizioni hanno fatto da base per le annotazioni (vedi sotto). Le linee guida per la trascrizione sono disponibili sulla piattaforma MERLIN (<<documentazione>>, solo in tedesco).

2.3 Rivalutazioni

Negli esami dai quali sono stati tratti i testi MERLIN, i candidati hanno ricevuto un punteggio che ha assunto un peso diverso a seconda dell'importanza attribuita alla produzione scritta in quel determinato esame, portando a una valutazione complessiva di promozione o bocciatura. Per MERLIN la procedura è stata necessariamente diversa: l'obiettivo era avere una relazione diretta tra i testi e le scale relative alla competenza linguistico-comunicativa presentate nel capitolo 5 del QCER. Pertanto, tutti i testi sono stati rivalutati in modo autonomo da valutatori professionisti.

L'affidabilità delle rivalutazioni è stata esaminata con l'aiuto della Teoria Classica dei Test e di un'analisi Multi-Facet Rasch. Quest'ultima è una procedura statistico-probabilistica, usata spesso nella verifica delle competenze linguistiche, che consente la correzione di tendenze di valutazione (ad es. indulgenza/severità) e rende possibile l'ottenimento di una valutazione media giusta per ogni testo. L'affidabilità dei/tra i valutatori è risultata in genere molto alta in MERLIN, con alcune eccezioni per l'italiano. Perciò l'intero processo di rivalutazione è stato ripetuto per l'italiano, portando infine ad una qualità di valutazione soddisfacente. Maggiori dettagli sono disponibili nella relazione tecnica (vedi la sezione <<documentazione>>).

In MERLIN è stata usata una scala olistica (basata sulla scala QCER relativa al repertorio linguistico generale) insieme a una griglia di valutazione analitica (criteri di valutazione: ortografia | correttezza grammaticale | ampiezza del lessico | padronanza lessicale | coerenza & coesione | appropriatezza sociolinguistica) con valori da A1 a C2. Entrambi gli strumenti possono essere scaricati dalla sezione MERLIN <<documentazione>>. La media giusta è calcolata in base alla scala olistica. Se si compila il proprio corpus in base ai livelli QCER, questi si basano anche sulle valutazioni medie imparziali. Nell'interfaccia puoi accedere anche a un profilo di valutazione con le valutazioni originali dei testi in base a questi criteri.

Tieni presente che molti candidati hanno sostenuto un esame che poi si è rivelato essere troppo difficile o troppo facile per loro. Per questo, sulla piattaforma MERLIN è stata fatta una distinzione tra il *livello QCER di un test* e il *livello QCER delle valutazioni*, i quali non necessariamente coincidono (vedi anche la tabella 9 di seguito). Entrambi possono essere ricercati separatamente (ad es. puoi cercare i testi di apprendenti tedesco che hanno sostenuto un test B1 ma che hanno ricevuto soltanto una valutazione A2).

2.4 Annotazione manuale

L'annotazione è uno degli aspetti centrali del progetto MERLIN. MERLIN ha due tipi di annotazioni: **“target hypotheses”** e annotazioni delle **caratteristiche del linguaggio degli apprendenti**. Dove possibile, sono state usate **procedure automatiche** (vedi 1.4), tuttavia la maggior parte delle annotazioni è stata portata a termine **manualmente**.

L'annotazione è stata organizzata in 2 blocchi:

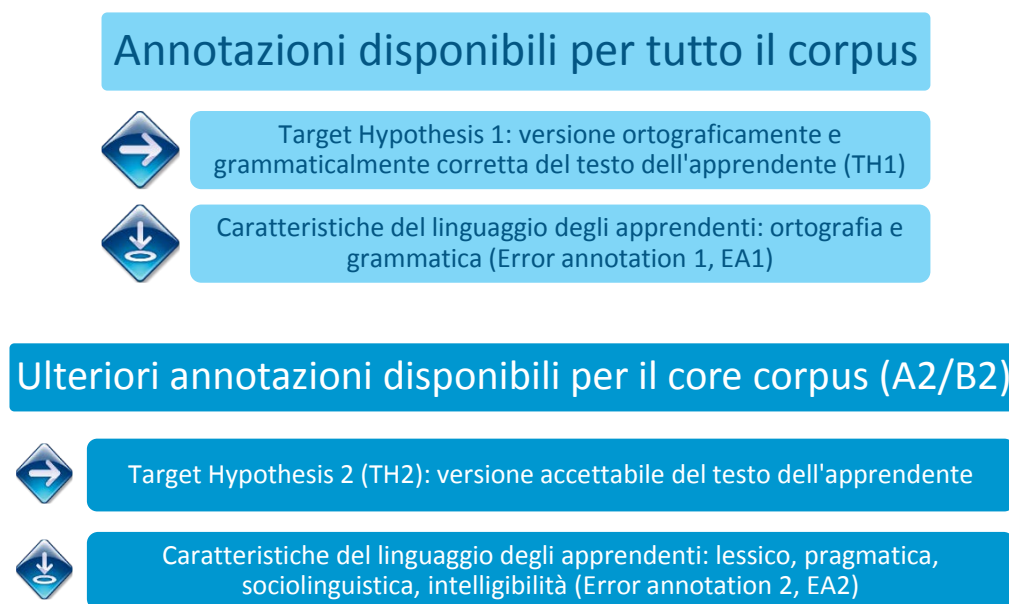


Grafico 1: Struttura delle annotazioni MERLIN

2.4.1 Target Hypotheses (TH)

La ricerca ha evidenziato come l'annotazione del linguaggio degli apprendenti sia uno sforzo complesso e in parte speculativo. Qualsiasi marcatura di un fenomeno del linguaggio degli apprendenti richiede un'interpretazione mentale da parte dell'annotatore. Per garantire la trasparenza, la coerenza e l'affidabilità delle annotazioni, è una buona idea formulare prima una **“target hypothesis”** (TH), cioè una ricostruzione corretta del testo dell'apprendente, sulla quale può basarsi una successiva annotazione degli errori (Reznicek/Lüdeling et al. 2012). Inoltre, le target hypotheses sono necessarie per la corretta applicazione di molte analisi automatizzate (Díaz-Negrillo et al. 2010, Hirschmann et al. 2009). Infine, possono aiutare futuri utenti della piattaforma MERLIN a comprendere le annotazioni. MERLIN collabora con il progetto Falko ²(Humboldt University, Berlino), che rappresenta una delle pochissime iniziative di corpus che si focalizza sulle target hypotheses e fornisce accesso gratuito ai dati. In MERLIN ci sono due tipi di TH (TH1 e TH2) di cui si tratterà brevemente nei prossimi paragrafi.

Target hypothesis 1

² <https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko>

Le target hypotheses per errori di ortografia e grammatica (TH1) sono state formulate per l'intero corpus MERLIN. Nella formulazione della TH1, agli annotatori è stato richiesto di modificare il meno possibile il testo originale dell'apprendente in modo da produrne una versione grammaticalmente e ortograficamente corretta (TH "minima"). Se ne può trovare un esempio in questa tabella:

learner	Caro	Daniele	,	io	sono	in	Italia	per	studio	bene	il	italiano	.
TH1	Caro	Daniele	,	io	sono	in	Italia	per	studiare	bene	l'	italiano	.
TH1Diff				CHA					CHA		CHA		

Tabella 2: Target hypothesis 1 (TH1), esempio

Il seguente esempio mostra come nella TH1 siano stati ignorati errori riguardanti altre aree linguistiche, per varie ragioni tecniche e di contenuto.

learner	È	possibile	di	incontrarci	?		Spero	!	Dai	mi	un	telefono	!
TH1	È	possibile		incontrarci	?		Spero	!	Dammi		un	telefono	!
TH1Diff									MERGE/CHA				
TH1Diff			DEL						DEL	DEL			

Tabella 3: Fenomeni non catturati dalla TH1, esempio

Mentre gli errori ortografici (limite di parola) e grammaticali (uso della preposizione) sono corretti in TH1, le forme erranee sul piano lessicale ("Dammi un telefono" invece di "Fammi una telefonata") non sono state sostituite con altri lessemi.

Target hypothesis 2

Su un secondo livello, le target hypotheses estese (TH2) si riferiscono ad aspetti sociolinguistici, lessicali e pragmatici che costituiscono deviazioni rispetto a ciò che ci si aspetterebbe da un parlante nativo. La TH2 quindi mira a produrre una versione *accettabile* del testo originale dell'apprendente. Per la TH2 sono presi in considerazione aspetti contestuali. La TH2 è un'estensione della TH1.

Rispetto alla TH1, la TH2 comporta maggiore soggettività e difficoltà nell'ottenere decisioni affidabili. Questo è un ulteriore motivo per cui i due livelli sono stati separati. L'esempio seguente può risultare utile per illustrare la differenza tra TH1 e TH2:

learner	Abitiamo	in	l'	albergo	"	PALAZZO	"	.
TH1	Abitiamo	nell'		albergo	"	PALAZZO	"	.
TH1Diff		MERGE/CHA						
TH2	Alloggiamo	nell'		albergo	"	PALAZZO	"	.
TH2Diff	CHA	MERGE/CHA						

Tabella 4: Target hypothesis 2 (TH2), esempio

L'esempio mostra chiaramente che la frase già annotata in TH1 necessita di un'ulteriore annotazione lessicale ("Alloggiamo" invece di "Abitiamo").

2.4.2 Annotazione delle caratteristiche del linguaggio degli apprendenti

Un principio guida importante in MERLIN è la visione del linguaggio degli apprendenti come un sistema a sé, il quale non può essere descritto in modo soddisfacente utilizzando solo tag di errore orientati al deficit. È importante sottolineare che l'annotazione MERLIN include molti tag di errore, tuttavia, soprattutto nella EA2, ci sono tag di annotazione che contrassegnano fenomeni che non costituiscono errori, ad es. la realizzazione dell'atto linguistico "RICHIESTA".

Lo schema di annotazione MERLIN rappresenta dunque una selezione di caratteristiche significative, valide e possibili ("tag") che vengono annotate manualmente e che sono supportate dal team di linguisti computazionali MERLIN. Lo schema contiene le seguenti caratteristiche:

G_ Grammatica

G_Agr	Accordo (soggetto e verbo)
G_Art	articolo
G_Clit	ITA: clitico
G_Conj	coniunzione
G_Infect_inexist	forma flessa inesistente (sostantivi, aggettivi, verbi)
G_Morphol_wrong	Forma flessa scorretta (nomi, pronomi, aggettivi)
G_Neg_negdoub	CZE: doppia negazione
G_Neg_neggen	Negazione generale
G_POS	Errore di parte del discorso
G_Prep	preposizione
G_Refl_pronrefl	pronome riflessivo
G_Refl_pronreflposs	CZE: pronome possessivo riflessivo
G_Valency_complnumb	valenza verbale: numero di argomenti obbligatori
G_Verb_asp	verbo: aspetto (CZE+ITA)
G_Verb_compl	verbo: predicati complessi
G_Verb_main	verbo principale
G_Verb_md	verbo: modo
G_Verb_tns	verbo: tempo
G_Verb_vc	verbo: voce
G_Wo_womaincl	ordine delle parole nella frase principale
G_Wo_wosubcl	ordine delle parole nella frase subordinata

O_ Ortografia

O_Abbrev	abbreviazione
O_Apostr	GER+ITA: apostrofo
O_Capit	uso delle maiuscole
O_Graph_act	CZE+ITA: segni diacritici

O_Graph_graphgen	errore generale di grafema
O_Graph_trans	grafema: trasposizione
O_Punct	punteggiatura
O_Wordbd	limite di parola

G_ Intelligibilità

G_Intelltxt	intelligibilità del testo
G_Intells	intelligibilità della frase

V_ Vocabulary

V_FS	sequenza formulare
V_Sequence	sequenza incomprensibile a causa di un'accumulazione di
_lexgrammerr_incompr	errori lessicali/grammaticali
V_form_nonexist	forma inesistente (parola/sequenza formulare)
V_FS_form_incompr	sequenza formulare: intelligibilità limitata
V_semdenot	errore semantico: denotazione (parola o sequenza formulare)
V_semcon_att	errore semantico: connotazione (parola o sequenza formulare)
V_Word_semimprec	errore semantico: precisione (parola o sequenza formulare)
V_Wordform_deriv	errore di formazione delle parole: derivazione
V_Wordform_comp	errore di formazione delle parole: composizione
V_FS_form	sequenza formulare: errore formale

C_ Coerenza/Coesione

C_Con_accur	connettivi: accuratezza
C_Coh_jump	salti di contenuto
C_Coh_ref	espressioni referenziali
C_Coh_txtstruct	connettivi metatestuali

S_ Appropriatazza sociolinguistica

S_Txt_grfw	formule di saluto
S_Txt_opcl	formule di apertura/chiusura
S_Form_gen	stile inappropriato (formalità)

S_Form_addr	allocuzione inappropriata (formalità)
S_Var_clit	ITA: verbi procomplementari
S_Var_duppron	ITA: duplicazione pronominale
S_Var_synstr	ITA: strutture sintattiche marcate
S_Var_che	ITA: 'che polivalente'
S_Var_woweil	GER: frase principale dopo 'weil' (ordine delle parole)
S_Var_partik	GER: particelle modali

P_ Pragmatica

P_Pol_dir	cortesia – forma linguistica troppo diretta
P_Request_direct	RICHIESTA diretta
P_Request_indirect	RICHIESTA indiretta

Tabella 5: Caratteristiche del linguaggio degli apprendenti annotate (tag e definizioni)

Questi tag sono stati scelti da una lista comprensiva delle caratteristiche e degli indicatori che sono stati messi insieme nell'ambito del progetto. Le possibili annotazioni **derivano...**:

- 1) dalla prospettiva degli utenti della piattaforma (in base a uno studio degli utenti e ad analisi di manuali ed esami di lingua)
- 2) dal QCER (attraverso l'operazionalizzazione degli elementi delle scale del capitolo 5, CoE 2001)
- 3) dalla ricerca sull'acquisizione della seconda lingua (in base a un esame approfondito della letteratura di ricerca)
- 4) dai testi degli apprendenti (in base a un'analisi induttiva del 10% di tutti i testi degli apprendenti)

Ci sono molte fonti di informazione riguardanti le annotazioni, alle quali chi utilizza MERLIN può accedere liberamente:

- la **bibliografia** che è stata usata in MERLIN per scegliere tag di annotazione significativi, nella sezione bibliografica di questo manuale
- lo **schema di annotazione** con tutti i tag di annotazione che sono stati implementati in MERLIN dopo un controllo di fattibilità di un elenco esaustivo di annotazioni (disponibile per il **download**)
- documenti come il **manuale degli annotatori MERLIN** (disponibile in inglese) e una **documentazione di ulteriori questioni di annotazione**, con decisioni molto specifiche riguardo singoli fenomeni di annotazione, per tutte e tre le lingue del progetto, e altri materiali che si possono trovare nell'interfaccia (disponibili per il **download**)
- funzioni di guida nell'interfaccia, come una **lista di tag di annotazione con esempi, screenshot** che chiariscono i livelli di annotazione, un **glossario**, e molto altro
- un **libro nero** che riassume le esperienze legate all'annotazione che progetti futuri in questo campo potrebbero voler evitare (disponibile per il **download**)

Tag di annotazione suggeriti dagli esperti

MERLIN è stato creato per i professionisti che lavorano con il QCER. Per questo è importante rendere le annotazioni il più possibile utili per loro. A questo scopo è stato portato a termine uno **studio degli utenti**, i cui dettagli sono descritti in due relazioni, disponibili nell'interfaccia (*<<documentazione>>*). La prima parte si concentra sull'utilità di aspetti relativi ai contenuti, mentre la seconda parte prende in considerazione il lato tecnico legato alla fruibilità della piattaforma MERLIN. Lo studio degli utenti ha fornito importanti informazioni sui tag di annotazione da includere. Lo schema di annotazione contiene informazioni riguardo ai tag derivati dallo studio degli utenti.

Un'ulteriore possibilità di comprendere le esigenze degli utenti è data dall'integrazione, nello schema di annotazione MERLIN, di aspetti dell'acquisizione della L2 che sono comunemente trattati in **libri di testo ed esami di lingua**. Tra i libri analizzati in MERLIN, ad esempio, vi erano "Tangram" per il tedesco (Dallapiazza 1998), "Rete!" per l'italiano (Mezzadri 2000) e "Brána jazyka českého otevřená" per il ceco (Hasil 2007). Inoltre, l'analisi di test linguistici somministrati da UJOP e telc ha evidenziato alcuni ben noti argomenti che hanno fornito informazioni per lo schema di annotazione.

Esempi di annotazioni derivate da queste analisi includono errori ortografici come l'uso errato delle maiuscole, l'uso scorretto dell'apostrofo in tedesco e in italiano, o ancora errori grammaticali come l'aspetto verbale in italiano. A livello lessicale, temi ricorrenti che sono riflessi nello schema di annotazione MERLIN sono i "falsi amici" o l'uso delle espressioni idiomatiche.

Testi degli apprendenti come fonte per tag di annotazione

Il 10% di tutti i testi degli apprendenti presenti nel corpus sono stati analizzati manualmente. Quest'approccio qualitativo e induttivo ha rivelato una serie di fenomeni che sono stati poi considerati degni di essere integrati nello schema di annotazione. Ne sono esempi gli errori di formazione delle parole, diversi tipi di problemi legati all'uso di sequenze formulari, problemi riguardanti la scelta di un registro appropriato in termini di formalità o cortesia. Curiosamente, la maggior parte di questi fenomeni appartiene a categorie diverse da grammatica e ortografia.

Annotazione basata sulla ricerca

Grande impegno è stato dedicato a un esame approfondito della letteratura di ricerca relativa alle diverse aree linguistiche interessate dalle annotazioni MERLIN. Non è possibile affrontare qui una discussione completa sui singoli tag, per questo ne viene offerta una breve panoramica. Le fonti citate (e molte altre) sono riportate nella bibliografia.

Le annotazioni di fenomeni grammaticali includono accordo, ordine delle parole, negazione, errori di parte del discorso e molto di più. Perciò, possono essere calcolati diversi parametri di accuratezza e complessità (vedi Lu 2010, 2011; Wolfe-Quintero et al. 1998, Ortega 2003, 2012, Housen/Kuiken 2009). **L'ortografia** è un'area sulla quale la ricerca non si concentra molto. Le annotazioni MERLIN permettono di accedere alla frequenza relativa/assoluta di diversi tipi di errori ortografici riguardanti, ad esempio, la punteggiatura, l'uso delle maiuscole o i segni diacritici (Al-Jarf 2009, Bredel 2010, Cook 2005,

Granger/Bestgen 2011, Perfetti/Rieben/Fayol 1997, Rimrott/Heift 2008, Sassoon 1995). Queste annotazioni sono disponibili per l'intero corpus.

Le annotazioni pilota del core corpus MERLIN (EA2)

Per un piccolo sottocorpus sono stati annotati anche aspetti relativi ad altre aree linguistiche (EA2, error annotation 2). I testi di questo sottocorpus hanno ottenuto TH1 ed EA1 e, in aggiunta, TH2 ed EA2. Sarebbe auspicabile, in futuro, che queste annotazioni esplorative pilota siano riesaminate per poter poi essere in grado di dotare l'intero database MERLIN di annotazioni TH2 ed EA2. Per ora, preghiamo gli utenti di impiegare con cautela le annotazioni del core corpus.

Tag di annotazione per il **lessico** prendono in considerazione le molteplici dimensioni della conoscenza lessicale come l'accuratezza, la profondità, l'ampiezza e la raffinatezza (Nation 2001, 2007, Read 2000). Qui, MERLIN pone un grande accento sulle sequenze formulari (Wray 2002), che assumono un ruolo particolare nell'apprendimento di lingue straniere (e.g., Pawley/Syder 1987, Schmitt et al. 2004). Diversi tag lessicali riguardanti le sequenze formulari non identificano errori, ma mirano a evidenziare strutture di particolare interesse nel processo di acquisizione linguistica.

La competenza sociolinguistica è definita come 'la capacità di riconoscere e produrre linguaggio socialmente appropriato al contesto' (Lyster 1994: 263). Come in generale per tutte le annotazioni EA2, l'annotazione di fenomeni attinenti a questa competenza (soggettività, affidabilità) è metodologicamente impegnativa. I tag applicati sono in linea con il progetto e non si attengono sempre a ciò che di solito viene analizzato in questo campo (ad es. abilità e quantità del contatto linguistico, o variazione diasistemica in testi di apprendenti confrontata con la variazione nella L1, Baker 2010, Baylea 2007, Bayley/Regan 2004, Biber/Finegan 1994, Hudson et al. 1005, Hymes 1974, Mougeaon/Dewaele 2004, Regan et al. 2009, Van Compernelle/Williams 2012, Yu 2012, Zuskun 1992). In MERLIN, è annotata l'appropriatezza di forme linguistiche con riferimento alla 'formalità' (ad es., forme al di sotto dello standard, uso di forme tipiche della comunicazione orale, linguaggio eccessivamente formale), e sono integrati tag che sono fatti su misura per la tipologia di testo, come saluti e formule di apertura e chiusura. I singoli aspetti di variazione scelti, specifici per ogni lingua, riflettono scelte di strutture che non riguardano la lingua scritta e/o che sembrano essere sulla via dell'accettazione come varianti dello standard, ma che non verrebbero normalmente accettate nelle tipologie di compiti nei quali possono essere trovate nei testi MERLIN.

Nell'area della **pragmatica, la coerenza/coesione** è annotata principalmente mediante una marcatura dei connettivi metatestuali, indipendentemente dalla correttezza, dai connettori e dall'annotazione di problemi sul piano referenziale del testo. (Bachmann 2002, Halliday/Hasan 1976, 1989, Castro 2004, Carlsen 2010, Chiang 2003, Cornish 2009, Fabricius-Hansen 2005, Louwerse/Graesser 2004, McNamara et al. 1996, McNamara/Kintsch 1996, Spooren/Sanders 2008). Inoltre è annotato l'atto linguistico di RICHIESTA (Al-Gahtani/Roever 2012, Bardovi-Harlig 2013, Barron 2003, Blum-Kulka 1987, 1991, Blum-Kulka/Olshtain 1984, Cho 2005, Held 1995, Nuzzo 2007, Trosborg 1995, Veddersen 2007).

Annotazione basata sul QCER

Per scoprire se le scale del QCER riflettono il linguaggio degli apprendenti è importante operazionalizzarne i descrittori senza utilizzare valutazioni umane, che spesso si sono

rivelate non essere basate su strumenti di valutazione anche quando questi sono affidabili. (Eckes 2008, Wisniewski 2010). Se i descrittori delle scale vengono espressi in una forma misurabile come questa, la relazione tra le scale selezionate del QCER e il linguaggio degli apprendenti diventa molto più chiara.

In questo processo di operazionalizzazione, si sono dovuti escludere i termini estremamente vaghi, autoreferenziali o soggettivi nelle descrizioni dei livelli (ad es. “È in grado di interagire con parlanti nativi senza [...] metterli nella necessità di comportarsi in modo diverso da come farebbero con un parlante nativo”, appropriatezza sociolinguistica, B2, CoE 2001: 122), ma anche aspetti chiaramente legati alla lingua parlata sono stati ignorati (ad es. “Riesce... a intervenire in una discussione prendendovi parte [...]”, scala dell’appropriatezza sociolinguistica, B2, CoE 2001: 122) (cfr. Wisniewski 2013, 2014). Se, tuttavia, la descrizione di un livello fa riferimento ai “saluti”, ai “salti nel contenuto”, all’“intelligibilità”, alle “espressioni idiomatiche” o alle “locuzioni” come caratteristiche di determinati livelli del QCER, queste caratteristiche sono state sottoposte a una verifica di fattibilità nello schema di annotazione MERLIN, anche se queste cosiddette “variabili delle scale” possono non avere un ruolo nella ricerca o sono spesso difficili da definire. Queste annotazioni permettono di controllare la rilevanza empirica delle scale QCER coinvolte. Sarebbe un segno di validità empirica se i contenuti delle scale fossero sufficientemente rilevanti e osservabili in modo affidabile nelle prestazioni degli apprendenti. MERLIN non può offrire una validazione completa delle scale QCER, ma si focalizza su una selezione di aspetti significativi.

Accedere alle annotazioni nell’interfaccia

Si può accedere direttamente alle annotazioni attraverso le funzioni di ricerca sull’interfaccia (<<ricerca avanzata>>, <<crea un sottocorpus>>), dove ogni loro occorrenza è mostrata nel contesto.

Un’altra possibilità per accedere alle annotazioni è usarle per misure statistiche. La possibilità più semplice è quella di contare il numero totale di tag di annotazione che ricorrono in un (sotto)corpus. In molti casi, tuttavia, ha più senso usare le annotazioni per il calcolo di parametri normalizzati (cioè, in MERLIN, per frase, T-unit, o token). Ciò permette, ad esempio, di confrontare il numero medio di errori morfologici per frase tra i testi degli apprendenti di ceco B1 e ceco B2. I parametri normalizzati dei fenomeni annotati manualmente sono calcolati sulla base di procedure automatizzate di segmentazione (vedi sotto). I parametri basati sulla frequenza sono disponibili nella sezione <<statistiche>> dell’interfaccia, insieme ad altri più complessi parametri di complessità.

2.5 Annotazioni automatizzate

L’annotazione automatica nel corpus MERLIN serve da supporto all’annotazione manuale e a rendere accessibile un’ampia gamma di caratteristiche linguistiche per il calcolo di indicatori e nelle ricerche dirette nel corpus.

Le annotazioni automatiche per il corpus MERLIN possono essere divise in quattro categorie:

1. Unità linguistiche necessarie per l’annotazione manuale:
 - a) token
 - b) frasi
2. Unità linguistiche necessarie per il calcolo dei parametri:

- a) t-unit
- b) vari tipi di proposizioni
- 3. Annotazioni linguistiche che utilizzano definizioni e strumenti esistenti
 - a) parti del discorso
 - b) lemmi
 - c) parse di costituenza e dipendenza
- 4. Annotazione linguistica con definizioni e strumenti specifici di MERLIN
 - a) ripetizioni all'interno dei testi
 - b) citazioni delle tracce dei compiti

Abbiamo impiegato strumenti automatici di annotazione sviluppati appositamente per le lingue target, al fine di ampliare la gamma di annotazioni linguistiche disponibili al di là di quanto sarebbe stato possibile con l'annotazione manuale, che è costosa e porta via molto tempo. Tuttavia, è importante tenere presente che l'annotazione automatica è particolarmente impegnativa nel caso del linguaggio degli apprendenti, dato che la lingua dell'apprendente spesso devia considerevolmente dalla lingua target in tutti i livelli di analisi linguistica, dall'ortografia alla semantica.

Per tutte e tre le lingue MERLIN sono stati usati i seguenti strumenti:

I testi sono stati tokenizzati usando il *tokenizzatore per le lingue indoeuropee* di LingPipe e la risultante tokenizzazione è stata corretta a mano. Le frasi sono state annotate con il *segmentatore di frasi OpenNLP*. Le ripetizioni sono state identificate usando la *biblioteca Saphre* sulla base dell'annotazione automatica di parti del discorso e lemmi.

Si prega di ricorrere alla sezione <<MERLIN per la ricerca>> per sapere di più sugli strumenti specifici per ogni lingua usati nell'annotazione automatica.

2.6 Controllo qualità

Per poter organizzare e controllare l'affidabilità dell'annotazione, sono state prese diverse misure. Tutti gli strumenti (regole di TH1 e TH2, schema di annotazione per EA1 e EA2) sono stati sottoposti ad una fase di piloting e revisionati prima della loro applicazione. Il piloting è stato organizzato in due stadi. Innanzitutto, la procedura di annotazione stessa è stata testata per avere una prima idea su come concretizzare e cambiare le linee guida e gli strumenti di annotazione. Poi, in un processo di piloting, un numero ristretto di testi è stato annotato da tutti gli annotatori di ogni team linguistico per poter ancora una volta controllare la praticabilità e l'adeguatezza delle linee guida per l'annotazione e rilevare possibili problemi tecnici.

In secondo luogo, tutte le annotazioni sono basate su linee guida (manuale per l'annotatore, vedi interfaccia). Le linee guida sono state arricchite con decisioni molto specifiche riguardo singoli aspetti dell'annotazione (documentazione relativa alle decisioni sulle annotazioni, vedi interfaccia). Infine, l'affidabilità delle annotazioni è controllata: il controllo è avvenuto per il 5% dei testi degli esami di ciascun livello, per le target hypotheses (1 e 2) e le error annotation (1 e 2). Sono stati applicati diversi metodi:

Secondo un approccio di tipo qualitativo, metà dei file sono stati annotati in modo autonomo dagli annotatori per poi essere discussi comunemente col fine di raggiungere un **consenso**. Questi testi sono serviti come riferimento nel corso dell'intero processo di annotazione. L'approccio qualitativo si è rivelato di fondamentale importanza per una comune interpretazione dello schema di annotazione. Mediante una procedura in doppio cieco, la seconda metà dei file sottoposti al controllo di affidabilità è stata annotata da tutti gli annotatori a loro insaputa. Per questi file è stata verificata l'affidabilità sia a livello qualitativo che quantitativo.

PARTE II: Guida per l'utente

3. La sezione di documentazione MERLIN

La macrostruttura della piattaforma MERLIN, liberamente accessibile, è organizzata in un'area di **documentazione** (posizionata verticalmente a sinistra nell'interfaccia, vedi grafico 2) e in un'area di **ricerca** (barre orizzontali di ricerca, vedi grafico 23). Il capitolo 3 si sofferma sui contenuti più importanti e le funzioni della sezione di documentazione.




Grafico 2: <<documentazione>>

3.1 <<Usare MERLIN>>



Grafico 3: <<usare MERLIN>>

Nella sezione <<usare MERLIN>>, puoi ottenere consigli generali sulle possibilità di applicazione di MERLIN in diversi ambienti professionali. Questa sezione è utile per farsi una prima idea dei contenuti del progetto.

Puoi aprire i capitoli (cliccando su ) per scoprire di più, ad esempio, su come usare MERLIN per l'insegnamento delle lingue o per l'elaborazione di materiali didattici.

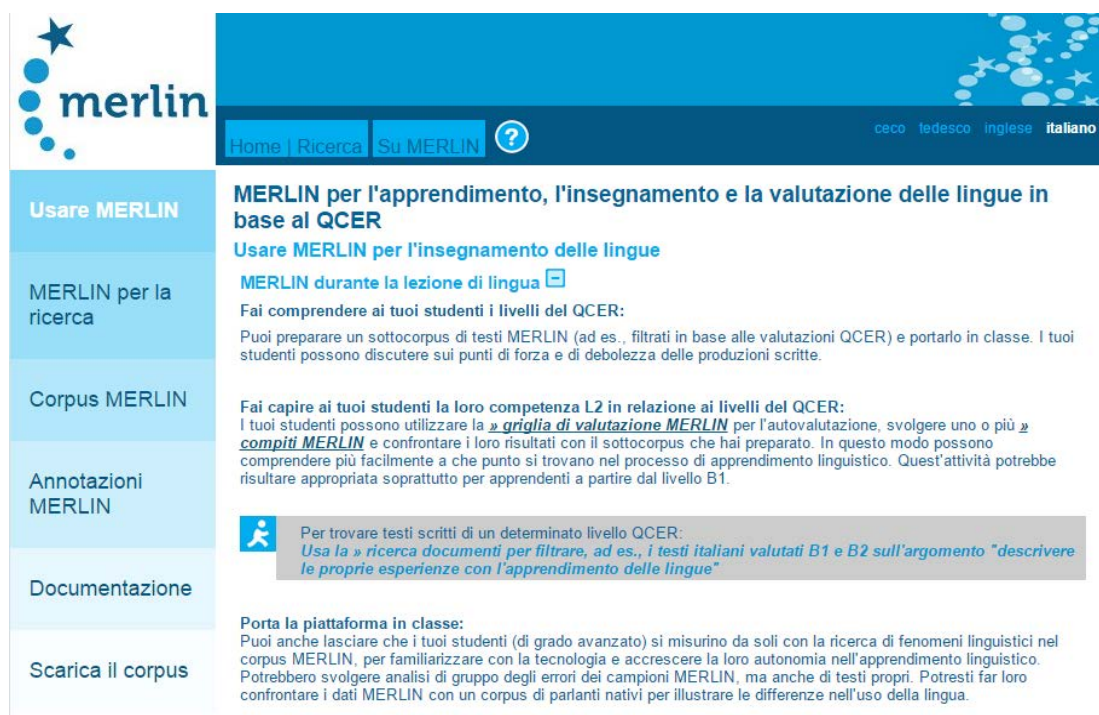


Grafico 4: <<usare MERLIN>>, <<Usare MERLIN per l'insegnamento delle lingue>>

3.2 <<MERLIN per la ricerca>>

La sezione <<MERLIN per la ricerca>> è destinata a chiunque sia interessato a informazioni di base riguardanti diversi aspetti del progetto. Se navigando nell'interfaccia vuoi scoprire di più sulle ragioni fondamentali del progetto, ti consigliamo di consultare la sezione <<MERLIN per la ricerca>>.

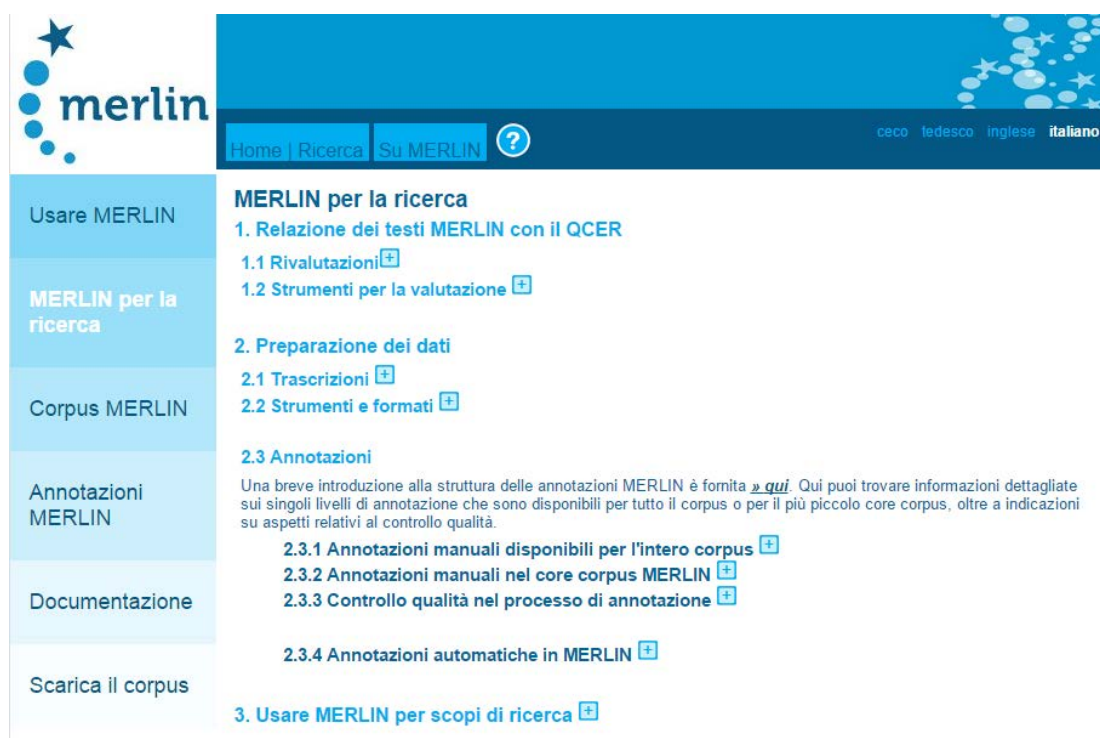


Grafico 5: <<MERLIN per la ricerca>>

Come mostrato nel grafico 5, c'è un gran numero di informazioni riguardo il **collegamento dei testi MERLIN al QCER**. Da qui sono accessibili tutti i documenti attinenti, come le griglie di valutazione, i compiti, la relazione tecnica sulla qualità delle valutazioni.

In questa sezione è inoltre delineato l'andamento del processo al quale i dati MERLIN sono stati sottoposti: puoi scoprire come sono state portate avanti la trascrizione e l'annotazione, e con l'ausilio di quali strumenti e schemi. Puoi anche ottenere informazioni su aspetti del controllo qualità, sia per le annotazioni manuali che per quelle automatiche.

Infine, ci sono informazioni riguardo le possibilità di **usare MERLIN in attività di ricerca** circa la validità delle scale QCER, l'acquisizione di lingue seconde e *l'Elaborazione del linguaggio naturale* applicata al linguaggio degli apprendenti. Puoi anche trovare una **lista di fonti pertinenti**.

3.3 <<Corpus MERLIN>>

The screenshot displays the Merlin website. On the left is a vertical navigation menu with the following items: 'Usare MERLIN', 'MERLIN per la ricerca', 'Corpus MERLIN' (highlighted in a darker blue), 'Annotazioni MERLIN', 'Documentazione', and 'Scarica il corpus'. The main content area has a blue header with the Merlin logo (a star and dots) and the text 'merlin'. Below the logo is a navigation bar with links 'Home | Ricerca | Su MERLIN' and a question mark icon. To the right of the links are language options: 'ceco', 'tedesco', 'inglese', and 'italiano'. The main content area is titled 'Il corpus MERLIN' and contains the following text: 'Il corpus MERLIN contiene 2,286 testi di apprendenti l'italiano, il tedesco e il ceco, tratti da esami scritti di enti certificatori riconosciuti. Gli esami mirano ad attestare le conoscenze linguistiche per i livelli A1-C1 del Quadro comune europeo di riferimento per le lingue (QCER)'. Below this are several links with expandable icons: 'Testi ed enti certificatori', 'Il collegamento al Quadro di Riferimento - la griglia di valutazione MERLIN', 'Compiti d'esame', 'Metadati disponibili', and 'Il corpus MERLIN in cifre'.

Grafico 6: << corpus MERLIN>>

La sezione <<corpus MERLIN>> fornisce molte informazioni sui dati MERLIN. Puoi scoprire di più sui testi e sugli enti certificatori, scaricare le griglie di valutazione che sono state usate e visualizzare una lista dei compiti d'esame:

Compiti d'esame

Di seguito è fornita una panoramica esauriente dei compiti d'esame, elencati per lingua d'arrivo e livello QCER. ...

Il livello del test può differire dal livello attribuito al testo dell'apprendente durante le rivalutazioni.

I compiti sono illustrati attraverso una » **griglia** che è stata appositamente sviluppata dall'ALTE (Associazione degli Esaminatori di Lingua in Europa, » www.alte.org). La griglia contiene informazioni dettagliate sui compiti e sulle caratteristiche specifiche del testo da produrre, ad es. riguardo all'argomento, al registro o al dominio (autore: Olaf Bärenfänger, in lingua inglese).

» Note generali sulle descrizioni dei compiti

Italiano


- A1 [E-mail informale: spostare un appuntamento](#) 
 [E-mail informale: aiutare un amico in cerca di lavoro](#) 
- A2 [Lettera informale: andare a trovare un amico](#) 
 [Lettera informale: contattare un amico dopo molto tempo](#) 
 [Lettera informale: informare amici sul corso di lingua](#) 
- B1 [Lettera formale: informarsi su un corso di lingua](#) 
 [Lettera informale: cucinare con l'insegnante](#) 
 [Lettera informale: rispondere ad un invito a un matrimonio](#) 
 [Lettera informale: aiutare un amico in cerca di lavoro dopo la maturità](#) 
- B2 [Lettera informale: aiutare qualcuno che ha problemi con le chat](#) 
 [Lettera formale: descrivere le proprie esperienze con l'apprendimento delle lingue](#) 
 [Lettera formale: reclamare contro un albergo](#) 
 [Lettera formale: chiedere informazioni su serate di cucina internazionale](#) 
 [Lettera formale: informarsi su un progetto di aiuto](#) 
 [Lettera formale: fare domanda per uno stage aziendale](#) 
 [Lettera formale: fare domanda per uno stage nel campo della moda](#) 

Grafico 7: compiti d'esame MERLIN per l'italiano (<<corpus MERLIN>>)

Cliccando su un compito, in una nuova finestra si aprirà un file pdf che contiene, oltre alla traccia, una descrizione dettagliata del compito basata su una griglia sviluppata dall'Associazione per la Valutazione Linguistica in Europa (ALTE, www.alte.org). La descrizione del compito fornisce maggiori informazioni sulla sua lunghezza, sul tipo di lingua dell'elaborato previsto, sulla sua difficoltà e su tanto altro.

i) Task input/prompt		
17	Language of input/prompt	German
18	CEFR level of input/prompt	A1
19	Time permitted or suggested for this task	n.a. minutes
20	Control/guidance	Semi-controlled
21	Content	Specified
22	Genre	Letter
23	Rhetorical function(s) of input	Describing, instructing
24	Imagined audience	Friend
25	Mode of input/prompt	Written
26	Topic or theme of input	Relations with other people
27	Integration of skills for input	Reading

Grafico 8: Dettaglio della descrizione di un compito (<<MERLIN corpus>>)

La sezione << corpus MERLIN>> fornisce anche informazioni sui metadati disponibili (come età, sesso o lingua materna), e contiene inoltre una tabella con il numero totale di testi disponibili in base al livello QCER del test e alla valutazione QCER complessiva ottenuta:

Il corpus MERLIN in cifre

Numero di testi per lingua e livello QCER dell'esame dal quale i testi sono stati estratti, a confronto con il numero di testi per livello QCER attribuito con le rivalutazioni MERLIN

	Livello QCER dell'esame		Livello QCER dopo le rivalutazioni (Fair Average)	
Ceco			A1	1
	A2	111	A2	189
	B1	143	B1	165
	B2	188	B2	81
			C1	2
Italiano	A1	207	A1	29
	A2	202	A2	378
	B1	201	B1	394
	B2	201	B2	2
Tedesco	A1	206	A1	57
	A2	209	A2	297
	B1	210	B1	331
	B2	204	B2	293
	C1	204	C1	42
			C2	4
Totale	2286		2265	

Grafico 9: Numero dei testi per ogni livello QCER & numero delle valutazioni per ogni livello QCER (<<corpus MERLIN>>)

E una tabella con informazioni riguardanti la profondità dell'annotazione dei testi MERLIN:

	ceco	tedesco	italiano
Testi	442	1033	813
TH1	440	1033	813
EA1	361	752	754
TH2	231	275	154
EA2	198	258	85

Grafico 10: Il corpus MERLIN in cifre: numero di testi con target hypotheses 1 & 2 e error annotation 1 & 2 (<<corpus MERLIN>>)



Grafico 11: << annotazioni MERLIN>>

Questa sezione contiene informazioni riguardo le annotazioni MERLIN, concentrandosi in particolare sulle annotazioni manuali.

Puoi visualizzare una panoramica dell'architettura di annotazione:

La struttura delle annotazioni

I dati MERLIN sono stati arricchiti con un'annotazione multilivello. ...

Anche se per la maggior parte delle caratteristiche del linguaggio degli apprendenti è stata necessaria un'annotazione manuale, è stato possibile effettuare, usando NLP (Natural Language Processing), anche **annotazioni automatiche**, come la tokenizzazione e la lemmatizzazione, il tagging di parti del discorso o la segmentazione in frasi o T-unit.

Annotazioni per l'intero corpus MERLIN

Annotazioni disponibili per tutto il corpus



Target Hypothesis 1: versione ortograficamente e grammaticalmente corretta del testo dell'apprendente (TH1)



Caratteristiche del linguaggio degli apprendenti: ortografia e grammatica (Error annotation 1, EA1)

Le annotazioni principali **disponibili per l'intero corpus MERLIN** sono le target hypotheses (**target hypotheses 1**) e le annotazioni di caratteristiche grammaticali e ortografiche del linguaggio degli apprendenti (**error annotation 1**):

Grafico 12: Struttura delle annotazioni MERLIN, corpus completo (<<Annotazioni MERLIN>>)

Inoltre, puoi avere accesso allo **schema di annotazione** sul quale si basano tutte le annotazioni MERLIN, con esempi per tutte e tre le lingue, un glossario per la terminologia meno comune e definizioni chiare per ogni tag. Inoltre, puoi scaricare una **documentazione di avanzamento** relativa a questioni difficili sorte attorno a singoli aspetti delle annotazioni. Se ti capita di imbatterti in un'annotazione che ti sembra discutibile, questo documento potrebbe esserti utile (è accessibile anche dalla sezione <<documentazione>>).

Se non vuoi leggere per intero le definizioni dei tag, invece di scaricare lo schema di annotazione completo, puoi anche consultare una lista con tutti i tag di annotazione corredati da esempi:

Lista di caratteristiche del linguaggio degli apprendenti corredate da esempi

[Grammatica](#) | [Ortografia](#) | [Intelligibilità](#) | [Lessico](#) | [Coesione/Coerenza](#) | [Appropriatezza sociolinguistica](#) | [Pragmatica](#)

TAG DI GRAMMATICA	Esempio*
ordine delle parole nella frase principale	*[Vielleicht du könntest mir bei meine Wohnungssuche helfen.] *[Sollst du Wasser und Bikini mitbringen.]
ordine delle parole nella frase subordinata	*[wenn haben Sie Zeit.] dann bitte sagen Sie mir.
negazione generale	*Ich habe [nicht] Zeit.; *Er wird dort arbeiten [nein].
CZE: doppia negazione	*[mám] žádný čas {nemám žádný čas}; *nikdo [volal] {nikdo nevolal}
valenza verbale: numero di argomenti obbligatori	CZE: *Petr vstává v 6 hodin. On nesnídá, protože [on] nemá hlad. GER: *Er hat uns nicht gesagt, ob {er} kommen will.
accordo (soggetto e verbo)	*Jana [hast] gelesen, *Jana [sind] müde
pronome riflessivo	CZE: *smála [si]; GER: *er [entschuldigt], *Laura und Ferdinand reden [sich] ITA: *[se] {si} lava ogni mattina
CZE: pronome possessivo riflessivo	*potřebuju [moji] knihu, *vidím [mého] otce
forma flessa inesistente (nomi, aggettivi, verbi)	adjective: *ein [blaus] Himmel {blauer}; [teuerer] {teurer}; [größen] {großen / größeren} noun: *das schöne [Hause], *[euche] [Fahrrade]

Grafico 13: Lista di tag di annotazione con esempi (<<Annotazioni MERLIN>>)



merlin

Home | Ricerca | Su MERLIN ? ceco tedesco inglese italiano

Usare MERLIN

Documentazione del progetto MERLIN
Download documenti relativi al progetto MERLIN.

MERLIN per la ricerca

Corpus MERLIN

Corpus: Test e preparazione dei dati
Linee guida per la trascrizione (versione online in via di pubblicazione)
Le tracce complete dei test, correlate da descrizione, sono disponibili per il download nella sezione » [Corpus MERLIN](#).
» [Griglia di valutazione MERLIN](#) 
» [Relazione tecnica](#) : Relazione sull'affidabilità e funzionalità delle scale delle valutazioni MERLIN di campioni di linguaggio scritto, di O. Bärenfänger

Annotazioni MERLIN

Annotazioni: schema di annotazione e processo di annotazione
Manuale di annotazione (versione online in via di pubblicazione)
» [Schema di annotazione MERLIN](#) 

Documentazione

Documentazione di ulteriori questioni di annotazione (versione online in via di pubblicazione)

Scarica il corpus

Interventi a convegni e pubblicazioni del team MERLIN
Katrin Wisniewski. *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen*. Language Testing and Evaluation vol.33, Frankfurt am Main 2014

Grafico 14: <<documentazione>>

Qui si possono trovare tutti i documenti riguardanti MERLIN e disponibili gratuitamente, tra cui:

- relazioni prodotte nell'ambito del progetto, ad es. sulla qualità delle valutazioni e sulla fruibilità della piattaforma (studi degli utenti)
- tutti i compiti d'esame utilizzati
- griglie usate per la valutazione
- linee guida e schemi usati per la trascrizione e l'annotazione (ad es., schema di annotazione)
- pubblicazioni, presentazioni e simili del team MERLIN
- ...

merlin

Home | Ricerca | Su MERLIN ? ceco tedesco inglese italiano

Usare MERLIN

MERLIN per la ricerca

Corpus MERLIN

Annotazioni MERLIN

Documentazione

Scarica il corpus

Scarica il corpus MERLIN

MERLIN offre il download gratuito di tutti i testi degli apprendenti inclusi nel corpus.

Scarica tutti i testi degli apprendenti, comprensivi di **metadati** (informazioni sull'apprendente, sul test e sulle valutazioni):

CECO	file PDF compressi	file TXT compressi
TEDESCO	file PDF compressi	file TXT compressi
ITALIANO	file PDF compressi	file TXT compressi

Scarica tutti i testi degli apprendenti, comprensivi di **metadati e target hypotheses**:

CECO	file PDF compressi	file TXT compressi
TEDESCO	file PDF compressi	file TXT compressi
ITALIANO	file PDF compressi	file TXT compressi

Grafico 15: <<scarica il corpus>>

Questa sezione ti offre un'opzione di download immediato del corpus MERLIN completo. Puoi scaricare file zippati per il ceco, l'italiano e il tedesco in formato .text o .pdf, scegliendo se scaricare solo i testi originali degli apprendenti, compresi di metadati (ad es. L1, età), oppure includere anche le target hypotheses.

Se sei interessato a una raccolta di testi più specifica, vai su <<crea un sottocorpus>>. Qui puoi creare il tuo sottocorpus personalizzato, scaricarlo, fare delle ricerche (<<ricerca semplice/avanzata>>) o visualizzare parametri statistici (<<statistiche>>).

4. Le funzionalità di ricerca MERLIN

merlin

Home | Ricerca | Su MERLIN ?

ceco tedesco inglese italiano

Usare MERLIN

Ricerca semplice | Ricerca avanzata | Crea un sottocorpus | Statistiche

Ricerca parole nei testi degli apprendenti e visualizzale nel contesto (ad es. nel testo completo)
esempi: *welchen, gli, rád*

Ricerca per parola

Cerca in:

Corpus:

Ancora nessun sottocorpus? Compila la tua personale raccolta di testi » [qui](#).

Il corpus MERLIN
MERLIN dà accesso a 2.286 testi scritti da apprendenti il ceco, l'italiano e il tedesco.
I testi degli apprendenti provengono da test linguistici standardizzati e sono stati collegati in modo affidabile ai livelli del QCER. » [Scopri di più](#)

Usa MERLIN ...
... per comprendere meglio i livelli del Quadro comune europeo di riferimento per le lingue (QCER).
» [Scopri di più](#)

Documentazione

Scarica il corpus

Grafico 16: Area di ricerca

Le opzioni di ricerca includono:

- una ricerca semplice
- una ricerca avanzata
- una sezione “crea un sottocorpus”
- una sezione di statistiche

4.1 <<Crea un sottocorpus>>

È consigliabile iniziare a usare le opzioni di ricerca creando una raccolta di testi ai quali si è interessati in <<crea un sottocorpus>> (vedi il grafico 17):

The screenshot shows the MERLIN website interface. On the left is a vertical navigation menu with the MERLIN logo and links: Usare MERLIN, MERLIN per la ricerca, Corpus MERLIN, Annotazioni MERLIN, Documentazione, and Scarica il corpus. The main content area has a top navigation bar with 'Home | Ricerca | Su MERLIN' and a language selector (ceco, tedesco, inglese, italiano). Below this is a sub-navigation bar with 'Ricerca semplice', 'Ricerca avanzata', 'Crea un sottocorpus' (highlighted), and 'Statistiche'. The 'Crea un sottocorpus' form contains the following fields:

- Lingua del test = (dropdown menu with options: tutti, A1, A2, B1)
- Livello QCER del test = (dropdown menu with options: tutti, A1, A2, B1)
- Valutazione QCER complessiva = (dropdown menu with options: tutti, A1, A2, A2+). To the right is a link 'Criteri di valutazione dettagliati' with an information icon.
- Compito = (dropdown menu with options: Per favor, Per favore, Per favore, Per favore)

Below the form are two filter links: 'Filtra in base alle informazioni sugli apprendenti' and 'Filtra per parola e in base alle caratteristiche del linguaggio degli apprendenti'. At the bottom is a 'Nome:' input field and a 'Crea sottocorpus e mostra testi' button with an information icon.

Grafico 17: Interfaccia di ricerca <<Crea un sottocorpus>>

Qui è possibile specificare i criteri secondo i quali una raccolta di testi di apprendenti (un cosiddetto sottocorpus) può essere adattata alle esigenze degli utenti. I criteri sono:

- lingua target del testo
- livello QCER originale dell'esame
- valutazione QCER media imparziale dei testi (vedi la sezione 2)
- Uno o più livelli QCER relativi a singoli criteri di valutazione (repertorio linguistico generale | correttezza grammaticale | ampiezza del lessico | padronanza lessicale | ortografia | coerenza & coesione | appropriatezza sociolinguistica)
- compito d'esame

Inoltre, è possibile classificare i testi in base alle informazioni sugli apprendenti:

- lingua materna (L1)
- età
- sesso

In più, i testi possono essere scelti selezionando fino a tre caratteristiche del linguaggio degli apprendenti e/o parole:

The screenshot shows a web interface titled "Filtra per parola e in base alle caratteristiche del linguaggio degli apprendenti". It includes a search bar labeled "Cerca in:" with a dropdown menu set to "testo dell'apprer" and an information icon. Below this are three rows for filtering by word ("Parola:") and three rows for filtering by characteristics ("Caratteristiche:"). Each row has a text input field, a greater-than-or-equal-to symbol (≥), a dropdown menu set to "1", and the word "occorrenze". Information icons are present next to the search bar and the first characteristic filter.

Grafico 18: *Specificazione delle caratteristiche di un sottocorpus, caratteristiche del linguaggio degli apprendenti (<<Crea un sottocorpus>>)*

Al sottocorpus deve essere assegnato un nome che puoi scegliere tu; cliccando su “crea sottocorpus e mostra testi”, questo sottocorpus rimarrà disponibile per ulteriori ricerche per 24 ore, dopodiché dovrai ridefinirlo.

Output di <<Crea un sottocorpus>>

Cliccando su “crea sottocorpus e mostra testi”, sarà visibile un risultato simile a quello rappresentato nel grafico 19.

[Ricerca semplice](#)
[Ricerca avanzata](#)
[Crea un sottocorpus](#)
[Statistiche](#)

Il tuo sottocorpus:
 lang=Italian; L1=all; CEFR level=all; overall rating=all;
[Sfoggia la ricerca semplice](#)
[Sfoggia la ricerca avanzata](#)

Opzioni di ricerca::
[Modifica la ricerca](#)
[Effettua una nuova ricerca](#)

Risultati
 totale risultati 813

[1](#)
[2](#)
[3](#)
[4](#)
[....](#)
[>>](#)

 Pagina 1 di 102 (8 risultati per pagina)

Italian A2 L1=French ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian B1 L1=Hungarian ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian A2+ L1=Polish ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian B1 L1=Hungarian ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian B1+ L1=German ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian A2 L1=German ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian B1 L1=German ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni
Italian B1 L1=Hungarian ⓘ	Visualizza testo dell'apprendente e TH	Visualizza informazioni sull'apprendente e valutazioni

Scarica documenti:
 Numero di documenti: ▾
 Includi metadati ☐
 Dati: ▾ [Scarica documenti](#)

Grafico 19: Esempio di output, <<crea un sottocorpus>>

In questo output è visualizzata una lista di tutti i testi che soddisfano i criteri usati per la definizione del sottocorpus. In questo caso, l'utente ha messo insieme tutti i testi che sono stati scritti in italiano con riferimento ad un determinato compito MERLIN (che non è nello screenshot).

La pagina di output specifica il numero di testi trovati ("totale risultati: 813"). Dall'ID nella prima colonna puoi capire quale sia la lingua target ("italiano"), la valutazione complessiva (che varia nell'esempio) e la L1 (nell'esempio: francese, ungherese, polacco e tedesco).

Cliccando su "scarica documenti" puoi **scaricare** i testi (o solo una selezione dei testi del sottocorpus) in formati differenti, con o senza metadati e/o target hypotheses.

Inoltre puoi cliccare su "**Visualizza testo dell'apprendente e TH**" per ottenere il testo originale e la target hypothesis 1, o 1 e 2 (come mostrato nell'esempio del grafico 20).

Torna alla lista dei documenti

Suggerimento: per visualizzare nel contesto le parole o le caratteristiche del linguaggio dell'apprendente che hai cercato, usa la *ricerca semplice* o la *ricerca avanzata*.

ID-autore:

1385_0001730 [Copia ID-autore](#)

target hypothesis 1:

Città X , 9 luglio 2011 Caro Daniele , come stai ? Io sto proprio male . È una settimana orribile . In ufficio c' è molto lavoro e giovedì pomeriggio non posso incontrarti . Mi dispiace moltissimo . Devo partecipare a una riunione con il nostro capo . Ma hai tempo sabato sera ? Ho sentito che c' è un ristorante thailandese nuovo a Città X . Vogliamo andare qui a cena sabato sera ? A presto e molti saluti Maria

testo dell'apprendente:

Città X, 9 luglio 2011

Caro Daniele,
come stai? Io sto proprio male. È una settimana orribile. In ufficio c'è molto lavoro e giovedì pomeriggio non posso incontrarti. Mi dispiace moltissimo. Devo partecipare a una riunione con il nostro capo. Ma hai tempo sabato sera? Ho sentito che c'è un ristorante thailandese nuovo a Città X. Vogliamo qui andare a cena sabato sera?
A presto e molti saluti
Maria

Grafico 20: Esempio di output per “Visualizza testo dell'apprendente e TH” in <<Crea un sottocorpus>>

Cliccando su “**Visualizza informazioni sull'apprendente e valutazioni**”, vengono visualizzati i metadati per un testo specifico (il grafico 21 mostra i metadati inerenti al testo del grafico 20):

[Ricerca semplice](#)

[Ricerca avanzata](#)

[Crea un sottocorpus](#)

[Torna alla lista dei documenti](#)

Suggerimento: per visualizzare nel contesto le parole o le caratteristiche del documento usa la *ricerca semplice* o la *ricerca avanzata*.

Generale:

ID-autore:	1385_0001730	Copia ID-autore
Lingua del test:	Italian	
Livello QCER del test:	A1	
Compito:	reschedule an appointment ⓘ	
Lingua materna:	German	
Età:	45	
Sesso:	female	

Valutazione:

Valutazione QCER complessiva:	B1
Correttezza grammaticale:	A2
Ortografia:	B1
Ampiezza del lessico:	A2
Padronanza lessicale:	B1
Coerenza/coesione:	A2
Appropriatezza sociolinguistica:	A2

Grafico 21: Esempio di output per “Visualizza informazioni sull'apprendente e valutazioni”, nell'output di <<crea un sottocorpus>>

Il sottocorpus può essere ulteriormente esplorato attraverso la ricerca semplice o avanzata o nella sezione statistiche.

4.2 <<Ricerca semplice>>

Nella <<ricerca semplice>>, è possibile cercare forme di una parola:

The screenshot shows the MERLIN website interface. On the left is a vertical menu with links: Usare MERLIN, MERLIN per la ricerca, Corpus MERLIN, Annotazioni MERLIN, Documentazione, and Scarica il corpus. The main content area has a top navigation bar with 'Home | Ricerca | Su MERLIN' and a language selector (ceco, tedesco, inglese, italiano). Below this is a sub-navigation bar with 'Ricerca semplice', 'Ricerca avanzata', 'Crea un sottocorpus', and 'Statistiche'. The 'Ricerca semplice' section contains the text: 'Ricerca parole nei testi degli apprendenti e visualizzale nel contesto (ad es. nel testo completo) esempi: *welchen, gli, rád*'. It features a search input field, a 'Ricerca per parola' button, and dropdown menus for 'Cerca in:' (set to 'testo dell'apprer') and 'Corpus:' (set to 'intera raccolta'). Below these are two informational boxes: 'Il corpus MERLIN' (explaining access to 2,286 tests) and 'Usa MERLIN ...' (explaining the purpose for understanding language levels).

Grafico 22: Interfaccia di ricerca <<ricerca semplice>>

La ricerca può essere eseguita ...

- nel testo dell'apprendente o nelle target hypotheses (TH1 o TH2),
- nell'intero corpus MERLIN o in un sottocorpus da te creato (<<crea un sottocorpus>>)

La ricerca semplice non è basata sui lemmi (tutte le forme flesse di una determinata entrata lessicale): si riferisce solo alla forma flessa digitata. Perciò, se digiti ad es., "lavorare" (come nell'esempio di seguito, grafico 23), non otterrai risultati per "lavori" o "lavorato".

Puoi usare la *tastiera virtuale*, se necessario. Nell'usare la funzione di ricerca semplice, tieni presente anche che MERLIN è basato su un numero limitato di compiti che producono una gamma ristretta di scelte lessicali.

output di <<ricerca semplice>>

L'output della ricerca semplice rende la parola cercata nel suo immediato contesto (la cosiddetta keyword in context, o KWIC):

The screenshot shows the MERLIN interface. On the left is a sidebar with links: Usare MERLIN, MERLIN per la ricerca, Corpus MERLIN, Annotazioni MERLIN, and Documentazione. The top navigation bar includes Home, Ricerca, Su MERLIN, and a help icon, along with language options (ceco, tedesco, inglese, italiano). The main content area is titled 'Ricerca semplice' and shows search results for the keyword 'lavorare'. It indicates 215 results found in 154 documents. A pagination bar shows 'Pagina 1 di 15 (15 risultati per pagina)'. Below this, a 'Parola chiave nel contesto' section displays several sentences containing the word 'lavorare'. To the right of each sentence is a blue button labeled 'Visualizza informazioni sull'apprendente e valutazioni'.

Grafico 23: Esempio di output, <<ricerca semplice>>

Cliccando sulla parola chiave nel contesto, verrà visualizzato il testo completo dell'apprendente. Cliccando su <<visualizza informazioni sull'apprendente e valutazioni>>, verranno mostrati i metadati del testo. Puoi anche scaricare direttamente un file .pdf del compito e copiare l'ID-autore se vuoi effettuare delle ricerche più specifiche:

This screenshot shows the detailed view of a learner and task. At the top are navigation links: Ricerca semplice, Ricerca avanzata, Crea un sottocorpus, and Statistiche. Below is a section titled 'Torna alla lista dei documenti'. A suggestion text reads: 'Suggerimento: per visualizzare nel contesto le parole o le caratteristiche del linguaggio dell'apprendente che hai cercato, usa la ricerca semplice o la ricerca avanzata.' The main content is divided into two sections: 'Generale:' and 'Valutazione:'. The 'Generale:' section lists various metadata fields and their values, with a 'Copia ID-autore' button next to the ID. The 'Valutazione:' section lists proficiency levels for different aspects of the learner's performance.

Generale:	
ID-autore:	1385_0000044 Copia ID-autore
Lingua del test:	Italian
Livello QCER del test:	A1
Compito:	help a friend who is looking for work ⓘ
Lingua materna:	German
Età:	48
Sesso:	female

Valutazione:	
Valutazione QCER complessiva:	A2+
Correttezza grammaticale:	B1
Ortografia:	B1
Ampiezza del lessico:	B1
Padronanza lessicale:	B1
Coerenza/coesione:	A2
Appropriatezza sociolinguistica:	B1

Grafico 24: Esempio di output per <<visualizza informazioni sugli apprendenti e valutazioni>> nell'output di <<ricerca semplice>>

4.3 <<Ricerca avanzata>>

Questa sezione ti consente un accesso più sofisticato al corpus MERLIN. È possibile combinare la ricerca di lemmi/parole con la ricerca di una varietà di annotazioni che sono disponibili in MERLIN.

Puoi cercare...

- nel testo dell'apprendente o nelle target hypotheses (TH1 or TH2),
- nell'intero corpus MERLIN o in un sottocorpus da te specificato (<<crea un sottocorpus>>)

The screenshot shows the 'Ricerca avanzata' (Advanced Search) interface of the MERLIN corpus. The interface is designed for complex queries, allowing users to filter results by language, word pairs, part of speech, and specific linguistic features. The search form includes dropdown menus for language selection and word pair distance, as well as checkboxes for part of speech and linguistic features. The interface is user-friendly, with clear labels and helpful icons.

Grafico 25: Interfaccia di ricerca <<Ricerca avanzata>>

In aggiunta, puoi decidere se vuoi cercare 1 o 2 parole o lemmi che ...

- sono direttamente **adiacenti** o compaiono separati da un determinato numero di parole
- appartengono a una specifica **classe lessicale** alla quale sei interessato (in base all'annotazione automatica delle parti del discorso). Se clicchi sul menù a comparsa dopo aver selezionato la lingua target, otterrai una lista con abbreviazioni delle parti del discorso e brevi spiegazioni (vedi grafico 26).

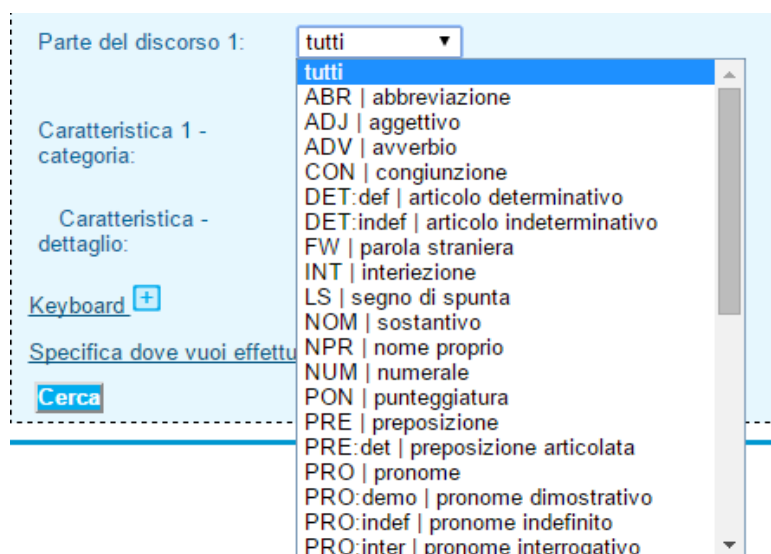


Grafico 26: Definizione dei tag delle parti del discorso (POS), <<ricerca avanzata>> (dettaglio, tedesco)

Puoi anche accedere alle **annotazioni manuali** disponibili in MERLIN:

- scegli innanzitutto <<caratteristica 1 - categoria>> per specificare a quale categoria di annotazione sei interessato, ad es. grammatica, vocabolario, ortografia
- Come opzione, puoi restringere la ricerca in <<caratteristica - dettaglio>>: otterrai così una lista con tutti i **tag di annotazione** che sono stati usati in una determinata categoria di annotazione (nel grafico sotto puoi trovare un esempio relativo alla grammatica). Se non scegli un tag specifico da questo elenco, scegliendo “grammatica” nel campo <<caratteristica 1 - categoria>> tutti i tag grammaticali verranno visualizzati.

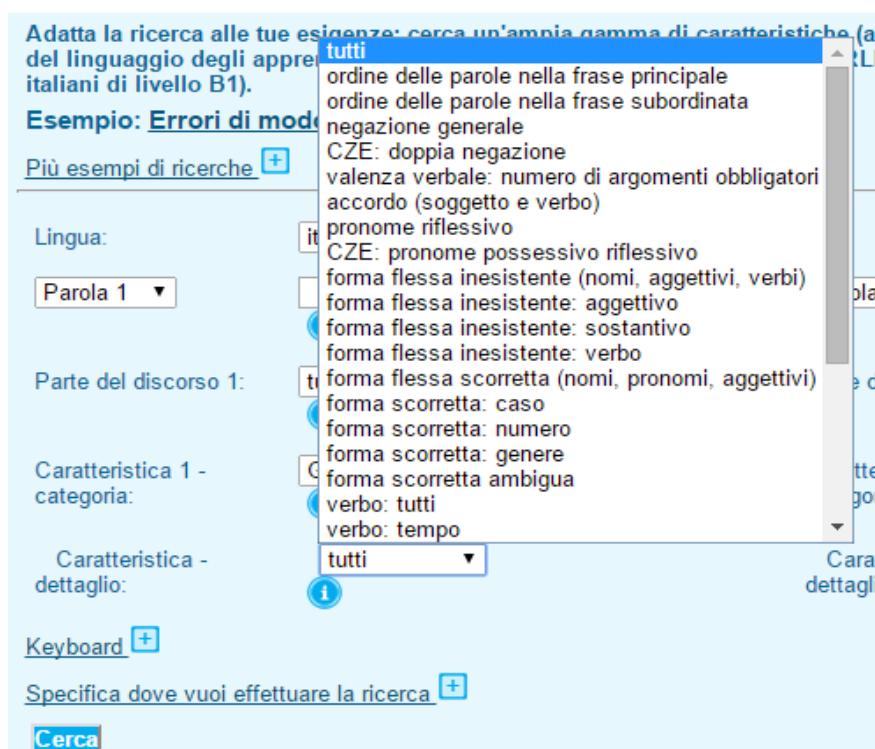


Grafico 27: Definizione: <<caratteristica - dettaglio>> nella <<Ricerca avanzata>> (dettaglio, grammatica)

Puoi attuare questa procedura per una parola/lemma o per una combinazione di due parole/lemmi. Attenzione: non è possibile cercare caratteristiche del linguaggio degli apprendenti senza prima specificare una parola o un lemma.

Se hai bisogno di maggiori informazioni riguardo le caratteristiche annotate, nella sezione <<annotazioni MERLIN>> puoi consultare una lista con tutti i tag e gli esempi, scoprire di più sulla struttura dell'annotazione MERLIN e scaricare lo schema di annotazione MERLIN.

output di <<ricerca avanzata>>

La sezione <<ricerca avanzata>> utilizza l'architettura di ricerca e visualizzazione open source ANNIS (www.annis-tools.org), motivo per il quale il suo output appare diverso rispetto a quelli delle altre sezioni dell'interfaccia MERLIN. L'output è illustrato nei seguenti screenshot.



Nel grafico 28, nella <<ricerca avanzata>> è stato digitato il lemma "venire".

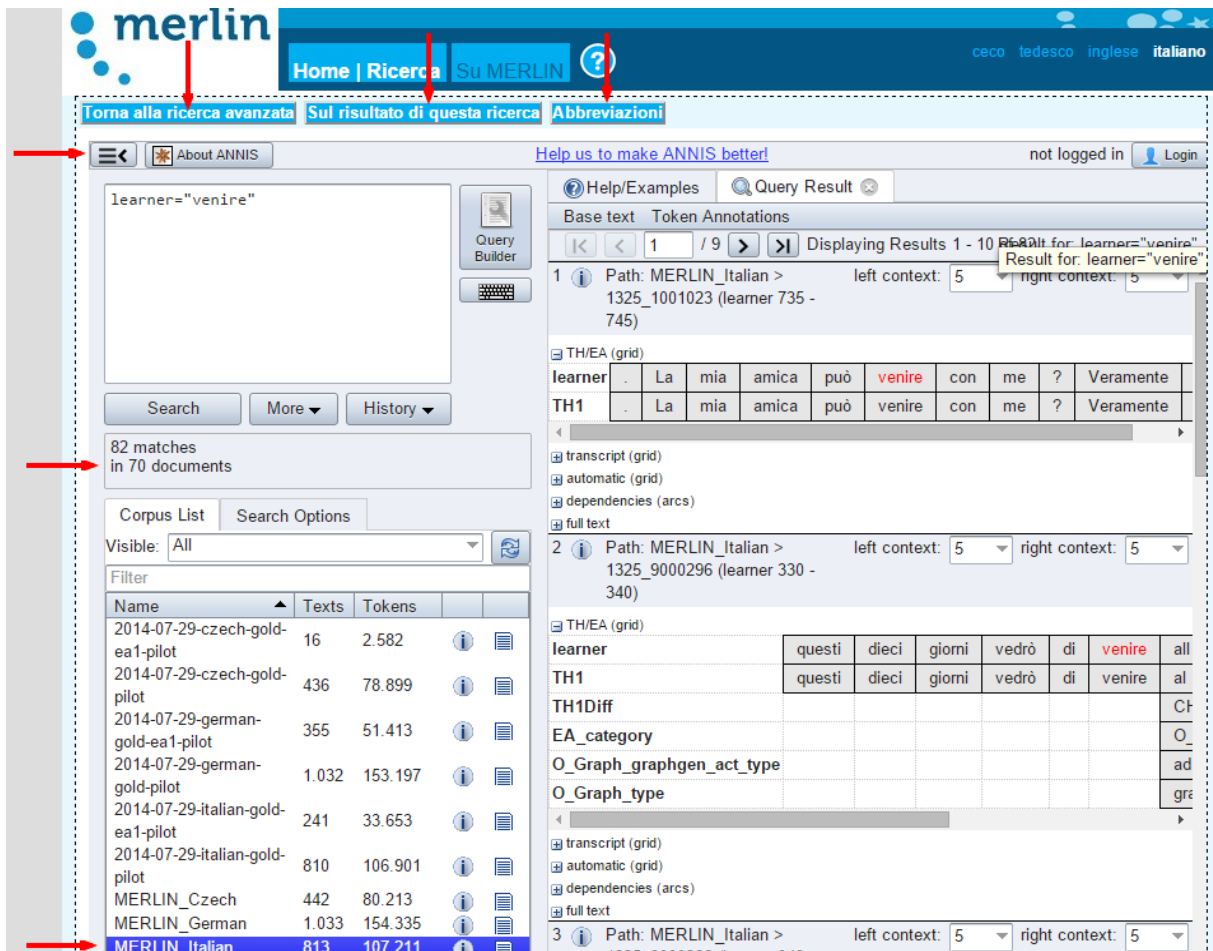
The screenshot displays the MERLIN web interface. On the left is a sidebar with a blue header and white text links: 'Usare MERLIN', 'MERLIN per la ricerca', 'Corpus MERLIN', 'Annotazioni MERLIN', 'Documentazione', and 'Scarica il corpus'. The main content area has a blue header with the MERLIN logo and navigation links: 'Home | Ricerca | Su MERLIN'. Below the header, there are tabs for 'Ricerca semplice', 'Ricerca avanzata' (which is active), 'Crea un sottocorpus', and 'Statistiche'. The 'Ricerca avanzata' section contains a text box with instructions: 'Adatta la ricerca alle tue esigenze: cerca un'ampia gamma di caratteristiche (ad es. classi di parole, caratteristiche del linguaggio degli apprendenti » visualizza ulteriori informazioni) in MERLIN o in un sottocorpus (ad es. testi italiani di livello B1).'. Below this is an example: 'Esempio: Errori di modo in testi di apprendenti l'italiano'. A link 'Più esempi di ricerche' is also present. The search filters are arranged in two columns. The first column includes 'Lingua:' (dropdown set to 'italiano'), 'Parola 1:' (dropdown set to 'venire'), 'Parte del discorso 1:' (dropdown set to 'tutti'), 'Caratteristica 1 - categoria:' (dropdown set to 'Nessuna se'), and 'Caratteristica 1 - dettaglio:' (dropdown set to 'Per favore'). The second column includes 'Parola 2:' (empty), 'distanza tra le parole:' (dropdown set to '0'), 'Parte del discorso 2:' (dropdown set to 'tutti'), 'Caratteristica 2 - categoria:' (dropdown set to 'Nessuna se'), and 'Caratteristica 2 - dettaglio:' (dropdown set to 'Per favore'). At the bottom, there are links for 'Keyboard' and 'Specifica dove vuoi effettuare la ricerca', and a 'Cerca' button.

Grafico 28: Esempio di ricerca di un lemma, <<Ricerca avanzata>>

Nell'output trovi

- (1) Il numero di risultati e il numero di documenti con risultati (lato sinistro)
- (2) Il corpus nel quale hai effettuato la ricerca (in basso a sinistra)

- (3) Un pulsante che nasconde il lato sinistro dell'output, importante per poter visualizzare completamente il testo dell'apprendente

- (4) La possibilità di ritornare alla <<ricerca avanzata>> - la ricerca che hai effettuato non sarà salvata
- (5) Un collegamento a ulteriori informazioni (<<Sul risultato di questa ricerca>>); si aprirà una nuova finestra con informazioni riguardanti i singoli livelli dell'annotazione (vedi il capitolo 5 di questo documento)
- (6) Un collegamento a una lista con brevi spiegazioni di tutte le abbreviazioni, principalmente tag di annotazione, che sono usati nell'output, si aprirà in una nuova finestra (<<abbreviazioni>>)
- (7) La funzione di aiuto MERLIN 



La screenshot mostra l'interfaccia web di MERLIN. Le annotazioni rosse indicano:

- Il pulsante "Torna alla ricerca avanzata" (punto 4).
- Il pulsante "Sul risultato di questa ricerca" (punto 5).
- Il pulsante "Abbreviazioni" (punto 6).
- Il pulsante "Help/Examples" (punto 7).
- Il pulsante "About ANNIS" (punto 3).

Il contenuto principale della pagina mostra i risultati della ricerca per il lemma "venire".

Corpus List

Name	Texts	Tokens
2014-07-29-czech-gold-ea1-pilot	16	2.582
2014-07-29-czech-gold-pilot	436	78.899
2014-07-29-german-gold-ea1-pilot	355	51.413
2014-07-29-german-gold-pilot	1.032	153.197
2014-07-29-italian-gold-ea1-pilot	241	33.653
2014-07-29-italian-gold-pilot	810	106.901
MERLIN_Czech	442	80.213
MERLIN_German	1.033	154.335
MERLIN_Italian	813	107.211

Query Result

Base text Token Annotations

1 Path: MERLIN_Italian > 1325_1001023 (learner 735 - 745) left context: 5 right context: 5


learner	La	mia	amica	può	venire	con	me	?	Veramente
TH1	.	La	mia	amica	può	venire	con	me	?

2 Path: MERLIN_Italian > 1325_9000296 (learner 330 - 340) left context: 5 right context: 5

learner	questi	dieci	giorni	vedrò	di	venire	all
TH1	questi	dieci	giorni	vedrò	di	venire	al

3 Path: MERLIN_Italian > 1325_9000296 (learner 343) left context: 5 right context: 5

Grafico 29: Esempio di output, ricerca di un lemma, <<Ricerca avanzata>>

Nascondendo l'area verticale a sinistra, nella quale sono contenute meta informazioni, cliccando su , è più facile visualizzare direttamente i testi degli apprendenti:

The screenshot shows the MERLIN interface with the following components:

- Header:** merlin logo, navigation links (Home | Ricerca | Su MERLIN), language options (ceco, tedesco, inglese, italiano), and a login status (not logged in).
- Navigation:** Links for "Torna alla ricerca avanzata", "Sul risultato di questa ricerca", and "Abbreviazioni".
- Search Bar:** "About ANNIS" button and a search input field.
- Results:** Two search results are displayed.
 - Result 1:** Path: MERLIN_Italian > 1325_1001023 (learner 735 - 745). The TH/EA grid shows the lemma "venire" in red. The transcript grid shows the full text: "La mia amica può venire con me ? Veramente non".
 - Result 2:** Path: MERLIN_Italian > 1325_9000296 (learner 330 - 340). The TH/EA grid shows the lemma "venire" in red. The transcript grid shows the full text: "questi dieci giorni vedrò di venire al quinto giorno è resterò".
- Annotations:** On the left side of each result, there are expandable sections for "transcript (grid)", "automatic (grid)", "dependencies (arcs)", and "full text".

Grafico 30: Esempio di output, ricerca di lemma, meta informazioni nascoste sul lato sinistro, <<Ricerca avanzata>>

Puoi vedere il lemma cercato nel suo immediato contesto, che può essere esteso fino a 25 token su entrambi i lati:

The screenshot shows a dropdown menu for extending the context of a lemma. The menu is open, showing options from 0 to 25. The current selection is 5.

left context:	right context:
5	5
0	
5	
10	
15	
20	
25	

Grafico 31: Modifica dell'estensione del contesto di un lemma nell'output di <<Ricerca avanzata>>

Puoi accedere alle annotazioni automatiche, a una vista completa, alle trascrizioni, e agli archi di dipendenza dei risultati di ricerca. Nel grafico 32, c'è un esempio nel quale l'annotazione automatica è mostrata in una griglia:

merlin Home | Ricerca Su MERLIN ? ceco tedesco inglese italiano

Torna alla ricerca avanzata | Sul risultato di questa ricerca | Abbreviazioni

Help/Examples Query Result not logged in Login

Base text Token Annotations

1 / 9 Displaying Results 1 - 10 of 82 Result for: learner="venire"

1 Path: MERLIN_Italian > 1325_1001023 (learner 735 - 745) left context: 5 right context: 5

TH/EA (grid)

learner	.	La	mia	amica	può	venire	con	me	?	Veramente	non
TH1	.	La	mia	amica	può	venire	con	me	?	Veramente	non

transcript (grid)

automatic (grid):

learner	.	La	mia	amica	può	venire	con	me	?	Veramente	non
tok_lemma	.	il	mio	amico	potere	venire	con	me	?	veramente	non
tok_pos	SENT	DET: def	PRO: poss	NOM	VER: pres	VER: infi	PRE	PRO: pers	SENT	ADV	AD
tok_pos_bohnet	.	DET	ADJ	NOUN	AUX	VERB	ADP	PRON	.	PNOUN	AD
sentence	Naturalmente se Lei offre qualcosa leggerò volentieri.	La mia amica può venire con me? Veramente non abbiamo mai lavorato in un campo di lavoro.									
repetition											id: 1 cou
repetition											id: 1 cou

dependencies (arcs)

full text

Grafico 32: Visualizzazione dell'annotazione automatica nell'output di <<Ricerca avanzata>>

Le righe della tabella di output sono spiegate in <<sul risultato di questa ricerca>>.

4.4 <<Statistiche>>

In questa sezione sono disponibili alcune fondamentali informazioni statistiche. Puoi cercare in tutti i testi MERLIN o nella tua selezione di testi (“sottocorpus”, vedi la sezione <<crea un sottocorpus>>.)

merlin

Home | Ricerca Su MERLIN ?

ceco tedesco inglese italiano

Usare MERLIN

MERLIN per la ricerca

Corpus MERLIN

Annotazioni MERLIN

Documentazione

Scarica il corpus

Ricerca semplice Ricerca avanzata Crea un sottocorpus **Statistiche**

Mostra la frequenza della caratteristica in gruppi di testi:

Attenzione: +

Corpus: tutti testi cechi tutti testi tedeschi tutti testi italiani

Testo (ID-autore)

Ancora nessun sottocorpus? Compila la tua personale raccolta di testi » qui.

Frequenza delle caratteristiche annotate + i

Frequenza relativa delle caratteristiche annotate + i

Linguaggio dell'apprendente senza errori + i

Parametri di complessità (disponibili solo per il tedesco) + i

Visualizza

Grafico 33: Interfaccia di ricerca <<Statistiche>>

Avvertenze per l'interpretazione dei parametri statistici

I parametri statistici MERLIN devono essere interpretati con estrema cautela. Possono fornire indicazioni e tendenze, ma non devono essere sovra-interpretati come prova per determinati percorsi di apprendimento linguistico. Ci sono diverse ragioni per questo, alcune delle quali sono elencate qui:

- 1) MERLIN è un corpus piccolo – non sono possibili semplici generalizzazioni
- 2) Il database per la EA1 è diverso dal database per la EA2; quest'ultimo è estremamente ridotto
- 3) Analisi contrastive dovrebbero essere trattate con cautela. Non si possono confrontare in modo diretto parametri complessivi di lingue diverse, poiché ci sono alcuni tag che sono specifici per determinate lingue, come ad esempio gli errori riguardanti l'aspetto verbale in ceco e italiano, ma non in tedesco.
- 4) I parametri dipendono strettamente dalla gamma di compiti usata in MERLIN; questo vale soprattutto per il lessico e per gli altri campi della EA2

Nella sezione <<statistiche>>, ci sono **quattro tipi di parametri** disponibili. Per ogni scelta che fai puoi selezionare più caratteristiche annotate tenendo premuto il tasto CTRL:

(1) *Frequenza delle caratteristiche annotate*

Numero assoluto delle caratteristiche annotate nel tuo database. Puoi scegliere una o più categorie di caratteristiche (ad es. grammatica, ortografia). Poi, puoi sia fare un calcolo di tutte le annotazioni riguardanti quella/e categoria/e (ad es., tutte le annotazioni di grammatica nei testi in italiano), sia specificare quali annotazioni ti interessano ad un livello più specifico (ad es., tutti gli errori morfologici nei testi in italiano):

Grafico 34: Esempio di calcolo di frequenza assoluta (tutti gli errori grammaticali, italiano), <<Statistiche>>

L'**output** ti dà una panoramica del numero totale delle caratteristiche che hai cercato:

[Usare MERLIN](#)
[MERLIN per la ricerca](#)
[Corpus MERLIN](#)
[Ricerca semplice](#)
[Ricerca avanzata](#)
[Crea un sottocorpus](#)
[Statistiche](#)
[Modifica la ricerca](#)
Frequenza della caratteristica in testi selezionati

Corpus	Italian
Numero di testi	813
Frequenza delle caratteristiche annotate	
GRAMMAR	4540

Grafico 35: Esempio di output, calcolo di frequenza assoluta (tutti gli errori grammaticali, italiano), <<Statistiche>>

Cliccando sul numero di risultati (nel nostro esempio, 4540), verrai diretto alla sezione <<Ricerca avanzata>> con accesso diretto a tutte le annotazioni attinenti alla tua ricerca:

[Torna alla ricerca avanzata](#)
[Sul risultato di questa ricerca](#)
[Abbreviazioni](#)

[Help/Examples](#)
[Query Result](#)

[Base text](#)
[Token Annotations](#)

1 / 748
 Displaying Results 1 - 10 of 745

Path: MERLIN_Italian > 1325_1001008 (learner 241 - 251)
 left context: 5 right context: 5

TH/EA (grid)	Signore	/	a	,	La
learner					
TH1	a				Le
TH1Diff	INS				CHA
EA_category	O_Graph				
EA_category					G_Morphol_Wro
G_Art_type					
G_Morphol_Wrong_type					case
G_Prep_type					
O_Graph_graphgen_act_type	o				
O_Graph_type	graphgen				

transcript (grid)
 automatic (grid)
 dependencies (arcs)
 full text

2 Path: MERLIN_Italian > 1325_1001008 (learner 243 - 253)
 left context: 5 right context: 5

TH/EA (grid)	/	a	,	La	scrivo	al	rif
learner							
TH1	/	a		Le	scrivo	in	rif
TH1Diff				CHA			CHA

Name	Texts	Tokens
2014-07-29-czech-gold-ea1-pilot	16	2.582
2014-07-29-czech-gold-pilot	436	78.899
2014-07-29-german-gold-ea1-pilot	355	51.413
2014-07-29-german-gold-pilot	1.032	153.197
2014-07-29-italian-gold-ea1-pilot	241	33.653
2014-07-29-italian-gold-pilot	810	106.901
MERLIN_Czech	442	80.213
MERLIN_German	1.033	154.335
MERLIN_Italian	813	107.211

Grafico 36: Tutti i risultati di un calcolo di frequenza (grafici 34-35), nell'output di <<Ricerca avanzata>>

(2) Frequenza relativa delle caratteristiche annotate

Il numero relativo delle caratteristiche annotate nel tuo database per frase o per token. Dal momento che la lunghezza del testo varia considerevolmente attraverso il corpus, è utile usare frequenze normalizzate, ovvero numeri di occorrenze di uno o più fenomeni con riferimento ad un'entità standardizzata, ad es., come in questo caso, frasi o token. Con questa funzione potresti, per fare un esempio, calcolare il numero medio di errori grammaticali in testi di apprendenti che sono stati valutati B1 in confronto a testi valutati B2 (crea prima i tuoi sottocorpora).

La procedura per visualizzare le frequenze relative è analoga alla procedura appena descritta per le frequenze assolute. Dall'output su *<<Statistiche>>*, cliccando sul risultato della tua ricerca, puoi accedere a tutti gli esempi nel corpus mediante l'output della *<<ricerca avanzata>>*.

(3) Linguaggio dell'apprendente senza errori

Questi parametri partono da una prospettiva positiva del linguaggio degli apprendenti, dal momento che viene visualizzata la percentuale di linguaggio che **non** contiene nessuna delle caratteristiche basate su errori che sono state annotate. Puoi calcolare la percentuale di frasi o token senza errori rispetto al numero totale di frasi o token. Ad un livello più specifico, puoi anche scoprire, ad esempio, qual è la percentuale di linguaggio senza alcun errore grammaticale, morfologico o relativo all'uso delle maiuscole.

(4) Parametri di complessità (solo per il tedesco)

Per il tedesco, è stato possibile includere nella sezione di statistiche parametri calcolati automaticamente e riguardanti aspetti di complessità morfologica, lessicale e sintattica. Per l'italiano e il ceco purtroppo non vi erano i presupposti tecnici.

La complessità è un aspetto importante del linguaggio degli apprendenti che ha una stretta relazione con la competenza. In molti studi, è stato mostrato come la complessità sia chiaramente distinguibile dall'accuratezza e dalla fluenza (i cosiddetti studi 'CAF', vedi la bibliografia per qualche riferimento). I parametri ai quali hanno accesso gli utenti MERLIN derivano dalla ricerca sulla complessità e sul giudizio di leggibilità nella L2. Sono stati applicati per la prima volta ai dati MERLIN da Hancke (2013) e Hancke & Meurer (2013) (*<<documentazione>>*).

5. Aiuto



merlin

Home | Ricerca | Su MERLIN ? ceco tedesco inglese italiano

Aiuto

Tutorial
Il video tutorial d'introduzione alle funzioni di ricerca MERLIN sarà presto disponibile.

FAQ
Per cosa posso usare MERLIN? [+](#)
Dove posso trovare ... ? [+](#)
Cosa posso fare con i risultati della mia ricerca? [+](#)
Cos'è ANNIS? [+](#)

Glossario
[A-H](#) [I-O](#) [P-Z](#) [Abbreviazioni nell'output di ricerca avanzata \(G_Infect, O_Punct, etc.\)](#)


A-H
ANNIS
ANNIS è un software open-source (in particolare un'architettura di ricerca e visualizzazione) che visualizza annotazioni multilivello come quelle del corpus MERLIN.
» per maggiori informazioni su ANNIS visita il [sito degli sviluppatori](#)

errore di accordo
Un errore di accordo in MERLIN include forme grammaticali scorrette nelle combinazioni di soggetto e verbo. *» [vedi lo schema di annotazione](#)*

annotazione
marcatura del testo dell'apprendente; ci sono diversi tipi di annotazioni nel corpus MERLIN: *metadati, caratteristiche del linguaggio degli apprendenti*, annotazioni di *parti del discorso*

ID-autore
numero che identifica univocamente un testo nel corpus MERLIN. Puoi trovarlo e copiarlo nei risultati di ricerca della *» Ricerca semplice* o in *» Crea un sottocorpus » Informazioni sugli apprendenti e sulle valutazioni*

Grafico 37: Aiuto

Ci sono molti documenti e funzionalità per aiutarti ad orientarti nell'interfaccia MERLIN. Puoi accedere alla sezione di aiuto () dovunque nell'interfaccia.

5.1 Manuale per l'utente

Questo documento, il manuale per l'utente, è disponibile nella pagina <<aiuto>> dell'interfaccia MERLIN in tedesco, italiano, inglese e ceco, e verrà continuamente aggiornato.

5.2 Familiarizzare con MERLIN: presentazione video

Puoi accedere a una videoregistrazione di un'introduzione del progetto MERLIN di circa 30 minuti (in inglese, di Katrin Wisniewski) per ottenere una prima panoramica del progetto.

5.3 Navigare nell'interfaccia MERLIN: il tutorial screencast

Il tutorial screencast è disponibile in 4 lingue (inglese, tedesco, italiano, ceco) e fa da guida attraverso le principali funzionalità dell'interfaccia MERLIN. Il tutorial ha due blocchi: una prima parte basilare (<<crea un sottocorpus>>, <<ricerca semplice>>) ti consente di iniziare a mettere insieme e/o scaricare testi/compiti ai quali sei interessato, e di cercare parole in essi. In una seconda parte sono fornite spiegazioni su come eseguire una <<ricerca avanzata>> su lemmi e annotazioni, e viene brevemente introdotta la sezione <<statistiche>>.

5.4 Come adoperare MERLIN: scenari d'uso

Nel dicembre 2014, il team MERLIN ha organizzato alcuni workshop a Linz, in Austria, durante i quali la piattaforma MERLIN è stata presentata con l'aiuto di scenari d'uso esemplificativi rivolti agli insegnanti di lingue, ai valutatori e agli educatori. Anche questi materiali sono disponibili gratuitamente per il download nella sezione <<Usare MERLIN>>, in ceco, italiano e tedesco.

5.5 Comprendere l'output della <<Ricerca avanzata>>


Come già accennato (capitolo 4.3), la <<ricerca avanzata>> usa l'architettura di ricerca e visualizzazione open source ANNIS. Nell'output di ricerca, in aggiunta alle caratteristiche incorporate in ANNIS, il team MERLIN ha inserito alcune funzionalità di guida per renderne più semplice la comprensione

The screenshot shows the MERLIN web interface. At the top, there's a blue header with the MERLIN logo and navigation links: Home, Ricerca, Su MERLIN, and a help icon. Language options (ceco, tedesco, inglese, italiano) are on the right. Below the header, there's a navigation bar with links: Torna alla ricerca avanzata, Sul risultato di questa ricerca, and Abbreviazioni. The main content area shows search results for the word 'casa'. It includes a table with columns for learner, TH1, TH1Diff, EA_category, and G_Infect_Inexist_type. The table shows the word 'casa' in red in the learner's text, and its corresponding annotations in the other columns. The interface also includes a search bar, a query result display, and a table of results.

learner	TH1	TH1Diff	EA_category	G_Infect_Inexist_type
, noi abbiamo visto una casa vecchia . Le parete sono	, noi abbiamo visto una casa vecchia . Le pareti sono		CHA	G_Infect_Inexist
				noun

Grafico 38: Funzioni di guida nell'output di <<Ricerca avanzata>>

Cliccando su <<sul risultato di questa ricerca>>, si apre una finestra esterna, in modo da permetterti di continuare ad analizzare l'output di ricerca. Tutti i livelli della griglia dell'output di ricerca avanzata sono spiegati qui (grafico 39).



Home | Ricerca
?

[ceco](#) [tedesco](#) [inglese](#) [italiano](#)

Usare MERLIN

 MERLIN per la ricerca

 Corpus MERLIN

 Annotazioni MERLIN

 Documentazione

 Scarica il corpus

Spiegazione dell'output della tua ricerca

Attenzione: questo testo è stato aperto in una nuova finestra (nuova scheda del tuo browser).

Il risultato della tua ricerca viene visualizzato in ANNIS. ANNIS è un software open-source (in particolare un'architettura di ricerca e visualizzazione) che è in grado di visualizzare annotazioni multilivello. ANNIS consente agli utenti del corpus di esplorare l'intero set delle varie annotazioni MERLIN: target hypotheses (TH1, TH2), annotazioni di caratteristiche del linguaggio degli apprendenti e annotazioni assegnate in automatico (ad es. parti del discorso, frasi, ecc.).

Nel campo di ricerca a sinistra (vedi [1] nello schema sottostante) puoi visualizzare la ricerca che hai avviato nell'interfaccia di ricerca MERLIN (ricerca avanzata) tradotta nel linguaggio di interrogazione ANNIS. Se vuoi cambiare la tua ricerca, puoi tornare all'interfaccia MERLIN ("torna alla ricerca avanzata") o modificarla usando il linguaggio di interrogazione ANNIS.

» Per ulteriori informazioni su ANNIS e sul linguaggio di interrogazione clicca qui » [Homepage ANNIS](#).

Informazioni di base sull'output della tua ricerca in ANNIS

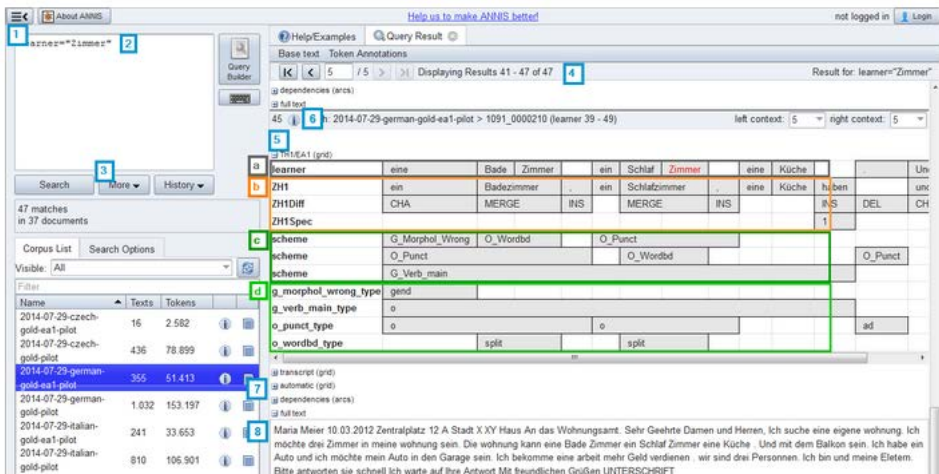


Grafico 39: Funzione di guida <<Sul risultato di questa ricerca>> nell'output di <<ricerca avanzata>>

Un'altra opzione di aiuto è quella di aprire una lista con le <<Abbreviazioni>> usate (grafico 40). I tag di annotazione con brevi definizioni sono visualizzati in un'altra finestra.

Usare MERLIN

MERLIN per la ricerca

Corpus MERLIN

Annotazioni MERLIN

Documentazione

Scarica il corpus

Glossario delle abbreviazioni (tag di annotazione)

Attenzione: questo testo è stato aperto in una nuova finestra (nuova scheda del tuo browser).

Il glossario si riferisce alla griglia TH1/EA1. Per ulteriori dettagli sulle annotazioni automatiche visualizzate nella griglia "automatic" (annotazioni di parti del discorso, lemmi, frasi e periodi) visita la sezione » [MERLIN per la ricerca](#).

Elementi "scheme" nella griglia TH1/EA1

Il livello scheme contiene informazioni sulla caratteristica del linguaggio dell'apprendente che è stata annotata. I seguenti schemi comprendono tutte le abbreviazioni e il loro significato. Le caratteristiche annotate sono illustrate da esempi nella sezione » [Annotazioni MERLIN](#). Per dettagli sulle regole di annotazione, sulle descrizioni dei tag e sui tag span, vai su: » [Schema di annotazione MERLIN](#).

G_Grammatica

G_Agr	accordo (soggetto e verbo)
G_Art	articolo
G_Clit	ITA: clitico
G_Conj	coniunzione
G_Infect_inexist	forma flessa inesistente (sostantivo, aggettivo, verbo)
G_Morphol_wrong	forma flessa scorretta (sostantivi, pronomi, aggettivi)
G_Neg_negdoub	CZE: doppia negazione
G_Neg_neggen	negazione generale
G_POS	errore di parte del discorso
G_Prep	preposizione
G_Refl_pronrefl	pronome riflessivo
G_Refl_pronreflposs	CZE: pronome possessivo riflessivo

Grafico 40: Lista di abbreviazioni usate nell'output di <<ricerca avanzata>>

5.6 Glossario

In questo glossario sono spiegati termini usati nell'interfaccia MERLIN, molti dei quali riguardano l'annotazione.

5.7 Domandi frequenti

Qui sono raccolte le domande degli utenti. Puoi trovare informazioni, ad esempio, su come adoperare MERLIN, su cosa fare con i risultati della tua ricerca, e sull'output della <<ricerca avanzata>>. Questa lista verrà continuamente aggiornata.

5.8 Contattaci

Non esitare a contattare il team MERLIN per qualsiasi domanda o commento (info@merlin-platform.eu). Siamo felici di aiutarti.

Bibliografia

- [ALTE 2001] = ALTE Working Group on the Code of Practice: *Principles of Good Practice for ALTE Examinations*. Revised Draft. http://www.testdaf.de/institut/pdf/ALTE/ALTE_good_practice.pdf, December 2014.
- [Consiglio d'Europa 2004a] = Trim, J./North, B./Coste, D.: *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. La Nuova Italia: Oxford.- A cura del Consiglio d'Europa.
- [Council of Europe 1975] = Van Ek, J. A.: *The Threshold Level in a European unit/credit system for modern language learning by adults*. Strasbourg: Council of Europe.
- [Council of Europe 1994a] = North, B.: *Scales of language proficiency: a survey of some existing systems*. Strasbourg: Council of Europe, CC-Lang (94) 24.
- [Council of Europe 1994b [1981]] = Galli de' Paratesi, N.: *Livello Soglia per l'insegnamento dell'italiano come lingua straniera*. Strasbourg: Edizioni del Consiglio d'Europa.
- [Council of Europe 1999 [1980]] = Baldegger, M./Müller, M./Schneider, G. (1999): *Kontaktschwelle Deutsch als Fremdsprache*. 4. Auflage. Berlin u.a.: Langenscheidt.- ed. by Council of Europe.
- [Council of Europe 2001a] = Trim, J./North, B./Coste, D.: *Common European Framework of Reference for Languages: Learning, teaching, assessment*. -Edited by the Council of Europe. Online-Dokument: www.coe.int/lang, December 2014.
- [Council of Europe 2001b] = Trim, J./North, B./Coste, D.: *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin u.a.: Langenscheidt.- Herausgegeben vom Europarat, Online-Dokument: <http://www.goethe.de/z/50/commeuro/i7.htm>, December 2014.
- [Europarat 2004] = Takala, S./Kaftandjieva, F./Verhelst, N./Banerjee, J./Eckes, T./van der Schoot, F.: *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.
- [Europarat 2009 [2003]] = North, B./Figueras, N./Takala, S./Van Avermaet, P./Verhelst, N.: *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Manual. Preliminary Pilot Version*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.
- Abel, A. / Wisniewski, K. / Nicolas, L. / Boyd, A. / Hana, J. / Meurers, D. (2014): A Trilingual Learner Corpus illustrating European Reference Levels. In: *Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne* 2 (1), 111-126. (<http://www.ojs.unito.it/index.php/ricognizioni>)
- Abel, A. / Glaznieks, A. / Nicolas, L. / Stemle, E. (2014): KoKo: an L1 Learner Corpus for German. In: Calzolari, N. / Choukri, K. / Declerck, Th. / Loftsson, H. / Maegaard, B. / Mariani, J. / Moreno, A. / Odijk, J. / Piperidis, St. (eds.): *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, 26–31 May 2014. Reykjavik: European Language Resources Association (ELRA). 2414-2421. (<http://www.lrec-conf.org/proceedings/lrec2014/index.html>)
- Abel, A. / Wisniewski, K. (2012): Sprechaktrealisierungen in der L2 und der GERS: Ein- und Aussichten für Sprachwissenschaft und Didaktik. In: Di Meola, Claudio / Hornung, Antonie / Rega, Lorenza (Hrsgg.): *Perspektiven Vier. Akten der 4. Tagung Deutsche Sprachwissenschaft in Italien*. Rom, 4.-6.Februar 2010. Frankfurt a.M. 311-325.
- Abel, A. (2010): Sprachtests und soziale Implikationen. In: *Deutsch als Fremdsprache*, 4/2010. 202-209.

- AERA/APA/NCME (1999): *Standards for educational and psychological testing*. Washington: AERA.
- Alderson, J.C. (1991): Bands and scores. In: Alderson, J.C./North, B. (eds.): *Language testing in the 1990s*. London: British Council/Macmillan, 71-86.
- Aguado, K. (2004): Evaluation fremdsprachlicher Wortschatzkompetenz: Funktionen, Prinzipien, Charakteristika, Desiderate. In: Tschirner (Hrsg.) 231- 250.
- Aijmer, K. (1996), *Conversational routines in English. Convention and creativity*. London/NY: Longman.
- Alderson, J. C./Figueras, N./Kuijper, H./Nold, G./Takala, S./Tardieu, C. (2006): Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of the Dutch CEFR Construct Project. In: *Language Assessment Quarterly* 3(1), 3-30.
- Alderson, J.C. (2007): The CEFR and the need for more research. In: *The Modern Language Journal* 91, 658-662.
- Al-Gahtani, S./Roever, C. (2012): Proficiency and Sequential Organization of L2 REQUESTs. In: *Applied Linguistics* 33/1, 42 –65.
- Al-Jarf, R. Spelling error corpora in EFL. In the proceedings of the International Conference on Multi Development and Application of Language and Linguistics, National Cheng Kung University, May 5-16 2009, Tainan City, Taiwan, 2009.
- Arnaud, P. J. L. (1984): The lexical richness of L2 written productionos and the validity of vocabulary tests: In: Culhane, T./Klein-Braley, C./Stevenson, D. K. (eds.): *Practice and Problems in Language*
- Arnaud, P.J.L. (1984): The lexical richness of L2 written productions and the validity of vocabulary tests. In: Culhane, T./Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and Problems in Language Testing*. Essex: Department of Language and Linguistics, University of Essex, 14-28.
- Arras, U. (2010): Subjektive Theorien als Faktor bei der Beurteilung fremdsprachlicher Kompetenzen. In: Berndt, A./Kleppin, K. (eds.): *Sprachlehrforschung: Theorie und Empirie - Festschrift für Rüdiger Grotjahn*. Frankfurt: Lang, 169-179.
- Bachman, L.F. (1990): *Fundamental Considerations in Language Testing*. New York: OUP.
- Bachman, L.F. (2004): *Statistical analyses for language assessment*. Cambridge: CUP 2004.
- Bachman, L.F./Palmer, A. (1996): *Language Testing in Practice*. New York: OUP.
- Bachman, L.F./Palmer, A. (2010): *Language Testing in Practice. Developing Language Assessment and Justifying their Use in the Real World*. Oxford: OUP.
- Bachmann, T. (2002): *Kohäsion und Kohärenz: Indikatoren für Schreibentwicklung: Zum Aufbau kohärenzstiftender Strukturen in instruktiven Texten von Kindern und Jugendlichen*. Innsbruck: Studienverlag.
- Baker, P. (2010): *Sociolinguistics and Corpus Linguistics*. Edingburgh: EUP.
- Bardovi-Harlig, K. & Bofman, T. (1989) Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11 (1), 17-34.
- Bardovi-Harlig, K. (2009): Conventional Expressions as a Pragmalinguistic Resource: Recognition and Productions of Conventional Expressions in L2 Pragmatics. In: *Language Learning* 59 (4), 755-795.
- Bardovi-Harlig, K. (2013): Developing L2 Pragmatics. In: *Language Learning* 63 (1): Suppl. 1, 66-86.
- Barron, A. (2003), Acquisition in Interlanguage Pragmatics. Learning How to do Things with Words in a Study Abroad Context. Amsterdam/Philadelphia: Benjamins.
- Bausch, K.-R./Christ, H./Königs, F.G./Krumm, H.-J. (eds.) (2003): *Der Gemeinsame Europäische Referenzrahmen für Sprachen in der Diskussion. Arbeitspapiere der 15. Frühjarskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr.

- Bayley, R. (2007): Second language acquisition: a variationist perspective. In: Bayley, R. /Lucas, C. (eds.) (2007): *Sociolinguistic Variation : Theories, Methods, and Applications*. Cambridge University Press, 133-144.
- Bayley, R. / Regan, V. (eds) (2004). *The acquisition of sociolinguistic competence*. Special Issue of the *Journal of Sociolinguistics*, 8 (3).
- Beebe, L. (1988). Five sociolinguistic approaches to Second Language Acquisition. In: L.Beebe (ed.), *Issues in Second Language Acquisition: Multiple Perspectives*. Cambridge, MA: Newbury House, pp. 43–75.
- Bestgen, Y./Granger, S. (2011): Categorising spelling errors to assess L2 writing. In: *International Journal of Continuing Engineering Education and Life Long Learning*, 21 (2), 235-252.
- Biber, D./Finegan, E. (eds.) (1994): *Sociolinguistic perspectives on register*. New York: OUP.
- Blum-Kulka, S. (1987), Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11, 1, 131-46.
- Blum-Kulka, S. (1991), Interlanguage pragmatics: The case of requests. In Phillipson, R./Kellerman, E./Selinker, L./Sharwood Smith, M./Swain, M. (eds.) (1991), *Foreign/second language pedagogy research: A*
- Blum-Kulka, S./House, J./Kasper, G. (eds.) (1989), *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.
- Blum-Kulka, S./Olshtain, E. (1984), Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics* 5, 3, 196-213.
- Bond, T. G./Fox, C. M. (2007): *Applying the Rasch model: Fundamental measurement in human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Botley, S. and Dillah, D. (2007) Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, Vol. 3, pp.74-93.
- Bredel, U. (2010) (ed.) : *Schriftsystem und Schrifterwerb: linguistisch – didaktisch – empirisch*. Berlin: de Gruyter.
- Bulté, B./Housen, A. (2012): Defining and operationalising L2 complexity. In: Housen, A./Kuiken, F./Vedder, I. (eds.): *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: Benjamins, 21-46.
- Burger, H. (2007): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. (3. Aufl.).Berlin: Erich Schmidt Verlag.
- Carlsen, C. (2010): Discourse connectives across CEFR levels: A corpus-based study. In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research* (Eurosla). 191-210. purl.org/net/Carlsen-10.pdf
- Carlsen, C. (2010); Linking a learner corpus to the Common European Framework of Reference. Manuscript submitted for publication.
- Carlsen, C. (ed.) 2013. *Norsk Profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus.
- Casanave, C. (1994) Language development in students' journals. *Journal of Second Language Writing*. 3, 179-201.
- Castro, C. D. (2004): Cohesion and the social construction of meaning in the essays of Filipino college students' writings in L2 English, in: *Asia Pacific Education Review*, 5, 215-225.
- Chastain, K. (1990) Characteristics of graded and undergraded compositions. *Modern Language Journal*, 74, 10-14.

- Chen, M., Zechner, K. (2011) Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, June 19-24, 2011. ACL, pp. 722–731.
- Chiang, St. (2003): The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning, in: *System*, Vol.31(4), 471-484.
- Cho, Y. (2005) :Grammatik und Höflichkeit im Sprachvergleich. Direktive Handlungsspiele des Bittens, Aufforderns und Anweisens im Deutschen und Koreanischen. Tübinge.
- Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.
- Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.
- Cook, V. (2005): Second Language writing systems. Clevedon: Multilingual Matters.
- Cook, V. J. (1997). L2 users and English spelling. *Journal of Multilingual and Multicultural Development*, 18(6), 474-488.
- Corder, S. P. (1993 [1973]): *Introducing Applied Linguistics*. Harmondsworth: Pelican.
- Dallapiazza, R.M./von Jan, E., Schönherr, T. (1998) (eds.): *Tangram: Deutsch als Fremdsprache. Kurs- und Arbeitsbuch 1 A*. Munich: Hueber.
- Cornish, F. (2009): Inter-sentential anaphora and coherence relations in discourse: a perfect match, in: *Language Science* 31 (2009), 572-592.
- Coulmas, F. (Hrsg.) (1986): *Direct and Indirect Speech*., Berlin, New York, Amsterdam.
- Crossley, S./Salsbury, T./McNamara, D. (2009): Measuring L2 Lexical Growth Using Hypernymic Relationships. In: *Language Learning* 59 (2) 307-334.
- Crossley, Sc.A./McNamara, D.S. (2011): Shared features of L2 writing: Intergroup homogeneity and text classification, in: *Journal of Second Language Writing* 20 (4) 271-285.
- Dale, E. (1965): Vocabulary measurement: Techniques and major findings. In: *Elementary English* 42, 895-901.
- Daller, H./Milton, J./Treffers-Daller, J. (eds.) (2007): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: CUP.
- Daller, H./van Hou, R./Treffers-Daller, J. (2003): Lexical richness in spontaneous speech of bilinguals. In: *Applied Linguistics* 24, 197-222.
- Dellapiazza, R.M./von Jan, E., Schönherr, T. (1998) (Hrsgg.): *Tangram: Deutsch als Fremdsprache. Kurs- und Arbeitsbuch 1 A*. Munich: Hueber.
- Dewaele, J.-M. (2004): Individual differences in the use of colloquial vocabulary. The effects of sociobiographical and psychological factors. In: Bogaards, P./Laufer, L. (eds.): *Vocabulary in a second language*. Amsterdam: John Benjamins, 127-154.
- Dewaele, J.-M. (2004): The acquisition of sociolinguistic competence in French as a foreign language: An overview. In: *French Language Studies* 14, 301–319.
- Díaz-Negrillo, A./Fernández-Domínguez, J. (2006): Error-coding systems for learner corpora. In: *RESLA* 19, 83-102.
- Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.
- Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.
- Eckes, T. (2008): Rater types in writing performance assessments: A classification approach to rater variability. In: *Language Testing* 25 (2) 155-185.

- Eckes, T. (2009): *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Section H: Many-Facet Rasch Measurement*. (http://www.coe.int/t/dg4/linguistic/manuel1_en.asp, December 2014.)
- Eisenberg, P. (2007): Sprachliches Wissen im Wörterbuch der Zweifelsfälle. über die Rekonstruktion einer Gebrauchsnorm. In: *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 3/2007: 209-228.
- Ellis, R. (1994): *The study of Second Language Acquisition*. Oxford: Oxford University Press.
- Fabrics-Hansen, C. (2005): Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics*, 43, 17-48.
- Fatemi, M. A. (2008) The relationship between writing competence, language proficiency and grammatical errors in the writing of Iranian tefl sophomores. Doctoral thesis. Universiti Sains Malaysia.
- Fender, M. Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language*, 2008, Volume 20, No. 1, pp. 19-42.
- Feng, L. (2010): Automatic readability assessment. PhD thesis, City University of New York (CUNY), <http://gradworks.umi.com/3426751.pdf> (February 2015).
- Foster, P./Tavakoli, P. (2009): Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. In: *Language Learning* 59 (4) 866-896.
- Fulcher, G. (2004): Deluded by Artifices? The Common European Framework and Harmonization. In: *Language Assessment Quarterly* 1 (4), 253-266.
- Fulcher, G./Davidson, F. (2007): *Language Testing and Assessment*. London/New York: Routledge.
- Gould, S.J. (1996): *The mismeasure of man*. London: Penguin.
- Gernsbacher, M. A./Givón, T. (Eds.) (1995): *Coherence in Spontaneous Text*. Benjamins, Amsterdam.
- Glaznieks A./Nicolas L./Stemle E./Lyding V./Abel A. (2012): Establishing a Standardised Procedure for Building Learner Corpora. In: *Apples - Journal of Applied Language Studies. Special Issue: Proceedings of LLLC2012*.
- Graesser, A. C./Millis, K. K./Zwaan, R. (1997): Discourse comprehension, in: *Annual Review of Psychology* 48, 163-189.
- Granger, S. (2002): A Bird's-eye view of learner corpus research. In: Granger S./Hung, J./ Petch-Tyson, St (eds.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.
- Granger, S. (2003): Error-tagged learner corpora and CALL: a promising synergy. In: *CALICO Journal* 20 (3). Special issues on error analysis and error correction in computer-assisted language learning, 465-480.
- Granger, S. (2008): Learner corpora. In: Lüdeling, A. / Kytö, M. (eds.): *Corpus linguistics: an international handbook* (Handbooks of linguistics and communication science; 29.1_ 29.2). Berlin - New York: de Gruyter. 259-275.
- Granger, S., Bestgen, Y (2011) Categorizing spelling errors to assess L2 writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21, 2-3, 2011, 235 - 252.
- Guiraud, P. (1954): *Les caractères statistiques du vocabulaire*. Paris: Presse Universitaires de France.
- Gyllstad, H./Granfeldt, J./Bernardini, P./Källkvist, P. (2014): Linguistic correlates to communicative proficiency levels of the CEFR. The case of syntactic complexity on written L2 English, L3 French and L4 Italian. In: *EUROSLA Yearbook* 14, 1-30.
- Halliday, M. A. K. /Hasan, R. (1989): *Language, context and text: a social semiotic perspective*. Oxford: Oxford University Press.

- Halliday, M. A. K./Hasan, R. (1976): *Cohesion in English*. London, Longman.
- Hana, J./Rosen, A./ Štindlová, B./Štěpánek, J. (2014): Building a learner corpus. In: *Language Resources and Evaluation* 8, SE
- Hana, J./Rosen, A./Skodová, S. Stindlová, B.: Error-tagged learner corpus of Czech. In: *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010, Uppsala, Sweden 2010*, 11-19.
- Hancke J./Meurers D./Vajjala S. (2012): Readability Classification for German using lexical, syntactic, and morphological features. In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, 1063-1080.
- Hancke, J. *Automatic Prediction of CEFR Proficiency Levels Based on Linguistic Features of Learner Language*. Master's thesis, Universität Tübingen, April 2013.
- Hancke, J./Meurers, D./Vajjala, D. (2012): Readability classification for German using lexical, syntactic, and morphological features. In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, 1063-1080, Mumbai, India.
- Hasil, J./Hájková, E./Hasilová, H. (2007): *Brána jazyka českého otevřená*. Prague: Karolinum.
- Hasko, V. (2013): Capturing the Dynamics of Second Language Development via Learner corpus research: a very long engagement. In: *The Modern Language Journal* 97, S1, 1-10.
- Hattingh, K. (2005) The syntactic development of grade 12 ESL learners. Dissertation, North-West University, Potchefstroom Campus.
- Hawkey, R./Barker, F. (2004): Developing a Common Scale for the Assessment of Writing. In: *Assessing Writing* 9, 122-159.
- Hawkins, J. A./Filipović, L. (2012): *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*. Cambridge: CUP.
- Held, G. (1995): Verbale Höflichkeit. Studien zur linguistischen Theorienbildung und empirische Untersuchung zum Sprachverhalten französischer und italienischer Jugendlicher in Bitt- und Dankessituationen. Tübingen: Narr.
- Holmes/Brown 1976: Developing sociolinguistic competence in a second language. *Tesol Quarterly* 10/4.
- Housen, A., Kuiken, F. (2009) Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4): 461-473
- Housen, A./Kuiken, F. (2009): Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics* 30 (4) 461-473.
- Housen, A./Kuiken, F./Vedder, I. (eds.) (2012): *Dimensions of L2 Performance and Proficiency Investigating Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins.
- Hovermale, DJ., Martin, S. Developing an Annotation Scheme for ELL Spelling Errors. *Proceedings of MCLC-5 (Midwest Computational Linguistics Colloquium)* East Lansing, Michigan, USA, May 10-11, 2008.
- Hudson, Tom/Detmer, Emily/Brown, J.D. (1992): A framework for testing cross-cultural pragmatics. Hawaii: University of Hawai'i at Manoa.
- Hudson, Tom/Detmer, Emily/Brown, J.D. (1995): *Developing prototypic measures of cross-cultural pragmatics*. Hawaii: University of Hawai'i at Manoa.
- Hulstijn, J. H. (2007): The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. In: *The Modern Language Journal* 91, 663-667.
- Hulstijn, J. H./Alderson, C./Schoonen, R. (2010): Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic development: intersections between SLA and language testing research*. Eurosla Monograph Series. (<http://eurosla.org/monographs/EM01/EM01home.html>)

- Hymes, D. (1974): *Foundations in Sociolinguistics*. Philadelphia: University of Pennsylvania Press.
- Ishikawa, S. (1995) Objective Measurement of Low-Proficiency EFL Narrative Writing. *Journal of Second Language Writing*, 4: 51 - 70.
- Jarvis, S. (2002): Short texts, best-fitting curves and new measures of lexical diversity. In: *Language Testing* 19 (1) 57-84.
- Jelínek, T., Barbora Štindlová, Alexandr Rosen, Jirka Hana (2012). Combining Manual and Automatic Annotation of a Learner Corpus. Text, Speech and Dialogue Lecture Notes in Computer Science Volume 7499, pp 127-134 <http://ufal.mff.cuni.cz/~hana/bib.html#rosen-et-al-2013-czesl-lre>, December 2014.
- Johns, T. (1988): Whence and whither classroom concordancing? In: Bongaarts, T./de Haan, P./Lobbe, S./Wekker, H. (eds.): *Computer Applications in Language Learning*. Dordrecht: Foris, 9-33.
- Johns, T. (1997): Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In: Wichmann, Anne/Fligelstone, Steven/McEnery, Tony/Knowles, Gerry (eds.) (1997), *Teaching and Language Corpora*. London: Longman, 100-115.
- Kaczmarek, C. M. (1980) Scoring and Rating 'essay tasks'. (& Oller, J.W. and Perkins, K. &. Research in Language Testing. Rowley, Massachusetts: Newbury House.)
- Kroll, B. (1990) What does time buy? ESL student performance on home versus class compositions. In B. Kroll, ed. *Second language writing: Research insight for the classroom*. Cambridge: CUP, pp. 140 – 154.
- Kuiken, F./Vedder, I./Gilbert, R. (2010): Communicative Adequacy and Linguistic complexity in L2 writing. *EUROSLA Monographs Series* 1, 81-100.
- Larsen-Freeman, D. (2009): Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics* 30 (4) 579-589.
- Laufer, B. (1995): Beyond 2,000. A measure of productive lexicon in a second language. In: Eubank, L./Selinker, L./Sharwood Smith, M. (eds.): *The Current State of Interlanguage*. Amsterdam/Philadelphia: John Benjamins, 265-272.
- Laufer, B./Nation, P. (1995): Vocabulary size and use: lexical richness in L3 written production. In: *Applied Linguistics* 16, 307-322.
- Linnarud, M. (1986): *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö: CWK Gleerup.
- Little, D. (2007): The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Languages Education Policy. In: *The Modern Language Journal* 91, 645-655.
- Lorenzo-Dus, N. (2007): The best of both worlds? Combined methodological approaches to the assessment of vocabulary in oral proficiency interviews. In: Daller/Milton/Treffers-Daller (eds.) 220-233.
- Louwerse, M.M./Graesser, A. C. (2004). Coherence in discourse, in: Strazny, P. (Ed.): *Encyclopedia of linguistics*. Chicago: Fitzroy Dearborn.
- Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics* 14, 3–28(26). URL <http://www.ingentaconnect.com/content/jbp/ijcl/2009/00000014/00000001/art00002>.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474-496.
- Lu, X. (2011): A corpus-based evaluation of syntactic complexity measures as indices of College-level ESL writers' language development. In: *TESOL Quarterly* 45 (1) 36-62.
- Lu, X. (2012): The relationship of lexical richness to the quality of ESL learners' oral narratives. In: *The Modern Language Journal* , 190-208.
- Lüdeling, A. (2008): Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In: Walter, M./Grommes, P. (eds.): *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, 119-140.

- Lüdeling, A./Walter, M./Kroymann, E./Adolphs, P. (2005): Multi-level Error Annotation in Learner Corpora. In: Hunston, S./Danielsson, P. (eds.): *Proceedings from the Corpus Linguistics Conference Series* (Corpus Linguistics 2005, Birmingham, 1415 July 2005). (<http://www.corpus.bham.ac.uk/PCLC>).
- Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15 (3) 263–287.
- Malvern, D./Richards, B./Chipere, N./Durán, P. (2008²): *Lexical Diversity and Language Development. Quantification and Assessment*. New York: Palgrave Macmillan.
- Matsuda, P. K. (1997): Contrastive rhetoric in context: A dynamic model of L2 writing, in: *Journal of Second Language Writing*, 6, 45-60.
- McCarthy, P.M./Jarvis, S. (2007): vocd: A theoretical and empirical evaluation. In: *Language Testing* 24 (4) 459-488.
- McNamara, D. S./Kintsch, E./Butler Songer, N./Kintsch, W. (1996): Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text, in: *Cognition and Instruction*, Vol. 14, No. 1 (1996), 1-43.
- McNamara, D.S./Kintsch, W. (1996): Learning from text: Effects of prior knowledge and text coherence, in: *Discourse Processes*, 22, 247-287.
- McNamara, D.S./Louwerse, M.M. /Graesser, A.C. (unpublished): Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Grant proposal.
- Mellor, A. (2011): Essay Length, Lexical Diversity and Automatic Essay Scoring. In: *Memoirs of the Osaka Institute of Technology*, Series B Vol. 55, No. 2 (2011), 1-14.
- Ménard, N. (1983): *Mesure de la richesse lexicale. Théorie et vérifications expérimentales. Études stylométriques et sociolinguistiques*. Genf/Paris: Slatkine-Champion.
- Meurers, D. (2012): Natural Language Processing and Language Learning. *Encyclopedia of Applied Linguistics*. Blackwell. purl.org/dm/papers/meurers-11.html.
- Mezzadri, M. (2000). *Rete! Book 1*. Perugia: Guerra Edizioni.
- Mezzadri, M. (2000): *Rete! Book 1*. Perugia: Guerra Edizioni.
- Mougeon, R./Dewaele, J.-M. (2004): Preface. In: *IRAL* 42 (4) (Special Issue: Variation in the interlanguage of advanced second language learners.), 295-301.
- Mougeon, Raymond/Nadasdi, Terry/Rehner, Katherine (2010): *The Sociolinguistic Competence of Immersion Students*.
- Müller, Ch./Strube M. (2006): Multi-Level Annotation of Linguistic Data with MMAX2. In: S. Braun, K. Kohn, J. Mukherjee (Eds.): *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*. Frankfurt: Peter Lang, 197-214.
- Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87, 261-276.
- Nation, P. (2001): *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. (2007): Fundamental issues in modelling and assessing vocabulary knowledge. In: Daller, H./ Milton, J./Treffers-Daller, J. (eds.): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2005): *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency*. Oxford: Peter Lang.
- North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency*. Oxford: Peter Lang.

- Nuzzo, E. (2007): *Imparare a fare cose con le parole. Richieste, proteste, scuse in italiano lingua seconda*. Perugia: Guerra.
- Nuzzo, E. (2009): "Buongiorno, ho bisogno dell'informazione per andara a barcellona": uno studio longitudinale sulle richieste di informazioni e suggerimenti in italiano L2. In: *Linguistica e Filologia* 28, 83–109.
- O'Loughlin, K. (1995): Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. In: *Language Testing* 12 (2) 217-237.
- Okada, T. A Corpus Analysis of Spelling Errors Made by Japanese EFL Writers.
- Ortega, L. (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24 (4), 492-518.
- Ortega, L. (2012): Interlanguage complexity: A construct in search of theoretical renewal. In: Szmrecsanyi, B./Kortmann, B. (Eds): *Linguistic complexity in interlanguage varieties, L2 varieties, and contact languages*. Berlin: Walter de Gruyter.
- Ott, N. (2009). Information Retrieval for Language Learning: An Exploration of Text Difficulty Measures. Master's thesis, University of Tübingen, Seminar für Sprachwissenschaft, Tübingen, Germany. URL <http://drni.de/zap/ma-thesis>.
- Paquot, M./Granger, S. (2012): Formulaic language in Learner Corpora. In: *Annual Review of Applied Linguistics* 32, 130-149.
- Perfetti, C. A., Rieben, L. & Fayol, M. (eds) (1997). *Learning to Spell: Research, Theory, and Practice across Languages*. Mahwah, NJ: Lawrence-Erlbaum Associates.
- Petersen, S.E./Ostendorf, M. (2009): A machine learning approach to reading level assessment. In: *Computer Speech and Language* (23) 86-106.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Pollitt, A./Murray, N.L. (1996): What raters really pay attention to. In: Milanovic, M./Saville, N. (eds.): *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium*. Cambridge: Cambridge University Press, 74-91.
- Preston, D. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbins, F. Morris, M. E. Anderson, C. E. Klee and E. Tarone (eds), *Social and cognitive factors in second language acquisition*. Somerville: Cascadia Press, pp. 3–30.
- Read, J. (2000): *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2007): Second language vocabulary assessment: current practice and new directions. In: *International Journal of English Studies* 7 (2) 105-125.
- Read, J./Chapelle, C. (2001): A framework for second language vocabulary assessment. In: *Language Testing* 18, 1-32.
- Read, J./Nation, P. (2004): Measurement of formulaic sequences. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 23-35.
- Regan, V. (1995). The acquisition of sociolinguistic native speech norms. In: B. Freed (ed.), *Second language acquisition in a study abroad context*. Amsterdam, Philadelphia: Benjamins, pp. 245–267. (vorgemerkt)
- Regan, Vera/Howard, Martin/Leme, Isabelle (2009): *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*.
- Reznicek, M./Lüdeling, A./Hirschmann, H. (in print): Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: Díaz-Negrillo, A./Ballier, N./Thompson, P. (eds.): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins (Series Studies in Corpus Linguistics).

- Reznicek, M./Lüdeling, A./Krummes, C./Schwantuschke, F./Walter, M./Schmidt, K./Hirschmann, H./Andreas, T. (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen*. Version 2.01. HU Berlin (http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/Falko-Handbuch_Korpusaufbau%20und%20Annotationen_v2.01)
- Rimrott, A., Heift, T. (2008). Evaluating automatic detection of misspellings in German. *Language Learning & Technology*.
- Römer, U. (2010): Using general and specialized corpora in English language teaching: past, present and future. In: Campoy-Cubillo, M. et al. (eds.): *Corpus-based approaches to English Language Teaching*. London: Continuum, 18-38.
- Römer, Ute. 2008. 7. Corpora and language teaching. In: Lüdeling, Anke & Merja Kytö (eds.). *Corpus Linguistics. An International Handbook (volume 1)*. [HSK series] Berlin: Mouton de Gruyter. 112-130.
- Römer, U. (2006): Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. In: *Zeitschrift für Anglistik und Amerikanistik* 54 (2) 121-134.
- Rose, K.R. (2000), An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition* 22, 1, 27-67.
- Rosen, A., Jirka Hana, Barbora Štindlová, and Anna Feldman (2013): Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation*, pages 1-28, April 2013. <http://ufal.mff.cuni.cz/~hana/bib.html#rosen-et-al-2013-czesl-lre>, December 2014.
- Sanders, T./Pander Maat, H. (2006): Cohesion and coherence: Linguistic approaches, in: Brown, K., et al. (Eds.), *Encyclopedia of Language and Linguistics*. Elsevier, London.
- Sassoon, R. (1995). *The Acquisition of a Second Writing System*. Oxford: Intellect.
- Schmitt, N. (ed.) (2004): *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Press.
- Schmitt, N./Carter, N. (2004): Formulaic sequences in action: An Introduction. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 1-21.
- Schneider, G./North, B. (2000): *Fremdsprachen können - was heißt das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme. Chur, Zürich: Rüegger.
- Schneider, J. G. (2013): Sprachliche ‚Fehler‘ aus sprachwissenschaftlicher Sicht. In: *Sprachreport* 1-2/2013, 30-37.
- Skehan, P. (2009): Modelling Second Language Performance: Integrating complexity, accuracy, fluency and lexis. In: *Applied Linguistics* 30 (4) 510-532.
- Škodová, S., Barbora Štindlová, Jirka Hana and Alexandr Rosen (2011). Víceúrovňová anotace českého žákovského korpusu. In: V. Petkevič and A. Rosen (Eds.), *Korpusová lingvistika Praha 2011: 3 - Gramatika a značkování korpusů*, 16. pp 208-225. Studie z korpusové lingvistiky. Nakladatelství Lidové noviny: Praha.
- Spinelli, B./Parizzi, F. (ed.) (2010): *Profilo della lingua italiana*. Firenze: La Nuova Italia.
- Stede, M. (2007): *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik*. Tübingen: Narr.
- Spooren, W./Sanders, T. (2008): The acquisition order of coherence relations: On cognitive complexity in discourse, in: *Journal of Pragmatics* 40 (2008), 2003-2026.
- Štindlová B., S. Škodová, A. Rosen and J. Hana (2012). Annotating foreign learners' Czech In: *Studies in Formal Slavic Linguistics. Contributions from Formal Description of Slavic Languages 8.5*, Peter Lang GmbH, Frankfurt am Main, Germany, pp. 205-219.
- Taguchi, N. (2012): *Context, individual differences and pragmatic competence*. Bristol: Multilingual Matters.

- Tarone, E. (2007): Sociolinguistic approaches to second language acquisition research, 1997-2007. In: *Modern Language Journal* 91, 837-848.
- Trosborg, A. (1995), *Interlanguage pragmatics: Requests, complaints, apologies*. Berlin: Mouton de Gruyter.
- UIMA Framework: <http://uima.apache.org> (July 2013)
- Ure, J.N. (1971): Lexical density and register differentiation. In: Perren, G.E./Trim, J.L.M. (eds.): *Applications of Linguistics: Selected papers of the Second International Congress of Applied Linguistics*. Cambridge: CUP, 443-452.
- Vajjala, S./Meurers, D. (2012): On improving the accuracy of readability classification using insights from second language acquisition. In: Joel Tetreault, Jill Burstein, and Claudia Leacock, editors, *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7) at NAACL-HLT*. Montreal, Canada, June 2012, 163-173. <http://aclweb.org/anthology/W12-2019.pdf> (February 2015).
- van Compernelle, Remi A. /Williams, Lawrence (2012): Teaching, Learning, and Developing L2 French Sociolinguistic Competence: A Sociocultural Perspective *Applied Linguistics* (2012) 33(2): 184-205 first published online January 24, 2012.
- Vaughan, C. (1991): Holistic assessment: What goes on in the rater's mind? In: Hamp-Lyons L. (ed.): *Assessing Second Language Writing in Academic Contexts*. Norwood: Ablex, 111.125.
- Vedder, I.(2007): Competenza pragmatic e complessità sintattica in italiano L2: l'uso dei modificatori nelle richieste. In: *Linguistica e Filologia* 25, 99-123.
- Vivanco, V. (2005): The absence of connectives and the maintenance of coherence in publicity texts, in: *Journal of Pragmatics* 37 (2005), 1233-1249.
- Warga, M. and U. Scholmberger (2007). The acquisition of French apologetic behaviour in a study abroad context. *Intercultural Pragmatics*, 4, 221-251.
- Wesche, M./Paribakht, T.S. (1996): Assessing second language vocabulary knowledge depth versus breadth. In: *The Canadian Modern Language Review* 53, 13-40.
- Wisniewski, K. (2010): Bewertungsvariabilität im Umgang mit GeRS-Skalen. Ein- und Aussichten aus einem Sprachtestprojekt. In: *Deutsch als Fremdsprache* 3, 143-150.
- Wisniewski, K. (2012): Lexikalische Kompetenzen in der Fremdsprache testen: Ein Modellierungsansatz. In: Abel/Vettori/Wisniewski (Hrsg.), Bd. 2, 24-49.
- Wisniewski, K. (2013): The empirical validity of the CEFR fluency scale: the A2 level description. In: Galaczi, E.D./Weir, C.J. (eds.): *Exploring Language Frameworks: Proceedings of the ALTE Krakow Conference*. Cambridge: Cambridge University Press, 253-272. *Studies in Language Testing*.
- Wisniewski, K. (2014): *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen*. Frankfurt: Peter Lang. *Language Testing and Evaluation Series*, 33.
- Wisniewski, K./Schöne, K./Nicolas, L./Vettori, C./ Boyd, A./Meurers, D./ Abel, A./Hana, J. (2013): MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data. In: *ICT for Language Learning, Conference Proceedings 2013*. Libreriauniversitaria.it Edizioni. (http://conference.pixel-online.net/ICT4LL2013/common/download/Paper_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf)
- Wolfe-Quinteiro, K., Inagaki, S., Kim, H.-Y. (1998) Second language development in writing: measures of fluency, accuracy, & complexity. Hawai'i: University of Hawai'i at Manoa.
- Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: CUP.
- Yang, W./Sun, Y. (2012): The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. In: *Linguistics and Education*, 23 (1), 31-48.
- Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

- Young, R. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics*, 19: 105–32.
- Yu, Ming Chung (2005): Sociolinguistic Competence in the Complimenting Act of Native Chinese and American English Speakers: A Mirror of Cultural Value. In: *Language and Speech* 48 (1) 91-119.
- Yu, Ming-chung (2011): Learning how to read situations and know what is the right thing to say or do in an L2: A study of socio-cultural competence and language transfer. *Journal of Pragmatics*, 2011, Vol.43(4), psp.1127-1147.
- Zeldes, A./Ritz J./Lüdeling A. et al. (2009): *Annis: A search tool for multi-layer annotated corpora*. In *Proceedings of Corpus Linguistics*, July 20-23. Liverpool. (<http://ucrel.lancs.ac.uk/publications/cl2009/>)
- Zipser, F./Romary, L./al. (2010). A model oriented approach to the mapping of annotation formats using standards. In: *Workshop on Language Resource and Language Technology Standards, LREC 2010*.
- Zuskin, Robin D. (1992): Assessing L2 Sociolinguistic Competence: In Search of Support from Pragmatic Theories. In: *Pragmatics and Language Learning* 4.