

Analytics dei testi riflessivi scritti dai docenti neoassunti nel portfolio digitale

Newly Qualified Teachers' reflective writing: analytics and preliminary results

Valentina Della Gala

Istituto Nazionale Documentazione, Innovazione, Ricerca Educativa (INDIRE), v.dellagala@indire.it

Giulia Chiriatti

Università di Pisa, giuliachiriatti@gmail.com

Felice Dell'Orletta

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) - ItaliaNLP Lab, felice.dellorletta@ilc.cnr.it

Maria Chiara Pettenati

Istituto Nazionale Documentazione, Innovazione, Ricerca Educativa (INDIRE), mc.pettenati@indire.it

Giulia Venturi

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) - ItaliaNLP Lab, giulia.centuri@ilc.cnr.it

187

In this paper, we present the preliminary results of the analysis performed on 50,000 texts written by the Newly Qualified Teachers to reflect on their practice. The Italian Induction Programme for NQT includes online activities to be carried out in the E-portfolio managed by Indire. As part of the training monitoring with the aim to evaluate our support and to improve instructional scaffolding to teachers, we analysed teachers' reflective writing mixing a new developed framework for teachers' reflective writing and Natural Language Processing (NLP) techniques. In providing a case study of a large-scale analysis of textual teachers' data in an institutional setting, this paper wants to contribute to the field of learning analytics. We will describe the context, the characteristics of the data analysed, how all of this has shaped the analytics design. We will describe its implementation and therefore the procedures, tools and metrics developed to represent the content of the teachers' reflective writing, and we will try to evaluate if and how these preliminary results can help us to reach the research goals and to understand the strengths and weaknesses of different methodologies.

Keywords: Teacher professional development; reflective writing; content analytics; learning content.


Presentiamo i risultati preliminari e l'analisi svolta su circa 50.000 testi scritti dai docenti neo nominati in ruolo per riflettere su due attività didattiche svolte con gli studenti, nel contesto del percorso dell'anno di formazione e prova 2016/17. Il percorso prevede attività in presenza e attività a distanza completate sul portfolio digitale, ospitato nell'ambiente online gestito dall'Indire. Nell'ambito del monitoraggio della formazione, con il fine di ottimizzare gli strumenti e il supporto fornito, abbiamo interrogato i dati testuali prodotti dai docenti nell'interazione con l'ambiente per capire se i testi presentassero evidenze riconducibili alle scritture riflessive. Obiettivi dell'indagine sono stati la definizione di uno schema per la classificazione dei testi sulla base del livello di riflessività evidenziato e l'im-

Valentina Della Gala è autrice dei §§ 1, 3, 4, 5, 6, 6.1 e 7; Giulia Chiriatti, Felice Dell'Orletta e Giulia Venturi dei §§ 5.1 e 6.2; Maria Chiara Pettenati del § 2.

piego di strumenti di Trattamento Automatico del Linguaggio (TAL) per l'analisi dell'intero corpus testuale prodotto dai docenti. Descriveremo il contesto scientifico e progettuale, le caratteristiche dei dati analizzati, come questo abbia determinato il disegno d'indagine; descriveremo inoltre la sua implementazione e dunque le procedure, gli strumenti e le metriche adottate o elaborate per rappresentare il contenuto dei dati; infine discuteremo i primi risultati e alcuni vantaggi e limiti dell'approccio adottato.

Parole chiave: formazione professionale dei docenti; scrittura riflessiva; analytics del contenuto; contenuto di apprendimento.

1. Introduzione



Dalla sua istituzione sperimentale nell'a.s. 2014/15, il percorso di formazione e prova ha coinvolto ad oggi circa 194.000 docenti neo nominati in ruolo, la cui interazione con l'ambiente *online* ha prodotto un'enorme quantità di dati di tracciamento delle azioni e di contenuto elaborato dai docenti, oltre al contenuto dell'ambiente stesso, che è variato, pur di poco, di edizione in edizione. La sovrabbondanza di dati è un fenomeno ampiamente discusso da anni in molti campi di ricerca. Nell'ambito educativo l'analisi di grandi quantità di dati è stata riconosciuta come una strada per nuovi modelli di scoperta (Siemens, 2013), con importanti implicazioni ad ogni livello, da quello istituzionale a quello del singolo studente. L'analisi del *learning content*, cioè dei dati, principalmente testuali ma in potenza anche in formato video o audio, prodotti dagli studenti (saggi, discussioni ecc.), dagli educatori (sillabi, documenti ecc.) o dagli editori (libri di testo ecc.), è una sotto area dei LA definita *content analytics* (CA), ad indicare i metodi automatici per esaminare, valutare, indicizzare, filtrare, raccomandare e visualizzare «different forms of digital learning content, regardless of its producer (e.g., instructor, student) with the goal of understanding learning activities and improving educational practice and research» (Kovanović et al., 2017). Con l'intenzione di ottenere informazioni sull'uso degli strumenti a supporto dei docenti in anno di formazione e prova e fornire evidenze per ottimizzarli e migliorare il supporto alla loro esperienza formativa, abbiamo analizzato il contenuto dei testi scritti da circa 25.000 docenti per riflettere su 2 attività didattiche svolte con gli allievi. Nei paragrafi che seguono presentiamo gli obiettivi dell'anno di formazione e prova, il disegno di indagine, le analisi compiute, gli strumenti impiegati e i risultati preliminari ottenuti.

2. Contesto: il percorso, l'attività oggetto di analisi e il ruolo della riflessione nel modello formativo dei neoassunti

Il sostegno agli insegnanti durante la fase iniziale della carriera o *induction*, è un elemento di indirizzo strategico per tutti i paesi, evidenziato anche nella Comunicazione della Commissione europea sullo sviluppo scolastico e l'eccellenza nell'insegnamento (Commissione Europea, 2018). L'Italia, come la grande maggioranza dei sistemi europei, mette in atto ogni anno un programma nazionale obbligatorio della durata di un anno per l'accompagnamento all'ingresso in servizio dei nuovi docenti. Fino al 2013/14 il modello italiano di *induction* prevedeva un impianto centrato su incontri formativi in presenza e un'attività online dedicata alla fruizione di contenuti in auto-formazione e alla redazione di un documento di sintesi finale, dal 2014/15 il modello è stato revisionato e consolidato nel suo nuovo formato con il DM 850/15. Recependo i risultati della ricerca educativa nazionale ed internazionale sulla formazione dei docenti, con particolare attenzione all'avvio alla professione, il nuovo modello poggia su alcuni principi fondanti (D'Amico, 2018), che possono essere così riassunti:

- *governance* distribuita tra autorità centrale (MIUR, che stanziamenti, definisce l'impianto, le priorità, attribuisce ruoli, funzioni e tempi), gli Uffici Scolastici Regionali (che recepiscono le indicazioni nazionali e le contestualizzano, attuando un governo territoriale della formazione, col supporto degli Uffici di Ambito Territoriale), le Scuole Polo (che implementano l'offerta formativa in presenza prevista dal modello di formazione secondo le priorità tematiche stabilite a livello nazionale e i bisogni formativi locali), la scuola di presa di servizio dei nuovi docenti (che accoglie e guida l'inserimento del docente nella comunità educante), Indire (Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa) che implementa e gestisce l'LMS a supporto all'anno di formazione e prova.
- accompagnamento guidato nella scuola di presa di servizio con il Dirigente Scolastico chiamato a stipulare un Patto Formativo col nuovo docente, osservarne l'attività didattica e nominare un Tutor accogliente, che svolge un'azione di affiancamento tra pari (*peer to peer*) che comprende una riflessione sulle competenze in ingresso, una co-progettazione dell'attività didattica, l'osservazione reciproca in classe.
- alternanza tra immersione e distanziamento dalla pratica didattica guidata da una serie di "dispositivi" che confluiscono nel portfolio formativo digitale (Rossi et al., 2015) implementato da Indire.



Uno dei dispositivi implementato nel portfolio formativo è quello relativo alle “Attività Didattiche” (le attività sono 2 identiche), oggetto di approfondimento nel presente studio. Attraverso questo dispositivo si dà supporto alla progettazione-documentazione-riflessione guidati dell’azione didattica del docente nel nuovo contesto educativo (Rossi et al., 2015), al fine di supportarne l’acquisizione di un atteggiamento di riflessività che – adottato sistematicamente come *habitus* professionale – favorisca la crescita professionale dell’insegnante in un’ottica di formazione continua, oltre l’anno di prova (Rossi et al., 2018).

3. Il corpus documentale analizzato

Il dispositivo tramite il quale al docente è stato richiesto di documentare 2 attività didattiche, si componeva nell’a.s. 2016/17 di 5 sezioni, (progettazione; riflessione sulla progettazione; documentazione; narrazione; riflessione post-azione). Analizziamo i testi scritti nella sezione sulla riflessione post-azione in cui era richiesto di riflettere sull’attività didattica svolta rispondendo, se lo si riteneva opportuno, ad 8 domande stimolo. La scelta di fornire un solo *form* per rispondere a un intero gruppo di domande è dovuta alla volontà di lasciare il docente libero di scrivere potendo scegliere a quali domande rispondere, in quale ordine o se avvalersi o meno delle domande. Il risultato sul piano della struttura dei dati è stato un data-set in cui ad ogni codice identificativo del docente corrispondono il numero dell’attività (1 e 2) e l’intero testo composto nella sezione.



Fig. 1: Screenshot della sezione del Portfolio digitale dedicata alla scrittura di un testo riflessivo post-azione

4. Analisi del contenuto e disegno dell'indagine

L'articolazione dell'indagine origina dalla *content analysis* (Krippendorff, 2004), di un campione di circa 200 testi che ci ha permesso di individuare alcuni tratti ricorrenti. I testi del campione sono risultati:

- composti da un numero di caratteri molto inferiore ai 6000 (limite massimo consentito);
- contenere risposte alle domande tendenzialmente disposte in modo ordinato;
- non contenere sempre le risposte a tutte le domande;
- presentare livelli diversi di complessità sintattica;
- presentare livelli diversi di ricchezza lessicale.

La complessità sintattica e l'aderenza alla realtà dell'esperienza si sono rivelati criteri discriminanti per distinguere tra segmenti di testo (o risposte) in cui vi si può riconoscere un ripensamento autentico dell'esperienza e segmenti non coerenti con le aspettative. Abbiamo dunque ipotizzato che questi, insieme ad altri attributi relativi al coinvolgimento cognitivo (Garrison, 2001) ed emotivo (Dewey, 1933; Chung, 2007), potessero fornirci indicazioni automatizzabili per indagare tutti i testi.

Una scelta critica in questi casi è quella dell'unità da considerare per l'analisi. Noi abbiamo scelto di ricorrere alla suddivisione, in prevalenza già presente nel *corpus*, in testi di risposta alle 8 domande stimolo. Scelta che ci consentiva anche di valutare singolarmente l'efficacia di ogni domanda. Abbiamo, quindi, strutturato l'indagine in 2 macro azioni per indagare:

1. cosa i docenti hanno scritto, cioè a quale sollecitazione/*topic* avanzata dalle domande stimolo hanno risposto.
2. come lo hanno scritto, i testi si risolvono in compilazioni burocratiche o mostrano evidenze di un ripensamento circostanziato dell'esperienza didattica e delle caratteristiche condivise in letteratura come proprie di un approccio riflessivo? Quali indicazioni forniscono sull'adeguatezza degli strumenti? Alcuni hanno funzionato meglio?

5. Segmentazione dei testi per argomento: individuazione delle unità di analisi

La prima azione di questa fase è stata l'annotazione manuale di un campione di testi per risposta. Cioè ad ogni segmento abbiamo attribuito



il codice della domanda cui risponde (Tab. 1). Il campione così annotato è andato a costituire il *corpus* per l'addestramento (*training corpus*) degli algoritmi di classificazione.



Codice	DOMANDE	Codice rivisto
12.1	L'attività si è sviluppata esattamente come avevi previsto? In caso contrario, cosa è stato modificato durante la sua realizzazione?	12.1.2
12.2	Quali sono state le cause delle variazioni? (es: eventi imprevisti, stimoli costruttivi emersi durante la lezione, domande degli allievi, una non adeguata previsione dei tempi, una non adeguata valutazione delle difficoltà che avrebbero incontrato gli allievi, ecc.)	
12.3	Quali credi siano state le tue scelte più efficaci?	12.3.4
12.4	Quali mediatori (consegne, organizzazione del lavoro, natura dei materiali, supporti digitali, ecc.) hanno funzionato meglio? E perché?	
12.5	Il modo in cui hai valutato l'andamento del lavoro ti ha fornito informazioni adeguate sugli apprendimenti? E ti ha aiutato a direzionare il percorso?	12.5
12.6	L'attività proposta potrebbe essere migliorata? Se sì attraverso quali interventi? (es. intervenendo su: i. progettazione iniziale, ii. rispondenza dell'attività alle linee guida/indicazioni nazionali, iii. sostenibilità, iv. bisogni e livelli di partenza della classe/sezione o dell'allievo)	12.6
12.7	Le osservazioni del tutor hanno contribuito a considerare sotto una luce diversa la tua azione?	12.7
12.8	Pensi che il modo in cui hai documentato possa influire sulla tua futura modalità di agire in classe/nella sezione?	12.8

Tab. 1: Elenco delle domande stimolo con il codice attribuito per l'analisi e quello rivisto dopo i primi cicli di annotazione

5.1 Classificazione con strumenti di TAL

Allo scopo di riconoscere all'interno di ogni testo i singoli segmenti di risposta scritti dai docenti alle domande stimolo, sono stati adottati algoritmi di classificazione automatica supervisionata (*machine learning*) basati su strumenti di TAL. Tali algoritmi sono capaci di associare ad ogni testo in analisi la classe di appartenenza (nel nostro caso la singola domanda: 12.1, 12.2 ecc.), assegnando al testo un punteggio di probabilità di appartenenza ad ogni classe. La classe con punteggio massimo è quella che l'algoritmo associa al segmento di testo. Il classificatore è capace di apprendere la strategia di classificazione "imparando" dal *training corpus*. Il classificatore è capace nella fase di addestramento di apprendere le caratteristiche specifiche dei testi appartenenti ad una

classe e di associare ad ognuna di esse un peso, che rappresenta il contributo della specifica caratteristica nel predire la classe di appartenenza del testo. Terminata la fase di addestramento, in fase di analisi il classificatore è capace di estrarre da un nuovo testo preso in esame (quindi sconosciuto al classificatore fino a quel momento) le varie caratteristiche che, associate ai pesi definiti in fase di addestramento, determinano la classe di appartenenza del testo. Pertanto, i passaggi dell'intero processo di classificazione sono stati: *i*) la scelta della dimensione del *training corpus*; *ii*) la selezione delle caratteristiche per classificare in modo adeguato i segmenti di testo.

In letteratura, è ben noto come una maggiore quantità di testi contenuti nel *training corpus* permette di raggiungere accuratezze di analisi maggiori. Questo è dovuto al fatto che più esempi annotati a mano costituiscono una base di apprendimento più ampia. Dal momento però che la costruzione del *training corpus* è un'operazione manuale, essa richiede un notevole investimento di tempo e di esperti del dominio. Per questo abbiamo deciso di procedere per gradi, incrementando progressivamente la quantità di segmenti di testo annotati manualmente. In questo processo, l'effetto di ogni nuovo incremento è stato verificato ripetendo la classificazione automatica e misurando le nuove accuratezze di volta in volta raggiunte. Siamo partiti da un numero iniziale di 762 frasi e siamo arrivati a un totale di 3.322 frasi. Da notare come oggetto delle nostre analisi siano state le singole frasi e non l'intero segmento. La motivazione di questa scelta è legata al fatto che i docenti non hanno sempre risposto in maniera sequenziale alle domande stimolo. La classificazione della singola frase ha pertanto permesso di superare questo problema, andando a rintracciare le diverse frasi che, a volte sparse nel testo, rispondevano alle domande poste. Un ulteriore aspetto di organizzazione interna del *training corpus* ha riguardato il numero di domande rispetto a cui classificare il *corpus*. Le 8 domande stimolo richiedevano risposte diverse. Tuttavia, i primi esperimenti di classificazione automatica hanno rivelato che la distinzione tra le risposte alle domande 12.1 e 12.2, e 12.3 e 12.4 era difficilmente individuabile. In entrambe le coppie, il contenuto della risposta richiesta dalla domanda è molto simile, poiché le due domande hanno uno stretto legame logico tra di loro. La mancanza di caratteristiche chiaramente distintive portava ad accuratezze di classificazione molto basse. È stato pertanto scelto di accorpare le domande, riducendo il numero delle classi a 6 (Tab. 1, "codice rivisto").

La selezione di quali siano le caratteristiche che contribuiscono a classificare le frasi contenute nei testi dei docenti è stato un secondo aspetto considerato. Le caratteristiche scelte includono elementi lin-



guistici estratti sia dalle risposte sia dalle domande stimolo, oltre ad elementi di contiguità testuale. Inizialmente, l'intero *corpus* è stato analizzato impiegando la catena di strumenti di annotazione linguistica automatica sviluppata dai ricercatori del gruppo ILC-CNR¹. Ogni risposta e ogni domanda è stata segmentata in singole frasi e all'interno di ogni frase sono state individuate le unità ortografiche che la compongono (*token*). Successivamente è stata fatta l'analisi morfo-sintattica e la lemmatizzazione del testo *tokenizzato*, infine l'analisi della struttura sintattica della frase in termini di relazioni di dipendenza sintattica. Una serie di esperimenti preliminari hanno permesso di individuare l'insieme di caratteristiche che hanno il maggiore potere discriminatorio. Le caratteristiche estratte dalle risposte fornite dai docenti includono: *i*) sequenze di unigrammi e bigrammi di caratteri, *token* e lemmi e *ii*) la rappresentazione vettoriale e compatta della semantica di ogni parola contenuta nelle risposte utilizzando informazioni sul contesto linguistico in cui è essa inserita. Per estrarre dal testo questa seconda caratteristica sono stati adottati metodi allo stato dell'arte noti come metodi di estrazione di *word embeddings* (Mikolov, 2013). In aggiunta, è stato considerato in fase di classificazione il lessico contenuto nelle domande stimolo fornite. L'intuizione è stata che alcune delle parole impiegate nelle domande fossero state utilizzate anche nelle risposte e che potessero dunque contribuire a classificarne il contenuto. Infine, gli esperimenti preliminari condotti hanno messo in luce che l'informazione sulla classe attribuita alla frase precedente o successiva a quella in analisi, in casi di dubbi sull'attribuzione, permette di migliorare l'accuratezza della classificazione. L'idea è che se il contenuto di una frase verte su di un determinato argomento di riflessione è molto probabile che la frase successiva tratti dello stesso argomento e condivida con essa il livello di riflessione. Al termine dell'intero processo, sono stati ottenuti i risultati di classificazione riportati nella Tab. 2, secondo i quali l'accuratezza complessiva della classificazione è del 70%, con estremi nelle risposte alle prime domande, che sono risultate anche le più lunghe e articolate, e alla domanda 12.7, sola domanda ad avere per oggetto il tutor, e per questo facilmente individuabile². Al contrario la

1 <http://www.italianlp.it/demo/linguistic-annotation-tool/>

2 In questo contributo l'accuratezza delle due tipologie di esperimenti condotti è stata calcolata come l'F-score, la media armonica tra precisione (calcolata come il numero di veri positivi diviso il numero di tutti i risultati positivi) e recupero (calcolato come il numero di veri positivi diviso il numero di tutti i test che sarebbero dovuti risultare positivi).

domanda 12.6 è quella classificata con meno accuratezza perché presenta elementi meno caratterizzanti. Inoltre i dati confermano il buon esito della scelta di accorpare le domande 12.1 e 12.2, la cui attribuzione è notevolmente migliorata.

Sotto-risposta	Accuratezza
12.1.2	80,9
12.3.4	70,3
12.5	61,4
12.6	58,1
12.7	83,5
12.8	65,8
TOTALE	70

*Tab. 2: Risultati della fase finale di classificazione automatica
Distribuzioni delle attribuzioni corrette per la classe della domanda.*



6. Analisi dei testi per livello di riflessione

A partire dall'analisi delle risposte e dagli studi sulle caratteristiche e il ruolo del pensiero riflessivo nell'ambito della formazione docente, abbiamo identificato 4 categorie in cui classificare i testi; selezionato gli attribuiti riconosciuti come specifici di uno scritto riflessivo ed elaborato una matrice che incrocia categorie e attributi per validare l'affidabilità della classificazione attraverso una codifica da parte di più valutatori. Il *corpus* codificato manualmente è stato poi analizzato con strumenti di TAL.

6.1 Cosa s'intende per riflessione e perché è stata categorizzata

«There is no such thing as the definition of reflection, but there is an on-going discourse about what constitutes reflection» (Ullman, 2015). Nell'ambito della formazione e della professione docente, riflettere ha un'accezione specifica e l'educazione alla riflessività un ruolo sostanziale. La letteratura sull'argomento è talmente ampia da coincidere quasi con l'intero ambito scientifico della formazione degli insegnanti. Tutti gli studi consultati fanno risalire a Dewey la prima esplicita e generatrice associazione tra pensiero riflessivo ed educazione. Dewey distingue il processo mentale riflessivo dagli altri (credenze, fantasticherie, flusso di coscienza), identificandolo in un processo del pensiero che a partire da una situazione percepita come ambigua, problematica, non nota,

indaga la realtà e la relaziona ad esperienze pregresse e ulteriori termini, al fine di eliminarne l'ambiguità e sciogliere la situazione problematica con l'azione e l'acquisizione di una nuova conoscenza (Dewey, 1933). Il pensiero riflessivo prende le mosse dall'esperienza, quindi, e si configura come un processo di costruzione di significati, attraverso la creazione di relazioni tra le dimensioni di un'esperienza, tra esperienze diverse e tra esperienze e conoscenze. Il fine di questo processo è trovare una soluzione a un problema/ambiguità e/o aumentare la propria base di conoscenze (Rodgers, 2012). La postura riflessiva è propria dei professionisti esperti. Una pratica in cui essa agisce è foriera di conoscenza professionale e risolve quello che Schon chiama il «dilemma tra rigore e pertinenza» (Schön, 2006), cioè consente l'integrazione tra ricerca scientifica e pratica didattica. Quando il processo riflessivo, ponendo in discussione i presupposti del problema, guida quello decisionale o l'azione, allora si configura come apprendimento dall'esperienza (Mezirow, 1990). Il ruolo della riflessione è stato analizzato in contrasto con gli obiettivi tecnico-procedurali delle formazioni docenti come progettate fino agli anni '80 negli Stati Uniti, in questo senso la "scoperta" della riflessività ha messo al centro dell'attenzione degli educatori il ruolo delle capacità cognitive, affettive, comunicative e critiche/emancipatorie del docente e dello studente nel processo di insegnamento/apprendimento. Sono stati dimostrati i benefici di una postura riflessiva (Boud, Walker 1991), e descritti metodi e strumenti per sostenerne lo sviluppo (p.e. *reflective conversation, storytelling, micro-teaching, action research* ecc.). Tra questi, la scrittura è il *medium* più ampiamente sperimentato. Come conseguenza negli ultimi 25 anni sono stati proposti molti schemi per valutare le "scritture riflessive". Alcuni hanno avuto l'obiettivo di valutare la profondità della riflessione; altri l'ampiezza, cioè i temi affrontati (Moon, 2006). L'approccio e i criteri di valutazione utilizzati sono tuttavia molto dipendenti dagli obiettivi dell'attività entro cui è analizzata, da cosa s'intende per riflessività in quel determinato contesto, dal contesto stesso e dagli input forniti. Al contrario di gran parte degli studi sull'argomento, nel nostro caso la redazione dei testi è avvenuta in un contesto "debole", di cui non è dato conoscere completamente il tipo di supporto e le indicazioni fornite al singolo docente, perché distribuite tra soggetti diversi; inoltre ha riguardato un numero altissimo di soggetti. Da questo punto di vista le sfide che ha presentato l'analisi sono state più vicine a quelle che affrontate nell'ambito dei *reflective writing analytics* (Gibson, Kitto 2015; Buckingham Shum et al., 2016; Buckingham Shum et al., 2017; Liu et al., 2019), che a quelle degli studi sulla valutazione della riflessività nella formazione docente, che generalmente coinvolge numeri bassi.

Da quest'ultimo ambito abbiamo però derivato le dimensioni che compongono il concetto di riflessività, mentre dal primo la legittimità per estendere l'esame all'intero *corpus* con metodi automatici.

Abbiamo assunto la complessità sintattica e l'aderenza alla realtà come criteri distintivi (Par. 4), e individuato le seguenti 4 classi per discriminare tra livelli di impegno/coinvolgimento diversi, e di cui solo in una, nella classe "riflessiva", c'è una compresenza delle caratteristiche considerate in letteratura come proprie di uno scritto riflessivo.

1. Non riflessiva: risponde con una semplice parafrasi della domanda e/o una descrizione dell'accaduto. Non viene fatto alcun riferimento esplicito o implicito a dubbi o problemi; non presenta evidenze di un approccio di ricerca. Assenza (o scarsa presenza e non significativa) di preposizioni subordinate.
2. Considerazione generale: il segmento mostra un legame debole con la realtà dell'esperienza didattica. I riferimenti ad essa, se ci sono, sono deboli. Si tratta di considerazioni generali o generiche che possono essere anche articolate in enunciati complessi.
3. Riflessiva descrittiva: presenta considerazioni direttamente connesse alla realtà dell'azione didattica; è possibile l'esplicitazione di conoscenze ed esperienze pregresse ma non emerge con chiarezza alcun problema o dubbio o questo non è portato alle sue conseguenze ultime, cioè non è individuata alcuna soluzione. L'enunciato è articolato.
4. Riflessiva: la questione affrontata è considerata problematica; sono riferiti dubbi relativi all'azione didattica; c'è evidenza di un cambiamento nell'attitudine del docente o che egli ha acquisito nuove conoscenze o a modificato la sua azione didattica. L'enunciato è articolato.

Abbiamo valutato l'affidabilità dello schema attraverso una procedura di validazione intersoggettiva. A questo fine abbiamo definito gli attributi di uno scritto riflessivo entro il nostro contesto di riferimento e sulla base della letteratura. Per ognuno abbiamo anche indicato come si può manifestare in forma scritta. Un segmento di testo è classificato come riflessivo se:

- I. contiene riferimenti espliciti all'esperienza didattica svolta. Presenza di termini concreti, ricchezza lessicale.
- II. include almeno un riferimento a: esperienze precedenti; conoscenze pedagogiche, di contenuto o di altro tipo con il fine di sostenere l'interpretazione e l'analisi dell'esperienza da parte del docente. Il risultato è una struttura sintattica articolata con presenza di verbi al gerundio, di avverbi e di preposizioni.



- III. discute un problema, un dilemma, un dubbio; reagisce allo stimolo.
- IV. genera un *output* (un risultato), che può essere la ri-modulazione dell'intervento didattico nel corso dell'azione, nel qual caso il testo si configura come la narrazione di una "riflessione in azione". Oppure può essere la previsione di possibili azioni didattiche modificate dalla conoscenza acquisita (in questo caso si ha una "riflessioni sull'azione"); la soluzione ad un dilemma ecc.
- V. presenta di verbi mentali.
- VI. presenta pronomi in prima persona.

Abbiamo elaborato una matrice per indicare la presenza o assenza nei segmenti di testo degli attributi. Come mostrato nella Tab. 3, combinazioni definite di presenza o assenza delle variabili indicano l'appartenenza a una delle 4 categorie di "riflessività" elaborate.



Livelli di riflessività/ attributi	Pronomi personali	Presenza di riferimenti	Esperienza	Termini astratti	Problema	Output	Verbi mentali
Non riflessiva	0, 1	0	0,1	0,1	0	0	0, 1
Considerazione generale	0,1	1	0	1	0,1	0,1	0, 1
Riflessiva descrittiva	0,1	1	1	0,1	0,1	1,0	0, 1
Riflessiva	1	1	1	0,1	1	1	1

Tab. 3: Matrice per la classificazione manuale dei testi sulla base dello schema elaborato

Sulla base dello schema e della matrice, 3 esperti di formazione docente e un membro del gruppo di progetto hanno annotato 75 segmenti di testo. Su 44 di questi c'è stato un accordo alto (3 su 4) o totale (4 su 4). L'accordo intersoggettivo misurato sulla base del Fleiss K (una misura statistica per valutare l'affidabilità dell'accordo nell'attribuzione di categorie, tra più di due soggetti) ha dato esiti positivi (Fleiss K: 0,66). Di seguito a titolo esemplificativo, un segmento di testo di risposta alle domande accorpate 12.1.2 (Tab.1), classificato unanimemente come riflessivo. Abbiamo omesso per ragioni di *privacy* i brani con riferimenti riconoscibili all'esperienza.

Quella di [...] è stata la prima fiaba prevista dal progetto proposta ai bambini in un momento dell'anno (fine settembre - ottobre) dove non avevo ancora la piena conoscenza della sezione e dove muovevo i primi passi nella scuola [...], realtà totalmente

diversa a quella a cui ero abituata nella [...]. Quindi il procedere delle varie attività non sempre si è sviluppato come avevo previsto e progettato inizialmente. Ma, uno dei punti essenziali di una buona didattica è proprio la sua flessibilità e soprattutto la possibilità, o meglio, il dovere di poter correggere il tiro anche in itinere. Proprio per questo ho modificato le attività che avevo previsto ma che si sono rivelate poco idonee non avendo valutato la tenera età dei bimbi e le loro ancora giovani competenze: ho ingrandito disegni predisposti inizialmente troppo piccoli, sostituendo la tecnica inizialmente proposta del colore a pastello o a pennarello al riempimento delle immagini con varie tecniche manipolative [...]. Un'altra cosa che non avevo valutato bene sono stati i tempi: infatti mi sono resa conto che, nonostante la sezione abbia un'alta percentuale di presenza scolastica, essendo bambini così piccoli ci sono stati dei picchi di assenze dovuti a malattie stagionali che non mi hanno permesso di portare a termine tutte le attività progettate nel tempo previsto. Ho operato quindi una scelta prediligendo le attività dove il bimbo si riusciva ad esprimere meglio e dove c'era più partecipazione, entusiasmo e cooperazione (la narrazione con la loro partecipazione attiva, la drammatizzazione, psicomotricità musicale legata alla storia). [...]



La matrice ha funzionato come strumento di orientamento e confronto per il giudizio umano, cui in molti casi spetta di interpretare la presenza e/o di valutare le eventuali diverse gradazioni di ogni attributo, un'abilità, quest'ultima, spesso decisiva nell'attribuzione del segmento di testo a una delle 4 categorie. Abbiamo annotato manualmente in cicli diversi di annotazione oltre 600 risposte, il *corpus* annotato è stato poi classificato con strumenti di TAL.

6.2 *Classificazione automatica del livello di riflessività*

Come per la classificazione rispetto al *topic* (Par. 5.1), anche in questo caso sono stati adottati algoritmi di classificazione supervisionata basati su strumenti di TAL. Primi esperimenti in questa direzione sono descritti in Chiaritti (2018), dove gli autori riportano i risultati condotti a partire da un nucleo iniziale di risposte impiegate come *training corpus*. L'ampliamento di tale nucleo ha permesso di migliorare progressivamente le accuratezze del sistema automatico originariamente sviluppato. I risultati così raggiunti sono contenuti nelle Tabelle 4 e 5 che riportano rispettivamente l'accuratezza nella distinzione tra risposte riflessive e non riflessive, e tra i diversi gradi di "riflessività".

	Accuratezza
Risposte non riflessive	81,06
Risposte riflessive	82,01
TOTALE	81,53

Tab. 4: Accuratezza della classificazione tra segmenti di testo (risposte) riflessivi e non riflessivi

	Accuratezza
Risposta non riflessiva	59,67
Considerazione generale	64,65
Risposta riflessiva descrittiva	37,86
Risposta riflessiva	58,89
TOTALE	55,27

Tab. 5: Accuratezza della classificazione per grado di "riflessività"



Nel primo caso, abbiamo usato 642 risposte annotate a mano come *training corpus* (bilanciato rispetto al numero di risposte categorizzate come riflessive e non riflessive); nel secondo caso, il *training* era composto da 158 risposte categorizzate a mano per ognuno dei tre livelli e un numero uguale di risposte non riflessive, per un totale di 632 risposte. Come ci potevamo aspettare, il compito di classificazione in due classi era più semplice e infatti abbiamo ottenuto risultati più alti. Distinguere le risposte rispetto al livello di riflessività è invece un compito più complesso e l'accuratezza raggiunta dipende da quanto sono distinte le classi considerate.

Come introdotto nel Par. 5.1, la metodologia di classificazione automatica da noi messa a punto si basa sul testo linguisticamente annotato rispetto a più livelli di descrizione linguistica. Questo ci ha permesso di classificare il livello di riflessività delle risposte tenendo in considerazione in fase di addestramento una ricca serie di caratteristiche linguistiche. Sono state in particolare considerate caratteristiche che catturano diversi aspetti dello stile di scrittura del testo:

- caratteristiche di base che includono la lunghezza *i*) della risposta, calcolata come numero medio di parole, *ii*) della frase, calcolata come numero medio di parole, *iii*) delle parole, calcolata come numero medio di caratteri;
- caratteristiche lessicali che includono *i*) la distribuzione del lessico delle risposte rispetto al «Vocabolario di Base» redatto da Tullio De Mauro (De Mauro, 2000) e ai repertori d'uso in cui è organizzato (Fundamentale, Alto Uso, Alta disponibilità) e *ii*) la ricchezza lessicale delle risposte calcolata come *Type/token ratio*;

- caratteristiche morfosintattiche che includono *i*) la distribuzione delle categorie morfosintattiche riconosciute in fase di annotazione morfosintattica dei testi e *ii*) la distribuzione dei verbi rispetto al tempo, modo e persona in cui sono coniugati nei testi;
- caratteristiche sintattiche: che includono una serie di informazioni estratte dal testo annotato, adottando il formalismo di rappresentazione sintattica a dipendenze, che permette di ricostruire le relazioni di dipendenza sintattica tra i singoli *token* di una frase. Tali informazioni includono *i*) l'altezza dell'intero albero sintattico di una frase; *ii*) la lunghezza delle relazioni di dipendenza sintattica, calcolata in numero di *token* presenti tra la testa sintattica e la dipendenza coinvolte nella relazione; *iii*) la distribuzione dei tipi di dipendenza sintattica (es. soggetto, oggetto, modificatore); *iv*) la profondità media e la distribuzione di sequenze di complementi preposizionali che modificano un sostantivo; *v*) una serie di informazioni che riguardano il comportamento sintattico di predicati verbali (es. l'arità verbale media di una frase, cioè il numero di dipendenti di un predicato verbale, e la distribuzione di predicati con soggetto ellittico); *vi*) l'uso della subordinazione (es. la distribuzione di frasi principali e subordinate e il grado di incassamento di frasi subordinate); *vii*) l'ordine canonico degli elementi all'interno di una frase (es. la posizione di soggetto e oggetto rispetto al verbo da cui dipendono e delle subordinate rispetto alla principale).




L'analisi di quali tra le caratteristiche considerate contribuisce di più alla classificazione del grado di riflessività delle risposte fornite ha permesso di condurre uno studio sullo stile distintivo della scrittura riflessiva. Un testo riflessivo rispetto a uno non riflessivo risulta essere innanzitutto più lungo (in termini di numero di parole e frasi) e lessicalmente meno ricco, rispetto ai valori di *Type/token ratio*. È contraddistinto da caratteristiche tipicamente associate in letteratura ad una maggiore complessità sintattica, quali: uso maggiore della modificazione verbale (distribuzione maggiore di ausiliari e verbi modali), strutture verbali complesse (verbi caratterizzati da un valore di arità maggiore), uso esteso della subordinazione, ordine non canonico degli elementi della frase (distribuzione maggiore di oggetti pre-verbali e soggetti post-verbali).

Sono in particolare le frasi classificate come “riflessiva descrittiva” e “riflessiva”, ad essere caratterizzate da elementi di complessità sintattica. Sono, ad esempio, mediamente più lunghe, contengono strutture verbali complesse, incassamenti profondi di frasi subordinate ecc. Al contrario, esse contengono un lessico più semplice (una maggiore quantità

di parole contenute nel “Vocabolario di Base” e in particolare nel repertorio delle parole più frequenti e più semplici, quelle cioè “Fondamentali”) e sono caratterizzate da una minore varietà lessicale, con valori di *Type/token ratio* inferiori. Interessante infine notare come il grande uso di pronomi dimostrativi e determinativi rifletta uno degli aspetti principali della prosa riflessiva, il riferirsi cioè ad esperienze di vita diretta.

7. Conclusioni



L'integrazione di metodi manuali e automatici ha comportato molteplici e impegnative fasi di analisi per l'ottimizzazione degli strumenti, di cui l'ultima è ancora in corso, ma ha mostrato senz'altro di poter fornire indicazioni per rispondere alle domande di ricerca. La classificazione dei testi sulla base delle risposte (con un'accuratezza del 70%) ha fornito alcune conferme dei risultati delle analisi manuali (le domande “preferite” dai docenti, ad esempio). Inoltre il classificatore automatico, restituendoci i gradi di “somiglianza” tra risposte, deducibili dall'analisi degli errori di attribuzione, ci ha confermato l'opportunità di ristrutturare le prime due domande, mentre ha confutato la scelta di accorpare le altre due domande (12.3 e 12.4), azione che non ha comportato un significativo miglioramento delle prestazioni del classificatore. Anche nel caso della classificazione sulla base del “livello di riflessività” l'analisi degli errori di attribuzione ci ha fornito evidenze utili a progettare interventi per una più efficace azione di *scaffolding* con una rielaborazione della struttura e del contenuto delle domande. I bassi valori di accuratezza nell'attribuzione delle classi “riflessiva” e “riflessiva descrittiva” suggeriscono, invece, l'opportunità di fare ricorso a metodi e strumenti per l'estrazione del contenuto semantico dei testi: un'azione che potrebbe contare sulla buona *performance* del classificatore nella distinzione tra testi riflessivi e non riflessivi e quindi su un *corpus* già annotato linguisticamente a diversi livelli di granularità (frase, segmento di testo/risposta e testo), per argomento (le domande), e per categoria di “riflessività”.

Riferimenti bibliografici

- Boud D., Walker D. (1991). *Experience and learning: reflection at work*. Victoria: Deakin University.
Buckingham Shum S., Sándor Á., Goldsmith R., Bass R., McWilliams M.

- (2017). Towards reflective writing analytics: Rationale, methodology and preliminary results. *Journal of Learning Analytics*, 4, 58-84.
- Buckingham Shum S., Sándor Á, Goldsmith R., Wang X., Bass R., McWilliams M. (2016). Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. New York: ACM.
- Chiriatti G., Della Gala V., Dell'Orletta F., Montemagni S., Pettenati M.C., Sagri M.T., Venturi G. (2018). A NLP-based Analysis of Reflective Writings by Italian Teachers. *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*. Torino. <http://ceur-ws.org/Vol-2253/>
- Chung C., Pennebaker J.W. (2007). The psychological functions of function words. *Social communication*, 1, 343-359.
- Commissione europea/EACEA/Eurydice (2018). *Teaching Careers in Europe: Access, Progression and Support. Rapporto Eurydice*. Lussemburgo: Ufficio delle pubblicazioni dell'Unione europea.
- D'Amico D., Mangione G., Pettenati M.C. (2018). Governing a State-wide Induction Program: Characteristics and Success Conditions of the Italian Model. *Journal of e-Learning and Knowledge Society*, 14, 15-37.
- De Liddo A., Buckingham Shum S., Quinto I., Bachler M., Cannavacciuolo L. (2011). Discourse centric learning analytics. *LAK 2011: Proceedings of the first International Conference on Learning Analytics & Knowledge*. New York: ACM.
- Dewey J. (1933). *How we think: a restatement of the relation of reflective thinking to the educative process*. Boston: Heath.
- Kember D., McKay J., Sinclair K., Wong F.K.Y. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education*, 33, 369-379.
- Kovanovic V., Joksimovic S., Gasevic D., Hatala M., Siemens G. (2017). Content Analytics: The Definition, Scope, and an Overview of Published Research. *Handbook of Learning Analytics. First Edition*. Society for Learning Analytics Research (SoLAR). <https://www.solaresearch.org/hla-17/>
- Krippendorff K. (2004). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks: Sage Publications.
- Garrison D.R., Anderson T., Archer W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15, 7-23.
- Gibson A, Kitto K. (2015). Analysing Reflective Text for Learning Analytics: An Approach Using Anomaly Recontextualisation. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Association for Computing Machinery*. New York: ACM.
- Liu M., Buckingham Shum B., Mantzourani E., Lucas C. (2019). Evaluating Machine Learning Approaches to Classify Pharmacy Students' Reflective Statements. *Conference proceedings International Conference on Artificial Intelligence in Education*. Chicago: AIED.
- Mezirow J. (1990). *Fostering critical reflection in adulthood: a guide to transformative and emancipatory learning*. San Francisco: Jossey-Bass Publishers.
- Mikolov T., Chen K., Corrado G., Dean J. (2013). Efficient estimation of word representations in vector space. *Proceedings of ICLR Workshops Track*. arxiv.org/abs/1301.3781

- Moon J.A. (2006). *Learning Journal: A Handbook for reflective practice and Professional Development*. New York: Routledge.
- Rodgers C. (2002). Defining Reflection: Another Look at John Dewey and Reflective Thinking. *Teachers College Record*, 104, 842-866.
- Rossi P.G., Pettenati M.C., Mangione G.R., Magnoler P., Giannandrea L., Fedeli L. (2018). Italian Induction Model Since 2014: Bridging Two Sides of Teachers Professional Development. *Proceedings of the AERA Annual Meeting April 13-17*.
- Rossi P.G., Magnoler P., Giannandrea L., Mangione G.R., Pettenati M.C., Rosa A. (2015). Il Teacher Portfolio per la formazione dei neo-assunti. *Pedagogia oggi*, 2, 223-242.
- Schön D.A. (2006). *Formare il professionista riflessivo. Per una nuova prospettiva della formazione e dell'apprendimento nelle professioni*. Milano: Franco Angeli.
- Siemens G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57, 1380-1400.
- Ullmann TD. (2015). *Automated detection of reflection in texts. A machine learning based approach*. PhD Thesis, The Open University. <http://oro.open.ac.uk/45402/>

