

**SIE-L** | Società Italiana di  
e-Learning

**UNIMORE**  
UNIVERSITÀ DEGLI STUDI DI  
MODENA E REGGIO EMILIA



**SIEL2011**

VIII Congresso Nazionale della  
Società Italiana di e-Learning

**Connessi!**  
**Scenari di Innovazione nella  
Formazione e nella Comunicazione**

a cura di  
**Tommaso MINERVA, Luigi COLAZZO**

**Ledizioni**   
*The Innovative LEDpublishing Company*

Reggio Emilia, 14-16 settembre 2011  
Università di Modena e Reggio Emilia

SIEL 2011 - VIII Congresso della Società Italiana di e-Learning

Connessi! Scenari di Innovazione  
nella Formazione e nella Comunicazione

A cura di  
Tommaso Minerva, Luigi Colazzo

Reggio Emilia, 14 al 16 settembre 2011  
Università di Modena e Reggio Emilia

ISBN: 9788895994765

Prima edizione: Settembre 2011

© 2011 Edizioni Ledizioni LediPublishing  
Via Alamanni 11 Milano  
<http://www.ledizioni.it>  
e-mail: [info@ledizioni.it](mailto:info@ledizioni.it)

## Analisi del lessico e individualizzazione del messaggio di apprendimento. La preparazione di un corso nell'ambito del progetto "adaptive message learning"

Andrea ZINI<sup>1</sup>

<sup>1</sup>Dipartimento di Educazione e Scienze Umane, Università degli Studi di Modena e Reggio Emilia

### Abstract

*"Adaptive-message learning" project (FIRB 2009 - 2012) provides a framework for improving the quality of e-learning in environments of individualized instruction by focusing on important aspects of learning: message perception, adaptation and understanding. Project's main goal is to find effective it-based solutions to assess individual's verbal skills and tailor texts' reading ease. At this time point, focus is on vocabulary breadth needed for adequate reading comprehension of specialized texts in university-level education. A trial adaptive LMS - including corpus analysis software, testing tool and adaptive tool - has been developed and tested on different academic courses. This paper reports findings from a field tryout at Modena and Reggio Emilia University.*

**Keywords:** individualized instruction, e-learning, reading comprehension, vocabulary knowledge, specialized texts.

### Introduzione

Il progetto "adaptive-message learning" (FIRB 2009-2012) affronta uno dei problemi teorici e operativi centrali nella ricerca in campo didattico - l'adeguamento del processo di istruzione alle caratteristiche dei soggetti che apprendono - sviluppando una linea originale che intende <definire strategie di individualizzazione praticabili solo tramite l'uso di sistemi tecnologici per la comunicazione e l'elaborazione dei dati> [1] che non intervengano solo a seguito del manifestarsi di una difficoltà di apprendimento durante il corso di studio, ma permettano di adattare il messaggio di istruzione alle esigenze di ciascun destinatario, ossia ad alcune caratteristiche del suo profilo rilevate in maniera continua nel contesto di un ambiente che preveda elevati livelli di interazione.

In estrema sintesi, <l'obiettivo scientifico primario della proposta progettuale è la messa a punto di un ambiente adattivo per l'istruzione in rete [...] che consenta la sperimentazione di una didattica fortemente individualizzata>. Fra gli obiettivi immediatamente correlati vi sono:

- 1) <l'elaborazione di modelli per uno strumentario valutativo innovativo utilizzabile sia per la precisazione della proposta di istruzione, sia per la verifica continua delle caratteristiche dei destinatari del messaggio di istruzione e dei cambiamenti che intervengono durante la fruizione di un corso. Si ritiene che lo strumentario valutativo possa in larga misura affidarsi all'analisi automatizzata del linguaggio utilizzato per la formulazione dei messaggi e prodotto dai fruitori nelle situazioni interattive promosse all'interno della procedura didattica;
- 2) la progettazione e la realizzazione di uno stimatore automatico di competenza verbale, come principale strumento per stabilire l'adeguatezza del messaggio alle caratteristiche del destinatario. Lo stimatore stabilirà l'intervallo di competenza linguistica esistente in un momento determinato tra un soggetto virtuale, le cui caratteristiche siano uguali a quelle supposte per la comprensione del messaggio, e i soggetti reali;
- 3) la progettazione e la realizzazione di un modulatore, ovvero di un automa in grado di operare l'adattamento di un testo, a partire dalle indicazioni ottenute dallo stimatore, per renderlo conforme alle esigenze di chi apprende>. [2]

<Si tratta, infatti, di un progetto che supera la tradizionale unicità del messaggio di istruzione, a favore della sua modulazione al fine di adeguarlo alla competenza verbale di chi lo riceve. Ciò comporta che si sia in grado di stimare tale competenza all'inizio e durante il percorso di istruzione e che alla stima segua l'aggiustamento nel messaggio perché soddisfi le esigenze di ciascuno studente>. [3]



**Figura 1** - Da B. Vertecchi [4], Procedura di individualizzazione del messaggio per l'istruzione a distanza centrata sulla stima del lessico disponibile.

<L'ideazione e lo sviluppo di soluzioni automatizzate che rispondano alle caratteristiche indicate consentono di realizzare sistemi per l'istruzione in rete di alta qualità, il cui intento è di ottimizzare l'apprendimento in vista del raggiungimento dei medesimi obiettivi da parte dell'intero gruppo degli allievi (individualizzazione)>. [2]

Partecipano al progetto tre unità di ricerca, rispettivamente insediate nel Dipartimento di Progettazione Educativa e Didattica dell'Università Roma Tre (capofila; responsabile scientifico il prof. Benedetto Vertecchi, coordinatore nazionale), nel Dipartimento di Fisiopatologia medica dell'Università "La Sapienza" (responsabile il prof. Andrea Lenzi) e nel Dipartimento di Educazione e Scienze Umane dell'Università di Modena e Reggio Emilia (responsabile il prof. Luciano Cecconi).

Durante la prima fase del progetto sono stati realizzati dal DiPED di Roma Tre due prototipi di dispositivi automatici in grado di produrre la stima della competenza verbale (LexMeter) e la modulazione del messaggio (Adapter), integrati all'interno di una nuova piattaforma LMS (OrbisDictus) [5] che ha reso possibile il collaudo del sistema.

### **Lessico e comprensione**

Della competenza lessicale (definita dalle dimensioni di ampiezza, profondità e automaticità) la letteratura attesta l'importante relazione sia con l'abilità di lettura (decodifica), sia con i fenomeni di comprensione del testo. La frequenza d'uso delle parole risulta sin dai primi studi ottocenteschi un importante criterio quantitativo utile a predire - sotto il riguardo lessicale - la leggibilità dei testi scritti per una determinata categoria di lettori. [6] <Un esame della frequenza delle parole mostra come venga soddisfatto un importante principio di un testo adattato per una determinata fascia d'età, e cioè la presenza di una maggioranza di parole familiari, ma anche l'inclusione di una percentuale, sia pur modesta, di parole rare di cui il contesto può aiutare a decodificare il significato. Brani con queste caratteristiche garantiscono una comprensibilità di fondo, ma anche un progresso nell'apprendimento del lessico e dei concetti implicati. Per valutare la difficoltà di un brano e la sua corrispondenza ad una certa fascia scolastica, bisogna tenere conto di molti altri indici> (Cornoldi et al., 2001). [7] Fra le

notevoli differenze che distinguono i primi spogli di frequenza, come quelli esposti nei *Teacher's Word Book* di E. L. Thorndike, dalle attuali risorse disponibili per la valutazione dell'aspetto lessicale della leggibilità dei testi, come sono per la lingua italiana il Vocabolario di Base e il Grande Dizionario dell'Uso di T. De Mauro, vi sono la considerazione della dispersione di una parola all'interno del corpus accanto alla sua frequenza assoluta (per non sovrastimare l'uso di parole che risutano peculiari di una limitata porzione del campione), l'attenzione alla composizione del campione (che non deve contemplare solo il linguaggio scritto), l'uso di criteri diversi dalla frequenza (per non sottostimare l'alta o comune disponibilità di vocaboli che raramente si scrivono o pronunciano ma sono ampiamente noti). [8]

P. E. Tressoldi e C. Zamperlin hanno recentemente costruito e validato una batteria di approfondimento per la valutazione delle difficoltà nella comprensione del testo, per la quale hanno selezionato cinque fra le dieci componenti cognitive e metacognitive considerate nella *Nuova guida alla comprensione del testo* (De Beni, Cornoldi e gruppo MT, 2004): estensione del vocabolario (sinonimi); inferenza lessicale; inferenza semantica; metacomprendione; strategie di studio. Nello studio condotto su un campione di 1272 studenti, suddiviso in vari livelli dalla terza elementare alla terza media, per ciascuna fascia di età una corrispondente prova di comprensione della batteria MT è stata utilizzata come <variabile dipendente> e le cinque prove della nuova batteria come <predittori>. Ne risulta che le prime tre variabili dimostrano complessivamente una più forte relazione con i risultati alla prova di comprensione e si evidenzia che <le prove che richiedono conoscenze e inferenze lessicali differenziano nettamente i due gruppi di abilità nella comprensione del testo [i lettori scarsi, sotto il 10° percentile, dai bravi lettori]. Le altre tre abilità considerate, inferenze semantiche, metacomprendione e strategie di studio, differenziano sostanzialmente allo stesso modo i due gruppi anche se in misura minore rispetto alle prime due [vocabolario/sinonimi e inferenze lessicali]>. [9]

Oltre ai dati raccolti in diverse indagini internazionali sul letteratismo funzionale nella popolazione adulta (es. OCSE-ALL 2003: età 16-65 anni), esistono studi recenti sulla comprensione della lettura di testi per l'apprendimento a livello universitario che confermano un basso livello generale e descrivono un quadro complesso delle diverse componenti del processo, dal quale risulta che l'abilità lessicale è la più critica per l'intero campione e in particolare per i lettori "poveri". La prova di comprensione utilizzata per lo studio condotto da E. Truffelli all'Università di Bologna ha inteso accertare quattro abilità, individuate e definite seguendo la classificazione presentata per l'indagine IEA-SAL (Corda Costa, Visalberghi, 1995): abilità lessicale (definita come conoscenza del significato di uno specifico termine o sintagma), localizzazione di informazioni, elaborazione di inferenze, riconoscimento dell'idea principale. <Si è scelto di inserire nella prova testi simili a quelli che gli studenti avrebbero incontrato nel corso del primo anno di università (ovvero espositivi e argomentativi) e che non richiedessero come presupposto per la comprensione una conoscenza specifica e preliminare dell'argomento trattato. Sulla base del punteggio sull'intera prova ottenuto da ciascuno studente è stata calcolata la soglia di accettabilità al fine di individuare un sottocampione di studenti che presentavano maggiori difficoltà nella comprensione del testo>. [10]

### **Dal modello al prototipo**

Il percorso didattico nell'ambiente adattivo prevede la somministrazione in ingresso, in itinere e in uscita da ciascuna unità didattica di due differenti tipi *cloze-test*, i cui criteri di costruzione e

valutazione sono codificati in una *routine* affidata a un dispositivo automatico. Da queste frequenti interazioni, che svolgono la funzione didattica di esercitare e misurare la comprensione del testo per l'apprendimento, il prototipo ricava una stima della <competenza verbale> del destinatario reale. Si tratta di *cloze* mirati - le lacune non sono disposte a intervalli regolari, non vengono cancellate parole vuote o grammaticali, è possibile determinare la gamma di frequenza entro la quale devono ricadere le parole nascoste - e a risposta chiusa. La parola da inserire deve essere scelta all'interno di una lista che, nel caso del primo tipo di *cloze*, impiegato come pre-post test, contiene tutte e soltanto le 15 parole nascoste nel brano, mentre nel caso dei test intermedi contiene anche un numero equivalente di distrattori. Il *cloze-test* somministrato in ingresso e in uscita si basa su un testo di 600-700 parole scelto dal docente. Per generare i test intermedi è invece il sistema a selezionare cinque frammenti (periodi) togliendoli dalla versione indifferenziata dell'ultima porzione di testo che lo studente ha letto. Nella costruzione degli *item* l'intervento del docente può essere nullo, ma non incontra limiti: egli può manipolare la difficoltà della prova, può intervenire sul testo e ovviamente può emendare le scelte operate dal sistema, fino al punto di produrre tutti gli *item*. La variabile che è possibile manipolare per aumentare o diminuire la difficoltà della prova è la frequenza d'uso dei termini che vengono cancellati nel *cloze passage*. Le sperimentazioni condotte durante il primo anno del progetto per tarare lo strumento hanno riscontrato una correlazione quasi perfettamente inversa tra l'indice di frequenza delle parole nascoste e la difficoltà della prova: se diminuisce la frequenza, cresce la difficoltà. Di questo aspetto il dispositivo tiene conto anche nel pesare gli *item*.

Un secondo dispositivo raffronta la stima di competenza verbale con quella presupposta dal messaggio nella sua formulazione originale e su questa base genera una versione individualizzata della porzione di testo successiva, interpolando nel testo (subito dopo la prima occorrenza) le definizioni dei termini che sono presenti nel glossario del corso e rientrano nella stessa fascia di frequenza per la quale si è diagnosticata l'incompetenza del lettore. <Non si tratta, infatti, di operare una semplificazione linguistica (lessicale e sintattica) che comporta anche un'attenuazione della densità di concetti che qualificano l'istruzione, ma di stabilire intervalli compatibili tra il linguaggio utilizzato per la formulazione del messaggio e il repertorio linguistico già disponibile nel destinatario>. [2] Si provoca, inoltre, una esposizione graduale all'uso naturale del linguaggio settoriale evitando di glossare tutte le parole che si stima il lettore reale non possieda in un dato momento, ma una parte di esse lievemente meno numerosa dell'insieme: si tende al graduale restringimento verso il basso della fascia di frequenza interessata dalla modulazione. L'adattamento precede la lettura, che dovrebbe quindi procedere in modo continuo e lineare.

Il *cloze-test* somministrato in uscita costituisce uno sbarramento nel percorso didattico: se il risultato è sufficiente lo studente può procedere e completare il percorso dell'unità didattica con la prova di valutazione finale, diversamente, prima di proseguire, deve svolgere un percorso compensativo che viene valutato attraverso domande a risposta aperta. Anche l'analisi testuale di questa prestazione produttiva contribuisce a definire il profilo dinamico dello studente.

Un questionario composto da 15 domande con risposte a scelta multipla costruito dal docente viene somministrato in uscita da ciascuna unità didattica. Si tratta di una classica prova oggettiva destinata, in sede sperimentale, a fornire la misura della variabile dipendente, mentre serve nel normale contesto didattico alla valutazione sommativa.

In uscita dal corso sperimentale si chiede allo studente di compilare un questionario volto a rilevare dati socio-anagrafici, comportamenti di studio e informazioni sull'esperienza del corso.

### **Il tryout a Reggio Emilia**

Durante il secondo anno di attività del progetto ciascuna unità di ricerca ha condotto un *tryout* del prototipo con l'obiettivo di:

- 1) verificare la funzionalità del sistema;
- 2) avviare la messa a punto di soluzioni tecniche;
- 3) raccogliere informazioni utili per impostare il disegno sperimentale.

Nel corso del primo semestre dell'A.A. 2010-2011 si è svolto il primo collaudo, nell'ambito del Laboratorio di Docimologia 2 rivolto agli studenti del CdL in Scienze della Formazione Primaria dell'Università di Roma Tre. Il percorso didattico è stato completato da 35 studenti, che hanno partecipato nello stesso periodo al laboratorio in presenza, assegnati casualmente al gruppo sperimentale (che ha usufruito dei testi adattati) e al gruppo di controllo. Durante il secondo semestre si è svolto il collaudo presso la Facoltà di Medicina e Chirurgia de La Sapienza, con l'erogazione interamente *online* di due corsi, ciascuno dei quali ha coinvolto circa 140 studenti. In questo caso è stata messa in opera anche l'attività compensativa e la valutazione dei risultati attraverso domande a risposta aperta (SOQs). Per il collaudo a Reggio Emilia, che ha coinvolto un numero più contenuto di studenti, si è adottato un disegno pre-sperimentale a gruppo unico con pre-test e post-test. Il *tryout* è stato ospitato dal Corso di Laurea in Fisioterapia (Facoltà di Medicina e Chirurgia) all'interno del primo modulo del Corso integrato di Biomeccanica e Cinesiologia applicata, tenuto dal Prof. Adriano Ferrari. L'attività di ricerca si è svolta in stretta collaborazione con i referenti del progetto, il prof. Adriano Ferrari, Presidente del Consiglio del Corso, la dott.ssa Luisa Montanari, Coordinatore, la dott.ssa Vittoria Mamoli, Tutor.

### **Materiali**

- 1) Collezionando 40 testi in formato digitale si è alimentato il preesistente corpus del linguaggio medico con una partizione rappresentativa del linguaggio speciale delle discipline della riabilitazione. Questo specifico "archivio lessicale" costituisce il corpus di riferimento dal quale il sistema estrae il lessico di frequenza che è necessario per il funzionamento dei dispositivi adattivi.
- 2) Un glossario compilato dagli esperti della materia raccoglie definizioni sintetiche o brevi parafrasi dei termini specialistici utilizzati nel corso. Si compone di 43 lessie semplici e 72 lessie complesse, per un totale di 115 ingressi.
- 3) Il lavoro di preparazione del testo base del corso (la sua versione indifferenziata) in funzione dei criteri stabiliti nelle linee guida del progetto, si è applicato a materiali autentici: l'archetipo del testo è infatti costituito da cinque lezioni trascritte di Cinesiologia tenute dal prof. Adriano Ferrari nell'A.A. 2009-2010. La trascrizione di un ciclo di lezioni accademiche risente dello scritto piuttosto che del parlato comune. Dal punto di vista della struttura del testo e dell'articolazione dell'informazione nell'enunciato, si tratta di un testo espositivo che presenta: chiaro ordine logico; alta frequenza di definizioni, descrizioni, elencazioni; complessità lessicale; semplicità sintattica. Si ha invece l'impressione di riconoscere l'impronta dell'occasione originaria, se non è dello stile individuale o di scuola, in alcuni passaggi scanditi

da allocuzioni e soprattutto domande; si può supporre cioè che la conversazione didattica, non trascritta, sia effettivamente avvenuta in aula.

### Preparazione del testo

Fra le operazioni compiute sul testo rivestono una particolare importanza:

- 1) la marcatura dei passi invariabili, sui quali il dispositivo di adattamento non deve intervenire (per esempio: elencazioni dense di termini specialistici; spiegazioni di termini diffuse e articolate, dunque non sostituibili da una definizione sintetica);
- 2) l'eliminazione delle definizioni di termini già presenti nel glossario (il testo base deve farne un uso naturale, rivolgendosi al destinatario virtuale).

Misure lessicometriche	Formule	Testo base	Archivio lessicale
totale delle occorrenze	N	22506	436797
totale delle forme grafiche	V	4577	32193
estensione lessicale	$(V / N) * 100$	20,33	7,37
percentuale di hapax	$(V_1 / V) * 100$	51,03	40,65
frequenza media generale	N / V	4,91	13,57
coefficiente G (di Guiraud)	$V / \sqrt{N}$	30,51	48,71
coefficiente a	$\log N / \log V$	1,19	1,25

**Tabella 1** - Misure lessicometriche sul vocabolario del testo base e dell'archivio lessicale.

Il contenuto del corso, testi e immagini, è suddiviso in due Unità didattiche: UD 1, Postura; UD 2, Cammino. Manipolazione.

Si presentano qui alcune misure lessicometriche (Tab. 1) sul vocabolario del testo base e del corpus di riferimento. Per quanto riguarda gli indicatori della ricchezza lessicale, vale l'ovvia avvertenza che il confronto è possibile soltanto fra quelli meno influenzati dalla taglia (qui si propone il coefficiente a). Tanto il corpus di riferimento quanto il glossario non raggiungono la misura prevista per il funzionamento ottimale del dispositivo, il che non ne ostacola l'uso ai fini del collaudo, che riguarda principalmente l'usabilità del sistema.

### Destinatari e modalità di erogazione del corso

Il corso si è rivolto a tutti gli studenti iscritti al primo anno del CdL, ventinove. Il piano degli studi prevede, nel secondo semestre del primo anno, questo insegnamento fondamentale. Gli studenti hanno frequentato le normali lezioni in presenza parallelamente al corso a distanza. Il corso sperimentale a distanza, che gli studenti hanno ricevuto come risorsa aggiuntiva destinata allo studio autonomo, è stato reso obbligatorio. Il percorso didattico all'interno di ciascuna unità del corso risulta così articolato (Tab. 2).



Pre test	Test a completamento (15 lacune – 15 alternative). Brano di 600-700 parole.
Testo sezione 1	Adattato in base alla prima stima del patrimonio lessicale del lettore reale.
Test intermedio 1	Test a completamento (5 lacune – 10 alternative). 5 frammenti del Testo 1
Testo sezione 2 [...]	Adattato in base alla stima aggiornata.
Post test	Riproposizione del pre test. Con feedback.
Prova di valutazione	15 quesiti a scelta multipla sui contenuti dell'unità. 4 alternative, una corretta.

**Tabella 2** - Percorso didattico.

In base ai valori modali riscontrati nelle risposte ad alcune domande del questionario conclusivo è possibile disegnare questo sintetico profilo dello studente:

- 1) età regolare o quasi all'immatricolazione (meno di 21 anni);
- 2) proveniente dal liceo scientifico;
- 3) uno dei genitori è laureato;
- 4) è impegnato in attività connesse allo studio per più di otto ore al giorno.

### Esiti del collaudo

Nello stesso modo si può ricostruire un quadro dell'esperienza in piattaforma descritta dallo studente-tipo.

- 1) Ha intuito senza difficoltà l'uso della piattaforma.
- 2) Ha apprezzato l'autonomia nella gestione dei tempi di studio.
- 3) Ritene di aver appreso grazie a questo percorso il significato di un maggior numero di parole.
- 4) Non ha avuto difficoltà nella lettura dei testi della prima UD. Ha trovato più difficile la seconda. La diversa percezione di difficoltà è confermata dai risultati delle prove (*cloze* e MCQs).
- 5) Ha percepito i *cloze-test* come più semplici rispetto ai quesiti a scelta multipla. Percezione confermata dai punteggi ottenuti nei due tipi di prove.

Dati gli obiettivi, le condizioni e il disegno non sperimentale del collaudo, i risultati delle prove non possono essere considerati validi. I dati, qui esposti per completezza, sono ulteriormente inficiati dall'esame del tempo impiegato per lo svolgimento delle prove e del lasso di tempo trascorso tra una prova e la successiva.

UD 1			UD 2		
N = 28			N = 23		
Punt. max = 15			Punt. max = 15		
Pre test	Post test	MCQs	Pre test	Post test	MCQs
12,06	12,73	12,31	10,88	11,51	9,17
2,11	2,4	2,53	2,5	2,16	2,41
4,46	5,76	6,41	6,27	4,68	5,79
		0,66	corr. pre-post test	0,78	
		0,12	corr. post test-MCQs	0,32	

**Tabella 3** - Risultati delle prove.

## Discussione

Una prima considerazione suggerita dall'esperienza condotta riguarda il trattamento automatico dei dati del linguaggio. Al fine di annullare una possibile variabile di disturbo, in sede di collaudo si è provveduto alla marcatura manuale delle lessie complesse presenti nel testo base e nel glossario, anche se il dispositivo include un componente dedicato all'analisi del testo. L'efficienza della procedura di classificazione automatica della terminologia riveste una grande importanza per il funzionamento del sistema, sia in termini di efficacia e sia di economicità. L'identificazione di forme complesse è un aspetto del problema, non indipendente da quello della disambiguazione di forme semplici. La soluzione può fondarsi su misure puramente lessicometriche (peculiarità della frequenza d'uso su base comparativa con l'uso generale; misure di tendenza all'associazione fra parole all'interno del corpus), oppure, come proposto da recenti ricerche, prevedere anche un pretrattamento di marcatura grammaticale del corpus e la lessicalizzazione di forme complesse che superino una determinata soglia di occorrenze, descritte con l'uso affinato di *regular expression* (es. <nome+preposizione+nome>). <Il riconoscimento automatico di tali espressioni favorisce la disambiguazione delle forme semplici che le compongono, anche se non si tratta di polirematiche o collocazioni. Il senso delle parole può essere infatti determinato dalle forme che le circondano. Pertanto vincolando i significanti al loro contesto, mediante il riconoscimento e la lessicalizzazione dei sintagmi nominali, si chiarisce l'uso effettivo dei singoli sostantivi nel testo> (P. Pavone, 2010). [11] La lista dei poliformi deve essere allineata anche con il glossario del corso, che a sua volta può costituire una risorsa per l'individuazione dei poliformi: le lessie semplici presenti nel glossario possono essere utilizzate per la ricerca di segmenti da lessicalizzare, contribuendo, per altra via, alla disambiguazione dei dati testuali. Inoltre, è importante che il glossario contenga di un singolo termine tutte le varianti occorrenti, come ad esempio l'acronimo e la traduzione (è il caso di: <COP>, <center of pressure>, <centro di pressione>). Peraltro, dal momento che il prototipo si basa su misure lessicometriche per la costruzione automatica delle prove diagnostiche e per la modulazione del testo, la procedura di classificazione automatica della terminologia è importante quanto la disponibilità di un corpus rappresentativo del linguaggio settoriale da cui estrarre il lemmario di frequenza. Il compito di collezionare testi pertinenti potrebbe essere in parte automatizzato attraverso procedure di *information retrieval* applicate a fonti accessibili in rete (es. banche dati bibliografiche).

Una seconda riflessione riguarda l'apparato valutativo. Nella sua lunga storia e la larga diffusione la *cloze procedure* ha conosciuto almeno tre principali applicazioni - prova di leggibilità dei testi, prova di comprensione della lettura, prova di conoscenza di una lingua o di suoi aspetti - e ha prodotto una fitta schiera di varianti ed un'ampia letteratura, della quale non è possibile dar conto qui, salvo segnalare che generalmente se ne apprezza la capacità di sollecitare un complesso di abilità che rappresentano globalmente le capacità linguistiche e di comprensione, soprattutto a livello infrafrastico. Nel nostro caso, della procedura si intende fare un uso mirato soprattutto alla competenza lessicale e, trattando di testi specialistici, si è in presenza di una forte componente di termini tecnici e sub-tecnici. La conoscenza della terminologia di una disciplina è legata in modo stretto alla conoscenza dei contenuti, per cui è da attendersi che questo tipo di test stimoli l'accesso alle conoscenze acquisite e quindi la memoria oltre alle competenze specifiche del processo di comprensione della lettura. L'incidenza di lacune dove la parola nascosta appartiene allo strato apicale di un lessico speciale (parole rare e di uso "iper-tecnico") risulta già opportunamente limitata dalla scelta di una fascia di frequenza sufficientemente ampia per le parole da nascondere; l'uso, combinato

all'indice di frequenza, anche indici di rilevanza delle parole nascoste, per costruire e pesare gli *item*, può scongiurare il problema. Ciò che rimane importante è limitare al massimo l'influenza di altri aspetti meno pertinenti, evitando risposte sulla base della concordanza grammaticale o di scorciatoie di pensiero che non mobilitano la competenza lessicale, nè la conoscenza pregressa. Trattandosi di *item* a risposta chiusa, la produzione automatica di distrattori plausibili riveste una importanza cruciale. L'annotazione grammaticale del testo potrebbe consentire di ridurre l'incidenza dei vincoli morfosintattici sul comportamento di risposta.

È inoltre possibile implementare l'apparato valutativo con altri strumenti automatizzati, che abbiano caratteristiche differenti dal *cloze* e siano comunque adatti a svolgere sia una diretta funzione didattica all'interno del corso, sia una funzione strumentale al dispositivo adattivo. Il *cloze* consiste nell'inserire (riconoscere) parole mancanti in un testo. Attraverso prove di definizione che riguardino lemmi presenti nel glossario del corso, si può invece stimolare il riconoscimento del significato di parole in contesto e fuori contesto. In questo secondo caso, il compito può consistere nell'abbinamento fra voci del glossario e definizioni (con o senza distrattori). Nel primo caso, si pensa ad una prova in cui si proponga la lettura di un brano e quindi si richieda di scegliere all'interno di una lista la corretta definizione di un termine lì utilizzato. Sostituendo la parola da definire con una parola inesistente è anche possibile testare la capacità di inferire correttamente dal contesto il significato di una parola sconosciuta, capacità che non viene specificamente testata dal *cloze* ma è di sicura rilevanza ai fini della comprensione.

### Conclusioni

È possibile sperimentare un sistema in larga parte automatico di supporto allo studente che intervenga su questo aspetto, proprio perchè la competenza lessicale è relativamente esplicita e quindi più adatta ai vincoli e alle potenzialità propri di un trattamento automatizzato. Peraltro, questa proposta apre la strada a ulteriori sviluppi. <It is possible to construct cloze tasks that comprise a coherent text and where the options include sentence-appropriate and story-appropriate replacement word options, to measure the ability to process text beyond the sentence level>. [12] La stessa tecnica *cloze* può essere impiegata, ad esempio, per misurare o esercitare la capacità di cogliere i nessi testuali e quindi intervenire su altre difficoltà di comprensione.

### Bibliografia

- [1] B. Vertecchi, *Nuove ipotesi per lo sviluppo della didattica in linea*, Journal of e-Learning and Knowledge Society, Vol. 6, No. 1, 2010, pp. 31 - 42.
- [2] Progetto di ricerca “am-Learning (adaptive message learning)”.
- [3] B. Vertecchi, *Premessa*, in A. Poce, C. Angelini, *Per una nuova cultura didattica. Riflessioni sul progetto am-learning*, Milano, FrancoAngeli, 2011.
- [4] B. Vertecchi, *Una nuova frontiera per l'istruzione a distanza: l'individualizzazione del messaggio di istruzione*, Istruzione a Distanza, Vol. 10, No. 10 - 11, 1998, pp. 83 - 115.
- [5] B. Vertecchi, A. Poce, C. Angelini, F. Agrusti, *Orbis dictus. Un ambiente adattivo multilingue per l'istruzione in rete / Orbis dictus. A Self-Adaptive Environment for Multi-language Teaching and Learning Opportunities*, Milano, FrancoAngeli, 2010.

- [6] M. Amizzoni, N. Mastidoro, *Linguistica applicata alla leggibilità: considerazioni teoriche e applicazioni*, Bollettino della Società Filosofica Italiana, No. 149, maggio - agosto 1993, pp. 49 - 63.
- [7] C. Cornoldi, R. De Beni, P. Palladino, F. Pazzaglia, *Lettori che non capiscono*, in G. Di Stefano, R. Vianello (a cura di), *Psicologia dello sviluppo e problemi educativi. Studi e ricerche in onore di Guido Petter*, Firenze, Giunti, 2001.
- [8] T. De Mauro, *La fabbrica delle parole. Il lessico e problemi di lessicologia*, Novara, UTET De Agostini, 2005.
- [9] P. E. Tressoldi, C. Zamperlin, *La valutazione della comprensione del testo: proposta di una batteria di approfondimento*, Psicologia clinica dello Sviluppo, No. 2, agosto 2007, pp. 271 - 290.
- [10] E. Truffelli, *Comprendere per migliorare lo studio: analisi e riflessioni a partire da un'esperienza biennale di sostegno alla matricole universitarie*, Giornale italiano della ricerca educativa, giugno 2010, pp. 91 - 104: p. 97.
- [11] P. Pavone, *Sintagmazione del testo: una scelta per disambiguare la terminologia e ridurre le variabili di un'analisi del contenuto di un corpus*, in JADT 2010: 10° International Conference on Statistical Analysis of Textual Data.
- [12] K. Cain, J. Oakhill, *Assessment matters: Issues in the measurement of reading comprehension*, British Journal of Educational Psychology (2006), No. 76, 697 - 708.