

CLUB

Working Papers in Linguistics

A cura di Emanuele Miola e Rosa Pugliese

Volume 5, 2021



CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA
ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

CLUB

Working Papers in Linguistics

A cura di Emanuele Miola e Rosa Pugliese

Volume 5, 2021

Collana
CLUB WORKING PAPERS IN LINGUISTICS
(CLUB-WPL)

Comitato di Direzione

Cristiana De Santis
Nicola Grandi
Francesca Masini
Fabio Tamburini

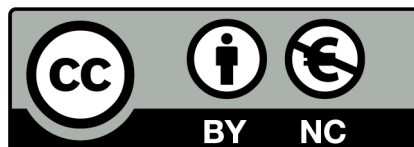
Comitato Scientifico

Marianna Bolognesi
Claudia Borghetti
Chiara Gianollo
Elisabetta Magni
Yahis Martari
Caterina Mauri
Marco Mazzoleni
Emanuele Miola
Rosa Pugliese
Mario Vayra
Matteo Viale

Il CLUB – Circolo Linguistico dell'Università di Bologna nasce nel 2015 con l'obiettivo di riunire coloro che, all'interno dell'Alma Mater, svolgono attività di ricerca in ambito linguistico.

Il CLUB organizza ogni anno un ciclo di seminari e pubblica una selezione degli interventi nella collana CLUB WORKING PAPERS IN LINGUISTICS.

I volumi, sottoposti a una procedura di peer-review, sono pubblicati online sulla piattaforma AMS Acta dell'Università di Bologna e sono liberamente accessibili.



CC BY-NC

CLUB Working Papers in Linguistics, Volume 5, 2021
ISBN: 9788854970649

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA
<http://corpora.ficlit.unibo.it/CLUB>

CLUB Working Papers in Linguistics

Volume 5

a cura di *Emanuele Miola e Rosa Pugliese*

Indice

Presentazione <i>Emanuele Miola e Rosa Pugliese</i>	5
Trattare gli errori a scuola: fra regolarità attese e sperimentazione linguistica <i>Cecilia Andorno e Anna Boario</i>	9
ParlaTO: la variazione (e il repertorio) in un corpus di parlato <i>Massimo Cerruti</i>	23
Prospettive di ricerca per valutare la competenza lessicale in una lingua seconda <i>Francesca Gallina</i>	41
Potenziali <i>marker</i> linguistici nello scritto di pazienti con Anoressia Nervosa: uno studio con tecniche di NLP <i>Giulia Minori</i>	57
The semantics behind the reanalysis of French articles in French-based creoles <i>Laura Tramutoli</i>	77
Sezione “Leggere per comprendere a partire dai dati INVALSI”	
Insegnare a leggere e capire <i>Mario Ambel</i>	103

Text comprehension: the role of individual differences	119
<i>Lucia Mason e Angelica Ronconi</i>	
Qualche osservazione sociolinguistica sulle prove INVALSI	135
<i>Emanuele Miola</i>	
Quattro caratteristiche dell'analisi logica. Considerazioni a margine delle prove INVALSI del grado 8	149
<i>Diego Pescarini</i>	
<i>Heri dicebamus</i> . Competenze linguistiche prima e dopo la pandemia	165
<i>Alberto A. Sobrero</i>	

Presentazione

Emanuele Miola

Università di Bologna

emanuele.miola@unibo.it

Rosa Pugliese

Università di Bologna

rosa.pugliese@unibo.it

1. Introduzione

I contributi compresi in questo fascicolo, il quinto della collana *CLUB Working Papers in Linguistics*, documentano le principali attività – ospitate nell’anno accademico 2019/2020 presso l’Università di Bologna – che sono state organizzate e promosse dal CLUB (Circolo Linguistico dell’Università di Bologna) o da alcuni suoi membri.

Si è trattato, come chiunque può immaginare, di un anno accademico profondamente diverso dagli altri, perché appena iniziata il secondo semestre la situazione sanitaria in Europa e nel mondo ha costretto l’Ateneo alla sospensione delle attività in presenza e, pertanto, anche degli eventi di carattere scientifico previsti, tra i quali le conferenze del Programma Ufficiale CLUB distribuite tra marzo e giugno. Fortunatamente, in questo torno di mesi, le piattaforme multimediali per videoconferenze hanno permesso di non cancellare almeno alcuni degli appuntamenti che erano stati organizzati.

Il volume si compone di due parti. I contributi che appaiono per primi sono quelli che, a vario titolo, sono stati presentati nell’ambito degli eventi CLUB 2019/2020. Ci piace segnalare che il penultimo saggio incluso nella prima parte è tratto dalla tesi di laurea magistrale che nel 2020 ha vinto il premio CLUB Day ‘Una tesi in linguistica’:¹ si tratta del lavoro di Giulia Minori. La seconda parte del fascicolo è invece dedicata alla pubblicazione di un largo numero di interventi presentati alla giornata di studio “Leggere per comprendere a partire dai dati INVALSI”, che si è tenuta a Bologna il 6 dicembre 2019 ed è stata organizzata dal GISCEL.

Ogni articolo è stato sottoposto alla revisione anonima di almeno due reviewer, sia facenti parte del comitato scientifico di questa collana sia esterni ad esso, che ringraziamo per la loro disponibilità e collaborazione.

Tutti i contributi raccolti in questa miscellanea sono idealmente connessi dal filo rosso dell’apprendimento/insegnamento della lingua e dell’analisi dell’uso concreto che di questa fanno le persone o, più in generale, le comunità sociali. Le discipline che sono rappresentate nel fascicolo spaziano dunque dalla glottodidattica alla linguistica clinica, passando per la sociolinguistica, la creolistica e la linguistica dei corpora.

In ciascuna delle due sezioni gli articoli si susseguono in ordine alfabetico.

¹ Si veda il sito <http://corpora.ficlit.unibo.it/CLUB/>.

2. I contributi di questo volume

Nel loro articolo, Cecilia Andorno e Anna Boario muovono dalla constatazione della distanza che tutt'ora separa la visione dell'errore nel dibattito scientifico (traccia visibile e transitoria di competenze in via di costruzione) da quella generalmente diffusa nella prassi glottodidattica (dato linguistico da sanzionare) per sviluppare alcune riflessioni sul tema, in chiave sia acquisizionale sia sociolinguistica, arricchita da esempi illustrativi. Dopo una chiarificazione terminologica e concettuale sull'errore (tipologia, cause, strategie), le autrici si soffermano sul suo trattamento a scuola, sottolineando il valore formativo di alcune procedure di spiegazione e di correzione, nonché dell'autocorrezione.

Massimo Cerruti presenta invece le caratteristiche del corpus per l'analisi del parlato ParlaTO, uno dei due moduli, assieme al KIP, del corpus KIParla (<https://kiparla.it>), e discute i risultati scaturiti dagli studi condotti sui materiali dei due sottocorpora. In particolare, le analisi mostrano l'infittirsi della circolazione di tratti regionali al di fuori dell'area di origine e perciò di produzioni linguistiche, quando non vere e proprie varietà, composite e la possibile affermazione, accanto alla varietà standard di impianto letterario, di un altro standard, coesistente con il primo, derivante dalla salita verso l'alto di fenomeni precedentemente giudicati sub-standard.

Il contributo di Francesca Gallina verte sulla valutazione della competenza lessicale in una lingua seconda, una pratica dalla triplice valenza – acquisizionale, didattica e certificatoria – che ne motiva il settore specialistico di studi (*Vocabulary assessment*) e la varietà di approcci al suo interno. L'autrice ripercorre le fasi di sviluppo di tali approcci, considerandone alcuni nodi critici (a partire dalla definizione stessa di *lessico*, *parola*, *competenza lessicale*). Tratta, poi, in dettaglio, le diverse dimensioni del lessico (l'ampiezza e la profondità, ad esempio), all'interno di test mirati alla valutazione della competenza lessicale, accennando anche alle modalità valutative di quest'ultima in prove centrate sulle abilità linguistiche.

Giulia Minori indaga i possibili biomarker linguistici dell'anoressia nervosa, estraendoli dalle produzioni scritte di un gruppo di pazienti (e rispettivi gruppi di controllo) mediante tecniche di *Natural Language Processing*. Minori verifica che, pur se il suo campione non è particolarmente numeroso, tra coloro che soffrono di anoressia nervosa vi è un aumento statisticamente significativo di semplificazione sintattica, compensata da una maggiore complessità lessicale. Si tratta di uno studio pilota che apre la strada al possibile riconoscimento di marcatori linguistici affidabili per la patologia in questione.

Il saggio di Laura Tramutoli rivolge l'attenzione ai pidgin e ai creoli e, tra questi ultimi, si concentra su quelli che, avendo come lingua lessificatrice il francese, presentano lessemi originatisi dalla rianalisi di un sintagma nominale francese formato da [articolo + nome], come in fr. *la plage* > cr. haitiano *laplaj* 'spiaggia'. L'autrice indaga quindi la compatibilità di questo fenomeno, detto agglutinazione dell'articolo, con il profilo semantico non oggettuale dei nomi su cui si produce, per mostrare come l'agglutinazione marchi specifici tipi semantici di nomi.

La didattica della comprensione, in particolare di testi scritti, è il tema dell'articolo di Mario Ambel, il cui obiettivo specifico è di sistematizzarne alcuni punti essenziali, in un

approccio strategico. L'argomento è contestualizzato dal punto di vista storico-pedagogico (le *Dieci Tesi*, negli anni '70, e il rinnovamento didattico negli anni '80), per rilevarne, in un'ottica comparativa e critica, alcuni aspetti legati alle politiche scolastiche, prevalentemente valutative, che negli ultimi decenni hanno segnato una battuta di arresto rispetto all'innovazione concettuale e operativa precedente, spostando il focus dal processo al prodotto della comprensione. L'autore specifica, dunque, cosa distingue le attività sulla comprensione (esemplificate poi in dettaglio) dalle prove intese a verificare tale abilità e indica, per quest'ultima, le cornici culturali programmatiche in cui essa si situa.

Lucia Mason e Angelica Roncon presentano innanzitutto una rassegna narrativa delle ricerche che, nel campo della psicologia, hanno gettato luce sui processi cognitivi implicati nella comprensione testuale, consentendo di chiarirne in particolare l'attività inferenziale (nei suoi diversi tipi) e le operazioni metacognitive. Dopo aver trattato alcuni aspetti rilevanti della *digital reading* (come l'affidabilità delle fonti nella lettura di testi in rete e il confronto tra lettura su schermo e lettura su pagina), le autrici esaminano in dettaglio le variabili (i fattori cognitivi, motivazionali, emotivi) che intervengono a livello individuale nella comprensione di testi, un'abilità la cui importanza travalica i contesti meramente educativi.

Emanuele Miola osserva le prove INVALSI delle scuole primarie e delle secondarie di primo grado con l'occhio del sociolinguista. Miola pone sotto analisi alcuni brani delle *Indicazioni nazionali* e alcuni tratti notevoli della varietà di italiano adoperata nei testi e nelle consegne delle prove, osservando, da un lato, se e in che modo sia presente l'attenzione al cambiare della lingua attraverso le dimensioni diatopica, diastratica e diafasico-diaemesica e, dall'altro, se e in che modo ne venga verificata la padronanza, concludendo che pur se la sociolinguistica non è trascurata nelle indicazioni, viene trattata prevalentemente, o soltanto, in relazione al lessico.

Partendo dall'assunto che l'analisi logica (intesa come analisi sintattica della frase semplice e complessa) sia tutt'ora un fondamento dell'educazione linguistica, Diego Pescarini, nel suo articolo, evidenzia i principi cardine che la qualificano, ovvero che la rendono: un'attività metacognitiva; un'analisi autonoma rispetto ad altre attività riflessive; razionale, cioè basata sulla verifica empirica di concetti chiari; trasversale e generalizzabile a tutte le lingue. L'autore discute questi quattro principi sullo sfondo di un contesto storico e pedagogico-sociale, delineato allo scopo, e attraverso una rilettura critica di alcune prove INVALSI per la scuola media. Da queste, infatti, trae spunto per la sua proposta, che è mirata ad offrire un valido orientamento alle/agli insegnanti.

Infine, Alberto Sobrero esamina i dati sulla competenza in lingua italiana ottenuti dai rilevamenti INVALSI degli ultimi anni. Il suo obiettivo è individuare alcune linee di tendenza nelle variazioni della competenza linguistica, attraverso un'analisi comparativa tra i dati del 2018 e quelli del 2019, compiuta secondo parametri quali: il divario tra aree geografiche, il livello socio-economico e culturale delle famiglie, la presenza di nativi e immigrati, le differenze tra scuole e classi. I dati ricavati da questa analisi sono poi discussi in rapporto a quelli più recenti, emersi dal confronto tra i dati INVALSI del 2019 e quelli del 2021, alla luce, quindi, dei cambiamenti (negativi) generati dal lockdown e dall'esperienza della didattica a distanza. Dall'analisi, nel suo insieme, l'autore trae alcuni spunti utili a riorientare la didattica.

Trattare gli errori a scuola: fra regolarità attese e sperimentazione linguistica

Cecilia Andorno

Università di Torino

ceciliamaria.andorno@unito.it

Anna Boario

I.I.S. Giulio, Torino

anna.boario@didalabgiulio.it

Abstract

Per quanto oramai ampiamente centrale nella riflessione scientifica, il trattamento e la considerazione dell'errore come indizio significativo delle competenze dell'allievo fatica a farsi strada nella pratica glottodidattica. Riteniamo che ciò possa essere in parte dovuto alla difficoltà di concettualizzare l'errore come una manifestazione della creatività che è alla base della capacità di apprendimento umana. Questo articolo propone, senza pretesa di esaustività, alcune riflessioni in merito alla natura e, di conseguenza, al trattamento degli errori in una prospettiva tanto acquisizionale quanto sociolinguistica, proponendo al contempo strumenti di consultazione che riteniamo possano costituire un ausilio per chi lavora in ambito glottodidattico: grammatiche di riferimento della lingua italiana, ma anche studi scientifici di riferimento che trattano l'errore in prospettiva psicolinguistica e storico-evolutiva.

1. Introduzione

“Gli errori sono necessari, utili come il pane e spesso anche belli: per esempio, la torre di Pisa”.

(Rodari 1964: 7)

L'assunto da cui partiamo in questo articolo è che, quando commettiamo un errore nel parlare o nello scrivere in una lingua, stiamo compiendo un'azione che spesso ci porta ad *errare* e non necessariamente a sbagliare.¹ L'ambiguità semantica della parola *errare*, dal

¹ Molte delle riflessioni qui contenute possono essere approfondite in altri lavori. Sul valore che hanno gli “errori” nelle lingue e nella descrizione linguistica si può vedere il volumetto a cura di Nicola Grandi (2015). Sul ruolo che l'errore ha per il processo di apprendimento individuale, della lingua materna come delle seconde lingue, e sulle conseguenti modalità di trattamento, si possono vedere invece il volume a cura di Roberta Grassi (2017) e il volume di Teresa Nesci e Anna Cattana (2004): si tratta di materiali pensati

latino *errare*, rimanda sia all'errore come distanza da ciò che è norma ('sbagliare'), sia all'erraticità del viandante ('vagare'), che attraverso un percorso non previsto si avvia alla scoperta di qualcosa di nuovo. Compiere un errore significa allontanarsi dalla regola o dalla norma stabilita ma anche, e contemporaneamente, compiere una sperimentazione linguistica. Se riflettiamo sulla lingua come sistema dinamico, caratterizzato da regole non sempre coerenti e spesso oscillante tra norma ed esperienza individuale del parlante, possiamo osservare che la maggior parte degli errori linguistici sono anche il risultato di anomalie della grammatica. Citando un esempio ben noto a chi lavora in ambito acquisizionale, un bambino in fase di apprendimento della sua lingua, che si trova cioè ancora immerso in una varietà di apprendimento, per analogia con la coniugazione dei verbi regolari della terza coniugazione dirà spontaneamente *io salo le scale*, e si adeguerà alla norma che indica come corretto l'uso di *io salgo le scale* solo quando ascolterà ripetutamente, nell'*input* linguistico con cui è a contatto, la forma irregolare cui altri parlanti si sono adeguati. È necessario che, alla luce delle conoscenze cui la ricerca in linguistica è giunta ormai da diversi anni, l'errore nel contesto scolastico più che essere qualcosa da sanzionare rappresenti, una volta per tutte, un'occasione per riflettere sulla lingua e sul suo funzionamento. Questo non vuol dire che gli errori non vadano corretti, ma che è necessario accompagnare alla correzione una riflessione su quali possano esserne le cause e quali strategie possano essere messe in atto per evitare di ripeterli. Dal punto di vista didattico, uno degli aspetti cui prestare molta attenzione è ovviamente il rapporto tra lingua parlata e lingua scritta. Il bambino, ma anche l'adulto, ha una naturale tendenza ad appoggiarsi alla percezione del parlato, che è la lingua con cui tutti abbiamo più familiarità, la nostra vera lingua madre, ma spesso questa strategia porta fuori strada perché lo scritto ha le sue regole, che per lo più riflettono la forma d'italiano standard stabilita dalla norma, non sempre manifestate dalla lingua parlata. Riflettere sull'errore può significare quindi, a volte, osservare questo scarto e sperimentare l'efficacia della norma.

2. Avere chiaro che cosa si considera errore

La nozione di errore implica quella di una regola(rità) attesa: un osservatore/ascoltatore che rileva un errore nella produzione linguistica di un altro parlante implicitamente compara una forma attesa o *target*, che conosce e reputa appropriata per quel contesto, con quella che è stata prodotta effettivamente dal parlante, e che lui reputa inappropriata. La rilevazione di un errore implica quindi il riferimento (eventualmente implicito, inconsapevole) a una forma *target* "normale", ovvero prevista dalla norma. Ora, è ben noto che la norma della lingua non è stabile in nessuna delle sue dimensioni di variazione: l'aspettativa di ciò che "è appropriato" in un certo contesto cambia a seconda del luogo, delle occasioni comunicative, delle persone coinvolte nella comunicazione, del tenore della comunicazione. Mentre forme come *salò* o *salisco* per *salgo* sono errate in qualsiasi contesto d'uso (anche se non "fuori dal sistema", come dimostra il fatto che le sentiamo usate da un bambino italofono, o anche da un adulto che sta imparando l'italiano come lingua straniera: torneremo su questo punto nel par. successivo), altre varianti potrebbero

per l'insegnamento dell'italiano come seconda lingua, con molte riflessioni utili anche in chiave di insegnamento a nativi. Adriano Colombo (2015) inquadra la questione dell'intervento didattico sugli errori in una prospettiva sociolinguistica, proponendo attività condivise di analisi e interventi sugli errori del tipo di quelle qui proposte. Riflette sul tema degli errori in chiave di spia di meccanismi di apprendimento anche Rosaria Solarino (2009).

prestarsi a maggior discussione; il problema più interessante sollevato dalla questione della variabilità della norma è probabilmente dato, in italiano, dalla tensione fra varietà *standard* e *neostandard*. Nell'esemplificare la sua proposta descrittiva nel noto volume sulla *Sociolinguistica dell'italiano contemporaneo*, Berruto (1987: 30) propone queste varianti dello stesso enunciato, tutte ugualmente proprie del sistema dell'italiano, perlomeno per un parlante di area settentrionale:

- (1) *Non sono affatto al corrente di ciò che è stato detto loro*
- (2) *Non so che cosa è stato detto loro*
- (3) *Non so cosa gli hanno detto*
- (4) *Non so mica cosa gli han detto*
- (5) *So mica cosa che ci han deto*

È evidente che è difficile, e non particolarmente proficuo per la nostra riflessione, stabilire semplicemente quale sia la (sola) variante “giusta”. Da un lato, qualunque parlante colto qualificherebbe immediatamente chi usa la variante 5 come parlante poco colto, soprattutto per la spia dell'uso di *deto*, che riproduce una parlata regionale con mancata realizzazione di consonanti doppie, e di *ci* come pronome dativo di III persona. D'altronde, probabilmente, gli stessi parlanti colti non sarebbero altrettanto concordi su quali delle altre varianti debbano essere ammesse e quali no. Ad esempio, l'uso di *loro* come pronome di III persona plurale (es. 1-2) è considerato l'unico accettabile da molte grammatiche didattiche, ma d'altronde nel parlato anche di persone colte è molto frequente *gli* (es. 3-4), al punto da passare del tutto inavvertito. Ecco in merito il parere di Luca Serianni in una delle principali grammatiche di riferimento della lingua italiana:

L'atono *gli* – largamente attestato in tutti i secoli di storia della nostra lingua – appartiene al registro familiare; il parlato formale e la massima parte dello scritto (tecnico-scientifico, letterario e in una certa misura anche giornalistico) preferiscono la forma *loro*. [...] Per la storia di *gli/loro* nell'italiano moderno, è significativo che il Manzoni nei *Promessi Sposi* accolga solo limitatamente *gli* in luogo della forma tradizionale, che resta la regola anche nelle parti dialogate. (Serianni 1989: 249)

Come si vede, il parere di Serianni è sfumato: tanto *gli* quanto *loro* hanno spazio d'uso nella norma dell'italiano, ed entrambe le forme sono usate anche da parlanti colti; la differenza d'uso è più da cogliere in termini di canale di comunicazione (scritto *vs.* orale) e, soprattutto, di occasione comunicativa (formale *vs.* informale).

Berruto dà un parere conforme, ma non identico, qualificando la forma *loro* come propria dello standard e la forma *gli* come propria del neostandard, cioè come una forma che si sta attestando in un nuovo standard, e che quindi viene sempre più spesso considerato accettabile (e non percepito come errore) anche in molti contesti formali. A testimonianza di questo, possiamo richiamare l'esperienza personale avuta con un insegnante che in un testo scritto prodotto durante un corso di formazione, nel descrivere la propria vigilanza sulla scrittura dei propri allievi, spiegava che preferiva sempre *consigliargli* le forme più accettate dallo standard: questo insegnante, che pure si descrive come un conservatore in fatto di lingua, evidentemente non percepisce più la forma *gli* per *loro* come un tratto informale, perlomeno non in maniera così vistosa da impedirgli di usarla in uno scritto di media formalità.

In merito alla scelta fra *cosa* e *che cosa*, ecco i riferimenti forniti da Serianni:

Che si alterna, tanto nelle interrogative dirette quanto nelle indirette, con *che cosa* e con l'ellittico *cosa*. Le tre forme sono tutte molto comuni nell'uso. [...] *Cosa* invece di *che cosa* fu spesso osteggiato dai grammatici del secolo scorso, ma si diffuse ampiamente grazie al largo impiego fattone dal Manzoni nell'edizione definitiva dei *Promessi Sposi*. (Serianni 1989: 325)

Anche in questo caso, Berruto attribuisce la forma *che cosa* allo standard e la forma *cosa* al neostandard. Rispetto al caso precedente, però, Serianni ritiene che l'evoluzione della lingua sia avanzata maggiormente, e che la forma del neostandard risulti sempre più comune (anche nello scritto e nelle varietà più formali), anche a causa del modello, in questo caso innovante, di Manzoni. In futuro, forse, si potrà dire lo stesso di *gli* per *loro*. D'altronde, sembrano già andare in questa direzione alcuni pareri autorevoli contenuti in testi più recenti della Grammatica di Serianni. La *Grammatica Italiana Treccani* online del 2012 ([www.treccani.it/enciclopedia/gli-o-loro_\(La-grammatica-italiana\)/](http://www.treccani.it/enciclopedia/gli-o-loro_(La-grammatica-italiana)/)) dichiara:

L'uso della forma pronominale atona *gli* in funzione di complemento di termine in riferimento non solo al maschile singolare, ma anche al maschile plurale è ormai da considerarsi accettabile in quasi tutti i livelli di lingua.

E l'Accademia della Crusca, sul suo sito di consulenza linguistica online (www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica), a una domanda su quello che evidentemente è un problema di oscillazione della norma molto sentito, risponde nel 2003 con una descrizione molto circostanziata della questione, per concludere:

L'uso di *gli* in luogo di *loro*, *a loro*, *a essi* e *a esse* è da considerare senz'altro corretto (*Ora vado dai tuoi amici e gli dico che la devono smettere di fare chiasso*), tranne che, forse, nel caso di registri altamente formali (*Il parroco espresse loro le sue più sentite condoglianze*).

Valutare adeguatamente gli errori, in definitiva, non implica (sol)tanto il saper identificare la (sola) forma *target* corretta, ma piuttosto l'essere consapevoli del quadro, mobile e sfumato, della norma linguistica, delle varianti esistenti e del loro valore sociale. Per una valutazione di questo tipo, non sempre la competenza naturalmente acquisita dal parlante nativo, sia pure colto, è sufficiente: nessun parlante, neppure il parlante colto, domina e possiede l'insieme delle varietà di una lingua e ne conosce nell'interezza le norme di riferimento. Per l'insegnante è utile quindi documentarsi su grammatiche di riferimento – che spesso offrono risposte più aggiornate e ragionate delle grammatiche scolastiche, non tanto in cerca di una risposta semplice (“è un errore o no?”), ma di una descrizione ragionata della norma che sia fondata su un'osservazione ampia del sistema linguistico, qual è quella fornita qui sopra dagli esempi di Serianni. È utile inoltre un approccio aperto, di osservazione e ricerca del dato linguistico continuamente disponibile nel mondo circostante: la lingua è in continuo movimento, e gli stessi linguisti che la descrivono perfezionano e adeguano costantemente le proprie osservazioni; la stessa cosa può fare l'insegnante, proponendo questa attività anche ai propri allievi.

3. Violare la norma, sfruttare le potenzialità del sistema

“In realtà il concetto di errore presuppone un parlante che parli una lingua non sua: nella propria lingua il parlante non può, per definizione, fare errori”.

(Cardona 1988: 124)

Nel sistema di una lingua rientra qualsiasi regolarità osservabile nel comportamento dei suoi parlanti: in questo senso, come illustra la citazione di Cardona, i parlanti di una lingua “non sbagliano mai”, cioè non si comportano mai in modo diverso da quanto previsto dal sistema, perché esso è descritto a partire da ciò che loro stessi fanno. Le regole linguistiche che si ricavano da un approccio di questo tipo sono di tipo descrittivo. La distinzione che si traccia in questa prospettiva è semmai legata alla distinzione fra parlanti nativi (che in qualche misura sono “il metro di paragone” di ciò che è proprio di una lingua e ciò che non lo è) e parlanti non nativi. Chi insegna una lingua si pone però solitamente (anche) in un’ottica prescrittiva, si pone cioè l’obiettivo di indicare ai propri allievi quali delle regole disponibili in un sistema sono appropriate in un dato contesto. Abbiamo visto che, in generale, il modello di italiano proposto dalle grammatiche, e che costituisce la “norma” sociale di riferimento, è l’italiano standard. In base a questo modello, è errore, benché attestato nell’uso, l’utilizzo dell’indicativo invece del congiuntivo in dipendenza da determinati verbi e congiunzioni (**credo che va bene; *prima che te ne vai*), o l’uso di *gli* come dativo femminile (**gli ho detto (a Carla)*); in contesti di più rilassata formalità, invece, può essere adottato il modello dell’italiano neostandard, che prevede l’uso rispettivamente dell’indicativo e di *gli*. Oltre alle scelte legate al modello di lingua che si intende prendere a riferimento, esistono norme che regolano l’adeguamento locale a caratteristiche della situazione comunicativa, pensiamo ad esempio alla scelta nell’uso di formule di saluto e allocutivi per rivolgersi all’interlocutore: “dare del lei” rispetto a “dare del tu”. Intendiamo qui come errore non tanto la non appartenenza di una struttura al sistema della lingua, ma piuttosto la sua non appropriatezza rispetto al contesto comunicativo. Errore, nel senso di inappropriata, può anche essere l’instabilità nel mantenere una scelta normativa adottata. Possiamo pensare al caso – frequente nei bambini o in parlanti non nativi che ancora non dominano i diversi registri di formalità – di chi si congedasse da una persona adulta e sconosciuta con le parole *grazie signora, ciao*. I casi appena visti rimandano quindi al problema di scegliere, fra le diverse possibilità del sistema, quelle appropriate per la comunicazione in corso: gli errori, in questo senso, sono errori di selezione fra le diverse forme disponibili.

Nello sfruttare le potenzialità del sistema, un errore può anche dare origine a una forma che non è propria di nessuna varietà del sistema stesso. Queste forme tuttavia spesso non si pongono al di fuori del sistema, nel senso che si producono seguendone – e sovraestendone – le regole. Un bambino (ma anche un adulto) che dice *prenduto* per *preso*, *rubatore* per *ladro*, *bevei* per *bevi* sta commettendo degli errori, dal punto di vista delle forme che il sistema della lingua italiana effettivamente attesta, ma lo fa proprio in quanto sfrutta delle regole presenti nel sistema, come nel caso già visto di **salo* per *salgo*. **Prenduto* è formato regolarmente come participio da *prendere* seguendo il modello di *vendere* > *venduto*; è il sistema a prevedere invece, in questo caso, una forma irregolare, *preso* (o, meglio, una forma che segue una regola più specifica, cui si attengono anche altri verbi simili come *rendere* > *reso*, *difendere* > *difeso*, ecc.). Da *bevo* si può produrre un passato

remoto regolare **bevei*, come da *parlo* > *parlai* e da *parto* > *partii*, ma il sistema prevede invece per il verbo *bere* una forma irregolare (o meglio, anche in questo caso, una regola più specifica, con raddoppiamento consonantico, *bevvi*, come accade ad esempio per *cadere* > *caddi*, *venire* > *venni*). Da *rubare* si può infine derivare **rubatore*, ovvero ‘colui che ruba’, secondo un modello derivativo molto produttivo in italiano, come da *saldare* > *saldatore*, ‘colui che salda’; ma l’esistenza della forma *ladro* nel lessico ha perlomeno finora bloccato la fortuna di una simile parola derivata. È importante osservare che questo tipo di errore per analogia, che sfrutta creativamente le regole del sistema, suggerisce una competenza della regola stessa da parte di chi lo commette: chi usa *rubatore* conosce sicuramente la relativa regola di derivazione, dato che non è possibile che questa parola sia stata semplicemente ricavata dall’*input* dei parlanti; in modo quasi paradossale, non potremmo fare la stessa affermazione per chi usa la parola *saldatore*, che potrebbe essere una parola appresa dall’*input*, e non costruita a partire da *saldare*. Gli errori di analogia, insomma, aprono una finestra preziosa sulla competenza del sistema da parte dei parlanti che li commettono, e andrebbero perciò trattati con la dovuta attenzione, soprattutto in quei parlanti per i quali il sistema di regole è ancora in formazione. Questo vale per i parlanti di italiano come seconda lingua, nei quali sono ancora in corso di apprendimento tratti essenziali della morfosintassi che nei parlanti nativi di pari età sono già consolidati. E vale per i parlanti nativi di italiano, per tutti quei tratti che sono oggetto di apprendimento tardivo, come può essere il passato remoto per parlanti di varietà settentrionali, per i quali il passato remoto non è di uso abituale ma appartiene soprattutto a registri scritti e formali che si incontrano in misura consistente solo a partire dagli anni di scolarizzazione.

Possiamo infine chiederci se si diano anche errori che infrangono le regole del sistema. Sulla base della definizione di errore e di sistema che discendono dalla citazione di Cardona che abbiamo posto ad inizio paragrafo, errori di sistema non dovrebbero essere possibili per definizione, per un parlante nativo. Possiamo però incontrarne nei parlanti non nativi, ovvero in coloro per i quali il sistema è ancora in corso di apprendimento dopo gli anni della prima scolarizzazione; così come ne possiamo incontrare nei parlanti nativi, quando il sistema è, anche per loro, ancora in corso di apprendimento, ovvero nei primi anni di vita. È importante infine osservare che, mentre il cuore del sistema, la sua norma in senso descrittivo, può essere piuttosto definito e uniforme fra i parlanti, i “confini” del sistema possono invece essere piuttosto sfumati. Per fare un esempio, le regole di accordo morfosintattico fra nome e aggettivo in italiano paiono da un lato parte di quel nocciolo duro della lingua italiana particolarmente stabile e non soggetto a variazione; d’altronde, nell’italiano popolare, è facile incontrare molte oscillazioni.² Nei cartelli di un banco di ortofrutta di un mercato rionale, di proprietà di un venditore sicuramente di origine italiana, abbiamo trovato attestate, ad esempio, le forme *cipolle rossi*, *cipolli bionde*, *cipolli freschi*, *carote pugliese*: queste oscillazioni mostrano che tanto l’appartenenza di nomi e aggettivi alle varie classi flessive, quanto le regole di accordo fra nome e aggettivo possono non fare parte del repertorio di regole possedute e condivise dalla totalità degli appartenenti alla comunità italoфона; inoltre, poiché riguardano soprattutto l’opposizione fra forme in *-e* e forme in *-i*, queste oscillazioni suggeriscono che la stessa opposizione fonematica fra le due vocali /e/ e /i/, perlomeno in posizione finale di parola, non sia parte

² In una prospettiva di approfondimento della nozione di italiano popolare, utile anche in chiave di riflessione didattica per la composizione del repertorio linguistico dei parlanti nativi di italiano, si può fare riferimento a De Mauro (1970), Berruto (1993); un’ottima rassegna bibliografica sull’argomento è offerta dalla relativa voce di D’Achille (2010) sull’Enciclopedia Treccani.

del repertorio condiviso da tutti gli italofoeni. Oltre che a una simile oscillazione di forme, il mancato possesso di una regola può dar luogo a fenomeni di ipercorrettismo una volta che si è divenuti consapevoli, di solito a seguito di un insegnamento esplicito, dell'inadeguatezza di una forma alternativa posseduta. Questi errori emergono specialmente in contesti formali, come può essere proprio il contesto scolastico, dove il parlante manifesta così il proprio sforzo di adeguamento a una norma non pienamente posseduta. Può essere di questo tipo l'errore di sovraestensione dell'uso di *lo/gli* su *il/i*, ad esempio la forma **gli dentisti* per *i dentisti*: un parlante che conosce l'esistenza delle due forme di articolo determinativo *il* e *lo* può non essere del tutto certo dei rispettivi ambiti d'uso, e quindi mostrare delle esitazioni che possono dar luogo ad errori di sovraestensione della forma più rara, soprattutto nello scritto, in cui i parlanti sorvegliano maggiormente le proprie scelte in uno sforzo di adeguamento a una norma non pienamente posseduta.

4. Sbagliando s'impara

La conoscenza è sempre fallibile ed è in questa dimensione di fallibilità che è possibile errare ed espandere gli orizzonti di ciò che ancora è ignoto. L'errore può essere inteso come processo in divenire, come momento transitorio nel terreno della conoscenza.

Perché possa esserci un dialogo sugli errori è molto importante considerare il momento della valutazione didattica come occasione formativa, sia per lo studente che per il docente. Consideriamo ad esempio l'attività della scrittura di un testo. Come osserva Adriano Colombo (2015: 122) "la correzione utile dovrebbe avvenire nel corso stesso dell'attività di scrittura". Si immagini ad esempio, come suggerisce Colombo, un contesto molto diverso da quello scolastico e si prenda in considerazione un corso di danza. Noi penseremo al tango: i passi del tango per essere appresi richiedono gradualità e per evitare che si sedimentino in chi balla dei movimenti e una postura sbagliati, che comprometterebbero il piacere di danzare in coppia, questi devono essere eseguiti davanti allo sguardo attento dell'insegnante che prontamente interviene e spiega perché quei movimenti non sono adeguati. In tal senso nessun ballo, durante le lezioni, viene considerato come la prova definitiva di ogni iscritto al corso ma, passo dopo passo, revisione dopo revisione del proprio incedere, i movimenti dei ballerini saranno sempre più armoniosi. In classe forse non si può osservare con la medesima attenzione l'allievo che scrive (lo scrivere è un gesto più intimo del ballare e sentirsi troppo osservati può essere molto inibente), tuttavia è altrettanto possibile permettere all'allievo di svolgere prove intermedie in cui gli errori non siano considerati come l'esito definitivo della sua prestazione ma come piccoli passi verso una maggiore competenza che lo porteranno, nel tempo, alla realizzazione di un testo armonico e corretto. Un'altra cosa che gli insegnanti di tango chiedono ai ballerini è la ripetizione di serie di movimenti e la composizione degli stessi su un intero brano musicale. Nel contesto scolastico questo potrebbe tradursi in esercizi che ripropongano varianti della medesima frase nei contenuti ma non nella struttura e "nella richiesta di riscrivere il proprio testo, in tutto o in parte, tenendo conto delle correzioni in un apposito quaderno" (Colombo 2015: 126).

Un altro parallelo utile con le lezioni di ballo è la considerazione che, mentre ad un corso di ballo i ballerini, benché inesperti, si esercitano in coppia, raramente gli insegnanti considerano l'opportunità di far scrivere e rileggere testi agli studenti in coppia. Così come in una coppia di ballerini ciascuno svolge ruoli diversi ma ascolta attentamente il proprio compagno per comprendere come muoversi e individuare gli sbagli commessi,

così due studenti potrebbero imparare l'uno dall'altro ascoltandosi e osservando i reciproci errori. Queste pratiche facilitano i processi di autocorrezione e possono prevedere osservazioni condivise da tutta la classe o la correzione reciproca in piccoli gruppi. Come osserva Serafini (1984), gli studenti sono più inclini a svolgere il ruolo di giudici dei lavori altrui piuttosto che a produrre testi propri e questa è sicuramente una risorsa da tenere in considerazione. Il dialogo fra pari sugli errori apre quindi finestre di spontaneità che con il docente sarebbero limitate e attiva quel clima sperimentale che amplifica la curiosità. Il docente che corregge in solitudine a casa propria non solo rischia di vivere in maniera alienante la fase di valutazione ma si priva di uno strumento molto utile. Lo scarto tra le proprie competenze e quelle dell'allievo non sempre gli permette di cogliere il piano creativo ed errante su cui si muove lo studente. Chiedere esplicitamente alla classe di spiegare perché sono state fatte determinate scelte da un determinato allievo permette invece di ascoltare la logica che sussiste dietro percorsi imprevedibili che aprono ulteriori scenari di scoperta e approfondimento.

Il perfezionamento della competenza attraverso un percorso per errori e correzioni è proprio di ogni apprendimento, ed è ormai accertato che, per quanto attiene all'apprendimento delle lingue, riguarda tanto l'apprendimento della lingua materna quanto di una lingua seconda o straniera. I parlanti nativi di una lingua diventano infine competenti nella loro lingua materna attraverso un lunghissimo allenamento che inizia nella socializzazione primaria con i genitori, i fratelli e le altre persone che interagiscono con loro nei primi anni di vita. Ecco ad esempio come il padre e il fratello di una bambina di tre anni, Luisa, replicano ad una forma che Luisa produce, segnalandole indirettamente che la forma non va bene (esempi in Fasulo & Pontecorvo 1999):

- (6) *Luisa: se questo- se questo si spegnava*
Fratello: si spegnava? ((ridendo))
- (7) *Luisa: se vorrebbe uscire il cane*
Padre: se?

Attraverso piccoli ma continui segnali di questo tipo (il riso del fratello; l'invito alla riformulazione del padre) i bambini imparano nel corso dei primi anni di vita la propria lingua materna. Una volta usciti dalla cerchia familiare ristretta (andando all'asilo, ai giardini, e più tardi a scuola, nei negozi, o in qualsiasi altro contesto che richieda attività di comunicazione) i bambini approfondiscono la propria competenza in ulteriori funzioni comunicative (raccontare, descrivere, chiedere il permesso, chiedere un favore, far valere le proprie ragioni...), ulteriori registri (rivolgersi a sconosciuti, scherzare, essere gentili, mostrarsi offesi ...), ed anche ulteriori forme linguistiche (il lessico si espande, la grammatica diviene più complessa). In questo lungo percorso i bambini imparano la propria lingua materna in forma negoziata e interattiva, chiedendo istruzioni e sperimentando; d'altronde, diverse figure si incaricano del ruolo di guidarli in questo percorso di apprendimento, eventualmente verificando e correggendo più o meno implicitamente il loro comportamento. Come suggerisce il titolo del libro di Fasulo e Pontecorvo, "Come si dice?" è una domanda a doppio senso: lo chiede il bambino che ha bisogno di istruzioni, e lo dice l'adulto che vuole richiamare l'attenzione del bambino alle competenze che già possiede. L'esperienza scolastica imprime nuovo slancio a questo apprendistato, esponendo il bambino a nuove forme di comunicazione verbale, anche ma non solo attraverso

il canale della letto-scrittura. Gli insegnanti, come figure preposte all'apprendimento, offrono costantemente un rimando agli allievi sulle loro produzioni linguistiche, non solo attraverso le pratiche valutative più istituzionali (la correzione, l'assegnazione di un voto), ma nel corso di ogni interazione (l'esempio che segue è di Grassi & Mangiarini 2010):

(8) *INS: Elsa, a suo fratello hai telefonato?*

STU: sì, lo ho telefonato

INS: lo?

A fronte del lungo, articolato e complesso percorso che porta l'allievo madrelingua ad avere alle soglie dell'adolescenza una competenza linguistica ampia (anche se non necessariamente nella varietà standard!), non deve stupire né essere fonte di scoramento o impazienza il constatare che l'allievo non madrelingua non riesce a “mettersi in pari” con i coetanei rapidamente e continua a commettere errori. Gli errori sono parte ineliminabile del percorso di apprendimento, e le tappe di sviluppo della competenza non possono essere saltate: importante è che il sistema continui a muoversi, e che non si fossilizzi ad uno stadio in cui non si vedono ulteriori progressi.

5. Adesso ti spiego che errori hai fatto

Con questa frase l'insegnante intende di solito dire “ti faccio vedere dove ci sono degli errori”. Individuare l'errore e il suo *target* non vuol dire però ancora aver individuato la causa (la spiegazione) dell'errore stesso.

Considerando ad esempio gli errori ortografici che possono presentarsi in un qualsiasi tipo di testo, possiamo interrogarci sulle loro cause: si tratta di problemi che hanno una causa ortografica, o di problemi che si vedono nella scrittura ma hanno altra fonte? Abbiamo già incontrato errori ortografici nella scrittura dei cartelli al banco del mercato: *cipolle rossi, cipolli bionde, cipolli freschi, carote pugliese*. In questo caso, abbiamo ipotizzato che l'errore ortografico sia dovuto a un problema morfologico, ovvero l'incerta conoscenza delle terminazioni nominali e della classe flessiva di appartenenza dei nomi, che può verificarsi nell'italiano popolare e manifestarsi anche nella scrittura. A sua volta, la difficoltà morfologica di apprendimento della flessione nominale, che vediamo specificamente legata all'alternanza fra le vocali <e> ed <i>, potrebbe essere dovuta a una difficoltà di discriminare i rispettivi suoni in posizione finale di parola, caratteristica questa propria di alcuni dialetti italiani, da cui l'italiano popolare può essere influenzato. In questo caso, abbiamo dei problemi di competenza fonetica e (di conseguenza) morfologica che si manifestano poi anche nella grafia.

Quando in una composizione scolastica incontriamo invece l'errore ortografico seguente, possiamo fare altre ipotesi (per il testo completo dell'esempio cfr. più avanti):

(9) *Dopo averlo chiamato più volte senza risposta decisi di alzarmi e di andare a vedere cosa stesse succedendo*

È difficile qui immaginare che l'errore ortografico sia causato da competenze limitate sul piano morfologico: è probabilmente una congiunzione scritta male, e difficilmente possiamo supporre che la scrivente lo consideri invece un verbo, dato che è usato al presente

in un contesto di passato e alla terza persona in un contesto di narrazione alla prima persona. Un ulteriore approfondimento sull'errore, ad esempio attraverso un'interazione con la scrivente, potrà chiarire ulteriormente se tale incompetenza sia dovuta a una mancata competenza della norma ortografica che richiede la forma *e* per la congiunzione e la forma *è* per il verbo, o a una scarsa competenza metalinguistica (nella capacità di discriminare esplicitamente fra le categorie di verbo e congiunzione), oppure ancora a scarsa attenzione (nel caso che la scrivente conosca la norma e sappia discriminare verbo e congiunzione e quindi, richiesta di rileggere il testo, sia in grado di correggersi).

Spiegare un errore significa quindi prendere in considerazione almeno due variabili: la tipologia superficiale cui è possibile ricondurlo, e le cause del fenomeno, che non appaiono in superficie. Durante la correzione è molto importante che queste due variabili siano tenute distinte e che possa essere l'allievo stesso a intuire, dopo un periodo di esercizio mirato e la creazione condivisa di una griglia di correzione, a quale categoria ricondurre il proprio errore.

Questo processo è necessario per ridurre lo scarto che esiste tra il punto di vista di chi usa la lingua, con le sue esigenze espressive e il percorso del proprio errare, e quello di chi ne delimita i confini. Si tratta di un'idea ormai ampiamente condivisa in ambito didattico: come osserva Colombo (2011: 43) “una classificazione eziologica cerca di raggruppare gli errori a seconda delle cause che li producono, si pone quindi dalla parte dello scrivente e dei suoi processi mentali” mentre “una classificazione formale raggruppa gli errori dal punto di vista linguistico, ponendosi piuttosto dalla parte del testo”.

La distinzione fra gli errori “dal punto di vista dello studente” e “dal punto di vista dell'insegnante” è rilevante. Gli errori “non esistono” come realtà totalmente consapevole per lo studente nel momento in cui li ha commessi (a volte è forse da costui percepibile l'eco di un'incertezza o la fragilità di un tentativo), ma esistono nell'osservazione dell'insegnante, come scarto fra ciò che lo studente fa e quello che l'insegnante si attende come comportamento ideale. Per dotare un errore di consapevolezza anche per lo studente, e quindi poter poi intervenire sulla sua correzione e valutazione di apprendimento, la correzione condivisa pare la procedura più efficace.

La procedura di correzione di un errore potrebbe quindi richiedere alcuni passaggi. Il primo di questi è la rilevazione dell'errore da parte del docente, che potrebbe procedere, già in questa sede, alla proposta di classificazione dell'errore adottando un simbolo specifico per ogni tipologia di errore, ortografia, sintassi, lessico, punteggiatura.

A questo punto si potrebbe chiedere allo scrivente di ricondurre ad una tassonomia più specifica quanto è stato segnalato. Il docente quindi potrebbe elaborare una griglia di analisi così composta:

- una casella per la forma errata da affiancare a una casella per la forma corretta (target);
- una casella per la descrizione dell'errore basata su una tassonomia di manifestazioni superficiali, ad esempio: scambio, aggiunta o sottrazione di grafemi, universioni o separazioni inappropriate, accenti e apostrofi.

Il docente potrebbe poi partire dalla casella descrittiva basata sulla tassonomia per ragionare con l'allievo (o ancora meglio, tutta la classe) sulle cause che hanno generato la forma errata. Esse potranno, a loro volta, essere di diverso tipo: ci sono diversi motivi per cui un grafema può essere omesso, ad esempio nel caso di *camica* per *camicia*, o per *ho*, *sbaliare* per *sbagliare*. Prima di richiedere allo studente un'analisi di questo tipo è

necessario fare esempi ed esercizi per apprendere la tassonomia stessa e in tale occasione riflettere sul perché questi errori accadano.

Una volta verificata l'esattezza della tabella, si potrebbe passare alla fase successiva che consiste nella vera e propria risoluzione dell'errore. Una parola, un segno d'interpunzione o un'intera frase potrebbero essere riscritti dallo studente sotto la guida dell'insegnante a margine del testo in forma coerente con la norma. Il lavoro svolto con la tabella permette quindi di aggiungere, alla fase in cui l'errore è stato rilevato, una fase in cui l'errore viene classificato, compreso e poi risolto con l'indicazione della forma corretta.

Un'attività di questo tipo richiede molto tempo per essere preparata e svolta, tuttavia parte della riflessione che essa attiva rientra pienamente nello svolgimento di attività previste dalla grammatica curricolare e quindi può rientrare a pieno titolo nella relativa programmazione didattica. Ritornando a quanto esposto nel paragrafo 3, la correzione condivisa può essere un ulteriore passaggio che permette di capire meglio quale percorso abbia compiuto lo studente prima di giungere alla forma indicata dalla norma. Si consideri ad esempio il seguente testo scritto da una studentessa frequentante il primo anno della scuola secondaria di primo grado:

- (10) *Era una notte calda quando io mi addormentai in tenda verso le 22:00, con la consapevolezza che mio papà e mio fratello erano andati al bar del campeggio in cui eravamo situati.*

Quando nel mezzo della notte non saprei dire con precisione l'ora, sentii un forte rumore; un rumore di foglie, vento e sbattere della tenda assieme.

Mi alzai di colpo e guardai alla mia destra, dove in teoria ci sarebbe dovuto essere mio fratello. Lui non c'era allora chiamai mio papà che era nello scompartimento di tenda successivo, ma dopo averlo chiamato più volte senza risposta decisi di alzarmi e di andare a vedere cosa stesse succedendo. Non vidi niente a parte molta sabbia che volava e spaventata decisi di tornare nel mio scompartimento di tenda. Per la paura che la tenda violasse via la tenei forte. Quando ad un certo punto mi venne una voglia girante di andare in bagno, ma la tenei. Dopo un po' arrivò mio padre dal bar dicendomi che ero nel centro di una tromba d'aria e di aspettare che finisse. Ci volle un po' prima che la tromba d'aria finì. Alla sua fine io non c'era facevo più DOVEVO ANDARE IN BAGNO!

Allora lo dissi a mio padre che mi accompagnò fino al bagno del campeggio. Dopo di che arrivò anche mio fratello e tutti assieme andammo a dormire. La mattina seguente girando per il campeggio mi accorsi che molte tende non c'erano più e che molti alberi erano sradicati, anche se la nostra tenda non aveva subito danni si vede che io pesavo talmente tanto che l'avevo tenuta ferma.

Da questa esperienza ho imparato prima di tutto che devo fare sempre la pipì prima di andare a dormire e per seconda cosa ma non meno importante, che anche quando credi che a sfortuna si tutta su di te ci sono persone che stanno peggio

Lavorando su un testo come questo, potrebbe essere importante sottolineare che ci sono diversi livelli di problemi, che non è opportuno trattare tutti insieme, durante la stessa lezione. Gli ambiti su cui cominciare a lavorare, in modo separato, potrebbero essere:

Formazione e uso del passato remoto

Si osservi che nel testo l'uso del passato remoto è corretto, a parte il caso di *prima che la tromba finì* in cui è la scelta del modo verbale ad essere sbagliata anche se in generale la

consecutio temporum (anche per condizionale e congiuntivo) è abbastanza ben padroneggiata; il problema è piuttosto, ma solo in qualche caso, la costruzione delle forme: ad esempio, *tenei*. Questa potrebbe essere un'ottima occasione per attivare un ripasso generale proprio sulla costruzione delle forme del passato remoto e magari attivare una riflessione sul perché sia così difficile da costruire rispetto ad altri modi verbali;

Forma è vs. forma e

Si tratta di una confusione che accade spesso: di solito è l'accento che indica il verbo al posto di una congiunzione ad essere dimenticato; in altri casi, come qui, accade, forse per ipercorrettismo, che l'accento venga inserito dove non serve;

L'uso della punteggiatura

Tutta la classe potrebbe avviare una riflessione sull'uso della punteggiatura nella sua funzione sintattico-semanticamente.

L'instabilità dei registri

Consapevolezza, eravamo situati, subire danni, verso le 22 e lo stesso uso del passato remoto sembrano voler emulare uno stile narrativo sostenuto, con sconfinamenti nell'italiano burocratico. Forse la scelta di un registro come questo pare più appropriata allo studente per un testo da presentare in un contesto formale. Tuttavia colloquialismi come *in teoria, girare per il campeggio, tenerla* (la pipì) e l'uso del maiuscolo, per sottolineare l'impellenza di una necessità, non sono coerenti con un registro formale. Questa oscillazione forse è dovuta al fatto che allo scrivente è stato chiesto di raccontare, in un contesto formale come quello scolastico, un'esperienza molto personale e con un notevole carico emotivo. Difficile quindi per lo studente mantenersi saldo su un unico registro. Anche in questo caso si potrebbe cogliere l'occasione per approfondire quali difficoltà possano sorgere nella scelta del registro adeguato da adottare quando si scrive in differenti contesti su differenti argomenti;

Mancanza di rilettura

Molti errori sembrano dovuti al fatto che l'allievo abbia consegnato il testo senza rileggerlo dopo averlo scritto. Può essere una buona pratica da proporre in classe la rilettura da parte di un compagno prima di consegnare il testo. Il compagno sottolinea a matita i punti in cui gli sembra di trovare degli errori e l'autore del testo, se lo ritiene opportuno, compie le dovute correzioni prima della consegna.

6. Le conseguenze dell'errore

Nel discutere che cosa può essere considerato come parte del "sapere una lingua", e nel constatare che sul tema di quali siano le competenze "minime" e "massime" di un parlante di una lingua specifica la linguistica applicata è ben lontana dall'aver dei punti di partenza latamente condivisi, Berruto (2012) richiama l'attenzione su un tema importante anche per il trattamento degli errori: la varietà dei motivi per cui è più e meno importante un certo aspetto della competenza linguistica. Berruto mette in evidenza diversi ordini di motivi: cognitivi e culturali, legati cioè all'uso della lingua per la manipolazione e la trasmissione di contenuti concettuali; sociali, legati cioè all'uso della lingua come forma di

rappresentazione sociale di sé e degli altri. Non ogni aspetto della competenza linguistica ha uguale rilevanza per ogni ordine di motivi, e analogamente gli errori hanno conseguenze diverse per diversi ordini di motivi: gli errori di ortografia (uso appropriato di maiuscole, accenti, apostrofi; corretta grafia di digrammi e trigrammi...), ad esempio, hanno solitamente scarso rilievo sul piano cognitivo-culturale, ma hanno sempre un elevato impatto sul piano sociale. Se spesso si interviene sugli errori degli alunni invocando motivi di tipo cognitivo e contenutistico (“gli errori impediscono di trasmettere correttamente il messaggio”), molto più spesso sono invece motivi di tipo sociale a rendere vistoso o “grave” un errore: il messaggio passerà lo stesso, ma a chi lo ha prodotto verrà attribuita una scarsa capacità di usare la lingua, venendo quindi qualificato come persona poco colta e scarsamente scolarizzata. La sanzione sociale non è una motivazione per un intervento correttivo su un errore meno rilevante di quanto non sia il fraintendimento, ma ha diversi ambiti di applicazione: si tratta infatti di un intervento che riguarda piuttosto le violazioni della norma e non del sistema.

Così, gli errori che, come cause, possono rientrare nella categoria delle “sviste”, i così detti errori di distrazione che tendono a non ripetersi e richiedono soltanto una rilettura attenta di quanto scritto, possono essere socialmente sanzionati, qualificando il testo come sciatto e poco curato, e su questo aspetto si può richiamare l’attenzione dell’allievo: dare una rilettura al testo è come guardarsi allo specchio prima di uscire.

Tuttavia, ci sono errori più significativi sia per le loro cause che per le loro conseguenze, e richiedono una maggiore riflessione. In questo senso è bene, nell’economia dell’energia e di capacità di attenzione dello studente, non dare troppo peso ai primi e dedicarsi attentamente ai secondi. In tal senso forse vale il motto “meno è meglio”, poiché dagli errori si può imparare poco alla volta, senza esagerare ad inserire nuovi contenuti nello stesso momento.

In definitiva, individuare le conseguenze di un errore, così come le sue cause, è rilevante per stabilire come e quando intervenire. Quando il docente si pone come primo obiettivo non quello di sanzionare l’errore, ma quello di far capire perché sia stato commesso e quali possano essere le sue conseguenze, dota lo studente di una consapevolezza che non è più legata tanto all’aspetto sanzionatorio ma al fenomeno nella sua complessità. Questo approccio da parte del docente può favorire nello studente una maggiore motivazione all’apprendimento e una maggiore disponibilità al riconoscimento dei propri errori. Si tratta di un modo di procedere che è più orientato ad un atteggiamento sperimentale verso la lingua, quello che Lucia Lumbelli (1981) e Maria G. Lo Duca (1997), in ambiti diversi della riflessione sulla lingua, raccomandano per stimolare il piacere della scoperta delle regole che sottostanno alla norma. Il docente potrà quindi porsi come obiettivo professionale quello di trasmettere ai propri studenti che la prima conseguenza di un errore è spesso una scoperta.

Bibliografia

- Bazzanella, Carla. 1994. *Le facce del parlare. Un approccio pragmatico all’italiano parlato*. Firenze: La Nuova Italia.
- Berruto, Gaetano. 1987. *Sociolinguistica dell’italiano contemporaneo*. Roma: La Nuova Italia Scientifica.

- Berruto, Gaetano. 1993. Varietà diamesiche, diastratiche, diafasiche. In Sobrero, Alberto (a cura di), *Introduzione all'italiano contemporaneo. Vol. 2°: La variazione e gli usi*, 37-92. Roma-Bari: Laterza.
- Berruto, Gaetano. 2012. Che cosa vuol dire 'sapere una lingua'? Dai fonemi alle espressioni idiomatiche. In Bernini, Giuliano & Lavinio, Cristina & Valentini, Ada & Voghera, Miriam (a cura di), *Competenze e formazione linguistiche. In memoria di Monica Berretta*, 27-53. Perugia: Guerra.
- Cardona, Giorgio Raimondo. 1988. *Dizionario di linguistica*. Roma: Armando.
- Cignetti, Luca & Demartini, Simona. 2016. *L'ortografia*. Roma, Carocci.
- Colombo, Adriano. 2015. *A me mi. Dubbi, errori, correzioni nell'italiano scritto*. Milano: Franco Angeli.
- D'Achille, Paolo. 2010. Italiano popolare. In *Enciclopedia dell'Italiano Treccani*, http://www.treccani.it/enciclopedia/italiano-popolare_%28Enciclopedia-dell%27Italiano%29/.
- De Mauro, Tullio. 1970. Per lo studio dell'italiano popolare unitario. In De Rossi, Alberto (a cura di), *Lettere da una tarantata*, 43-75. Bari: De Donato.
- Fasulo, Alessandra & Pontecorvo, Clotilde. 1999. *Come si dice? Linguaggio e apprendimento in famiglia e a scuola*. Roma: Carocci.
- Franciolini, Letizia. 2011. *Dalla correzione alla condivisione dell'errore: riflessione analitica sull'interazione verbale in classe attorno all'errore*. Università degli studi di Milano Bicocca: Tesi di dottorato non pubblicata.
- Grandi, Nicola (a cura di). 2015. *La grammatica e l'errore. Le lingue naturali tra regole, loro violazioni ed eccezioni*, Bologna: Bononia University Press.
- Grassi, Roberta & Mangiarini, Claudia. 2010. Feedback 'implicito' e feedback 'esplicito' a confronto: uno studio sperimentale per l'italiano L2. In Grassi, Roberta & Piantoni, Monica & Ghezzi, Chiara (a cura di), *Interazione didattica e apprendimento linguistico*, 129-162. Perugia: Guerra.
- Grassi, Roberta (a cura di). 2017. *Il trattamento dell'errore nella classe di italiano L2*. Firenze: Cesati.
- Lo Duca, Maria Giuseppina. 1997. *Esperimenti grammaticali*, Firenze: La Nuova Italia.
- Lumbelli, Lucia. 1981. *Quando insegnare è capire e far capire*. Milano: Emme Edizioni.
- Nesci, Teresa & Cattana, Anna. 2004. *Analizzare e correggere gli errori*. Perugia: Guerra.
- Rodari, Gianni. 1964. *Il libro degli errori*. Torino: Einaudi.
- Serafini, Maria Teresa. 1985. *Come si fa un tema in classe*. Bologna: Bompiani.
- Serianni, Luca. 1989. *Grammatica italiana. Italiano comune e lingua letteraria*. Torino: UTET Libreria.
- Solarino, Rosaria. 2009. *Imparare dagli errori*. Napoli: Tecnodid.

ParlaTO: la variazione (e il repertorio) in un corpus di parlato

Massimo Cerruti

Università di Torino

massimosimone.cerruti@unito.it

Abstract

ParlaTO è un corpus di parlato spontaneo raccolto a Torino, di recente allestimento. Insieme al corpus KIP, a cui è per molti aspetti complementare, confluisce nel più ampio KIParla. È costituito da interviste semi-strutturate a informatori con caratteristiche sociodemografiche differenti e comprende prevalentemente produzioni in italiano, pur includendo anche brani in dialetto e lingue d'immigrazione. In questo contributo se ne illustrano le caratteristiche essenziali e si discutono i risultati dei primi studi condotti sui materiali di ParlaTO e KIP, collocandoli nel quadro di alcune tendenze in atto nell'italiano: la crescente circolazione di tratti regionali oltre i confini d'origine, il progressivo ridursi delle differenze linguistiche fra parlanti con collocazione sociale diversa e la presenza sempre più regolare di fenomeni propri del parlato colloquiale anche nell'uso di norma.

1. Introduzione

ParlaTO (www.corpusparlato.com) è un corpus di parlato spontaneo raccolto a Torino, che viene ad aggiungersi ai (pochi) corpora di parlato esistenti per l'italiano.¹ Il corpus consente di indagare l'italiano parlato in città, per come questo si presenta in gruppi sociali differenti e in parlanti con provenienza geografica diversa, talvolta in alternanza con dialetti e lingue d'immigrazione. Un'introduzione al corpus è fornita in Cerruti & Ballarè (2020), a cui si rimanda per la discussione dei criteri di allestimento, della metodologia di raccolta e trattamento dei dati e delle modalità di consultazione.

In questo articolo si presentano le caratteristiche essenziali del corpus, commentando alcuni brani esemplificativi delle varietà d'italiano e delle interazioni bilingui presenti in esso (cfr. paragrafo 2); si illustrano i rapporti fra ParlaTO e KIP, due corpora per molti versi complementari che confluiscono nel KIParla (cfr. paragrafo 3); e si prendono in

¹ Fra i quali, mette conto ricordare per lo meno LIP (badip.uni-graz.at/it, www.parlaritaliano.it/index.php/it/volip), CLIPS (www.clips.unina.it), C-ORAL-ROM (Cresti & Moneglia 2005) e KIP (cfr. paragrafo 3). Per una rassegna recente, comprensiva anche di corpora in via di allestimento, si veda Gorla & Ciccolone (2020).

esame i risultati delle prime ricerche condotte su questi due corpora, discutendoli alla luce di alcune tendenze attuali nella variazione sociolinguistica dell'italiano (cfr. paragrafo 4). Si conclude con qualche cenno agli sviluppi futuri del ParlaTO (cfr. paragrafo 5).

2. ParlaTO

2.1 Il corpus

ParlaTO è composto essenzialmente da una serie di conversazioni registrate a Torino fra il 2018 e il 2020 per mezzo di interviste semi-strutturate. Le interazioni hanno coinvolto più di un centinaio di parlanti d'età diversa (giovani, adulti e anziani), alcuni di origine italiana (piemontese e non), altri di origine straniera, e con gradi d'istruzione e tipi di occupazione differenti. Le interviste si sono concentrate su esperienze di vita personale in città (studio, lavoro, attività nel tempo libero o in pensione, ricordi del passato, ecc.) e hanno visto all'opera più intervistatori, ciascuno dei quali quasi sempre appartenente alla rete sociale o familiare dell'intervistato. A una stessa intervista, inoltre, hanno spesso preso parte più intervistati; il che ha dato luogo di frequente a vere e proprie discussioni collettive. Condizioni come queste hanno favorito modalità di interazione tipiche del comportamento *in-group*, caratterizzate dall'uso di varietà spontanee e, in casi specifici, dall'alternanza di lingue diverse.

Il corpus è costituito da due sezioni, contenenti l'una le conversazioni con parlanti di origine italiana e l'altra le conversazioni con parlanti di origine straniera. Al momento è interrogabile online soltanto la prima sezione, che ammonta a circa 50 ore di parlato e comprende produzioni linguistiche in italiano, piemontese e altri dialetti (in particolar modo di area meridionale). L'uso dell'italiano è largamente prevalente in tutte le conversazioni, ed esclusivo nella maggior parte di esse, mentre il piemontese e gli altri dialetti compaiono soltanto occasionalmente, secondo la nota fenomenologia del discorso bilingue. La seconda sezione, attualmente in preparazione, avrà anch'essa una dimensione stimabile in circa 50 ore di parlato e sarà incentrata sull'italiano di nativi e non nativi e sulle lingue di recente immigrazione. Le Tabelle 1 e 2 riportano alcune informazioni essenziali riferite alla prima sezione del corpus.

INFORMATORI	88 parlanti
RACCOLTA DATI	Torino, 2018-2020
METODO	Interviste semi-strutturate (individuali e di gruppo)
LINGUE	Italiano, dialetto piemontese, altri dialetti
METADATI	Parlanti: classe d'età, sesso, regione di nascita, titolo di studio, occupazione, lingua materna, competenza (attiva o passiva) di dialetto/i italo-romanzo/i, competenza (attiva o passiva) di altre lingue, città e quartiere di residenza, città di nascita del padre e della madre. Interazioni: numero di partecipanti, lingue impiegate
ORE DI PARLATO	Giovani ($18 \leq x \leq 30$): 17:33:20
	Adulti ($30 < x \leq 60$): 14:49:53
	Anziani ($60 < x \leq 89$): 16:15:31
DIMENSIONI	552.461 tokens
ACCESSO	www.corpusparlato.com

Tabella 1. Prima sezione del corpus ParlaTO.

Il corpus è dotato di una serie di metadati relativi alle caratteristiche sociodemografiche degli informatori e ad aspetti specifici dell'interazione (cfr. Tabella 1). Alcuni metadati, quali ad esempio la classe d'età, il sesso, la regione di nascita, il titolo di studio e l'occupazione del parlante, sono usabili come filtri di ricerca per l'interrogazione (cfr. Tabella 2);² altri, quali il quartiere di residenza dell'informatore o la città di nascita del padre e della madre, sono accessibili soltanto come informazioni supplementari (contenute in tabelle Excel scaricabili dalle pagine del sito web). Oltre a questi, la seconda sezione del corpus avrà alcuni metadati riguardanti più direttamente i parlanti di origine straniera, come il tempo di permanenza e gli anni di studio in Italia.

PARLANTI	Classe d'età:	16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, 81-85, 85-90
	Sesso:	M, F
	Regione di nascita:	Abruzzo, Basilicata, Friuli-Venezia Giulia, Lazio, Lombardia, Piemonte, Puglia, Sardegna, Sicilia, Trentino-Alto Adige, Veneto
	Titolo di studio:	<i>elem</i> (diploma di scuola elementare), <i>medie</i> (licenza media), <i>it</i> (diploma di istituto tecnico o professionale), <i>lic</i> (diploma di liceo), <i>laurea</i> (laurea triennale, magistrale e a ciclo unico), <i>phd</i> (dottorato di ricerca)
	Occupazione:	<i>artig</i> (artigiani, operai specializzati e agricoltori), <i>comm</i> (professioni qualificate nelle attività commerciali e nei servizi), <i>disocc</i> (disoccupati), <i>impr</i> (legislatori, imprenditori e alta dirigenza), <i>intell</i> (professioni intellettuali, scientifiche e di elevata specializzazione), <i>nonq</i> (professioni non qualificate), <i>oper</i> (conduttori di impianti, operai di macchinari fissi e mobili e conducenti di veicoli), <i>pens</i> (pensionati), <i>stud</i> (studenti)
INTERAZIONI	Partecipanti:	2, 3, 4, 5, 6
	Lingue:	italiano, italiano e dialetto

Tabella 2. Prima sezione del corpus ParlaTO: metadati usabili come filtri di ricerca.

Allo stato attuale, l'interrogazione del corpus può avvenire soltanto per forme; è dunque possibile estrarre singole parole, oppure insieme e sequenze di parole, in formato KWIC e/o ottenere liste di frequenza (PoS-tagging e lemmatizzazione sono tuttavia in fase di sviluppo; cfr. Bosco et al. 2020). A mo' d'esempio, si riporta in Figura 1 il risultato di una ricerca semplice. Ogni occorrenza della forma richiesta (in questo caso, *tipo*) è inserita in una riga di contesto ed è associata al codice della conversazione in cui essa compare. Cliccando sul codice della conversazione (es. *PTA013*), appare un riquadro che fornisce una serie di informazioni; tra le altre: i link alla trascrizione dell'intera conversazione, in formato ortografico (*doc.full_conversation*) e nel 'sistema Jefferson'

² Per le denominazioni dei tipi di occupazione si è fatto riferimento alla classificazione delle unità professionali proposta dall'Istituto nazionale di statistica (<http://professioni.istat.it>), aggiungendovi tre categorie (*disoccupati*, *pensionati*, *studenti*).

(*doc.full_conversation_Jefferson*);³ il link alla traccia audio (*annotation.audio_file*) della porzione di dialogo individuata dall'occorrenza; e i metadati usabili come filtri di ricerca, che danno conto delle caratteristiche sociodemografiche del parlante e di alcuni aspetti dell'interazione (cfr. Tabella 2).

The screenshot shows the NoSketch Engine search interface. The search term is 'tipo', with 467 results (1,004.69 per million). The results are displayed in a list format, with each entry showing a document ID (e.g., PTD020, PTA013) and a snippet of text containing the search term. A metadata panel is visible at the bottom, listing various fields such as 'Token number', 'Document number', 'doc.doc_number', 'doc.full_conversation', 'doc.full_conversation_jefferson', 'conversation.type', 'conversation.participants_number', 'conversation.languages', 'annotation.audio_file', 'annotation.participant_occupation', 'annotation.participant_sex', 'annotation.files_in_which_participant_appears', 'annotation.participant_birth_region', 'annotation.participant_age_range', 'annotation.participant_degree', and 'annotation.speaker_id'.

Figura 1. Un esempio di ricerca semplice.

2.2 Varietà d'italiano e discorso bilingue

Come già accennato al paragrafo 2.1, le condizioni dell'intervista hanno favorito l'uso di varietà spontanee, caratterizzate anche da tratti sub-standard. Si veda ad esempio il brano (1),⁴ in cui si registrano diversi fenomeni del parlato dialogico; alcuni tipici dell'italiano colloquiale, altri più frequenti nei registri informali. Oltre all'emergere di disfluenze testuali, come autocorrezioni (es. *nella tribu- nel settore ospiti*), cambi di pianificazione (es. *c'è da dire che il terz- eh Milano*) e anacoluti (es. *il Delle Alpi non potevi vedere le partite*), si può facilmente riscontrare la presenza di segnali discorsivi e interiezioni (es.

³ Le convenzioni usate per la trascrizione conversazionale rappresentano in realtà una versione semplificata del sistema proposto da Jefferson (2004); cfr. Cerruti & Ballarè (2020), Ballarè & Mauri (2020).

⁴ Tutti i brani nell'articolo sono in trascrizione conversazionale. Si segnalano qui alcune convenzioni: [*ciao*] sovrapposizione tra parlanti, (*ciao*) sequenza di difficile comprensione, *cia-* parola interrotta, >*ciao*< pronuncia veloce, ((*ride*)) comportamento non verbale, (.) pausa breve, : suono prolungato, *x* sequenza incomprensibile. A ciascun partecipante è assegnato un codice alfanumerico (es. *TOI002*, *TOR001*), il terzo carattere del quale consente di distinguere gli informatori (*I*) dai raccoglitori (*R*).

cioè, vabbò), elementi deittici, anche in impieghi allusivi (es. *sei lì*), colloquialismi espressivi (es. *te lo giuro*), diftemismi (es. *cazzo, minchia*), usi non nominali di *tipo* (es. *è fatto tipo a teatro*), costruzioni negative marcate (es. *non è che si allarga*), dislocazioni a sinistra (es. *il gol non l'ho visto*) e a destra (es. *io le ho viste un paio di partite*), ecc. Alcuni di questi tratti tendono in realtà a perdere, o hanno già perso, l'originario valore sub-standard, essendo ormai in via di accoglimento, o entrati a tutti gli effetti, nell'italiano neo-standard (cfr. ad es. Berruto 2012: 67–126; 2017):

- (1) TOI002: no il delle alpi non potevi vedere le partite
 TOI002: io le ho viste un paio di partite quando ero bambino minchia
 TOI002: cazzo era (.) no eri lontanissimo [perche' poi fra l'altro c'era:]
 TOR001: [c'era la (.) pista]
 TOI002: [c'era la pista ciclabile] e la ram- minchia [x]
 TOI003: [e' un po' co]me l'olimpico eh io quando sono andato all'olim[pico:]
 TOI002: [vai a vedere adesso]
 TOR001: [mh mh]
 TOI002: [minchia adesso:]
 TOI002: [adesso e' (sfruttato)]
 TOI003: [e' un po' come il nostro terzo] anello
 TOI003: pero' c'e' da dire che il terz- eh milano
 TOI003: e' fatto pro(i)o b- eh e' fatto tipo a teatro [non e' che si allarga sei li']
 TOI002: [vabbo' io l'ho vista x]
 TOR001: [si' non s'allarga (almeno)]
 TOI003: [sei altis]simo eh vedi minchia [(vedi) delle]
 TOI002: [io l'ho vis]to in terza e non la vedi bene come (.) meglio del delle alpi al secon[do]
 TOI003: [no]: meglio del delle [alpi sicuro]
 TOR001: [davvero? cosi' male::]
 TOI002: [eh te lo giuro:] min[chia (.) io]
 TOI003: [il delle alpi era male]
 TOI002: [sono venuto a inter napoli]
 TOI002: no un toro napoli al delle alpi da bambino in serie b
 TOI002: e ero in nella tribu- nel: settore ospiti
 TOI002: io il gol non l'ho visto del e::h del toro (quello [che avevano fatto])
 TOI003: [anche all'olimpico vedi m]ale
 TOI002: no no ma il delle alp- c(io)e' lo juventus stadium no
 (ParlaTO, PTA003)

Nel corpus compaiono inoltre più varietà sociali e (socio-)geografiche. Nel brano (2),⁵ ad esempio, sono presenti tratti diagnostici dell'italiano popolare (cfr. D'Achille 2010 e Berruto 2014: 287), come la sovraestensione di forme dell'articolo determinativo (*i stracci*) e gli scambi di ausiliare (*aveva cambiato*, con *cambiare* inaccusativo), e tratti sub-standard particolarmente frequenti in parlanti semicolti, come le reduplicazioni pronominali (*dovesse capitarmi a me*); alcuni dei quali diffusi soprattutto in varietà meridionali (come la forma plurale *melenzane*). A questi si aggiunge un certo numero di

⁵ L'intervistato, TOI017, è un parlante anziano (fascia d'età: 85-90 anni) di origine pugliese con sola licenza elementare.

fenomeni in comune con l'italiano colloquiale, come la sovraestensione di *le* per *gli* (*le ho detto*), il dativo etico col pronome di prima persona (*lui m'ha abbassato la testa*) e il rafforzamento dei dimostrativi (*questo qua*); fenomeni che tendono a penetrare, o sono già penetrati, nell'italiano neo-standard (cfr. sopra). Nel brano (3), invece, si notano almeno tre tratti morfosintattici dell'italiano del Piemonte, ben appoggiati al dialetto di sostrato: uno più propriamente locale, il focalizzatore esclusivo e scalare *solo più* (cfr. Regis 2006), e due attestati anche in altre varietà regionali: la costruzione *aver male* + SN (*una ha male la gamba*) e il mancato accordo nel costrutto locativo-esistenziale-presentativo (*c'è tre o quattro campi, ce n'è solo più due che giocano*; cfr. Berruto & Cerruti 2015):

- (2) TOI017: questo qua m'ha risposto male
 TOR004: mh
 TOI017: allora io l'ho preso per i stracci e le ho detto, per il braccio e le ho detto dovesse capitarmi a me
 TOI017: io non, no qua dentro, fuori
 TOI017: ti faccio gli occhi come melanzane, del colore della melanzana
 TOI017: capito?
 TOR004: si' si'
 TOI017: qua non ti faccio niente ma quando ti prendo fuori
 TOI017: [allora lui m'ha abbassato] la testa e se n'e' andato e non m'ha risposto piu' pero' (aveva) cambiato da quella sera
 TOR004: [te le do di santa ragione]
 (ParlaTO, PTB002)
- (3) TOI024: quando possiamo (.) giochiamo tra di noi [tanto c'e'] (.) tre o quattro campi
 TOI024: e il posto c'e'
 TOI024: pero' purtroppo date le eta' eh non stan tutti bene come me
 TOR005: eh cer[to]
 TOI024: [ma]gari una ha male la gamba l'altra non sta bene e sono mancate del-
 due tre amiche sono manca[te]
 TOR005: [mh mh]
 TOI024: e allora adesso:
 TOI024: ce n'e' solo piu' due che giocano
 (ParlaTO, PTB008)

In certi casi, e il più delle volte *a latere* dell'italiano, sono impiegati anche il piemontese e altri dialetti italo-romanzi. Si veda ad esempio il brano (4), in cui intervengono alcune commutazioni di codice, inter- e intra-frasali, fra italiano e piemontese. L'occorrenza di porzioni in dialetto, e più in generale in lingue diverse dall'italiano – quando questi codici appartengano al repertorio linguistico del parlante –, è segnalata da un 'cancellotto' (#) posto all'inizio dell'unità intonativa. Per la trascrizione delle parti in dialetto si è fatto ricorso a un'ortografia di servizio, basata di fatto sulla grafia dell'italiano:

- (4) TOI118: m- mio padre (.) era nato in francia
 TOI118: #mi am ricordu nen perche': ((*non mi ricordo perché*))
 TOI118: #a l'e' mort mi l'avia ses ani ((*è morto che avevo sei anni*))

TOI118: e mamma
 TOI118: veneta di noventa vicentina
 TOR002: m:h m:h
 TOI118: #e::h cume sun surti da:: ((*come sono uscito da*))
 TOI118: #da le scole ((*dalle scuole*))
 TOI118: #truva' un parent (a l'a) dime ven cun mi ven cun mi:: ((*ho trovato un parente mi ha detto vieni con me vieni con me*))
 TOI118: #a fe el tule' l'idr[aulico] ((*a fare l'idraulico*))
 TOR002: [mh mh]
 TOI118: e cosi' ho iniziato
 TOI118: a far l'idraulico (.) nei cantieri: eh
 (ParlaTO, PTB026)

Le interazioni che confluiranno nella seconda sezione del corpus vedranno altresì la presenza di varietà d'italiano di parlanti di origine straniera, di prima o seconda generazione, e l'uso alternato di italiano e lingue di recente immigrazione. Alcuni esempi commentati si possono trovare in Cerruti & Ballarè (2020).

3. ParlaTO e KIP

Il corpus ParlaTO si inserisce nel più ampio KIParla (www.kiparla.it, cfr. Mauri et al. 2019), che ospita già un altro corpus linguistico di recente allestimento, denominato KIP. Quest'ultimo è costituito da circa 70 ore di parlato e comprende un insieme di interazioni raccolte a Bologna e Torino fra il 2016 e il 2019 in contesto universitario. Le interazioni hanno coinvolto studenti e professori provenienti da varie regioni d'Italia e hanno avuto luogo in situazioni comunicative di tipo diverso, differenti tra di loro principalmente per il tipo di rapporto fra i partecipanti (simmetrico vs. asimmetrico), e quindi per il grado di formalità: lezioni, esami orali, colloqui in orario di ricevimento, interviste semi-strutturate e conversazioni libere.⁶ Il corpus KIP dà conto dunque di una gamma di tipi di interazione più ampia rispetto al ParlaTO, ma considera un campione di parlanti più omogeneo da un punto di vista sociodemografico (cfr. Figura 2).

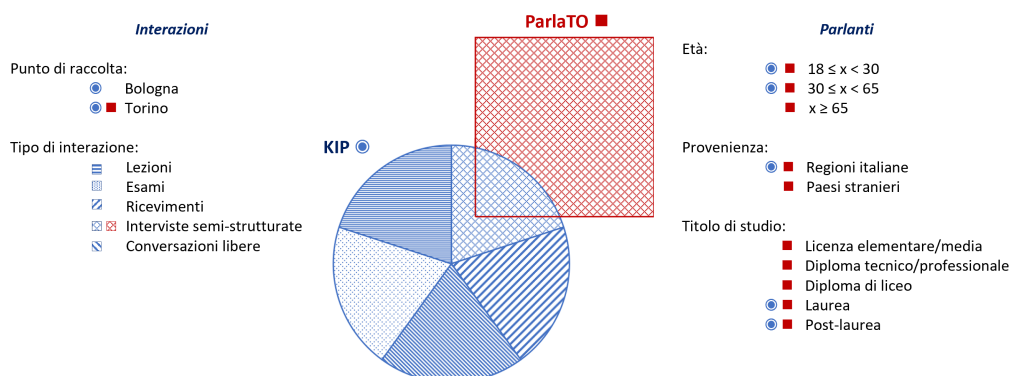


Figura 2. ParlaTO e KIP.

⁶ Le interviste semi-strutturate e le conversazioni libere vedono coinvolti soltanto studenti; il rapporto fra i partecipanti è perciò da intendersi come simmetrico.

ParlaTO e KIP hanno evidentemente obiettivi diversi. Il corpus ParlaTO offre in primo luogo l'opportunità di esplorare aspetti di diversificazione diastratica dell'italiano parlato, permettendo oltretutto di condurre ricerche mirate su categorie sociali specifiche (anziani dialettofoni, giovani non universitari, parlanti di origine straniera, ecc.). Il corpus KIP, invece, offre innanzitutto la possibilità di indagare fenomeni di variazione diafasica, e in modo particolare di registro, dell'italiano nel parlato di soggetti colti (studenti e professori universitari). Nondimeno, entrambi i corpora includono produzioni di parlanti con provenienza geografica diversa, e perciò consentono di osservare almeno alcuni elementi di differenziazione diatopica dell'italiano parlato.

ParlaTO e KIP condividono inoltre uno stesso nucleo di metadati (fra i quali, la classe d'età, il sesso, la provenienza geografica e il tipo di occupazione dei parlanti) e lo stesso sistema di raccolta, archiviazione e trascrizione dei dati (cfr. Mauri et al. 2019). Al momento sono consultabili soltanto separatamente, come due corpora distinti, ma, proprio in virtù di queste caratteristiche comuni, saranno presto interrogabili anche unitamente, come un solo corpus (cfr. Cerruti & Ballarè 2020). Si avrà quindi la possibilità di indagare insieme aspetti di differenziazione geografica (KIP e ParlaTO), sociale (ParlaTO) e, limitatamente a parlanti colti, situazionale (KIP) dell'italiano parlato.

4. Primi studi

Le prime ricerche condotte sui materiali dei due corpora si sono concentrate sui comportamenti linguistici di parlanti giovani (segnatamente, di età compresa fra 18 e 30 anni). Nei paragrafi successivi si discuteranno i risultati di questi studi nel quadro di alcuni cambiamenti in atto nella variazione sociolinguistica dell'italiano, prestando particolare attenzione al progressivo ridursi della gamma di variazione geografica e sociale e all'ampliamento della sfera d'influenza di tratti del parlato colloquiale e informale.

4.1 *Variazione diatopica*

La riduzione generale della differenziazione geografica dell'italiano è ascrivibile in primo luogo a due dinamiche di verso contrario (cfr. Berruto 2017; 2018): da un lato, la progressiva scomparsa dei tratti diatopicamente più marcati, dall'altro la circolazione inter-regionale di tratti di rango tradizionalmente regionale o locale. Questa seconda dinamica è foriera, in particolare, dello sviluppo di varietà regionali 'composite' (cfr. Canepari 1983; Cerruti 2011), che tendono ovvero a costituirsi non soltanto di tratti autoctoni, riconducibili, per lo meno originariamente, al contatto con i rispettivi dialetti di sostrato, ma anche di tratti alloctoni, provenienti da aree geografiche differenti e diffusisi oltre i confini originari essenzialmente per effetto della mobilità interna della popolazione.⁷

Alcune prime ricerche condotte su porzioni dei corpora ParlaTO e KIP offrono nuovi spunti per la comprensione di questi fenomeni. Fontanot (2019), ad esempio, fornisce un chiaro riscontro empirico della penetrazione di tratti centro-meridionali nell'italiano parlato da giovani torinesi. Il lavoro pone a confronto la distribuzione di una serie di fonemi consonantici in tre insiemi di dati raccolti a Torino: ca. 150 minuti di interviste semi-

⁷ Pur senza trascurare ragioni di altra natura, come ad esempio l'influenza dei modelli proposti dai mezzi di comunicazione (cfr. D'Achille 2011).

strutturate a dieci giovani lavoratori, alcuni con licenza media, altri con qualifica tecnica o professionale (ParlaTO); ca. 150 minuti di interviste semi-strutturate a dieci studenti universitari (KIP); ca. 45 minuti di lettura di frasi da parte di sedici studenti universitari (corpus CLIPS). A comparire con una certa regolarità sono soprattutto alcune varianti standard tradizionalmente alloctone per l'area settentrionale e non indotte dal modello scritto, come le geminate intrinseche (Bertinetto & Loporcaro 2005) [t:s], [d:z] e [ʎ:] (es. [dispera't:sjɔ:ne], [a'd:zjenda], [bi'ʎ:et:o]) e il raddoppiamento fonosintattico (es. [a'v:ɔlte], [ɔk:a'pi:to], [e'n:ɔj]). Tra le varianti alloctone non standard, gode di una certa diffusione soltanto la pronuncia intensa di [b] in posizione intervocalica (anche in confine di parola, es. ['kwɛl:a 'b:ɛl:a]).

Inoltre, il confronto tra le interviste del ParlaTO e quelle del KIP pone in luce come la presenza di pronunce alloctone sia significativamente più frequente nei giovani con licenza media o qualifica tecnica o professionale, e manifesti quindi un certo grado di marcatezza diastratica. Per quanto riguarda invece gli studenti universitari, il confronto tra le interviste del KIP e le letture di frasi del CLIPS mostra come le produzioni con pronunce alloctone siano più frequenti nel contesto meno sorvegliato, l'intervista, e rivelino dunque un certo grado di marcatezza diafasica.

Si può aggiungere che la diffusione di tratti alloctoni è fortemente variabile da parlante a parlante, e perciò ampiamente soggetta a variazione interindividuale; e gli stessi comportamenti individuali sono spesso incoerenti, si caratterizzano cioè per un continuo “bricolage” (cfr. Eckert 2008) di tratti alloctoni e tratti autoctoni. Si vedano ad esempio gli enunciati (5) e (6), nei quali due diversi allofoni di /ʎ:/ intervocalica, la variante alloctona [ʎ:] e la variante autoctona [lj], coesistono nelle produzioni di uno stesso parlante (in relazione oltretutto alla pronuncia di una stessa parola):

- (5) TOI005: quando lavori con la [fa'miʎ:a] e' b:ellissimo
 TOR001: mh mh
 TOI005: pero' devi essere una [fa'mi:lja] come la nostra
 (ParlaTO, PTA005; cfr. Fontanot 2019: 39)

- (6) TO041: e' completamente un altro mondo rispetto all'[i'ta:lja] a torino [...]
 TO041: le persone si comportano in modo totalmente diverso,
 TO041: perche' non e' come in [i'taʎ:a] o nel sud europa
 (KIP, TOD2005; cfr. Fontanot 2019: 57)

Altri studi vertono sulla presenza di fenomeni di italiano “composito” ai livelli del lessico e della morfosintassi. Cerruti (2018), ad esempio, si concentra sulle produzioni linguistiche di 43 studenti dell'Università di Bologna, provenienti complessivamente da dodici regioni italiane (corpus KIP), e prende in esame un certo numero di *types* morfosintattici ascrivibili ad aree geografiche specifiche, confrontandone l'occorrenza in tre diverse situazioni comunicative: conversazioni libere (ca. 4 ore), interviste semi-strutturate (ca. 4 ore), esami orali (ca. 4 ore). Particolarmente frequente è la presenza di tratti centro-meridionali nel parlato di giovani di origine settentrionale; tra i più ricorrenti: l'apocope di forme verbali, come *so'* e *anda'* in (7), e di altre parti del discorso, come *du' balle* in (8), e la sovraestensione di *stare* in funzione di copula, come *sta praticamente fuori corso* in (7). Negli esempi citati, a produrre questi tratti sono un parlante trentino (BO046) e un parlante emiliano (BO003):

- (7) BO046: questa in realta' sta praticamente fuori corso, una media bassissima, eppure l'hanno presa.
 BO047: il tirocinio, con l'universita', no?
 BO046: mh mh
 BO021: ma presa cioe' per meriti suoi, (.) non per?
 BO046: no no non e' raccomandata (ne') niente
 BO046: [ha fatto domanda e l'hanno presa.]
 BO047: [eh, che stavo dicendo?]
 BO046: oddio, e' dal giudice di pace, (.) io non ci andrei mai [...]
 BO046: cioe', a me, gia' non interessa il civile perche' so' cosucce
 BO046: se poi devo anda' dal giudice di pace
 (KIP, BOA3004)
- (8) BO003: a me piaceva ugoli:no
 BO002: [il conte ugolino]
 BO003: [quel]lo si':, [(.) cioe' proprio quelli catti:vi paolo e france]sca du balle
 BO002: [e' meraviglioso]
 (KIP, BOA3001)

La Tabella 3 riporta poi il numero di occorrenze complessive dei *types* esaminati in ciascuna delle tre situazioni comunicative, distinguendo fra *tokens* alloctoni e *tokens* autoctoni (il carattere alloctono o autoctono di ogni *token* è definito in relazione alla provenienza geografica dei singoli parlanti). Si può notare non soltanto come il rapporto fra *tokens* alloctoni e *tokens* autoctoni sia maggiore nelle conversazioni libere e abbia un valore nullo negli esami orali, coerentemente con la marcatezza diafasica di alcuni tratti regionali e, più nello specifico, di produzioni composite (cfr. sopra); ma soprattutto come vi sia una differenza di rapporto (oltretutto statisticamente significativa: $F=0,0054$, $p<0,01$) tra conversazioni libere (0,191) e interviste semi-strutturate (0,064), due contesti interazionali in cui predominano funzioni comunicative diverse: nelle conversazioni libere i parlanti sono intenti soprattutto a gestire la relazione con l'interlocutore e a manifestare o costruire la propria identità sociale, mentre nelle interviste semi-strutturate i partecipanti sono focalizzati principalmente sulla trasmissione di contenuti (cfr. Biber & Conrad 2009: 44–46, 64–69 e *passim*).

	conversazioni libere	interviste semi-strutturate	esami orali
tratti alloctoni (<i>tokens</i>)	64	7	0
tratti autoctoni (<i>tokens</i>)	335	109	3
rapporto	0,191	0,064	0

Tabella 3. Tratti regionali e contesti interazionali nel corpus KIP (cfr. Cerruti 2018: 21).

Il prevalere di tratti alloctoni nelle conversazioni libere (cfr. Tabella 3) potrebbe dunque stare a indicare come le produzioni “composite” siano, almeno in certi casi, correlate a funzioni comunicative specifiche, e in particolare ai vari modi di espressione linguistica dell'identità sociale. Studi precedenti hanno del resto mostrato come l'uso di tratti alloctoni, deprivato dei potenziali connotativi propri delle varietà regionali originarie, possa assumere per i membri di un certo gruppo sociale un valore simbolico di

affermazione identitaria; si veda, a questo proposito, la diffusione del raddoppiamento fonosintattico (Boario 2017, cfr. sopra) e del suffisso di origine meridionale *-uso* (es. *incazzuso, stiluso, punkettuso*; Cortinovis & Miola 2009) in gruppi giovanili torinesi.

4.2 Variazione diastratica

Per quanto riguarda la variazione diastratica, una delle principali dinamiche in atto consiste nella riduzione progressiva delle differenze linguistiche fra parlanti con collocazione sociale diversa (cfr. Cortelazzo 2001; Sobrero 2005; Berruto 2012: 157–159), anche in relazione alla tendenziale obsolescenza delle varietà popolari (cfr. Berruto 2005; 2014). Alcuni tratti dell'italiano popolare sono ormai largamente diffusi pure nel sub-standard di parlanti colti e, benché possano essere comunque più frequenti nelle produzioni di semicolti, hanno perso in buona parte il proprio carattere diagnostico. Altri invece, tuttora marcati come “bassi” in diastratia, sono usati sempre meno di generazione in generazione e sembrano destinati a scomparire. È certamente possibile riscontrare ancora la presenza di tratti esclusivi dell'italiano popolare, specialmente in parlanti anziani; si vedano ad esempio i fenomeni individuati in (2) e l'uso del condizionale nella protasi di un'ipotetica libera in (9). Ma le differenze maggiori tra parlanti colti e semicolti tendono a essere sempre meno categoriche e sempre più di carattere graduale:

- (9) TOI017: bravo m'ha detto, bravo ma
TOI017: e allora:
TOI017: se dovessi prendere una,
TOI017: una:: una camera, una stanza
TOI017: e dovesti riempirla di carbone,
TOI017: e difatti io ho detto, noi andiamo avanti con la legna li'
(ParlaTO, PTB002)

Si possono collocare in questo quadro, ad esempio, i risultati di un lavoro recente sui fenomeni di variazione osservabili nel settore delle costruzioni relative (Ballarè et al. 2019). Lo studio pone a confronto produzioni linguistiche di parlanti con collocazione sociale diversa, tra cui ca. 10 ore di interviste semi-strutturate a giovani lavoratori con licenza media o qualifica tecnico-professionale (ParlaTO) e ca. 10 ore di interviste semi-strutturate a studenti universitari (KIP). Le differenze riscontrabili fra i due gruppi di parlanti risiedono soprattutto nella diversa frequenza d'uso di costruzioni relative sub-standard (o neo-standard, cfr. paragrafo 4.3): pur presenti in entrambi i gruppi, queste costruzioni risultano più diffuse (con significatività statistica $p < 0,01$ o $p < 0,05$, cfr. Ballarè et al. 2019: 77–78) nei giovani con licenza media o qualifica tecnico-professionale, siano esse (seguendo la terminologia di Comrie & Kuteva 2013) del tipo *gap*, come *Rino Gaetano che ho comprato il CD* in (10), *pronoun retention*, come *uno che non gli piace il disordine* in (11), o *relative pronoun*, come *una generazione in cui il lavoro va bene* in (12):

- (10) TOI001: se no:: qualsiasi genere
TOI001: da:: rino gaetano che ho comprato il cd: l'altro ieri,
TOI001: anche lou reed ho comprato il cd l'altro ie-
TOI001: pero'::

TOI001: mina: e::h
(ParlaTO, PTA001)

- (11) TOI005: io non sopporto il disordine sono uno che non gli piace il disordine
TOI005: anzi son fatto cosi'
TOI005: c'e' il disordine che io non posso mettere (a posto)
TOI005: vado in palla, tientelo
(ParlaTO, PTA005)

- (12) TOI012: piu' che altro[: s-] siamo una generazione in cui
TOI013: [per me si']
TOI012: il lavoro::: va bene. punto.
TOI012: perche' o prendi quello o stai a casa a dormire
(ParlaTO, PTD021)

Lo studio mostra inoltre come uno stesso processo di grammaticalizzazione possa trovarsi a uno stadio più avanzato nella varietà di un certo gruppo sociale e meno in un'altra. È il caso, in particolare, della grammaticalizzazione di *dove*, che segue grosso modo un percorso di questo tipo: a partire dall'impiego come relativizzatore di locativi, *dove* può acquisire una funzione di *concern* o *aboutness* (cf. Ballarè & Micheli 2018; Ballarè & Inglese 2021), come in (13), e in una fase successiva può sviluppare la capacità di relativizzare il soggetto di frase (cfr. Brook 2011). Dai dati riportati in Ballarè et al. (2019: 85–86), l'ampliamento funzionale di *dove* raggiunge quest'ultimo stadio soltanto nel parlato di giovani con licenza media o qualifica tecnico-professionale, come in (14). Si può constatare, del resto, come le fasi più avanzate di vari processi di grammaticalizzazione compaiano proprio nelle varietà sociali dei parlanti meno istruiti (cfr., ad esempio, Berretta 1989; Ramat 2006; Berruto & Cerruti 2015; Amenta 2020):

- (13) BO013: i comunisti, sono gente tinta assa' ((legge))
BO013: tinto, vuol dire (.) cattivo eh?
BO013: vuol di- vorrebbe dire oscu::ri insomma eh?
BO013: pero', i pensieri tinti sono pensieri brutti
BO013: dove la gente tinta e' la gente cattiva e cosi' via
(KIP, BOD1001)

- (14) TOI012: comunque::: no bello e' una bella zona:: lati negativi:: di venaria:: [...]
TOI012: i ragazzini under sedici dove stan li' a far casino al giardinetto sotto casa
TOI012: ma quello e' normale: eh son ragazzi
(ParlaTO, PTD021)

Analogamente, è soltanto nelle produzioni di giovani non universitari che si riscontrano casi di accordo semantico nella relativizzazione del soggetto; come, ad esempio, *son di quelli che [...] non guarda molto cosa paga* in (15) ed *ero il pivello che facevo queste cose qui* in (16). La relativizzazione del soggetto con accordo semantico riflette bene certi aspetti della natura linguistica delle varietà sub-standard e, più in particolare, dell'italiano popolare: una varietà caratterizzata dal prevalere della semantica sulla sintassi, dalla denotazione di significati essenziali e dall'indicazione di referenti “concreti” (cfr. Berruto

1990). Sono fenomeni come questo, dunque, che, pure nel quadro di una generale riduzione della variazione diastratica, rivelano ancora la natura “popolare” di alcune varietà sociali:

- (15) TOI005: a me pia- e' sempre piaciuto andare a mangiar fuori
TOR001: mh mh mh
TOI005: son di quelli che quando va a mangiar fuori no:n non guarda:: (.) molto
[cosa paga]
TOR001: [mh mh mh]
(ParlaTO, PTA005)
- (16) TOI010: ho passato anche queste fasi qui nel senso che sono
TOI010: cresciuto (.) e:ro il pivello che facevo queste cose qui le facevo male
TOI010: adesso (.) mi- (.) credo di farle un po' meglio
(ParlaTO, PTD019)

4.3 Variazione diafasica

Per ciò che concerne la variazione diafasica, una delle maggiori tendenze in atto nell'italiano è l'ampliamento del raggio d'impiego dei registri bassi (Berruto 2012: 196); che porta con sé, fra l'altro, l'uso sempre più regolare di tratti del parlato colloquiale e informale anche in situazioni sorvegliate e il conseguente ingresso nella norma di tratti tradizionalmente sub-standard. Si tratta evidentemente della dinamica alla base del processo di ristandardizzazione dell'italiano, e quindi dello sviluppo di una varietà neo-standard (Berruto 2012; cfr. anche Cerruti, Crocco & Marzo 2017). Più in particolare, questo processo dà luogo alla coesistenza di due varietà standard: l'italiano neo-standard e l'italiano standard letterario; la prima più usata in certi domini, come ad esempio in vari ambiti della comunicazione mediatica, la seconda più presente in altri, come ad esempio nell'istruzione scolastica (cfr. anche Antonelli 2011: 51).

La coesistenza delle due varietà standard d'italiano è sottoposta a verifica empirica in uno studio recente (Cerruti & Vietti in stampa), condotto su una parte di materiali del corpus KIP. Lo studio prende in esame la distribuzione di 23 tratti lessicali e morfosintattici, propri alcuni dell'italiano neo-standard e altri dell'italiano standard letterario,⁸ in tre tipi diversi d'interazione: conversazioni libere (ca. 4 ore), interviste semi-strutturate (ca. 4 ore), esami orali (ca. 4 ore). Il lavoro si propone, sostanzialmente, di accertare quali di quei tratti tendano a co-occorrere in maniera statisticamente significativa in ciascuna delle tre situazioni comunicative indagate. La tecnica statistica usata è l'analisi delle componenti principali (PCA, *Principal Component Analysis*), che consente di rilevare l'esistenza di rapporti di correlazione tra variabili.

⁸ Fra i primi, ad esempio, si sono considerati alcuni ordini marcati dei costituenti (frasi scisse, dislocazioni a sinistra, dislocazioni a destra), costruzioni relative di tipo *gap* (cfr. paragrafo 4.2), le forme aferetiche del dimostrativo prossimale, *questo/quello* in funzione di dimostrativo neutro, *gli* come dativale sincretico, *ci* come locativo e l'uso dei pronomi personali *lui, lei, loro* con valore riflessivo. Fra i secondi, la costruzione passiva, le relative del tipo *relative pronoun* (cfr. paragrafo 4.2), le forme non aferetiche del dimostrativo prossimale, *ciò* come dimostrativo neutro, *le* come dativale femminile (singolare), *vi* come locativo e la presenza del pronome riflessivo *sé*.

L'analisi mostra che tutti i tratti dell'italiano neo-standard considerati (con una sola eccezione) correlano con il carattere informale dell'interazione, definito essenzialmente dal rapporto simmetrico fra i partecipanti. I fenomeni neo-standard tendono cioè a co-occorrere regolarmente nel corso di interviste semi-strutturate e conversazioni libere (cfr. paragrafo 3). Nel brano (17), ad esempio, tratto da una conversazione libera, si può osservare la co-occorrenza di una dislocazione a sinistra (*Grey's Anatomy la guardano tutti*), una relativa di tipo *gap* (*American Gigolò che mi pare ci sia sempre lui*) e un costrutto locativo con *ci* (*mi pare ci sia sempre lui*). I tratti neo-standard tendono poi a presentarsi insieme a fenomeni del parlato colloquiale e informale; quali ad esempio, in (17), colloquialismi espressivi (es. *tanta roba*), imprecazioni (es. *porca miseria*), e termini dal significato generale (come *fare*, es. *quello che ha fatto Derek*):

- (17) BO003: l'ultimo cavaliere
 BO002: non l'ho [mai visto]
 BO003: [mia mamma lo guar]da se:mpre li' c'e' lui (.) giovane. col capello lungo
 BO002: mi pare: di aver visto american gigolo' (.) che mi pare ci sia sem[pre lui]
 BO003: [si'] >porca miseria< (.) anche pretty woman
 BO002: si'. (.) >(proprio)< e'
 BO003: si'. li' lui si'
 BO002: no no era tanta roba [anche li']
 BO003: [si' si'] madonna [...]
 BO003: poi anche
 BO003: >non so< se lo conosci
 BO003: ehm (.) patrick dempsey: quello che ha fatto derek in gre[y's anatomy]
 anche lui
 BO002: [ah: si':]
 BO002: si' [piace-]
 BO003: [nei suoi momen]ti d'o:ro [era]
 BO002: [pia]ce a ta::n[ti (.)] si'
 BO003: [davvero]
 BO003: anche lui era bello. grey's anatomy tutti
 BO002: si'. grey's anatomy, la guardano tu:tti
 (KIP, BOA3001)

La stessa analisi mostra inoltre che tutti i tratti dell'italiano standard letterario (anche in questo caso con una sola eccezione) correlano invece con il carattere formale della situazione comunicativa, determinato fondamentalmente dalla relazione asimmetrica fra i parlanti. I tratti dello standard letterario tendono ovvero a co-occorrere regolarmente durante gli esami orali. In (18), ad esempio, si può notare la co-occorrenza di una frase passiva (*Anselmo fu cacciato dall'abbazia*), una costruzione con un pronome relativo (*re Desiderio [...] del quale era oppositore*) e un costrutto locativo con *vi* (*prima vi era [...] la raccolta di libri dall'esterno*). Fenomeni come questi tendono poi a presentarsi insieme a tratti dell'italiano formale; si veda ad esempio, in (18), l'uso frequente della subordinazione frasale e la ricorrenza di termini specifici (es. *oppositore*) e dal significato astratto (es. *attività*):

- (18) BO088: un episodio molto importante e:h si ebbe quando e:hm anselmo fu cacciato
 BO088: eh da- dall'abbazia

- BO088: eh dall'abazia, (.) ehm da da re desiderio, al quale appunto anselmo era::
era (.) del quale era oppositore diciamo,
BO088: e ehm si rifugio' a monte cassino [...]
BO088: e fu appunto e:hm una volta che carlo magno conquisto' l'italia, fu
permesso ad anselmo di tornare eh a mantova,
BO088: nel settecentosettantaquattro,
BO088: e::hm::e cosi' po- pote' riprendere la sua attivita', prima e::h di creazione
della biblioteca,
BO088: perche' secondo appunto l'uso eh delle biblioteche=e::h
BO088: medioev- medievali diciamo, prima e:h vi era
BO088: m:h la insomma la raccolta di libri dall'esterno
(KIP, BOC1001)

È perciò possibile identificare due *cluster* di tratti standard co-occorrenti: l'uno di tratti neo-standard, l'altro di tratti dello standard letterario. Italiano neo-standard e italiano standard letterario paiono quindi effettivamente manifestarsi come due varietà separate, autonome e coerenti, nei comportamenti linguistici dei parlanti (per lo meno nel gruppo osservato, piuttosto omogeneo e ben definito in termini di collocazione sociale, e in certe situazioni comunicative).⁹ Occorre inoltre mettere in risalto che questi due *cluster* di tratti correlano, come si è detto, con il carattere informale vs. formale dell'interazione verbale. Ciò è rilevante perché svela l'esistenza di *pattern* sistematici nella variazione di registro, determinati non soltanto dall'alternanza fra singoli tratti ma – più significativamente – dall'alternanza fra insiemi di tratti co-occorrenti; è noto infatti come la variazione di registro sia più spesso caratterizzata da co-occorrenze invece poco sistematiche di tratti linguistici (cfr. ad es. Ghyselen & De Vogelaer 2018: 2).

Si può infine sottolineare che i tratti dell'italiano neo-standard tendono a co-occorrere in interazioni verbali di tipo diverso, nelle quali prevalgono funzioni comunicative differenti: le interviste semi-strutturate, incentrate sulla trasmissione di contenuti, e le conversazioni libere, dominate dalla relazione fra gli interlocutori e dalla costruzione di identità sociale (cfr. paragrafo 4.1). Ciò è interpretabile come un riflesso del carattere “multi-funzionale” dell'italiano neo-standard, varietà d'uso medio di una lingua con funzioni di distanza e immediatezza comunicativa (la “multi-funzionalità” è, più in generale, un carattere comune a tutte le varietà neo-standard di lingue europee; cfr. ad es. Mattheier 1997 e Coupland & Kristiansen 2011). Appare invece più circoscritto l'uso dello standard tradizionale, coerentemente con il suo essere “lingua di livello letterario” (Berruto 2012: 26).

5. Conclusioni

I primi studi condotti sui materiali di ParlaTO e KIP forniscono riscontri empirici e nuovi spunti su alcune tendenze in atto nella gamma di variazione sociolinguistica dell'italiano; in particolare, sulla crescente circolazione interregionale di fenomeni propri di aree geografiche diverse, col conseguente sviluppo di produzioni linguistiche, se non varietà di lingua, “composite”; sul progressivo venir meno delle differenze di natura categorica

⁹ Può forse valer la pena chiarire che le etichette *neo-standard* e *standard letterario* non sono assegnate in seguito ai risultati dello studio, che in sé non dimostra il carattere neo-standard o standard letterario dei tratti considerati, ma sono attribuite a priori in base alla letteratura esistente.

fra gruppi sociali, dettate cioè dalla presenza di tratti esclusivi di certe varietà diastratiche; e sull'affermarsi della coesistenza di due varietà standard, l'una di impronta letteraria e l'altra derivante dalla promozione di tratti sub-standard nell'uso di norma.

La gamma delle varietà rappresentate nel ParlaTO sarà poi utilmente ampliata, come si è detto, con la realizzazione di una sezione dedicata alle produzioni di parlanti di origine straniera, che offrirà tra l'altro l'opportunità di indagare una serie di aspetti dell'italiano di non nativi. Inoltre, le possibilità di interrogazione di ParlaTO e KIP saranno potenziate sia dalla consultazione congiunta dei due corpora sia dalla ricerca per lemmi e parti del discorso. L'auspicabile allestimento di nuovi moduli del KIParla (cfr. Mauri et al. 2019) potrà infine arricchire la rappresentatività, anche in termini di varietà di lingua, dell'intera risorsa.

Bibliografia

- Amenta, Luisa. 2020. *Strutture tempo-aspettuali nel siciliano e nell'italiano regionale di Sicilia*. Palermo: Centro di studi Filologici e Linguistici Siciliani.
- Antonelli, Giuseppe. 2011. Lingua. In Aferro, Andrea & Zinato, Emanuele (a cura di), *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi*, 15–52. Roma: Carocci.
- Ballarè, Silvia & Cerruti, Massimo & Gorla, Eugenio. 2019. Variazione diastratica nel parlato di giovani: il caso delle costruzioni relative. In Moretti, Bruno & Kunz, Aline & Natale, Silvia & Krakenberger, Etna (a cura di), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6-8 settembre 2018)*, 75–94. Milano: Officinaventuno.
- Ballarè, Silvia & Mauri, Caterina. 2020. La creazione del corpus KIParla: criteri metodologici e prospettive future. *Rivista italiana di dialettologia* 44. 53–69.
- Ballarè, Silvia & Micheli, M. Silvia. 2018. Usi di *dove* nell'italiano contemporaneo: costruzioni relative e dinamiche di ristandardizzazione. *Linguistica e Filologia* 38. 29–56.
- Ballarè, Silvia & Inglese, Guglielmo. 2021. The development of locative relative markers: from typology to sociolinguistics (and back). *Studies in language*. (<https://www.jbe-platform.com/content/journals/10.1075/sl.20013.bal>) (Consultato il 15.06.2021.)
- Berretta, Monica. 1989. Tracce di coniugazione oggettiva in italiano. In Foresti, Fabio & Rizzi, Elena & Benedini, Paola (a cura di), *L'italiano fra le lingue romanze. Atti del XX Congresso Internazionale di Studi della SLI (Bologna, 25-27 settembre 1986)*, 125–150. Roma: Bulzoni.
- Berruto, Gaetano. 1990. Semplificazione linguistica e varietà sub-standard. In Holtus, Günter & Radtke, Edgar (a cura di), *Sprachlicher Substandard*, Vol. III, 17–43. Tübingen: Niemeyer.
- Berruto, Gaetano. 2005. Intervento alla tavola rotonda. In Lo Piparo, Franco & Ruffino, Giovanni (a cura di), *Gli italiani e la lingua*, 332–342. Palermo: Sellerio.
- Berruto, Gaetano. 2012. *Sociolinguistica dell'italiano contemporaneo. Nuova edizione*. Roma: Carocci.
- Berruto, Gaetano. 2014. Esiste ancora l'italiano popolare? Una rivisitazione, in Danler Paul & Konecny, Christine (a cura di), *Dall'architettura della lingua italiana*

- all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, 277–290. Frankfurt am Main: Peter Lang.
- Berruto, Gaetano. 2017. Dinamiche nell'architettura delle varietà dell'italiano nel ventunesimo secolo. In Caprara, Giovanni & Marangon, Giorgia (a cura di), *Italiano e dintorni. La realtà linguistica italiana: approfondimenti di didattica, variazione e traduzione*, 7–31. Frankfurt am Main: Peter Lang.
- Berruto, Gaetano. 2018. Tendenze nell'italiano del Duemila e rapporti fra varietà standard e sub-standard. *Aggiornamenti* 13. 5–15.
- Berruto, Gaetano & Cerruti, Massimo. 2015. Un esercizio di analisi variazionista: l'accordo verbale nel costrutto locativo-esistenziale-presentativo. In Busà, Maria Grazia & Gesuato, Sara (a cura di), *Lingue e contesti. Studi in onore di Alberto M. Mioni*, 609–620. Padova: CLEUP.
- Bertinetto, Pier Marco & Loporcaro, Michele. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of The Phonetic Association* 35. 131–151.
- Biber, Douglas & Conrad, Susan. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.
- Boario, Anna. 2017. Native speaker–non-native speaker interaction: More than one type of accommodation. *Sociolinguistic studies* 11(2-3-4). 291–312.
- Bosco, Cristina & Ballarè, Silvia & Cerruti, Massimo & Gorla, Eugenio & Mauri, Caterina. 2020. KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech tagging. In Basile, Valerio & Croce, Danilo & Di Maro, Maria & Passaro, Lucia C. (a cura di), *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. (CEUR.org, <http://ceur-ws.org/Vol-2765/paper154.pdf>) (Consultato il 15.06.2021.)
- Brook, Marisa. 2011. One of those situations where a relative pronoun becomes a complementizer: A case of grammaticalization in progress... again. In Armstrong, Lisa (a cura di), *Proceedings of the 2011 Annual Meeting of the Canadian Linguistic Association*, 1–7. (<http://homes.chass.utoronto.ca/~cla-acl/actes2011/actes2011.html>) (Consultato il 15.06.2021.)
- Canepari, Luciano. 1983. *Italiano standard e pronunce regionali*. Padova: CLEUP.
- Cerruti, Massimo. 2011. Regional Varieties of Italian in the Linguistic Repertoire. *International Journal of the Sociology of Language* 210. 9–28.
- Cerruti, Massimo. 2018. Il parlato regionale oggi: un italiano composito? *Lid'O Lingua italiana d'oggi* 15. 15–31.
- Cerruti, Massimo & Ballarè, Silvia. 2020. ParlaTO: corpus del parlato di Torino. *Bollettino dell'Atlante Linguistico Italiano* 44. 13–38.
- Cerruti, Massimo & Crocco, Claudia & Marzo, Stefania (a cura di). 2017. *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian*. Berlin-New York: Mouton de Gruyter.
- Cerruti, Massimo & Vietti, Alessandro. In stampa. Identifying language varieties: Coexisting standards in spoken Italian. In Beaman, Karen & Guy, Gregory R. (a cura di), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*. London: Routledge.
- Comrie, Bernard & Kuteva, Tania. 2013. Relativization Strategies. In Dryer, Matthew S. & Haspelmath, Martin (a cura di), *The World Atlas of Language Structures Online*. (<http://wals.info/chapter/s8>) (Consultato il 15.06.2021.)
- Cortelazzo, Michele A. 2001. L'italiano e le sue varietà: una situazione in movimento.

- Lingua e stile* 34(3). 417–430.
- Cortinovis, Enrica & Miola, Emanuele. 2009. Chiamami *rappuso*: il suffisso *-uso* dai dialetti meridionali al linguaggio giovanile torinese. *Rivista italiana di dialettologia* 33. 195–218.
- Coupland, Nikolas & Kristiansen, Tore. 2011. SLICE: Critical perspectives on language (de)standardization. In Kristiansen, Tore & Coupland, Nikolas (a cura di), *Standard languages and language standards in a changing Europe*, 11–35. Oslo: Novus.
- Cresti, Emanuela & Moneglia, Massimo. 2005. *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins.
- D’Achille, Paolo. 2010. *Italiano popolare*. In Simone, Raffaele (a cura di), *Enciclopedia dell’italiano*, Vol. I, 723–725. Roma: Istituto dell’Enciclopedia italiana.
- D’Achille, Paolo. 2011. *Roma, italiano di*. In Simone, Raffaele (a cura di), *Enciclopedia dell’italiano*, Vol. II, 1262–1265. Roma: Istituto dell’Enciclopedia italiana.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12. 453–476.
- Fontanot, Arianna. 2019. *Aspetti di italiano composito nella pronuncia di giovani torinesi: un’indagine su corpora*. Torino: Università di Torino. (Tesi di laurea magistrale.)
- Goria, Eugenio & Ciccolone, Simone (a cura di). 2020. Corpora di parlato nel panorama italiano: verso l’individuazione di pratiche condivise. Sezione monografica di *Rivista italiana di dialettologia* 44. 17–169.
- Ghyselen, Anne-Sophie & De Vogelaer, Gunther. 2018. Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the “Variety” Concept. *Frontiers in Psychology* 9. (<https://doi.org/10.3389/fpsyg.2018.00385>) (Consultato il 15.06.2021.)
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Lerner, Gene H. (a cura di), *Conversation Analysis: studies from the first generation*, 13–31. Amsterdam-Philadelphia: John Benjamins.
- Mattheier, Klaus J. 1997. Über Destandardisierung, Umstandardisierung und Standardisierung in modernen europäischen Standardsprachen. In Mattheier, Klaus J. & Radtke, Edgar (a cura di), *Standardisierung und Destandardisierung europäischer Nationalsprachen*, 1–11. Frankfurt am Main: Peter Lang.
- Mauri, Caterina & Ballarè, Silvia & Goria, Eugenio & Cerruti, Massimo & Suriano, Francesco. 2019. KIParla Corpus: A New Resource for Spoken Italian. In Bernardi, Raffaella & Navigli, Roberto & Semeraro, Giovanni (a cura di), *Proceedings of the Sixth Italian Conference on Computational Linguistics*. (<http://ceur-ws.org/Vol-2481/paper45.pdf>) (Consultato il 15.06.2021.)
- Ramat, Paolo. 2006. Italian negation from a typological/areal point of view. In Grandi, Nicola & Iannàccaro, Gabriele (a cura di), *Zhì. Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*, 355–370. Cesena-Roma: Caissa.
- Regis, Riccardo. 2006. Breve fenomenologia di una locuzione avverbiale: il *solo più* dell’italiano regionale piemontese. *Studi di lessicografia italiana* 23. 275–289.
- Sobrero, Alberto A. 2005. Come parlavamo, come parliamo. Spunti per una microdiacronia delle varietà dell’italiano. In Lo Piparo, Franco & Ruffino, Giovanni (a cura di), *Gli italiani e la lingua*, 209–220. Palermo: Sellerio.

Prospettive di ricerca per valutare la competenza lessicale in una lingua seconda

Francesca Gallina

Università di Pisa

francesca.gallina@unipi.it

Abstract

La valutazione del lessico è una pratica didatticamente rilevante nel contesto educativo, ma è altresì una pratica assai diffusa nel sistema certficatorio e, negli ultimi decenni, anche negli studi dedicati alla descrizione dei processi di sviluppo della competenza lessicale. In ambito anglosassone si è sviluppato un vero e proprio filone di studi noto come *Vocabulary Assessment*, che nel corso del tempo ha visto alternarsi diversi approcci e metodi per osservare il lessico e la competenza lessicale, in relazione al quadro teorico di volta in volta dominante. Il contributo intende ricostruire sinteticamente l'evoluzione di tali approcci per poi prendere in esame le prospettive che si aprono quando si valuta il lessico, così come le prospettive di ricerca ad oggi esplorate per poter valutare un costrutto multidimensionale e ancora per certi versi non sufficientemente esplorato quale il lessico e la competenza lessicale in L2. Verranno discussi anche alcuni esempi di test lessicali utilizzati per valutare le molteplici dimensioni della competenza lessicale e come il lessico sia usato nelle abilità comunicative.

1. Introduzione

Nell'ambito del *Language Testing and Assessment* (LTA) i paradigmi teorici di riferimento, gli approcci alla valutazione e le tecniche impiegate per valutare la competenza in una L2 sono cambiati nel corso del tempo. Tradizionalmente la storia del LTA viene scandita in tre periodi (Spolsky 2017): l'era prescientifica, l'era strutturalista e l'era psicosociolinguistica e comunicativa. In ciascuna di queste tre fasi il lessico ha avuto un qualche ruolo, pur nella varietà delle modalità valutative e della definizione del costrutto di riferimento.

Nell'ambito dell'epoca prescientifica, che si estende fino agli anni Quaranta del Novecento, veniva praticata una valutazione intuitiva, basata su tecniche come la traduzione e la produzione libera, di cui veniva valutata la correttezza delle scelte lessicali rispetto alla norma, senza però alcun fondamento teorico linguistico o psicometrico. Il

lessico non rappresentava un costrutto autonomo.

Durante l'epoca strutturalista, che avvia l'era scientifica del LTA verso gli Cinquanta e in cui le istanze psicometriche dominavano l'approccio alla valutazione, si affermarono con forza test a punti discreti. Il lessico veniva assunto quale costrutto indipendente da altre dimensioni della lingua, con un focus soprattutto sul legame forma/significato e utilizzando, ad esempio, tecniche di riconoscimento o scelta multipla.

Nel periodo successivo, che possiamo definire integrativo e sociolinguistico, il LTA introdusse test integrati e pragmatici, come il dettato o il cloze test lessicale, in cui il lessico viene valutato in contesto e come parte di un costrutto più ampio, con un focus anche in questo caso sul legame forma/significato.

Infine, nell'epoca comunicativa, con l'affermarsi negli anni Ottanta di task di produzione e comprensione scritta e orale, il lessico assunse ruoli differenti, venendo talvolta trattato come costrutto autonomo e in altri casi integrato, così come potendo essere esplicitato o meno tra i criteri e i descrittori delle griglie di valutazione.

Tale breve sintesi della storia del LTA mostra chiaramente come l'attenzione alla dimensione lessicale in tale ambito sia una costante, pur con tutte le distinzioni necessarie a seconda del paradigma teorico e operativo assunto nelle diverse ere del LTA. Se la presenza del lessico nel LTA attraverso approcci e modelli valutativi diversi certamente dà sostegno a una riflessione interamente dedicata alla valutazione della competenza lessicale in una seconda lingua, essa non costituisce tuttavia l'unico motivo per il quale è opportuno riflettere sulla valutazione di tale competenza.

Il lessico occupa un posto privilegiato sia nella prospettiva di chi apprende una seconda lingua e vede nelle parole che via via impara a riconoscere e usare un segno tangibile dei propri progressi, sia nella prospettiva di chi la insegna e vi dedica larga parte del proprio tempo. Di recente il lessico ha assunto centralità anche tra chi si occupa di acquisizione, così come tra chi si occupa di valutazione delle competenze linguistico-comunicative. Tale centralità è ben sintetizzata da Read (2013: 1) che vede nel lessico una componente fondamentale della capacità di usare una lingua e pertanto, in un certo qual modo, ogni forma di valutazione è assumibile come una misura della competenza lessicale dell'apprendente. Se, dunque, il lessico è un fattore centrale nello sviluppo delle abilità comunicative, conseguentemente esso può giocare un ruolo di primo piano anche nel LTA in generale.

A tali fattori si somma anche il fatto che molti studi hanno messo in rilievo un'alta correlazione tra i test lessicali e i test di competenza più generali, tanto che è possibile anche ipotizzare una certa capacità predittiva della valutazione lessicale rispetto alla più generale competenza linguistico-comunitiva (Miralpeix 2020: 189).

Obiettivo del presente lavoro è, dunque, quello di avviare i primi passi di una riflessione sulle prospettive di ricerca e sui metodi per valutare la competenza lessicale in una seconda lingua, partendo dall'inquadramento storico del *Vocabulary Assessment* e sollevando alcune questioni critiche che sorgono nel momento in cui si ci propone di valutare il lessico e la competenza lessicale. Verrà successivamente proposta una sintesi degli approcci e dei metodi finalizzati alla valutazione del lessico.

2. Il *Vocabulary Assessment*: un breve profilo storico

Prima di entrare nel vivo di ciò che significa valutare il lessico e la competenza lessicale e discutere le questioni che conseguentemente si aprono o le modalità che abbiamo a

disposizione per valutare, ci pare utile tracciare un breve profilo storico dello sviluppo della valutazione del lessico. La pratica di valutare il lessico nel LTA non è certo cosa nuova; come abbiamo visto nel paragrafo precedente, tuttavia, uno specifico interesse del LTA verso il lessico ha portato soprattutto in ambito anglosassone allo sviluppo di un filone di testing, il *Vocabulary Assessment*, che ha una storia più recente.

Secondo Read (2012), anche se il primo articolo di *Vocabulary Assessment* risale al 1935, è a partire dagli anni Ottanta e Novanta che si hanno le prime pubblicazioni relative a misure lessicali e procedure finalizzate a valutare aspetti delle conoscenze e delle competenze lessicali. Negli anni Ottanta fioriscono i primi studi sui test lessicali, che in una prima fase erano però collocati all'interno di ricerche sulla comprensione del testo e sulla relazione tra comprensione e competenza lessicale. Nella decade successiva, anche a seguito delle modellizzazioni teoriche e descrittive dei processi di sviluppo della competenza lessicale proposte in quegli anni, nella discussione su come valutare il lessico entrano temi come la multidimensionalità della competenza lessicale (su cui torneremo nel paragrafo 3), l'ampiezza delle conoscenze lessicali e la differenza tra la valutazione del lessico in contesto e in isolamento. Nello stesso periodo appaiono anche i primi studi di validità sui test lessicali e i primi test standardizzati. Dopo il 2000 vengono pubblicate le prime monografie dedicate alla valutazione del lessico (Milton 2009; Read 2000) o che riservano alla misurazione e alla valutazione del lessico ampio spazio (Bardel et al. 2013; Daller et al. 2007; Meara & Miralpeix 2007; Schmitt 2010), accompagnate da molti lavori su test che misurano la dimensione produttiva e ricettiva, i diversi aspetti delle conoscenze lessicali, la definizione del costrutto del test e la relazione tra competenza lessicale, competenza linguistico-comunicativa e abilità.

Uno dei limiti del *Vocabulary Assessment* è dato dal fatto che per lungo tempo molti ricercatori che se ne sono occupati hanno sviluppato i propri studi senza alcuna collaborazione o interazione con gli studi acquisizionali o il più ampio filone del LTA. Tale mancanza di dialogo ha avuto come conseguenza da un lato il fatto che non venisse elaborato un modello teorico di competenza lessicale capace di diventare punto di riferimento per la definizione del costrutto nella prospettiva del LTA e, d'altro canto, una carenza di un'ampia e solida validazione degli studi di valutazione della componente lessicale. I test che sono stati impiegati per misurare il lessico negli studi acquisizionali hanno impiegato formati come le *checklist*, gli accoppiamenti parola/definizione, il completamento di una parola, ecc. Ciò che accomuna il testing lessicale sia nell'ambito del LTA che della linguistica acquisizionale è il fatto di essersi entrambi focalizzati principalmente sulle conoscenze lessicali, anziché sulle competenze lessicali, ovvero di aver posto maggior attenzione sulla conoscenza di elementi lessicali a discapito di una valutazione della competenza linguistica, una valutazione fondata su un costrutto più ampio e su un contesto più esteso rispetto a test basati su item lessicali isolati.

3. Alcune questioni critiche

Nel momento in cui ci si propone di valutare il lessico, è opportuno discutere alcune questioni preliminari, che hanno un certo grado di criticità e che pertanto necessitano di essere sollevate a priori rispetto al processo di progettazione e costruzione di un test lessicale. Alcune di tali questioni sono fondanti, dal momento che parlare di valutazione del lessico e della competenza lessicale presuppone che sia chiaramente definito cosa si intende per parola, per lessico e per competenza lessicale.

Secondo il GRADIT (De Mauro 1999) *parola* nel suo uso fondamentale è intesa come unità distinguibile all'interno di una frase, mentre nell'uso tecnico-specialistico è un'entità costituita da uno o più morfi, con un accento in rilievo che ne garantisce l'unità, un certo margine di libertà di dislocazione nella frase, la non intercalabilità di altre entità simili dentro il significante, una certa rigidità della successione dei morfi. In tale prospettiva possiamo includere nell'insieme costituito dalle parole una base tematica radicale con eventuali formanti, prefissi, infissi e suffissi, la composizione di più basi tematiche come nei composti, l'unione di più lessemi come nei lessemi complessi, le polirematiche, ecc. Decidere dunque quali di queste possibilità costituiscono delle parole ha delle implicazioni rilevanti per la valutazione, poiché significa, ad esempio, ammettere, o escludere, le espressioni polirematiche dai test lessicali di una L2.

Sempre prendendo come riferimento il GRADIT, il lessico può essere interpretato come l'insieme delle parole di una lingua, ma anche delle regole di formazione di altre parole, ovvero dei vocaboli usati dai parlanti e dei vocaboli di uso potenziale. Inoltre, il fatto che le parole non vivano in isolamento, ma che in prospettiva saussuriana esse siano legate al tempo, allo spazio e alla massa parlante, ci porta a intendere il lessico come un sistema aperto, creativo, variabile, sia a livello collettivo che individuale, fondato sull'uso linguistico da parte dei parlanti, articolato in una rete di relazioni tra parole, che a loro volta assumono in sé diverse dimensioni, da quella fonetica a quella sintattica.

La scelta nel definire concetti come *parola* e *lessico* ha evidentemente delle ricadute notevoli sul LTA, poiché implica una scelta su quale unità di misura adottare. Se in ambito internazionale il *Vocabulary Assessment* ha sempre utilizzato come riferimento la famiglia di parole, che comprende tutte le forme flesse e derivate da una singola parola, ultimamente la tendenza è stata quella di preferire il lemma come unità di misura per il lessico, come del resto già in uso negli studi sul lessico di ambito italiano (Kremmel 2016). Il vantaggio nell'uso del lemma rispetto alla forma di parola o alla famiglia lessicale dipende sia dal fatto che nel lessico mentale si registrano le forme base di un lemma cui leghiamo a seconda del concreto contesto d'uso i suffissi necessari di volta in volta, almeno per ciò che concerne le lingue indoeuropee (Aitchinson 2003), sia perché didatticamente si ha un minor carico di apprendimento nell'imparare la forma base cui applicare le proprie conoscenze morfologiche, sintattiche e pragmatiche. A ciò aggiungiamo che nell'interpretazione dei risultati di un test lessicale costituito da lemmi è più chiaro cosa ci dice una risposta costituita da un lemma rispetto a una famiglia di parole di cui l'apprendente potrebbe conoscere un solo elemento, come ad esempio *società*, ma non tutte le altre parole come *socializzazione*, *socializzare*, *asociale*, *socialmente*, ecc.

Tale complessità determina naturalmente alcune ricadute sulla valutazione del lessico e della competenza lessicale, soprattutto alla luce delle considerazioni di Alan Davies (2008: 109):

Le parole assumono il loro significato quando sono inserite in un contesto: al di fuori di esso sono, per così dire, prive di significato. Per questo motivo, se dobbiamo valutare il lessico, questo deve essere fatto attraverso una misurazione che viene realizzata in un contesto. *Ma sta proprio qui il paradosso: più contesto forniamo, più il lessico perde la sua lessicalità. E in questo caso possiamo continuare ad affermare che stiamo testando il lessico?*

Anche la definizione di ciò che intendiamo per competenza lessicale è fondamentale, poiché da questa dipende la definizione del costrutto stesso di un test. Nel corso degli

ultimi anni si sono affermati modelli descrittivi basati sulla convinzione che la competenza lessicale sia multidimensionale (Henriksen 1999; Read 2000; Wesche & Paribakht 1996).

La prima dimensione è di tipo quantitativo (ampiezza del lessico) e si riferisce al numero di parole conosciute, attivamente e passivamente, indipendentemente dal livello di conoscenza che di esse ha l'apprendente.

La seconda dimensione si fonda su una conoscenza qualitativa (profondità delle conoscenze), anche se in letteratura non è uniforme e condiviso il modo di intenderla, poiché essa viene intesa come conoscenza, ricettiva e/o produttiva, dei diversi aspetti di una parola (forma, significato, usi, frequenza, registro, legami associativi e collocazioni sulla base della proposta di Nation 2001), ma in questo modo si trasforma in un oggetto di difficile misurazione tramite un solo strumento.

La terza dimensione riguarda invece la velocità con cui accediamo alle conoscenze lessicali nelle abilità linguistiche (accessibilità o automaticità), per cui dipende essenzialmente dall'organizzazione del lessico e in un certo senso può essere interpretata anche come fluenza, secondo il modello di *lexical space* di Daller et al. (2007: 8).

La complessa e multidimensionale natura del lessico e della competenza lessicale porta necessariamente a interrogarsi su come e quanto tale natura si rifletta nei processi di misurazione e valutazione del lessico. Come è possibile valutare aspetti tanto diversi tra loro? È possibile tenere in considerazione le diverse dimensioni in contemporanea? Come vedremo nel paragrafo successivo, gli approcci e i metodi elaborati finora hanno optato principalmente per test focalizzati su una sola dimensione alla volta.

4. Approcci e metodi per valutare la competenza lessicale

Nel momento in cui ci sia appresta a svolgere una qualsiasi operazione di valutazione è innanzitutto necessario esplicitare con chiarezza quale scopo ha tale operazione, ovvero quali obiettivi si pone il test. Le ragioni per le quali decido di valutare il lessico possono infatti essere di natura diversa, anche a seconda del contesto in cui ci muoviamo, il che determina che di un test è possibile fare usi diversi, e che possiamo interpretarne i risultati con tutte le conseguenze che ne derivano per coloro che in qualche modo sono interessati dal processo valutatorio, siano essi apprendenti, insegnanti, istituzioni scolastiche, ricercatori, famiglie, ecc. Un test ha lo scopo di raccogliere informazioni affidabili che aiutino il processo decisionale basato sulla valutazione, per cui si rende necessario innanzitutto definirne con precisione lo scopo.

Un test lessicale utile per approfondire le conoscenze attuali su come si acquisisce il lessico e come si sviluppa la competenza lessicale non avrà la stessa validità e utilità se applicato in un contesto certificatorio o educativo. I test lessicali che sono utilizzati nell'ambito della ricerca linguistica vengono proposti a campioni non estesi di informanti, hanno una durata piuttosto lunga per consentire di raccogliere il maggior numero di dati possibile, spesso si accompagnano a interviste agli informanti per approfondire l'indagine ecc., per cui non hanno alcuna rilevanza applicativa, ma sono finalizzati solo ad ampliare lo stato delle conoscenze su come si sviluppa la competenza lessicale. Anche i test di valutazione lessicale di ambito clinico hanno obiettivi e destinatari diversi rispetto ai test lessicali che vengono abitualmente impiegati nella valutazione del lessico in contesto educativo o certificatorio.

I test standardizzati su larga scala possono prevedere delle prove lessicali, all'interno

di un set di prove su altre dimensioni linguistiche e altre abilità, come ad esempio accade negli esami di certificazione linguistica o nelle prove INVALSI. In tali casi l'obiettivo non è, come avviene in contesto educativo, quello di condizionare in qualche modo il comportamento degli apprendenti, ma può essere, ad esempio, quello di valutare l'efficacia di programmi di apprendimento (Read 2000: 152). Nel caso delle certificazioni il fatto che il lessico entri nel costrutto di riferimento sottolinea il ruolo da esso assunto all'interno della modellizzazione teorica di ciò che è la lingua come oggetto di valutazione e serve per completare il quadro che dipinge lo stato delle competenze linguistico-comunicative del candidato all'esame, determinando se è in possesso di quei requisiti che vengono fissati in base agli standard di riferimento.¹ I test focalizzati sul lessico che possono essere utilizzati in contesto educativo hanno evidentemente finalità, modalità e destinatari diversi, potendo essere *placement test* focalizzati sulle parole che verranno insegnate in un corso o che sono ritenute un prerequisito per il corso, *achievement test* per valutare l'apprendimento delle parole insegnate, test diagnostici sulle parole che ci attendiamo un apprendente conosca a un certo livello, test per incentivare la motivazione sulle parole insegnate di recente o più utili per l'apprendente, test di *proficiency* su un ampio spettro di parole.

4.1 Le dimensioni dei test lessicali

Abbiamo visto come nelle diverse ere del testing siano stati impiegati formati differenti per osservare la competenza lessicale, da quelli discreti a quelli comunicativi. Secondo Read (2000: 7–8) non si tratta di una vera e propria dicotomia, poiché entrambe le prospettive rientrano in un quadro unitario da cui selezionare il test più appropriato per i propri obiettivi e il proprio contesto. Per poter dunque analizzare e distinguere i test lessicali in base alle loro caratteristiche lo stesso Read propone tre dimensioni:

1. *Discrete – Embedded*
2. *Selective – Comprehensive*
3. *Context-independent – Context-dependent*

Le tre dimensioni vanno intese come dei *continua*, su cui si collocano i diversi tipi di test che possiamo utilizzare per la valutazione del lessico. La prima dimensione ha a che vedere con il costrutto del test, oscillando tra i test lessicali discreti, che assumono il lessico come costrutto indipendente e quindi lo valutano separatamente rispetto ad altre dimensioni, e i test lessicali integrati, che inseriscono il lessico in un costrutto più ampio e lo osservano, ad esempio, all'interno di task come una produzione scritta. I test discreti vengono percepiti come test lessicali per eccellenza, ma va precisato che la distinzione rispetto al test lessicale integrato non sta semplicemente nelle modalità di presentazione del lessico (parole isolate vs parole in contesto), ma nel modo di intendere il costrutto di riferimento della valutazione.

La seconda dimensione considera l'estensione del lessico che si vuole misurare, ovvero conduce a una scelta tra test che selezionano un certo numero di parole per costruire gli item e i test comprensivi che invece non si basano su una selezione di parole. La selezione delle parole da rendere oggetto di valutazione è una questione di importanza

¹ Sulla valutazione del lessico nelle certificazioni ci sia consentito rimandare a Gallina (2019).

capitale per il testing lessicale, qualunque sia il criterio in base al quale viene effettuata, come ad esempio la frequenza d'uso di una parola o la necessità di testare un certo lessico di ambito specialistico. Nei test omnicomprensivi la scelta del lessico oggetto di valutazione non è fatta a priori, ma sono prese in considerazione per la valutazione tutte le parole che vengono usate ad esempio nei task di produzione.

La terza e ultima dimensione gioca invece sul ruolo del contesto, collocando a un estremo del continuum i test con gli item in isolamento e all'opposto i test in cui la soluzione deriva dalla capacità di utilizzare le informazioni contestuali, andando oltre il significato della singola parola. I test indipendenti dal contesto possono, a prima vista, sembrare quasi test lessicali prototipici dal momento che si focalizzano su di un'unità lessicale, ma d'altra parte nell'uso linguistico una parola non è mai isolata, per cui anche in contesto valutatorio vale la pena proporre dei test lessicali dipendenti dal contesto.

Volendo dunque sintetizzare i possibili modi in cui un test può essere progettato sulla base delle dimensioni proposte da Read, possiamo immaginare un continuum, lungo il quale collocare i test in base a quanto contesto è necessario per eseguire il task. A un estremo, come da Figura 1, troviamo i test con gli item isolati, come ad esempio i task di abbinamento parola/immagine, le *checklist*, ecc. in cui non è necessario riferirsi al contesto d'uso di una unità lessicale per risolvere il test. All'estremo opposto si hanno i task comunicativi, in cui il lessico viene osservato e valutato tramite le abilità comunicative. Nel mezzo possiamo invece trovare dei test lessicali in cui il lessico è presentato in contesto, ma le strategie di risoluzione non sempre fanno leva su un contesto molto esteso, come accade ad esempio nel cloze test o nei completamenti di frase.

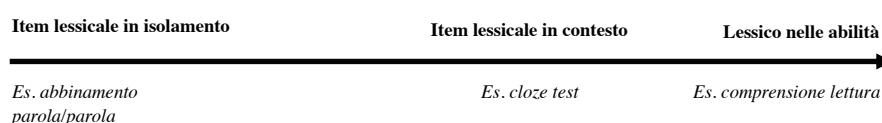


Figura 1. Il continuum dei test lessicali.

La scelta di un test posto in uno dei punti di questo continuum dipende essenzialmente dallo scopo che ci diamo: se voglio indagare la conoscenza del legame forma/significato anche un test in isolamento può essere sufficiente, come ad esempio la traduzione da L2 a L1 di un'unità lessicale, se invece voglio verificare la capacità di usare un certo target lessicale dovrò necessariamente ricorrere a test che si spingono verso il lato destro del nostro continuum.

4.2 Metodi per valutare il lessico: tra quantità, qualità, accessibilità, organizzazione e abilità

Se la competenza lessicale viene descritta come un oggetto multidimensionale, è evidente che non è facile individuare un test lessicale capace di informare contemporaneamente su questa molteplicità di aspetti. Per tale ragione nel *Vocabulary Assessment* sono stati elaborati e realizzati soprattutto test che di volta in volta hanno guardato a una sola dimensione della competenza lessicale, o che hanno messo insieme una batteria di prove per cercare di ricavare informazioni sullo stato di conoscenze quantitative, qualitative e organizzative che un apprendente ha del lessico.

4.2.1 I test sull'ampiezza del lessico

La maggior parte dei test lessicali si propone di valutare l'ampiezza delle conoscenze lessicali per la centralità che la dimensione quantitativa assume nello sviluppo di una L2, dal momento che maggiore è il numero di parole che un apprendente conosce, migliori sono le sue competenze linguistico-comunicative (Meara 1996: 37). Non è ovviamente possibile misurare tutte le parole che un individuo conosce, per cui di fatto i test quantitativi si basano sulla selezione di un certo numero di parole di cui si valuta la conoscenza da parte dell'apprendente.

La selezione delle parole viene fatta principalmente sulla base della loro frequenza, per cui nei livelli iniziali possiamo inserire tra gli item soprattutto le parole più frequenti di una lingua, mentre nei livelli avanzati possono essere selezionate parole appartenenti anche ai lessici di uso tecnico-specialistico a seconda del target di ciascun test.

Rispetto alla lunghezza del test è evidente che la numerosità degli item dipende da vari fattori, tra cui i seguenti: tempo a disposizione per lo svolgimento del test, grado di rappresentatività del lessico che si vuole ottenere, formato del test, poiché una *checklist* richiede una minor quantità di tempo rispetto ad esempio a una scelta multipla.

Ma che cosa valutano esattamente i test di taglio quantitativo? Quando si valuta l'ampiezza delle conoscenze lessicali di fatto si mira ad indagare la conoscenza del legame tra forma e significato di un'unità lessicale (*form-meaning link*), che rappresenta il costrutto minimo di qualsiasi test lessicale. In termini semplicistici, se le parole veicolano significati, è evidente che la conoscenza, anche parziale, della forma e del significato di una parola rappresenta il cuore delle conoscenze lessicali, la soglia minima perché si possa parlare di competenza lessicale. Tuttavia, definire in tal modo il costrutto dei test lessicali rischia di sminuire la capacità di un test di osservare il lessico di un apprendente. Valutare se un apprendente conosce il legame forma/significato di una parola non consente infatti di verificare se conosce anche altri aspetti, più di natura qualitativa, soprattutto nei casi in cui una parola non veicola un solo significato e non ha un unico uso nelle situazioni comunicative reali, così come non fornisce alcuna garanzia che possa e sappia usarla effettivamente in contesto, in ricezione e in produzione. Se, dunque, i test sull'ampiezza del lessico hanno trovato larghissima diffusione e applicazione in contesti anche molto diversi tra loro perché sono ritenuti test oggettivi, pratici, facili da somministrare e da valutare, adatti per valutare un ampio numero di parole, non è inutile tenere sempre presente anche i loro limiti. Kremmel & Schmitt (2016: 379) individuano infatti una serie di criticità, tra cui il fatto che la loro validità è a rischio poiché spesso chi li sostiene adotta delle strategie di indovino che rischiano di falsare i risultati e tali prove inoltre danno un quadro molto parziale delle conoscenze lessicali di un apprendente.

Proviamo dunque a vedere qualche esempio di prove focalizzate sull'ampiezza del lessico. Le liste di parole (*checklist*) che richiedono all'apprendente di indicare se conosce o meno una parola presente nella lista costituiscono forse la forma più elementare e pratica di valutazione del lessico, i cui limiti stanno da un lato nel fatto che si basano su autovalutazioni, dall'altro nel fatto che se non sono accompagnate da un task di riutilizzo di una parola non dicono molto su quali significati, che può assumere quella parola, sono effettivamente noti.

Anche le prove che richiedono di abbinare una parola a un altro elemento, come un'immagine, un'espressione equivalente, un sinonimo, una definizione, la traduzione nella L1, ecc. costituiscono test sull'ampiezza lessicale, che possono essere basati anche su un formato come la scelta multipla, e sono finalizzate a verificare la capacità di riconoscimento del significato di una parola, come nell'esempio seguente pensato per apprendenti di italiano L2 di madrelingua spagnola:

- (1) *cane* *a. ratòn* *b. oveja* *c. perro* *c. cerdo*

Anche in questo caso le strategie di indovino possono giocare un ruolo, così come accade ad esempio nei test di abbinamento parola/definizione come il *Vocabulary Size Test – VST* (Nation & Beglar 2007), che è ugualmente un test di riconoscimento del significato. In tutti questi test le parole vengono valutate in isolamento, per cui non garantiscono una capacità d'uso reale.

Anche le prove di riconoscimento della forma rientrano tra le prove quantitative, avendone gli stessi limiti. Esse sono focalizzate sulla capacità di individuare la forma corrispondente a un item che può essere nella L1 dell'apprendente, oppure in L2, come nell'esempio seguente tratto dal *Vocabulary Levels Test – VLT* (Nation 1983):

- (2) 1. *birth*
 2. *dust* _____ *game*
 3. *operation* _____ *winning*
 4. *row* _____ *being born*
 5. *sport*
 6. *victory*

Un altro tipo di prove di natura quantitativa è rappresentato dai test di recupero della forma o del significato di un'unità lessicale. Nel primo caso all'apprendente viene data una parola in L1, o la sua definizione, a partire dalla quale deve produrre l'equivalente in L2, oppure gli viene chiesto di completare una frase con la parola mancante, il che costituisce evidentemente un task più complesso rispetto a quelli di riconoscimento, che implica un grado maggiore di conoscenza di una parola per essere svolto. Nel secondo caso l'item viene presentato in L2 e l'apprendente deve darne una definizione o l'equivalente in L2 per dimostrare di saperne recuperare il significato.

Come hanno ben evidenziato Laufer & Goldstein (2004), esiste in realtà una gradazione di conoscenze del legame forma/significato, legata al fatto che quando affermiamo di conoscere una parola in realtà non esiste un'equivalenza perfetta rispetto alla stessa affermazione da parte di un altro apprendente. In particolare si ha una difficoltà crescente secondo il seguente schema: 1. riconoscimento passivo del significato di una forma; 2. riconoscimento attivo della forma di una parola; 3. recupero passivo del significato di una parola partendo ad esempio dalla sua definizione; 4. recupero attivo della forma di una parola partendo dal suo significato. Tale gradazione di difficoltà nei diversi task richiama il fatto che il lessico ricettivo è più ampio di quello produttivo, per cui possiamo presumere che vi sia un meccanismo implicazionale, che fa sì che se sono in grado di recuperare il significato di una parola lo posso anche riconoscere. In base a tali considerazioni e ai limiti dei test quantitativi cui abbiamo accennato, è doverosa una carta cautela nell'interpretazione dei risultati dei test.

4.2.2 I test sulla profondità delle conoscenze lessicali

Vediamo ora i test destinati a valutare la conoscenza qualitativa del lessico, con una precisazione: la dimensione della profondità delle conoscenze lessicali non viene intesa in maniera univoca da chi se ne occupa, poiché per alcuni viene interpretata come conoscenza di un certo numero di tratti di una parola, mentre per altri non si tratta di una conoscenza componenziale, ma globale, costituita dalla capacità di costruire e gestire una rete di parole organizzata, in cui sono le relazioni e le associazioni tra parole che consentono di dare profondità alle conoscenze lessicali. A ciò si aggiungono anche altri due modi di intendere la dimensione qualitativa della competenza lessicale, di cui il primo ha a che vedere con la capacità lessicale ricettiva e produttiva e il secondo invece con la velocità con cui accediamo al lessico, con cui riconosciamo o richiamiamo una parola quando leggiamo, ascoltiamo, scriviamo o parliamo.

Il costrutto della dimensione qualitativa è dunque molto ampio e difficile da rendere operativo in un solo test. Per tali ragioni sono stati proposti nel corso del tempo diversi approcci e test per valutare la qualità delle conoscenze lessicali.

Un primo approccio è quello delle scale di sviluppo, che identificano diversi gradi di conoscenza di una parola, sebbene questi sollevi numerosi nodi teorici: spesso tali scale hanno una valenza più didattica che acquisizionale, non è così immediata l'identificazione di tappe di apprendimento ben marcate e differenziate, ecc. Sul continuum si va da una competenza nulla, quando l'apprendente non conosce affatto una parola, alla padronanza ricettiva e produttiva, in tutti i suoi usi, di una parola. La *Vocabulary Knowledge Scale* (Wesche & Paribakht, 1996) è uno degli esempi più diffusi di tale tipo di test in contesto educativo, richiede all'apprendente di dichiarare quanto bene conosce una parola, ma anche di dimostrare che la conosce fornendo un sinonimo o la traduzione e producendo una frase facendo uso di quella parola.

Un secondo approccio è quello componenziale, che può valutare più aspetti in una batteria di prove o che può concentrarsi di volta in volta su un unico aspetto di ciò che significa conoscere una parola. Read (2000: 24) sottolinea come tale approccio trovi maggior utilità di applicazione nel contesto della ricerca lessicale, mentre abbia dei limiti, dettati in primis dal fatto che consente di valutare un numero piuttosto esiguo di parole, nell'utilizzo in contesto educativo. Concentrandosi su uno o pochi aspetti della conoscenza lessicale è evidente che si tratta di prove piuttosto limitate, che non restituiscono un'immagine ad ampio raggio del lessico noto a un apprendente.

Sempre nell'ambito della valutazione della profondità delle conoscenze lessicali segnaliamo i test che propongono le associazioni lessicali. Il *Word Associates Format test* di Read (2000) richiede all'apprendente di associare a una parola input quelle opzioni di risposta che hanno con la parola target una relazione paradigmatica (ad esempio un sinonimo), sintagmatica (frequenza di co-occorrenza) o analitica (la risposta rappresenta un tratto della parola target) come nell'esempio (3) sono rispettivamente le risposte *accordo, firma, formale*:

- (3) *contratto*
formale accordo avviso firma confidenziale

Anche i test sulla capacità di riconoscere le parti che formano una parola rappresentano un esempio di test lessicale sulla profondità delle conoscenze. Nel caso del *Test for*

English Derivatives – TED (Schmitt & Zimmerman 2002) l'apprendente deve completare delle frasi con la forma corretta della parola target come nell'esempio (4):

- (4) *philosophy*
Noun *She explained her _____ to me.*
Verb *She was known to _____ about her life.*
Adjective *She was known as a _____ person.*
Adverb *She discussed her life _____ .*

Abbiamo già accennato al fatto che la conoscenza degli aspetti collocazionali di una parola costituisce uno degli elementi che concorrono a definire la conoscenza qualitativa del lessico, per cui anche su tale tratto sono stati proposti dei test, come ad esempio il test COLLEX (Gyllstad 2007) o il test CONTRIX (Revier 2009). I primi test sulle collocazioni prevedevano un task di traduzione o un cloze test, mentre oggi si preferiscono formati come gli abbinamenti tra parole che sono parte di una collocazione, come nell'esempio (5):

- (5) *pazienza*
a. *dare*
b. *mandare*
c. *prendere*
d. *portare*

Ci pare evidente che i test che abbiamo qui segnalato assumano come oggetto della valutazione tratti della competenza molto eterogenei, per quanto tutti intimamente connessi alla competenza lessicale nel suo complesso.

4.2.3 I test sull'accessibilità e l'organizzazione

Concludiamo questo excursus sui test lessicali presentando alcuni test finalizzati alla valutazione della capacità di accesso al lessico e al grado di organizzazione del lessico di un apprendente. È evidente come una maggior velocità nell'accedere al lessico, sia per riconoscere una forma o un significato, sia per recuperarli, consenta di usare la lingua in modo più fluente e dipenda dal livello di competenza.

Le nuove tecnologie consentono di misurare con una certa precisione la velocità di accesso e di processamento di una parola o i tempi di reazione per rispondere a uno stimolo, ma la loro applicazione è ancora lontana dai contesti educativi e si limita ai contesti di ricerca lessicale. In tali contesti sono stati proposti test in cui il task è piuttosto lontano da compiti linguistici reali, come quando si richiede al soggetto di indicare in modo rapido se una parola esiste o meno, oppure di abbinare una parola al suo significato scelto in un set di quattro opzioni, o ancora quando viene misurato il tempo di reazione e l'accuratezza della denominazione di un apprendente chiamato a indicare il nome di ciò che vede in un'immagine. Tuttavia, è possibile ricorrere a task più vicini alle abilità comunicative, ad esempio misurando il tempo necessario all'apprendente per richiamare una parola da usare in un contesto o riconoscendola in un contesto dato.

Sotto questo profilo sono altresì interessanti i test sull'organizzazione delle conoscenze lessicali, che non prendono in considerazione singole unità lessicali, ma la

loro rete di rapporti, la natura e l'intensità di tali relazioni, che contribuiscono al complesso delle conoscenze e delle competenze lessicali. Oltre ai test sulle associazioni di parole di cui abbiamo già discusso, ci sono state proposte di utilizzare la teoria dei grafi per formalizzare la metafora del network per descrivere i legami del lessico mentale (Wilks & Meara 2007). Anche in questo caso si tratta di test difficilmente utilizzabili al di fuori del contesto della ricerca lessicale.

4.2.4 *Valutare il lessico tramite le abilità linguistiche*

Un ultimo cenno merita la valutazione del lessico non in prove specifiche come quelle che abbiamo passato sinteticamente in rassegna, bensì attraverso le abilità, ovvero ascoltare, leggere, parlare, scrivere. Mentre svolgiamo un'attività linguistica più ampia facciamo ricorso a conoscenze e competenze lessicali che, dunque, possono essere in qualche modo osservate e valutate nell'uso linguistico.

Per quanto concerne le abilità ricettive, che possono essere osservate solo in modo indiretto, ciò che possiamo valutare è la capacità di comprendere una parola all'interno di un contesto, in un testo letto o ascoltato. I test di comprensione sono, sotto il profilo lessicale, prove comprensive e dipendenti dal contesto. In letteratura non mancano lavori che hanno evidenziato lo stretto legame tra comprensione di un testo e componente lessicale (Bonk 2000; Buck 2001; Laufer & Aviad-Livitzky 2017; Nassaji 2004; Stæhr 2009; Wei 2020). Tra l'altro nella selezione del testo input pesano naturalmente anche le caratteristiche lessicali del testo, che devono essere ponderate per proporre un testo adeguato al livello di competenza di chi si sottopone al test.

Nella comprensione dell'ascolto, che differisce notevolmente dalla comprensione della lettura per una serie di fattori legati al tempo, alla memoria di lavoro, al contesto, al rumore, ecc., il lessico rappresenta un predittore importante della capacità di comprensione orale (Matthews & Cheng 2015: 9). Alcune prove di ascolto, soprattutto nei livelli di competenza iniziali, propongono una scelta multipla per verificare che l'apprendente abbia compreso un passaggio del testo, giocano molto sulla comprensione di unità lessicali o di espressioni presenti nel testo e nelle opzioni di risposta, come accade ad esempio nei test di comprensione orale di stampo strutturalista. Anche il dettato cloze e i test di completamento di una frase chiamano in causa la competenza lessicale e le conoscenze relative a una parola, tra cui la sua forma orale e quella scritta.

Nella comprensione della lettura, processo complesso e diversificato su cui non abbiamo qui il tempo di riflettere in modo approfondito, la capacità di riconoscere le parole e l'automaticità con cui ciò avviene sono componenti essenziali di una lettura fluente (Alderson 2000: 122). Tra i formati più diffusi per valutare la comprensione della lettura c'è la scelta multipla, che dipende in larga misura dalla conoscenza lessicale delle parole presenti sia nel testo input che nelle opzioni di risposta. Allo stesso modo i cloze test e i test di completamento rappresentano prove che, oltre a valutare la comprensione generale di un testo, così come sono stati inizialmente ideati i cloze test, consentono di osservare anche la componente lessicale, come avviene ad esempio in molti esami di certificazione di italiano L2. Sempre per osservare il lessico a partire dalla lettura di un testo è possibile richiedere anche di individuare un sinonimo di una parola nel testo, o di fornirlo.

Per quanto concerne le abilità produttive, Read (2013: 5) definisce tali prove come test d'uso, in cui la capacità di usare le parole è rappresentata in modo molto più vicino a ciò

che succede nella comunicazione reale rispetto ai test lessicali di recupero. Anche in questo caso, nella prospettiva lessicale, si tratta di test comprensivi, dipendenti dal contesto e di cui è difficile predire con esattezza quali saranno le scelte lessicali dell'apprendente, che possono essere solo indirizzate dal task proposto e dalla selezione del topic su cui svolgere la produzione.

Nelle prove di produzione il lessico è parte integrante di un costrutto ben più ampio rispetto a quello dei test focalizzati sul lessico e viene valutato secondo due prospettive principali: l'applicazione di misure statistico-lessicali e l'uso di scale di valutazione ad uso dell'insegnante o del valutatore. Le prime hanno come riferimento le unità lessicali presenti nel testo, di cui valutano la densità, ricchezza, la sofisticatezza lessicale, ecc. e trovano maggior applicazione nell'ambito della ricerca lessicale. Le scale per la valutazione invece possono essere proficuamente usate sia in contesto educativo che certificatorio.

La produzione orale, che ha sotto il profilo ontogenetico e filogenetico, ma anche didattico, una centralità indiscutibile, è un oggetto complesso da valutare, non solo perché più volatile rispetto allo scritto, ma anche perché in essa viene coinvolto anche l'ascolto dell'interlocutore e perché molti sono i fattori che influenzano il giudizio di chi valuta una prova orale. Alcuni studi hanno messo in evidenza come il lessico determini il livello di competenza in produzione più di altre dimensioni linguistiche (De Jong et al. 2012; Koizumi & In'nami 2013). Indipendentemente dalla natura dialogica o monologica di una prova di produzione, il lessico viene abitualmente valutato in base alla sua ampiezza, ma anche alla sua profondità intesa come precisione lessicale, e fa parte di un costrutto in cui rientrano anche la morfosintassi, la pronuncia, la fluenza. Lo stesso vale per le prove di produzione scritta.

A seconda della natura delle scale di valutazione, olistiche o analitiche, il lessico può essere inserito tra i criteri di valutazione sia in forma autonoma sia in modalità congiunta con altre dimensioni della lingua. La scelta di considerare il lessico come costrutto autonomo o come parte integrante del costrutto dipende essenzialmente dal modello di competenza che viene adottato da chi progetta il test. È evidente che renderlo un criterio autonomo non consente solo di dargli maggior visibilità, ma soprattutto di dargli un maggior peso, al netto delle variazioni che si possono stabilire nel valutare il rilievo della componente lessicale tra un livello e l'altro.

5. Conclusioni

Come abbiamo avuto modo di vedere, seppur nei limiti della sintesi del presente lavoro, la valutazione del lessico è una costante del LTA, anche se con approcci e metodi diversi a seconda del quadro teorico di riferimento. Ad oggi, restano aperte numerose questioni fondamentali: la scelta dell'unità di misura, la definizione del costrutto, la selezione delle parole da valutare, le modalità più opportune per elicitarle le conoscenze degli apprendenti, la validazione dei test lessicali. A ciò aggiungiamo il difficile equilibrio tra test lessicale e test più ampio di cui però valutiamo il lessico, la questione della multidimensionalità della competenza lessicale e le possibilità di una valutazione che si basi su una batteria di prove che consentano di scattare una fotografia da più punti di vista di quello che è lo stato delle competenze lessicali in L2 di un apprendente.

L'ampiezza e la numerosità di tali questioni spingono a sottolineare quanto sia importante un percorso ulteriore di riflessione e di ricerca per consentire di valutare la

competenza lessicale nei modi più opportuni, soprattutto nell'ambito dell'italiano L2, in cui non esiste allo stato attuale una solida tradizione di *Vocabulary Assessment* come quella di ambito anglofono.

Bibliografia

- Aitchinson, Jean. 2003. *Words in the mind*. Oxford: Blackwell.
- Alderson J., Charles. 2000. *Assessing reading*. Cambridge: Cambridge University Press.
- Bardel, Camilla & Lindqvist, Christina & Laufer, Batia (a cura di). 2013. *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis*. Eurosla Monograph Series 2. <http://www.eurosla.org/monographs/EM02/TOC.pdf>.
- Bonk, William. 2000. Second language lexical knowledge and listening comprehension. *International Journal of Listening* 14. 14–31.
- Buck, Gary 2001. *Assessing listening*. Cambridge: Cambridge University Press.
- Daller, Henry & Milton, James & Treffers-Daller, Janine (a cura di). 2007. *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Davies, Alan. 2008. Che cosa ci dicono i test lessicali riguardo alla padronanza linguistica?. In Barni, Monica & Troncarelli, Donatella & Bagna, Carla (a cura di). *Lessico e apprendimenti. Il ruolo del lessico nella linguistica educativa*, 99–110. Milano: Franco Angeli.
- de Jong, Nivja & Steinel, Margarita & Florijin, Arjen & Schoonen, Rob & Hulstijn, Jan. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34(1). 5–34.
- De Mauro, Tullio. 1999. *Grande Dizionario Italiano dell'Uso*. Torino: UTET.
- Gallina, Francesca. 2019. Valutare le parole: la dimensione lessicale nelle certificazioni di italiano L2. In Aldinucci, Benedetta & Carbonara, Valentina & Caruso, Giuseppe & La Grassa, Matteo & Nadal, Celia & Salvatore Eugenio, *PAROLA. Una nozione unica per una ricerca multidisciplinare*, 539–549. Edizioni Università per Stranieri di Siena.
- Gyllstad, Henrik. 2007. *Testing English collocations*. Lund: Lund University. (Tesi di dottorato.)
- Henriksen, Birgit. 1999. Three dimensions of vocabulary development. *Studies on Second Language Acquisition* 21(2). 303–317.
- Koizumi, Rie & In'nami, Yo. 2013. Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching & Research* 4(5). 900–913.
- Kremmel, Benjamin. 2016. Word Families and Frequency Bands in Vocabulary Tests: Challenging Conventions. *TESOL Quarterly* 50(4). 976–987.
- Kremmel, Benjamin & Schmitt, Norbert. 2016. Interpreting Vocabulary Test Scores: What do Various Item Formats Tell Us About Learner's Ability to Employ Words?. *Language Assessment Quarterly* 13(4). 377–392.
- Laufer, Batia & Aviad-Livitzky, Tami. 2017. What Type of Vocabulary Knowledge Predicts Reading Comprehension: Word Meaning Recall or Word Meaning Recognition?. *The Modern Language Journal* 101(4). 729–741.
- Laufer, Batia & Goldstein, Zahava. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54(3). 399–436.

- Meara, Paul. 1996. The dimensions of lexical competence. In Brown, Gillian & Malmkjaer, Kirsten & Williams John (a cura di), *Performance and Competence in Second Language Acquisition*, 35-53. Cambridge: Cambridge University Press.
- Meara, Paul & Miralpeix, Imma. 2017. *Tools for researching vocabulary*. Bristol: Multilingual Matters.
- Matthews, Joshua & Cheng, Junyu. 2015. Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System* 52. 1–13.
- Milton, James. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Miralpeix, Imma. 2020. L1 and L2 Vocabulary Size and Growth, in Webb, Stuart (a cura di), *The Routledge Handbook of Vocabulary Studies*, 189–206. London-New York: Routledge.
- Nassaji, Hossein. 2004. The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review* 61(1). 107–134.
- Nation, Paul. 1983. Testing and teaching vocabulary. *Guidelines* 5. 12–25.
- Nation, Paul. 2001. *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.
- Nation, Paul & Beglar, David. 2007. A vocabulary size test. *The language teacher* 31(7). 9–13.
- Read, John. 2000. *Assessing Vocabulary*, Cambridge: Cambridge University Press.
- Read, John. 2012. Second language vocabulary assessment. *Language Teaching* 46(1). 41–52.
- Read, John. 2013. Assessment of Vocabulary. In Chapelle, Carol (a cura di), *The Encyclopedia of Applied Linguistics*, 251–257. Oxford: Blackwell Publishing.
- Revier, Robert Lee. 2009. Evaluating a new test of whole English collocations. In Barfield, Andy & Gyllstad, Henrik (a cura di), *Researching collocations in another language: Multiple interpretations*, 125–138. Basingtoke: Palgrave Macmillan.
- Spolky, Bernard. 2017. History of language testing. In Shohamy Elana & Iair, G. Or & May, Stephen (a cura di), *Language Testing and Assessment*, 375–384. Cham: Springer.
- Schmitt, Nobert. 2010. *Researching Vocabulary. A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, Norbert & Zimmerman, Cheryl Boyd. 2002. Derivative Word Forms: What Do Learners Know?. *TESOL Quarterly* 36(2). 145–171.
- Stæhr, Lars Stenius. 2009. Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal* 36(2). 139–152.
- Vermeer, Anne. 2001. Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of output. *Applied Psycholinguistics* 22. 217–234.
- Wei, Xuerong. 2020. How does Vocabulary Knowledge Relate to Reading Comprehension?. In Dovigovic, Marina & Agustín-Llach, Maria Pilar (a cura di). *Vocabulary in Curriculum Planning. Needs, Strategies and Tools*, 57–74. Basingtoke: Palgrave Macmillan.
- Wesche, Marjorie & Paribakht, Sima. 1996. Assessing second language vocabulary knowledge: breadth vs depth. *Canadian Modern Language Review* 53. 13–39.
- Wilks, Clarissa & Meara, Paul. 2007. Implementing graph theory approaches to the exploration of density and structure in L1 and L2 word association networks. In Daller,

Harry & Milton, James & Treffers-Daller, Jeanine (a cura di). *Modelling and assessing vocabulary knowledge*, 167–181. Cambridge: Cambridge University Press.

Potenziali *marker* linguistici nello scritto di pazienti con Anoressia Nervosa: uno studio con tecniche di NLP

Giulia Minori

Università di Bologna

giulia.minori@studio.unibo.it

Abstract

Studi recenti hanno mostrato la potenziale rilevanza dell'applicazione di tecniche di Natural Language Processing (NLP) per l'individuazione di condizioni cliniche. Il lavoro qui proposto si inquadra in questo filone di ricerca e si pone l'obiettivo di individuare dei *marker* linguistici potenzialmente caratteristici dell'uso del linguaggio dei soggetti con Anoressia Nervosa (AN) attraverso tecniche di NLP. Nello studio è stato reclutato un campione di 51 partecipanti di sesso femminile, composto da un gruppo affetto dalla patologia di 17 ragazze con diagnosi di AN e da un gruppo di controllo di 34 coetanee normopeso. Ad entrambi i gruppi è stato chiesto di produrre tre brevi testi scritti. Successivamente, sono stati estratti automaticamente dai testi numerosi tratti linguistici con tecniche di NLP ed è stata misurata la significatività statistica con cui questi indici sono potenzialmente in grado di discriminare i soggetti con AN dai controlli. I risultati preliminari mostrano delle alterazioni latenti nell'uso del linguaggio dei soggetti con AN, soprattutto a livello lessicale e sintattico, e potrebbero rappresentare un *marker* affidabile della patologia.

1. Introduzione

In questo lavoro ci si è occupati dell'individuazione di potenziali *marker* linguistici caratteristici del linguaggio di soggetti affetti da Anoressia Nervosa (AN).¹ Lo studio si

¹ Il contributo riporta i risultati preliminari di uno studio ancora in corso e approvato dal Comitato Etico dell'Azienda Ospedaliero-Universitaria di Bologna, Policlinico Sant'Orsola-Malpighi (prot.683/2019/Oss/AOUBo). Questi stessi risultati, già pubblicati in altri articoli (Minori et al. 2020; Cuteri et al. 2021), sono qui riassunti in prospettiva prettamente linguistica.

Non solo per aver reso possibile la realizzazione di questa ricerca, ma anche per i momenti di confronto e riflessione condivisa, i miei ringraziamenti vanno a Gloria Gagliardi (Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna), Vittoria Cuteri e Antonia Parmeggiani (Dipartimento di Medicina e Chirurgia, Università di Bologna e Centro Regionale per i Disturbi del Comportamento Alimentare, Unità di Neuropsichiatria Infantile, Policlinico Sant'Orsola Malpighi di Bologna), Fabio Tamburini

inquadra nell'ambito di ricerca della linguistica clinica, ovvero quella "branca della linguistica che applica teorie e metodi linguistici all'analisi e al trattamento dei disturbi della lingua parlata, scritta e segnata" (Cummings 2016: 87). Oltre ai veri e propri deficit del linguaggio e della comunicazione (Adornetti 2018), generalmente legati a patologie di tipo neurologico, negli ultimi anni la linguistica clinica si è anche occupata dell'individuazione di alterazioni di natura morfosintattica e semantico-lessicale associate a diversi disturbi psichici (Anderson et al. 2008; Arntz et al. 2012; Affronti et al. 2015; Brockmeyer et al. 2015; Bambini et al. 2016; Bernard et al. 2016). Poco considerate, però, sia in contesto italiano che in ambito internazionale, sono le eventuali alterazioni presenti nell'uso del linguaggio di soggetti affetti da disturbi del comportamento alimentare e, nello specifico, da Anoressia Nervosa (Bruch 1962; Buhl 2002; Bydlowski et al. 2005; Lyons et al. 2006; Espeset et al. 2011; Espeset et al. 2012; Wolf et al. 2013; Dirkse et al. 2014; Dakanalis et al. 2016; Spinczyk et al. 2018). Si è quindi scelto di prendere in esame questa patologia.

Sulla base delle evidenze sperimentali raccolte per altre patologie e della fenotipizzazione clinica dei soggetti con Anoressia Nervosa, sembra infatti possibile ipotizzare che l'alterata percezione dell'immagine corporea, la rigidità di pensiero, l'ossessione per i dettagli e la modificazione del tono dell'umore (Pollatos et al. 2008; Vocks et al. 2010; Konstantakopoulos et al. 2012; Bambini et al. 2016; Dakanalis et al. 2016), associate a questa psicopatologia, possano trovare un correlato nell'alterazione delle verbalizzazioni dei pazienti, e che tali modificazioni siano rilevabili mediante analisi del linguaggio.

2. L'Anoressia Nervosa

In questa ricerca si è quindi scelto di prendere in esame proprio l'Anoressia Nervosa. Come si legge nel DSM-5 (American Psychiatric Association 2013), si tratta di una psicopatologia la cui eziologia è a tutt'oggi sconosciuta e che è sempre più diffusa e con esordio sempre più precoce in età evolutiva: solitamente in età puberale dai 9 ai 18 anni (Parmeggiani 2013) e coinvolge maggiormente i soggetti di sesso femminile rispetto a quelli di sesso maschile (in rapporto 10:1).

Inoltre, Parmeggiani (2013) scrive che la mortalità collegata all'Anoressia Nervosa è elevata, ossia circa il 10%, ed è spesso dovuta a suicidio; come si legge nel DSM-5: "suicide risk is elevated in anorexia nervosa, with rates reported as 12 per 100,000 per year" (American Psychiatric Association 2013: 343). Muir et al. (2017) specificano che l'Anoressia Nervosa e i disturbi del comportamento alimentare registrano il più alto tasso di mortalità tra tutti i disturbi di tipo psichico. L'intervento precoce è quindi essenziale, ma gli autori riportano che il 44% delle persone con disturbi del comportamento alimentare non si sottopone a trattamenti psicoterapici per la patologia.

(Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna) e Paola Gualandi, Elisabetta Malaspina, Francesca Rossi, Filomena Mancuso e Valentina Francia (Centro Regionale per i Disturbi del Comportamento Alimentare, Unità di Neuropsichiatria Infantile, Policlinico Sant'Orsola Malpighi di Bologna). Ringrazio infine la Chiara Gianollo (Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna) per i preziosi suggerimenti nella creazione del questionario relativo al plurilinguismo.

2.1. Il profilo linguistico dei soggetti con AN

Sebbene, come si è detto, gli studi relativi al profilo linguistico dei soggetti con AN non siano molto numerosi, è possibile riassumere alcuni tratti interessanti che sono emersi da alcune ricerche portate avanti in merito a questo disturbo.

Due studi operati su corpora di testi raccolti su internet (in particolare blog) hanno riscontrato un uso atipico dei riferimenti alla prima persona singolare da parte di soggetti che affermavano di essere affetti da Anoressia Nervosa. Nel primo di questi i riferimenti alla prima persona singolare sono stati trovati superiori nei blog di soggetti che si definivano in guarigione da AN o in quelli di coloro che dicevano di sostenere l'Anoressia Nervosa come uno stile di vita (pro-ana) rispetto ai blog di controllo (Wolf et al. 2013); nel secondo, sono stati trovati più riferimenti alla prima persona singolare nei blog di soggetti pro-ana rispetto ai blog di soggetti che si definivano in guarigione (Lyons et al. 2006). Come risultato complementare, in entrambe le indagini è stata rilevata una bassa frequenza d'uso dei riferimenti a persone diverse dalla prima e delle espressioni legate alla socialità da parte dei soggetti che si definivano affetti da AN. Questi elementi sono stati interpretati dagli autori come segni di deficit relazionali nei confronti del mondo esterno e degli altri.

Tali risultati appaiono in linea con quelli delle ricerche operate su altre psicopatologie, come la depressione (Brockmeyer et al. 2015; Bernard et al. 2016) o il disturbo d'ansia (Anderson et al. 2008; Brockmeyer et al. 2015), che hanno messo in evidenza il fatto che chi è affetto da questi disturbi tende a utilizzare in media più riferimenti a sé (*self-focused attention*).

Inoltre, sempre nello studio di Wolf et al. (2013), i blog dei soggetti pro-ana, rispetto a quelli scritti da autrici che si definivano in guarigione e a quelli riferibili al gruppo di controllo, presentano in media più punti esclamativi e meno punti interrogativi. Quest'ultimo dato statistico è stato interpretato dagli autori come riflesso, oltre che di una riduzione della complessità da un punto di vista meramente sintattico, anche di una più forte affermazione di sé tipica delle autrici di blog pro-ana da un punto di vista psicologico.

Un'altra caratteristica linguistica che sembra distinguere i soggetti affetti da questa psicopatologia è l'utilizzo maggiore di riferimenti linguistici alle emozioni negative (es. *cattivo, triste*) rispetto a quelli legati alle emozioni positive (es. *buono, felice*) (Brockmeyer et al. 2013; Spinczyk et al. 2018), tuttavia, sembra che i soggetti pro-ana tendano a utilizzare più riferimenti alle emozioni positive, forse come strategia compensativa (Lyons et al. 2006; Wolf et al. 2013). Un'altra peculiarità individuata è l'uso preponderante dei tempi presenti dei verbi rispetto a quelli passati da parte dei soggetti pro-ana rispetto ai soggetti che si definivano in guarigione da AN (Wolf et al. 2013).

Alcune ricerche hanno poi rilevato che l'uso linguistico dei pazienti con AN (soprattutto pro-ana) sembra mostrare un loro scarso ricorso a parole legate ai meccanismi cognitivi (es. *causa, realizzare*) e all'astrazione (Lyons et al. 2006; Wolf et al. 2013). Relativamente a quest'ultimo punto, uno studio specificamente volto a indagare la capacità di astrazione dei soggetti con AN ha ulteriormente confermato dei deficit in questo campo, individuando nell'eloquio dei pazienti la presenza di metafore concretizzate (Skårderud 2007a; 2007b), ossia metafore che non funzionano più come semplici rappresentazioni capaci di contenere un'esperienza, ma in qualità di rappresentazioni di esperienze concrete. In altre parole, lo psicoterapeuta ha riscontrato una traduzione quasi immediata delle emozioni descritte dalle pazienti in elementi corporei (es. l'esperienza concreta di avere lo stomaco vuoto si identifica nella lucidità mentale).

3. Le domande di ricerca

Come spiegato nei paragrafi precedenti, la conoscenza relativa ai processi che portano allo sviluppo e al mantenimento dell'Anoressia Nervosa è a tutt'oggi limitata ed è quindi particolarmente stringente la necessità di individuare nuovi approcci per comprendere meglio e poter quindi intervenire in modo più efficace sui soggetti affetti da AN (Munro et al. 2016).

I risultati ottenuti dagli studi brevemente riassunti nel paragrafo 2.1 hanno permesso di individuare delle tendenze nell'uso linguistico di soggetti con AN, tuttavia sono stati presi in esame testi e verbalizzazioni di parlanti non aventi come L1 l'italiano. Date le peculiarità della lingua italiana (es. a livello morfosintattico), quei risultati non possono essere direttamente generalizzati.

Per queste ragioni, sono state poste le seguenti domande all'origine della ricerca: esistono degli indicatori di tipo linguistico in grado di caratterizzare l'uso del linguaggio di soggetti di lingua italiana e di sesso femminile affetti da Anoressia Nervosa? Il linguaggio da essi utilizzato è assimilabile a quello dei coetanei o presenta delle peculiarità che sarebbero potenzialmente in grado di aprire la strada per un'indagine approfondita, finalizzata al miglioramento degli strumenti diagnostici per l'individuazione della patologia?

4. Lo studio

Per provare a rispondere a queste domande si è scelto di verificare, nell'uso linguistico di soggetti con AN, la presenza di eventuali *marker* linguistici che si discostino in maniera statisticamente significativa rispetto all'uso linguistico di soggetti normotipo. Per raggiungere questo scopo si è ritenuto opportuno concentrarsi sulla produzione scritta e non su quella orale, poiché la registrazione del parlato, ad esempio in sedute cliniche o in interviste strutturate, semi-strutturate o aperte (es. Skårderud 2007a; 2007b; Espeset et al. 2011; Brockmeyer et al. 2013), avrebbe potuto portare a ottenere dei dati linguistici probabilmente distanti dall'uso linguistico quotidiano, in quanto prodotti dai partecipanti in un contesto poco spontaneo e potenzialmente in grado di metterli in soggezione.

In secondo luogo, si è scelto di non raccogliere i dati linguistici dalle piattaforme social, poiché non è possibile avere un riscontro sicuro relativamente ai dati socio-anagrafici degli autori dei testi reperibili online.

4.1. I partecipanti allo studio

I partecipati allo studio, per quanto riguarda il gruppo dei soggetti con AN, sono quindi stati reclutati presso l'U.O.S. per i Disturbi del Comportamento Alimentare della U.O. di Neuropsichiatria Infantile dell'Azienda Ospedaliero-Universitaria Policlinico Sant'Orsola-Malpighi di Bologna. Si tratta di 17 soggetti di sesso femminile affetti dalla patologia (GP). Le loro produzioni scritte sono poi state confrontate con quelle di un gruppo di controllo di 34 soggetti di sesso femminile (GC), reclutato in due scuole secondarie di secondo grado del territorio bolognese (un liceo e un istituto tecnico-professionale) e

bilanciato per età e tipo di scuola rispetto al primo. I due gruppi campionari GP e GC sono in rapporto 1:2, come riportato in Tabella 1.

Età	GP		GC	
	I.T.P.	LICEO	I.T.P.	LICEO
14 anni	1	1	2	2
15 anni	1	5	2	10
16 anni	1	1	2	2
17 anni	1	3	2	6
18 anni	1	2	2	4

Tabella 1. Il bilanciamento del campione.

4.1.1. I criteri di inclusione ed esclusione

Tenendo presenti i dati sopra riportati relativamente all'età di esordio e al sesso, si è scelto di selezionare un campione di soggetti di sesso femminile e di età compresa tra i 14 e i 18 anni. Ai soggetti che in prima persona, o i cui genitori, non hanno acconsentito alla partecipazione allo studio non è stato sottoposto alcun test.

Per quanto concerne il GP tra i criteri di inclusione è stata considerata la presenza di una diagnosi certa di AN, secondo i criteri del DSM-5. Invece, dal GC sono stati esclusi i test di coloro che presentavano un $BMI \leq 18,5$. È stato scelto questo come criterio di esclusione per il GC, poiché, rientrando tra i parametri per la diagnosi di AN secondo il DSM-5, è stato considerato come un possibile indicatore di Anoressia Nervosa non diagnosticata. Il BMI è inoltre un dato facilmente ricavabile sulla base delle informazioni relative a peso e altezza, richieste agli studenti al momento della somministrazione del test.

Dal campione sono poi stati esclusi tutti i soggetti che non presentavano una buona padronanza della lingua italiana, testate tramite un questionario sul plurilinguismo (cfr. paragrafo 4.2). In Tabella 2 sono riportati in sintesi i criteri di inclusione ed esclusione utilizzati per il reclutamento del campione.

	GP	GC
Criteri di Inclusione	<ul style="list-style-type: none"> - Età compresa tra i 14 e i 18 anni, - ottenimento del consenso informato, - diagnosi di Anoressia Nervosa secondo il DSM 5, - buona padronanza della lingua italiana, verificata con il questionario sul plurilinguismo. 	<ul style="list-style-type: none"> - Età compresa tra i 14 e i 18 anni, - ottenimento del consenso informato, - buona padronanza della lingua italiana, verificata con il questionario sul plurilinguismo.
Criteri di Esclusione	<ul style="list-style-type: none"> - Diagnosi non certa, - diagnosi non formulata, - età inferiore ai 14 anni o superiore ai 18, - mancato ottenimento del consenso informato, - difficoltà nella lingua italiana, verificate con il questionario sul plurilinguismo. 	<ul style="list-style-type: none"> - Età inferiore ai 14 anni o superiore ai 18 anni, - mancato ottenimento del consenso informato, - difficoltà nella lingua italiana, verificate con il questionario sul plurilinguismo, - $BMI \leq 18,5$.

Tabella 2. Sintesi dei criteri di inclusione ed esclusione per il campione.

4.1.2. Le modalità di svolgimento dello studio

Il test è stato sottoposto al GP nell'ambito delle normali attività ospedaliere e al GC durante le ordinarie attività scolastiche. A tutti i soggetti maggiorenni è stata preventivamente richiesta la firma di un regolare consenso informato, a seguito della consegna delle carte informative relative alle modalità e agli obiettivi dello studio. Per i soggetti minorenni è invece stata richiesta l'autorizzazione da parte dei genitori, sempre a seguito della consegna delle informative. Il tutto nel rispetto delle vigenti normative sulla privacy. Lo studio è stato condotto in accordo con la Dichiarazione di Helsinki ed è stato approvato dal Comitato Etico dell'Azienda Ospedaliero-Universitaria di Bologna, Policlinico Sant'Orsola-Malpighi (prot.683/2019/Oss/AOUBo).

4.2. Il metodo d'indagine

Per quanto riguarda il metodo d'indagine, ci si è basati sulle ricerche in ambito psicopatologico precedentemente svolte in relazione all'Anoressia Nervosa e ad altri disturbi psichici, oltre che sulle osservazioni degli psicologi e degli psichiatri dell'U.O.S. per i Disturbi del Comportamento Alimentare. Nella creazione del test di produzione scritta, l'obiettivo è stato quello di approntare uno strumento che permettesse di indagare l'uso del linguaggio nel modo più ecologico possibile, per poter avere accesso ai reali *pattern* di utilizzo del linguaggio dei partecipanti allo studio, non viziati ad esempio da artifici stilistici (tipici, ad esempio, delle poesie e dei testi delle canzoni, spesso analizzati in studi precedenti) e possibilmente non influenzati dallo stato emotivo dei soggetti.

In merito a quest'ultimo punto, infatti, sono stati creati dei *task* indipendenti dai classici test utilizzati in ambito clinico per diagnosticare l'Anoressia Nervosa, al fine di evitare che i pazienti (e gli studenti) potessero sentirsi a disagio e produrre così dei testi in qualche modo alterati. I modelli, per quanto riguarda il paradigma dell'ecologia del metodo d'indagine in linguistica clinica, sono stati, tra gli altri, lo studio di Arcara & Bambini (2016) sulla schizofrenia e quello di Beltrami et al. (2018) sulle patologie di natura dementigena.

Il prodotto è un test linguistico articolato in tre *task*, scelti per elicitare la produzione scritta dei soggetti coinvolti nello studio:

- A. Parla di te: come sei, fisicamente e caratterialmente, cosa ami e non ami fare (*task* autobiografico personale);
- B. Cosa fai solitamente quando incontri i tuoi amici (*task* autobiografico neutro);
- C. Descrivi la figura sottostante (*The cookie theft* in Figura 1).²

Per tutti e tre i punti si è richiesto di scrivere indicativamente 10/15 righe.

Sia nel test proposto ai soggetti del GP che in quello proposto ai soggetti del GC, è stata prevista, inoltre, una sezione volta a isolare gli effetti del plurilinguismo, in quanto tale aspetto caratterizza in modo pervasivo l'ambiente scolastico in questo specifico momento storico. Si tratta di un breve questionario finalizzato ad accertare il livello di competenza dei soggetti nella lingua italiana. Questa parte introduttiva ha due scopi: in primo luogo, condurre a una relativa omogeneità linguistica tra i partecipanti reclutati per i due

² L'immagine in bianco e nero è stata ripresa dalla batteria BDAE (Goodglass et al. 2001).

gruppi; in secondo luogo, permettere di ottenere un campione di soggetti aventi l'italiano come L1.

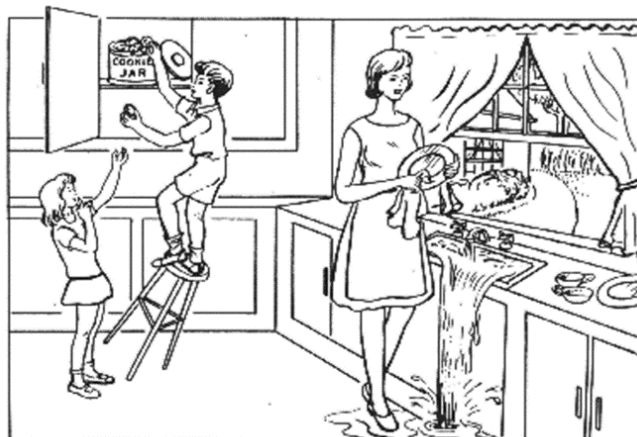


Figura 1. *The cookie theft.*

4.3. *L'analisi dei dati con tecniche di NLP*

Come hanno sottolineato Munro et al. (2016), date le poche conoscenze relative alle cause che portano allo sviluppo e al mantenimento di questa psicopatologia, è oggi stringente la necessità di sperimentare nuovi approcci per poter intervenire in modo sempre più efficace sui soggetti affetti da AN. Facendo quindi riferimento ai proficui contributi che il *Natural Language Processing* (NLP) ha apportato negli ultimi anni agli studi di linguistica clinica, si è deciso di servirsi di strumenti simili anche in questo lavoro.

Le tecniche di NLP, applicate ai vari livelli dell'analisi linguistica, si sono infatti rivelate efficaci non solo nell'individuazione di patologie neurodegenerative, quali ad esempio la demenza di Alzheimer e il *Mild Cognitive Impairment*, ma anche nella ricerca di eventuali alterazioni dell'uso del linguaggio caratteristiche di soggetti affetti da disturbi psichici. Come ha dimostrato lo studio di Beltrami et al. (2018) sulle demenze, l'applicazione del NLP alla linguistica clinica permette un'analisi dei testi sistematica ed ecologica, oltre che a basso costo e applicabile su larga scala. Anche gli studi in ambito psicopatologico (es. Pennebaker et al. 2003; Tausczik & Pennebaker 2010) hanno reso evidenti i vantaggi dell'utilizzo di tecniche di NLP (es. *Linguistic Inquiry and Word Count* e *Sentiment Analysis*), quali ad esempio la possibilità di un'analisi quantitativa, trasparente e più oggettiva dei testi.

Nello specifico, il corpus di testi raccolti è stato arricchito esplicitando informazioni linguistiche riguardanti il livello lessicale, morfosintattico e sintattico. Lo strumento computazionale che è stato scelto per l'annotazione è il *Turin University Linguistic Environment* (TULE)³ (Lesmo 2007), attraverso cui tutte le frasi complesse sono state sottoposte al *PoS tagging*, ossia annotazione rispetto alle parti del discorso, alla lemmatizzazione e al *parsing* sintattico, cioè un'analisi sintattica a dipendenze nel quadro teorico introdotto da Tesnière (De Santis, 2016). TULE è un *parser* per la lingua italiana che utilizza il *tagset* di TUT, *Turin University TreeBank* (Bosco et al. 2000). A seguito dell'analisi linguistica operata con TULE (cfr. Figura 2) i file, corretti manualmente con

³ <http://corpora.ficlit.unibo.it/COPA/TULEParser.html>.

il software DGA, *Dependency Grammar Annotator* (cfr. Figura 3),⁴ sono stati sottoposti a un'analisi dei parametri multidimensionale: l'algoritmo che è stato utilizzato analizza quantitativamente i testi scritti, calcolando i tratti lessicali, morfosintattici e sintattici.

```

1  Il  IL  ART  ART  DEF|M|SING  6  RMOD
2  sabato  SABATO  NOUN  NOUN  COMMON|M|SING  1  ARG  -  -
3  sera  SERA  NOUN  NOUN  COMMON|F|SING  2  RMOD  -  -
4  molto  MOLTO  ADV  ADV  QUANT  5  RMOD  -  -
5  spesso  SPESSE  ADV  ADV  TIME  6  RMOD  -  -
6  andiamo  ANDARE  VERB  VERB  MAIN|IND|PRES|INTRANS|1|PL  0  TOP  -  -
7  in  IN  PREP  PREP  MONO  6  INDCOMPL
8  discoteca  DISCOTECA  NOUN  NOUN  COMMON|F|SING  7  ARG  -  -
9  .  #\  PUNCT  PUNCT  -  6  END  -  -

```

Figura 2. Screenshot dell'annotazione linguistica di una frase, operata da TULE.

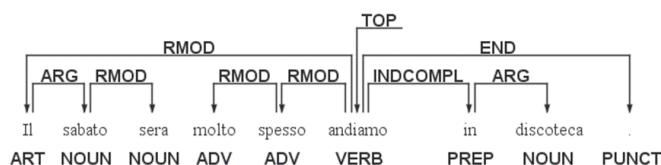


Figura 3. Screenshot dell'annotazione linguistica visualizzata con DGA.

Un altro strumento che è stato utilizzato per l'analisi computazionale dei testi è il software LIWC (*Linguistic Inquiry and Word Count*).⁵ La versione che è stata utilizzata in questo lavoro è quella del 2015 (Pennebaker et al. 2015a; Pennebaker et al. 2015b), tuttavia, il dizionario e le categorie disponibili per la lingua italiana risalgono alla versione del 2007, LIWC2007 (Agosti & Rellini 2007). LIWC lavora sfruttando due componenti: il programma stesso, che può aprire differenti tipi di file, in diversi formati ed è in grado di analizzarli parola per parola;⁶ e un dizionario con cui ogni parola (*target word*) individuata viene confrontata. Il dizionario è suddiviso per categorie all'interno delle quali sono raccolte le parole (*dictionary words*) che le caratterizzano. In Figura 4 è riportato un esempio di output restituito dal software LIWC che, nella maggior parte dei casi, restituisce la frequenza relativa (in percentuale e rispetto al numero totale dei *token* di ogni testo) con cui ogni categoria (prima riga della tabella in Figura 4) presente nel dizionario compare nei singoli testi analizzati (prima colonna della tabella in Figura 4).

Filename	Segment	WC	WPS	Sixtr	Dic	pronomi	Io	Noi	Se	Tu	Negazio	Affett	Sen_Pos	Emo_Pos	Emo_Neg
AMB 10_2A...	1	77	9,62	18,18	63,64	5,19	3,90	0,00	0,00	0,00	1,30	7,79	5,19	5,19	2,60
AMB 10_2B....	1	45	11,25	22,22	57,78	6,67	4,44	2,22	4,44	0,00	0,00	4,44	4,44	0,00	0,00
AMB 10_2C...	1	87	17,40	25,29	56,32	5,75	0,00	1,15	1,15	0,00	0,00	0,00	0,00	0,00	0,00
AMB 12_2A....	1	69	17,25	34,78	55,07	10,14	4,35	0,00	0,00	0,00	0,00	4,35	2,90	2,90	1,45
AMB 12_2B....	1	32	16,00	25,00	56,25	6,25	3,12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AMB 12_2C...	1	53	17,67	20,75	58,49	11,32	1,89	0,00	0,00	0,00	1,89	3,77	1,89	1,89	1,89
AMB 13_2A....	1	21	10,50	33,33	61,90	9,52	4,76	0,00	0,00	0,00	0,00	9,52	9,52	9,52	0,00
AMB 13_2B....	1	25	12,50	40,00	60,00	4,00	0,00	4,00	4,00	0,00	0,00	0,00	0,00	0,00	0,00
AMB 13_2C...	1	39	13,00	25,64	66,67	10,26	0,00	2,56	2,56	0,00	0,00	2,56	2,56	0,00	0,00
AMB 14_2A....	1	110	13,75	24,55	70,00	10,00	7,27	0,00	1,82	0,00	4,55	3,64	3,64	1,82	0,00
AMB 14_2B....	1	73	14,60	30,14	58,90	10,96	5,48	4,11	6,85	0,00	0,00	5,48	5,48	2,74	0,00
AMB 14_2C...	1	76	15,20	31,58	56,58	2,63	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Figura 4. Esempio di output generato da LIWC per il gruppo delle pazienti.

⁴ <http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/>.

⁵ <https://liwc.wpengine.com/>.

⁶ In relazione al software LIWC per *parola* s'intende un'unità compresa tra due spazi bianchi (Pennebaker et al. 2015b).

Poiché, come si è detto, la letteratura esistente circa le caratteristiche dell'uso del linguaggio di soggetti con Anoressia Nervosa è scarsa, si è scelto di includere nella ricerca, oltre agli indici già indagati in altri studi precedenti e riassunti nel paragrafo 2.1, anche diversi indici lessicali e sintattici estratti da un algoritmo sulla base delle annotazioni. Questi tratti non erano mai stati considerati in relazione ad AN, ma erano già stati indagati nell'ambito della linguistica clinica, seppure nel contesto di altre patologie (es. Beltrami et al. 2018). Nelle tabelle sottostanti (Tabelle 3 e 4) sono elencati e descritti tutti i parametri lessicali e sintattici che sono stati presi in considerazione (Beltrami et al. 2018).

TRATTI	DESCRIZIONI	SIGLA
<i>Densità lessicale</i>	Il rapporto tra le parole appartenenti a una classe aperta (parole lessicali) e quelle appartenenti a una classe chiusa (parole grammaticali). La misura è calcolata sulla base dei tag delle PoS (<i>Part of Speech</i>), considerando come appartenenti a una classe aperta i nomi, i verbi, gli aggettivi e gli avverbi, come appartenenti a una classe chiusa le espressioni caratterizzate da altre PoS.	LEX_ContDens
<i>Frequenza delle parti del discorso</i>	Si tratta di una classe di indici che indaga la frequenza media di occorrenza di ogni categoria di PoS: aggettivi, avverbi, articoli, congiunzioni, interiezioni, nomi, numerali, parole che esprimono intere frasi (es. no), predeterminanti, preposizioni, pronomi e verbi.	LEX_PoS_ADJ, LEX_PoS_ADV, LEX_PoS_ART, LEX_PoS_CONJ, LEX_PoS_INTERJ, LEX_PoS_NOUN, LEX_PoS_NUM, LEX_PoS_PHRAS, LEX_PoS_PREDET, LEX_PoS_PREP, LEX_PoS_PRON, LEX_PoS_VERB
<i>Frequenza dei riferimenti alla realtà</i>	Il rapporto tra il numero totale dei nomi e il numero totale dei verbi.	LEX_RefRReal
<i>Frequenza della deissi personale, spaziale e temporale</i>	Questi indici restituiscono la frequenza relativa delle espressioni deittiche nei testi scritti. I principali tipi di deissi sono i seguenti: deissi personale (es. io, tu, nostro, te); deissi spaziale (es. qui, là, questo, quello); deissi temporale (es. ora, oggi, presto).	LEX_PDEIXIS, LEX_SDEIXIS, LEX_TDEIXIS
<i>Ricchezza lessicale: Rapporto Type-Token, W Indice di Brunet e R Statistica di Honoré</i>	Questa classe di misure quantifica la ricchezza del vocabolario e la varietà del lessico: - TTR, <i>Rapporto Type-Token</i> : il rapporto tra il numero delle parole (vocabolario - V) e la lunghezza totale del testo. TTR dipende quindi dalla dimensione del testo: è maggiore se i testi sono brevi e minore se sono lunghi; - W, <i>Indice di Brunet</i> : quantifica la ricchezza lessicale indipendentemente dalla lunghezza del testo e si calcola mediante la seguente equazione: $W = \frac{V}{N} \cdot V^{(-.165)}$ dove N è la lunghezza totale del testo e V è il vocabolario totale utilizzato dal soggetto. Questa misura oscilla generalmente tra 10 e 20: più è basso il valore più il testo è lessicalmente ricco; - R, <i>Statistica di Honoré</i> : calcola la ricchezza lessicale ponendo l'accento sulla proporzione tra le parole utilizzate una sola volta e il numero totale di parole del testo: maggiore è il numero di parole utilizzate nel testo una sola volta (<i>hapax legomena</i>) e più ricco è il lessico. $(R=100\log N) / (1 - V1 = V)$ dove V1 sono le parole scritte una sola volta, V è il vocabolario totale e N è la lunghezza totale del testo. Un alto valore di R corrisponde all'uso di un vocabolario ricco da parte del soggetto.	LEX_TTR, LEX_BrunetW, LEX_HonorER
<i>Frequenza dei Verbi di Azione</i>	Frequenza dei verbi di azione, ossia verbi indicanti azioni fisiche (es. correre, mangiare).	LEX_ACTVRB
<i>Identificazione della Frequenza d'uso</i>	Peso medio della frequenza d'uso delle parole estratte dalla lista di frequenza di De Mauro.	LEX_DM_F
<i>Densità di concetti proposizionali</i>	Questo tratto indica il numero di proposizioni (es. fatti o nozioni contenute in un testo) diviso per il numero delle parole (token). È una misura del grado con cui il soggetto compie asserzioni invece che semplicemente riferirsi a delle entità. In questo contesto, le proposizioni corrispondono a verbi, aggettivi, avverbi, preposizioni e congiunzioni. I nomi non sono considerati tali, perché il verbo principale e tutti i suoi argomenti contano come un'unica proposizione.	LEX_IDEAD
<i>Numero medio di parole negli enunciati</i>	Numero medio di parole in una frase.	LEX_NW

Tabella 3. Trattati lessicali.

TRATTI	DESCRIZIONE	SIGLA
<i>Lunghezza del Sintagma Nominale</i>	Numero di elementi dipendenti collegati al nome. Questo tratto indaga la complessità del sintagma nominale, contando il numero di elementi dipendenti collegati alla testa (es. aggettivi, frasi relative). Per questa misura sono tenute in considerazione sia la media che la deviazione standard.	SYN_NPLENM SYN_NPLENSD
<i>Distanza Globale delle Dipendenze</i>	Questo indice quantifica la difficoltà dell'elaborazione sintattica delle dipendenze a lunga distanza. Per questa misura sono tenute in considerazione sia la media che la deviazione standard.	SYN_GRAPHDISTM, SYN_GRAPHDISTSD
<i>Complessità Sintattica</i>	La complessità sintattica è definita contando i token linguistici che possono essere considerati come indicatori di subordinazione e di radicamento grammaticale: congiunzioni subordinanti (es. perché, quando, se), pronomi interrogativi (es. chi), forme verbali, sia finite che non finite, e sintagmi nominali. Poiché i primi e i secondi sono maggiormente indicativi di un'alta complessità, a essi è dato un peso maggiore rispetto agli ultimi due. (2*cong + 2*pron + nomi + verbi) / parole	SYN_ISynCompl
<i>Radicamento sintattico: massima profondità della struttura</i>	La complessità sintattica è individuata anche valutando nello specifico il radicamento, ad esempio la profondità massima di una struttura sintattica. Per questa misura sono tenute in considerazione sia la media che la deviazione standard.	SYN_MAXDEPTM SYN_MAXDEPTHSD
<i>Lunghezza dell'Enunciato</i>	La lunghezza media della frase corrisponde al numero medio di parole per frase. Viene calcolata dividendo il numero di parole in ogni frase per il numero totale di frasi. Per questa misura sono tenute in considerazione sia la media che la deviazione standard.	SYN_SLENM SYN_SLENSD

Tabella 4. Tratti sintattici.

Nella Tabella 5 sono invece riportate tutte le categorie di LIWC che sono state prese in considerazione in questa ricerca. Nella prima colonna sono presenti le categorie originali basate sulla lingua inglese, nella seconda le corrispondenti categorie italiane, basate sul dizionario LIWC2007 (Agostini & Rellini 2007) con indicato tra parentesi il nome dell'etichetta presente nel software, nell'ultima colonna si trovano invece alcuni esempi di parole del dizionario abbinate a ogni categoria.

5. Risultati dell'analisi statistica dei dati

L'ipotesi generale, basata sulla letteratura esistente e sulla fenotipizzazione dell'Anorexia Nervosa, è che esistano degli indici linguistici in grado di distinguere l'uso del linguaggio (nello scritto) dei soggetti con patologia (AN) da quello dei soggetti senza patologia. Le singole ipotesi formulate in relazione alle differenze tra il GP e il GC riguardano la differenza tra i valori medi relativi a ogni indice presente nelle tabelle 3, 4 e 5. Tali ipotesi sono state testate con l'applicazione del test non parametrico *Kolmogorov-Smirnov* (*ks.test*).

CATEGORIE ORIGINALI	CATEGORIE ITALIANE (sigle)	ESEMPI
Linguistic processes	Processi linguistici	
<i>Total words count (WC)</i>	Conteggio delle parole totali (WC)	
<i>Words per sentence (WPS)</i>	Parole per frase (WPS)	
<i>Words > 6 letters (Sixltr)</i>	Parole con più di sei lettere (Sixltr)	
<i>Dictionary words count (Dic)</i>	Conteggio delle parole presenti anche nel dizionario LIWC2007 (Dic)	
Function Words	Parole funzione	
<i>1st pers singular (i)</i>	1a pers sing (Io+IoVer=Io Tot)	io, me, mi, mio, amo
<i>1st pers plural (we)</i>	1a pers plur (Noi+NoiVerb=Noi Tot)	noi, ci, nostro, abbiamo
<i>2nd person (you)</i>	2a pers sing (Tu+TuVerbo)	tu, facesti, sarai
	2a pers plur (Voi+VoiVerb)	voi, vi, vostro, veniste
<i>3rd pers singular (he, she, it)</i>	3a pers sing (Luilei+LuiVerb)	lui, glielo, voleva
<i>3rd pers plural (they)</i>	3a pers plur (Loro+LoroVer)	loro, hanno
<i>Negations</i>	Negazioni (Negazio)	no, non, niente, negaz*
<i>Past tense</i>	Tempo passato (Passato)	feci, offerto, ieri
<i>Present tense</i>	Tempo presente (Presente)	descrive, oggi
<i>Future tense</i>	Tempo futuro (Futuro)	vincerete, domani
	Gerundio (gerundio)	difendendo, discutendo
	Condizionale e congiuntivo (Condizio)	mangerei, apriate
	Participio passato (Ppass)	arrivato, caduto
	Transitivi (Transiti)	occupare, odio
Affective processes	Processi emozionali (Affett)	
<i>Positive emotions</i>	Emozioni positive (EmoPos)	rallegrare, amore, cari
<i>Negative emotions</i>	Emozioni negative (EmoNeg)	esasper*, gravemente
<i>Anxiety</i>	Ansia (Ansia)	impaurit*, inadeguat*
<i>Anger</i>	Rabbia (Rabbia)	ira, litigare
<i>Sadness</i>	Tristezza (Tristez)	mancanza, piangere
Cognitive Processes	Meccanismi cognitivi (MecCog)	
<i>Insight</i>	Introspezione (Intros)	ragion*, penso
<i>Cause</i>	Causa (Causa)	perché, quindi, risultat*
<i>Discrepancies</i>	Discrepanze (Discrep)	speranza, desiderat*
<i>Tentativeness</i>	Possibilità (possib)	se, sembra, tentiamo
<i>Certainty</i>	Certezza (Certez)	veramente, cert*, convint*
<i>Inhibition</i>	Inibizione (Inibiz)	problem*, rifiutare, rinunc*
<i>Inclusive</i>	Inclusione (Inclusi)	dentro, anche, complessiv*
<i>Exclusive</i>	Esclusione (Escusi)	escludere, ma, dividere
Perceptual processes	Processi sensoriali (ProcSen)	
<i>See</i>	Vista (Vista)	vede, leggere, sguard*
<i>Hear</i>	Udito (Udito)	chiacchier*, ascolto, telefono
<i>Feel</i>	Sentimento (Sentim)	afferr*, sentiamo, tocco, stretto
Biological processes	Processi biologici	
<i>Body</i>	Corpo (Corpo)/Fisico (Fisico)	ammalat*, grass*, anoressi*
<i>Health</i>	Cura (Curacor)	medico, capell*, bagn*
<i>Ingestion</i>	Mangiare (Mangiare)	sovrappes*, cucin*, fame
Personal concerns	Interessi personali	
<i>Work</i>	Lavoro (Lavoro)/Scuola (Scuola)	occuparsi, profession*, compit*
<i>Achievement</i>	Raggiungimento (Raggiun)	vincent*, soddisfacente
<i>Leisure</i>	Svago (Svago)	party, jazz, golf
<i>Home</i>	Casa (Casa)	vivere, camera, cucin*
Psychological processes	Processi psicologici	
<i>Family</i>	Famiglia (Famigl)	figli, mamm*, patern*
<i>Friends</i>	Amici (amici)	amica, compagno, fidanzat*
<i>Humans</i>	Esseri umani (Umano)	individuale, persona
<i>Social processes</i>	Processi sociali (Social)	sua, tutti, aiuta, utile

Tabella 5. Categorie di LIWC2007.

In questo lavoro sono stati considerati come potenziali *marker* dell'uso del linguaggio di soggetti con AN solo quegli indici linguistici per i quali i dati quantitativi sono risultati distribuiti in maniera differente per il GP e il GC e per i quali questa differenza può essere definita almeno statisticamente significativa sulla base del *p-value* restituito in seguito

dell'applicazione del *ks.test*. In Tabella 6 sono riportati gli indici statisticamente significativi.⁷

	Task A, B, C	Task A	Task B	Task C
LEX_PoS_CONJ	D = 0.235 p = 0.046* GP < GC			
LEX_PoS_ADV		D = 0.412 p = 0.043* GP < GC		
LEX_HonoreR	D = 0.255 p = 0.022* GP > GC			D = 0.412 p = 0.038* GP > GC
LEX_ContDens				D = 0.441 p = 0.024* GP > GC
LEX_PDEIXIS			D = 0.412 p = 0.043* GP < GC	
SYN_SELNM	D = 0.284 p = 0.008** GP < GC	D = 0.412 p = 0.043* GP < GC		
SYN_SLENSD	D = 0.245 p = 0.034* GP < GC	D = 0.412 p = 0.038* GP < GC		
SYN_NPLENSD			D = 0.411 p = 0.043* GP < GC	
SYN_GRAPHDISTM	D = 0.284 p = 0.008** GP < GC			
WPS	D = 0.245 p = 0.034* GP < GC		D = 0.412 p = 0.043* GP < GC	
Sixltr	D = 0.333 p = 0.001** GP > GC			D = 0.441 p = 0.024* GP > GC
ProcSen			D = 0.412 p = 0.043* GP < GC	
Present				D = 0.412 p = 0.043* GP < GC
Dic			D = 0.441 p = 0.024* GP < GC	D = 0.588 p = 0.0008*** GP < GC

Tabella 6. Indici statisticamente significativi.

6. Discussione dei risultati

Le domande che hanno infine guidato l'osservazione e la discussione dei risultati ottenuti sono due: è possibile, in ottica clinica, rilevare dei tratti potenzialmente caratteristici dell'uso del linguaggio di soggetti con Anoressia Nervosa, in grado di discriminare in modo statisticamente significativo le loro produzioni scritte da quelle dei soggetti

⁷ Sotto ogni *p-value* è indicato il livello di significatività mediante asterischi: * (*p-value* significativo, ossia $0,01 < p < 0,05$), ** (*p-value* molto significativo, ossia $0,001 < p < 0,01$), *** (*p-value* altamente significativo, ossia $p < 0,001$).

normotipo? In prospettiva linguistica, queste eventuali tendenze riflettono dei *pattern* in qualche modo tipici e già riscontrati in altri contesti?

Considerando la prima domanda, innanzitutto, prima di entrare nel merito dei risultati relativi ai singoli indici, è interessante notare che il numero maggiore di indici per i quali si è riscontrata una differenza statisticamente significativa tra il GP e il GC fa riferimento ai tre *task* considerati insieme. Questi sono infatti sette, dei quali, inoltre, due sono molto significativi (indici SYN_GRAPHDISTM e SYN_SLENM) e uno è altamente significativo (indice Sixltr). Il test nel suo complesso sembra dunque essere uno strumento efficace e si può constatare che permette di attenuare in parte gli effetti della scarsità dei dati, legata alla ridotta dimensione del corpus e intrinseca alla linguistica dei corpora in generale.

Facendo poi riferimento ai tre *task* considerati singolarmente, il numero maggiore di indici significativi si trova in relazione ai dati linguistici del *task C*, ossia la descrizione di figura complessa, e del *task B*, cioè quello autobiografico neutro. Infatti, in Tabella 6, è possibile vedere che gli indici significativi per il *task C* sono cinque, così come sono cinque per il *task B* e sono invece solo tre per il *task A*, ossia quello autobiografico personale. Inoltre, l'indice Dic, comune al *task C* e al *task B*, è risultato nel primo altamente significativo ($p = 0.0008$), mentre nel secondo è risultato solo significativo ($p = 0.024$). Ciò permette di richiamare la tesi di Chung & Pennebaker (2007), i quali sostengono che il modo più adatto per indagare l'uso del linguaggio sia quello di chiedere ai partecipanti a uno studio non tanto di parlare direttamente dei propri stati personali (come esplicitamente richiesto nel *task A*), ma di descrivere, ad esempio, immagini ambigue e non caratterizzate da una coerenza interna (es. *The cookie theft* nel *task C*), in modo che i soggetti siano meno portati a controllare il loro uso linguistico.

In relazione agli studi precedenti su soggetti con AN, riassunti nel paragrafo 2.1, è poi interessante notare che, relativamente al solo *task C*, è risultata statisticamente significativa la differenza tra il GP e il GC nell'uso dei riferimenti al tempo presente (indice Present). Laddove la ricerca di Lyons e colleghi (2006) aveva individuato un uso maggiore di questi da parte delle autrici di blog pro-ana rispetto alle autrici che si definivano in guarigione dall'Anoressia Nervosa, in questa ricerca è risultato invece che i soggetti del gruppo di controllo utilizzano in media più espressioni linguistiche riferite al tempo presente rispetto alle pazienti con AN. Invece, relativamente all'uso dei riferimenti alla prima persona singolare, non è stata trovata alcuna differenza statisticamente significativa. Dagli esiti di questa ricerca sembra dunque che la *self-focused attention*, non sia preponderante nei soggetti con AN testati.

Inoltre, a differenza degli studi precedenti, non sono state trovate differenze statisticamente significative nemmeno per quanto riguarda l'uso di pronomi personali non di prima persona singolare o di espressioni riferibili alla sfera della socialità, né per quanto concerne l'uso di riferimenti alle emozioni positive e negative o ai processi cognitivi e di astrazione, né in merito all'utilizzo della punteggiatura.

A livello lessicale, dai risultati ottenuti sembra che i soggetti con AN utilizzino in media meno elementi linguistici appartenenti ad una classe di parole chiusa, soprattutto per quanto riguarda le congiunzioni (indice LEX_PoS_CONJ) e la deissi personale (indice LEX_PDEIXIS), e, quindi, più elementi linguistiche appartenenti a una classe di parole aperta, il che comporta la presenza di una maggiore densità lessicale nelle produzioni scritte del GP (indice LEX_ContDens). Sembra poi che i testi delle pazienti presentino in media una ricchezza lessicale maggiore, per quanto riguarda l'utilizzo di *hapax legomena* (indice LEX_HonoreR), un quantitativo maggiore di parole con più di sei

lettere (indice Sixltr) e un numero inferiore di parole contenute anche nel dizionario italiano (Agostini & Rellini, 2007) abbinato al software LIWC (indice Dic). Un ultimo indice significativo di tipo semantico-lessicale è poi quello relativo alle espressioni legate ai processi sensoriali e percettivi (indice ProcSen), per cui è risultato che i soggetti con AN utilizzano in media meno espressioni linguistiche di questo tipo.

A livello sintattico, invece, è possibile osservare che le pazienti hanno prodotto frasi mediamente più brevi (indici SYN_SELNM e WPS) e caratterizzate da relazioni di dipendenza non a lunga distanza (indice SYN_GRAPHDISTM). Inoltre, i testi prodotti dai soggetti del GP presentano frasi molto simili in termini di lunghezza, mentre le studentesse del GC hanno scritto frasi di lunghezza più eterogenea; questo si deduce dall'indice SYN_SLENSD, il quale rivela infatti che i dati relativi al GP, per quanto riguarda la lunghezza media delle frasi, sono meno dispersi rispetto a quelli del GC. Analogamente, il GP è risultato più omogeneo in termini di lunghezza media dei sintagmi nominali (SYN_NPLENSD). Tuttavia, a differenza della lunghezza media delle frasi, non è possibile affermare che il GP abbia la tendenza a creare dei sintagmi nominali mediamente più brevi rispetto al GC: la lunghezza risulta infatti mediamente inferiore nei testi delle pazienti ma non in modo statisticamente significativo (l'indice SYN_NPLENM non è infatti presente in Tabella 6).

Considerando poi la seconda domanda, ossia se, in prospettiva linguistica, sia possibile riscontrare nei dati ottenuti dei *pattern* in qualche modo tipici e già riscontrati in altri contesti, è stato possibile notare in essi delle tendenze già evidenziate in relazione ai dibattuti concetti della semplicità e della complessità. I risultati statistici ottenuti in questo studio sono infatti stati confrontati sia con alcuni dei criteri di semplificazione linguistica riassunti da Berruto (1990) e citati, tra gli altri, anche da Voghera (2001) e Fiorentino (2009), sia con le fasi dello sviluppo linguistico dei bambini (Caselli & Casadio 1995; Caselli et al., 1999; Gagliardi 2019). Da questi confronti è emerso che il linguaggio dei testi scritti dai soggetti con AN sembra essere più semplice a livello sintattico per vari motivi.

Innanzitutto, sono in esso presenti meno elementi linguistici di tipo grammaticale, o parole funzione, (ossia quegli elementi a cui è primariamente affidato il compito di dare coesione morfosintattica ai testi), e più elementi linguistici di tipo referenziale, o parole contenuto. Berruto (1990) definisce le parole contenuto più semplici delle parole funzione, poiché le prime “consentono contemporaneamente un qualche appoggio all’extralinguistico e al situazionale, e sono più importanti semanticamente per il passaggio essenziale dell’informazione” (Berruto 1990: 26). Sembra inoltre che gli elementi linguistici di tipo referenziale, per la loro maggiore semplicità, vengano imparati dai bambini prima di quelli grammaticali. Infatti, la fase dell’acquisizione delle parole funzione (*grammar*) è l’ultima delle quattro individuate da Caselli et al. (1999), mentre quella di acquisizione delle parole contenuto (*reference*) è la seconda.

In secondo luogo, i testi prodotti dalle pazienti sono caratterizzati da frasi mediamente più brevi e non caratterizzate da relazioni di dipendenza sintattica a lunga distanza e questi due dati si prestano a essere letti come indici di una relativa semplificazione sintattica. Da un lato, come scrive Gagliardi (2009: 251), “numerosi studi hanno dimostrato che una crescente complessità delle strutture sintattiche prodotte dal bambino è associata a una maggiore lunghezza degli enunciati”; dall’altro, come afferma Berruto (1990), la contiguità sintagmatica di due elementi semanticamente legati e/o sintatticamente dipendenti è più semplice rispetto alla loro distanziamento, poiché “la contiguità di elementi

semanticamente o sintatticamente ‘vicini’ permette una più immediata recuperabilità e accessibilità dei rapporti” (Berruto 1990: 29).

Questa relativa semplificazione sintattica, però, sembra essere accompagnata da elementi che paiono dare complessità al lessico utilizzato dalle pazienti. Basti pensare che – come ha scritto Voghera (2001) e come si può osservare dai risultati di questa ricerca riportati sopra – la diretta conseguenza di una diminuzione degli elementi linguistici di tipo grammaticale, ossia appartenenti a una classe di parole chiusa, è l’aumento della densità lessicale dei testi. Oltre a questo dato, sembra poi che le pazienti utilizzino in media un numero maggiore di parole con più di sei lettere e più *hapax legomena*, che rendono i testi lessicalmente più ricchi. Di contro, nei testi scritti da loro il software LIWC ha individuato meno parole abbinabili a quelle contenute nel dizionario LIWC2007 e ciò, secondo Tausczik & Pennebaker (2010), sarebbe indice di un registro più formale e di un lessico più tecnico-specifico. Riprendendo sempre Berruto (1990), la presenza di un lessico più tecnico-specifico sarebbe indice di una maggiore complessità lessicale: lo studioso sostiene infatti che i termini generali e il lessico comune siano rispettivamente più semplici dei termini specifici e del lessico specialistico, in quanto i primi richiedono un minore sforzo di concettualizzazione.

7. Conclusione

In prospettiva linguistica e relativamente al campione qui considerato, è possibile avanzare l’ipotesi che i soggetti affetti da Anoressia Nervosa utilizzino un linguaggio più semplice dal punto di vista sintattico, semplificazione che sembra però essere compensata da una maggiore complessità lessicale.

Per via della ridotta numerosità dei due gruppi (GP e GC) non è ovviamente possibile generalizzare i risultati qui ottenuti, in quanto il campione, per quanto bilanciato, non è sufficientemente ampio da poter essere definito rappresentativo della popolazione generale. Un altro svantaggio di questa ricerca sta nel fatto che i soggetti reclutati nelle scuole, per motivi etici, non sono stati sottoposti ad alcun test specifico per la diagnosi dell’Anoressia Nervosa. L’unico criterio di esclusione è stato quello di un $BMI \leq 18,5$. Si tratta certamente di un criterio diagnostico presente nel DSM-5, ma non costituisce un parametro esaustivo e non permette di escludere con certezza la presenza latente della patologia all’interno del gruppo di controllo.

Dall’altro lato, un vantaggio non secondario di questa ricerca è che – a differenza di diversi studi precedentemente condotti per individuare gli eventuali *marker* caratteristici dell’uso del linguaggio dei soggetti con AN – le informazioni socio-anagrafiche e cliniche relative alle pazienti sono certe e verificabili, poiché ricavate dalle loro cartelle cliniche. Allo stesso modo, sono affidabili anche le informazioni socio-anagrafiche degli studenti, seppure con un margine di dubbio in quanto frutto di un’autodichiarazione dei partecipanti.

Vantaggiosi sono anche il metodo d’indagine scelto e gli strumenti computazionali adottati per l’analisi dei dati, entrambi sono infatti ecologici, a basso costo e potenzialmente applicabili su larga scala. Inoltre, si sono rivelati efficaci nell’individuazione di specifici *pattern* potenzialmente caratteristici dell’uso del linguaggio dei soggetti con Anoressia Nervosa testati. Gli strumenti di NLP, in particolare TULE, hanno infatti permesso un’analisi sistematica del linguaggio a livello morfosintattico e semantico-lessicale, non praticabile con i metodi diagnostici tradizionali utilizzati in ambito clinico. Per

quanto riguarda i livelli dell'analisi linguistica, in futuro sarebbe poi opportuno considerare anche il livello pragmatico. Come ha suggerito lo studio di Skårderud (2007a; 2007b), infatti, il linguaggio metaforico dei soggetti con AN sembra evidenziare il concretismo linguistico di questi (simile a quello trovato nei pazienti con schizofrenia). In questa ricerca non è stato possibile indagare ciò, sia per l'inadeguatezza dei *task* proposti, sia poiché un limite di uno strumento come LIWC, è che, trattandosi di un sistema probabilistico, non è in grado di cogliere le sfumature di significato legate al contesto, che caratterizzano ad esempio il linguaggio metaforico.

Nel complesso, comunque, i risultati ottenuti in questo studio pilota sembrano promettenti e, in riferimento alla letteratura disponibile sull'argomento, costituiscono le prime evidenze sperimentali relative alla lingua italiana in questo ambito. È dunque auspicabile che indagini scientifiche successive approfondiscano la ricerca di eventuali *marker* linguistici potenzialmente caratteristici dell'uso del linguaggio dei soggetti affetti da Anoressia Nervosa. Come si è detto, sarebbe opportuno riproporre il test qui utilizzato ad un campione più ampio, sia in termini numerici che per quanto riguarda il range dell'età. Infatti, l'esordio della patologia è sempre più precoce in età evolutiva, tra i 9 e i 18 anni come specifica Parmeggiani (2013), pertanto, è necessario includere nelle ricerche future anche soggetti di età compresa tra i 9 e i 13 anni, che in questa ricerca non sono stati considerati. In futuro, si potrebbe anche pensare di includere nel campione dei soggetti di sesso maschile, sebbene questa psicopatologia abbia un'incidenza nettamente superiore nei soggetti di sesso femminile (rapporto di 1 a 10).

Se i risultati preliminari qui descritti e discussi saranno confermati dagli studi a venire, in prospettiva linguistica sarà possibile offrire un ulteriore banco di prova ai dibattiti concetti di semplicità e complessità; mentre in ottica clinica sarà forse possibile ipotizzare la costruzione di un sistema automatico in grado di analizzare e classificare le produzioni scritte, utilizzabile come supporto per l'individuazione di condizioni patologiche e sub-cliniche e potenzialmente applicabile anche in contesto scolastico.

Bibliografia

- Adornetti, Ines. 2018. *Patologie del linguaggio e della comunicazione*. Roma: Carocci.
- Affrunti, Nicholas W. & Geronimi, Elena M.C. & Woodruff-Borden, Janet. 2015. Language of perfectionistic parents predicting child anxiety diagnostic status. *Journal of Anxiety Disorders* 30. 94–102.
- Agosti, Alberto & Rellini, Alessandra. 2007. *The Italian LIWC dictionary*. Technical Report 1, Austin, TX, <http://liwc.wpengin.com/> (Consultato il 12.02.2020.)
- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders, Fifth edition, (5th ed.; DSM-5)*. Washington, D.C.-London: American Psychiatric Publishing.
- Anderson, Barrett & Goldin, Philippe R. & Kurita, Keiko & Gross, James J. 2008. Self-representation in social anxiety disorder: Linguistic analysis of autobiographical narratives. *Behaviour Research and Therapy* 46. 1119–1125.
- Arcara, Giorgio & Bambini, Valentina. 2016. A Test for the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS): Normative Data and Psychometric Properties. *Frontier in Psychology* 7(70). 1–13. DOI: 10.3389/fpsyg.2016.00070

- Arntz, Arnoud & Hawke, Lisa D. & Bamelis, Lotte & Spinhoven, Philip & Molendijk, Marc L. 2012. Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behaviour Research and Therapy* 50. 191–202.
- Bambini, Valentina & Arcara, Giorgio & Bechi, Margherita & Buonocore, Mariachiara & Cavallaro, Roberto & Bosia, Marta. 2016. The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry* 71. 106–120.
- Beltrami, Daniela & Gagliardi, Gloria & Rossini Favretti, Rema & Ghidoni, Enrico & Tamburini, Fabio & Calzà, Laura. 2018. Speech analysis by Natural Language Processing techniques: a possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience* 10. 369. DOI: 10.3389/fnagi.2018.00369
- Bernard, Jared D. & Baddeley, Jenna L. & Rodriguez, Benjamin F. & Burke, Philip A. 2016. Depression, Language, and Affect: An Examination of the Influence of Baseline Depression and Affect Induction on Language. *Journal of Language and Social Psychology* 35(3). 317–326.
- Berruto, Gaetano. 1990. Semplificazione linguistica e varietà sub-standard. In: Holtus G., & Radtke E. (a cura di), *Sprachlicher Substandard III: Standard, Substandard Und Varietätenlinguisti*. 16–43. Tübinga: Max Niemeyer Verlag.
- Bosco, Cristina & Lombardo, Vincenzo & Vassallo, Daniela & Lesmo, Leonardo. 2000. *Building a treebank for italian: a data-driven annotation schema. Proceedings of the Second International Conference on Language Resources and Evaluation (LREC-2000)*. Athens, Greece: European Language Resources Association (ELRA).
- Brockmeyer, Timo & Zimmermann, Johannes & Kulesa, Dominika & Hautzinger, Martin & Bents, Hinrich & Friederich, Hans-Christoph & Herzog, Wolfgang & Backenstrass, Matthias. 2015. Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontier in Psychology* 6. 1564. DOI: 10.3389/fpsyg.2015.01564
- Bruch, Hilde. 1962. Perceptual and Conceptual Disturbances in Anorexia Nervosa. *Psychosomatic Medicine* 24(2). 187–194.
- Buhl, Charlotte. 2002. Eating disorders as manifestations of developmental disorders: Language and capacity for abstract thinking in psychotherapy for eating disorders. *European Eating Disorders Review* 10(2). 138–145.
- Bydlowski, Sarah & Corcos, Maurice & Jeammet, Philippe & Paterniti, Sabrina & Berthoz, Sylvie & Laurier, Catherine & Chambry, Jean & Consoli, Silla M. 2005. Emotion-Processing Deficits in Eating Disorders. *International Journal of Eating Disorders* 37(4). 321–329.
- Caselli, Cristina & Casadio, Paola & Bates, Elizabeth. 1999. A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* 26. 69–111.
- Caselli, Cristina & Casadio, Paola. 1995. *Il primo vocabolario del bambino. Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milano: Franco Angeli.
- Chung, Cindy K. & Pennebaker, James W. 2007. The Psychological Functions of Function Words. In Fiedler, Klaus (a cura di), *Social Communication*, 343–359. New York: Psychology Press.
- Cummings, Louise. 2016. Il contributo della linguistica clinica. In Daloisio, Michele (a cura di), *I Bisogni Linguistici Specifici. Inquadramento teorico, intervento clinico e didattico delle lingue*. 85–116. Trento: Erickson.

- Cuteri, Vittoria & Minori, Giulia & Gagliardi, Gloria & Tamburini, Fabio & Malaspina, Elisabetta & Gualandi, Paola & Rossi, Francesca & Moscano, Milena & Francia, Valentina & Parmeggiani, Antonia. 2021. Linguistic feature of anorexia nervosa: a prospective case-control pilot study. *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity*. DOI: 10.1007/s40519-021-01273-7
- Dakanalis, Antonios & Clerici, Massimo & Carrà, Giuseppe & Riva, Giuseppe. 2016. Dysfunctional bodily experiences in anorexia nervosa: where are we? *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity* 21(4): 731–732.
- De Santis, Cristina. 2016. *Che cos'è la grammatica valenziale*. Roma: Carocci.
- Edwards, To'Meisha & Holtzman, Nicholas S. 2017. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* 68. 63–68.
- Espeset, Ester M.S. & Gulliksen, Kjersti S. & Nordbø, Ragnfried H.S. & Skårderud, Finn, & Holte, Arne. 2012. Fluctuations of Body Images in Anorexia Nervosa: Patients' Perception of Contextual Triggers. *Clinical Psychology and Psychotherapy* 19(6). 518–530.
- Espeset, Ester M.S. & Nordbø, Ragnfried H.S. & Gulliksen, Kjersti S. & Skårderud, Finn & Geller, Josie, & Holte, Arne. 2011. The Concept of Body Image Disturbance in Anorexia Nervosa: An Empirical Inquiry Utilizing Patients' Subjective Experiences. *Eating Disorders: The Journal of Treatment & Prevention* 19(2). 175–193.
- Fiorentino, Giuliana. 2009. Complessità linguistica e variazione sintattica. *Studi Italiani di Linguistica Teorica e Applicata* 38(2). 281–312.
- Gagliardi, Gloria. 2019. *Linguistica per le professioni sanitarie*. Bologna: Patron Editore.
- Goodglass, Harold & Kaplan, Edith & Barresi, Barbara. 2001. *Boston Diagnostic Aphasia Examination (3rd ed.) (BDAE-3)*. Austin, Texas: Pro-Ed.
- Konstantakopoulos, George & Varsou, Eleytheria & Dikeos, Dimitris & Ioannidi, Nikoleta & Gonidakis, Fragiskos & Papadimitriou, George & Oulis, Panagiotis. 2012. Delusionalism of body image beliefs in eating disorders. *Psychiatry Research* 200. 482–488.
- Lesmo, Leonardo. 2007. Il parser basato su regole del Gruppo NLP dell'Università di Torino. *Intelligenza Artificiale* 4(2). 46–47.
- Lyons, Elizabeth J. & Mehler, Matthias R. & Pennebaker, James W. 2006. Pro-anorexics and recovering anorexics differ in their linguistic Internet self-presentation. *Journal of Psychosomatic Research* 60. 253–256.
- Minori, Giulia & Gagliardi, Gloria & Cuteri, Vittoria & Tamburini, Fabio & Malaspina, Elisabetta & Gualandi, Paola & Rossi, Francesca & Moscano, Filomena & Francia, Valentina & Parmeggiani, Antonia. 2020. Linguistic Markers of Anorexia Nervosa: Preliminary Data from a Prospective Observational Study. *3rd RaPID Workshop: Resources and Processing of Linguistic, Para-linguistic and Extra-linguistic Data from People with Various Forms of Cognitive/Psychiatric/Developmental Impairments*. 34–44.
- Muir, Sarah & Newell, Ciaràn & Griffiths, Jess & Walker, Kathy & Hooper, Holly & Thomas, Sarah & Thomas, Peter W. & Arcelus, Jon & Day, James & Appleton, Katherine M. 2017. MotivATE: A Pretreatment Web-Based Program to Improve Attendance at UK Outpatient Services Among Adults With Eating Disorders. *JMIR Research Protocols* 6(7). e146.

- Munro, Calum & Randel, Louise & Lawrie, Stephen M. 2016. An Integrative Bio-Psycho-Social Theory of Anorexia Nervosa. *Clinical Psychology and Psychotherapy* 24(1). 1–21.
- Parmeggiani, Antonia. 2013. *Pediatria generale e specialistica: 2. Neuropsichiatria Infantile*. Bologna: CLUEB.
- Pennebaker, James W. & Mehl, Matthias R. & Niederhoffer, Kate G. 2003. Psychological aspects of natural language use: Our word, Our selves. *Annual Reviews of Psychology* 54. 547–577.
- Pennebaker, James W., Booth, R.J., Boyd, Ryan L. & Francis, M.E. 2015a. *Linguistic Inquiry and Word Count: LIWC2015*. Austin, Texas: Pennebaker Conglomerates.
- Pennebaker, James W. & Boyd, Ryan L. & Jordan, Kayla & Blackburn, Kate. 2015b. *The development and psychometric properties of LIWC2015*. Austin, Texas: University of Texas at Austin.
- Pollatos, Olga & Kurz, Anne-Lene & Albrecht, Jessica & Schreder, Tatjana & Kleemann, Anna M., Schöpf, Veronika & Kopietz, Rainer & Wiesmann, Martin & Schandry, Rainer. 2008. Reduced perception of bodily signals in anorexia nervosa. *Eating Behaviors* 9. 381–388.
- Skårderud, Finn. 2007a. Eating One's Words, Part I: 'Concretised Metaphors' and Reflective Function in Anorexia Nervosa – An Interview Study. *European Eating Disorders Review* 15. 163–174.
- Skårderud, Finn. 2007b. Eating One's Words, Part II: The Embodied Mind and Reflective Function in Anorexia Nervosa – Theory. *European Eating Disorders Review* 15. 243–252.
- Spinczyk, Dominik & Nabrdalik, Karolina & Rojewska, Katarzyna. 2018. Computer aided sentiment analysis of anorexia nervosa patients' vocabulary. *BioMedical Engineering OnLine* 17. 19.
- Voghera, Miriam. 2001. Riflessioni su semplificazione, complessità e modalità di trasmissione: sintassi e semantica. In Dardano, Maurizio (a cura di), *Scritto e parlato. Metodi, testi e contesti, Atti del Colloquio Internazionale di Studi*, 65–78. Roma: Aracne.
- Vocks, Silja & Busch, Martin & Grönemeyer, Dietrich & Schulte, Dietmar & Herpertz, Stephan & Suchan, Boris. 2010. Neural correlates of viewing photographs of one's own body and another woman's body in anorexia and bulimia nervosa: an fMRI study. *Journal of Psychiatry Neuroscience* 35(3). 163–176.
- Wolf, Markus & Theis, Florian & Kordy, Hans. 2013. Language Use in Eating Disorder Blogs: Psychological Implications of Social Online Activity. *Journal of Language and Social Psychology* 32(2). 212–226.

The semantics behind the reanalysis of French articles in French-based creoles

Laura Tramutoli

Università di Bologna

laura.tramutoli@unibo.it

Abstract

Most French-based creoles have a number of nouns that have originated from the reanalysis of the French noun phrase [article + noun] into a new monomorphemic lexeme, where the former article is agglutinated to the noun; for instance French *la plage* > Haitian Creole *laplaj* ‘beach’.

The present paper proposes to link the agglutination of the article in creoles to nouns that exhibit a low prototypical profile and claims that their meaning matches peculiar semantic types according to a criterion which resembles the functioning of noun classes in other languages, where the agglutinated article works as a classifier. It is argued that most French nouns reanalysed as agglutinated in creoles are associated with a semantic profile to which natural languages allow reference on a different conceptual basis than that required for an object (Carlson 1977), and this triggers a morphological marking (alternation) when the context entails reference to an object.

1. Introduction: the article agglutination

The reanalysis into one monomorphemic word of two or more pieces of information in a lexifier/source language is a widespread feature in many pidgin and creole languages, regardless of their different origins and parent languages.¹ This feature usually involves elements that are frequently patterned together in the lexifier language, for instance nouns/verbs and prepositions (in collocations), like the word *komot* ‘leave’ in Nigerian English Creole, deriving lexically from the English phrasal verb *come out*,² or the word *diswà* ‘evening’ in Haitian Creole, from the French prepositional phrase *de soir* ‘at night’.

Since it is so pervasive in contact languages (all kinds of contact languages indeed, not only pidgin and creoles),³ this type of reanalysis is also often explained as an etymological feature of systems that have undergone an intense process of relexification or morphological restructuring.

¹ Cfr. Holms (1988) and Parkvall (2000) for an exhaustive list of contact languages which host the phenomenon.

² Examples from Parkvall (2000: 81).

³ These features are typically registered in acquisitional varieties, too.

[The] reinterpretation of morpheme boundaries is fairly common in language restructuring. [...] There is thus nothing remarkable about morpheme boundary reinterpretation in contact languages as such – on the contrary, it is precisely what we would expect from non-native parsing of utterances produced by speakers of another language (Parkvall 2000: 81).

One of the features of reinterpretation of boundaries is the ‘article agglutination’,⁴ which consists in the reanalysis of the whole [(in)definite/partitive article + nominal lexeme] NP of the lexifier language into a single noun, or alternatively into a pronoun. The examples in (1) and (2) show that the noun *diri*, ‘rice’ in Mauritian Creole, lexically derives from the partitive French noun phrase *du + riz*, as much as the noun *azàgua* ‘rainy seasons’ in Capoverdian Creole derives from the Portuguese plural and feminine form of the definite article *las* + noun *águas*:

- (1) French *du riz* > Mauritian Creole *diri*
‘some rice’ > ‘rice’ (Ladhams 2012: 33)
- (2) Portuguese *as águas* > Capoverdian Creole *azàgua*
‘the waters’ > ‘rainy seasons’ (Ladhams 2012: 33)

This phenomenon, particularly characteristic of the lexicon of French-based pidgins and creoles (Baker 1984; Grant 1995; McWhorter & Parkvall 2002; Chaudenson 2003), was broadly investigated by past research, which has however concentrated mainly on relevant aspects of phonetic and phonological order, the morphological patterns of reanalysis,⁵ or the substrate influence of African languages in the acquisition of the articles in creoles.

This paper focusses instead on the lexical semantics of the article-agglutinated nouns in French creoles, claiming it to be an essential aspect in the creole reanalysis of French NPs. It is shown that article-agglutinated nouns’ referents deviate systematically from a prototypical profile (Lyons 1977; Hengeveld 1992), and have a non-referential profile, instead.

Data for this study has been first collected from consulting previous both descriptive and critical literature and the lexical resources available for all the languages considered. The references taken into account are Kihm (1995), Ehrhart (1993), Corne (1989) on Tayo or Saint Louis Creole; Bakker (1997) for Michif; Targète & Urciolo (1993) for Haitian Creole; Bernabé (1983) for Martinican and Guadeloupean Creole; Tourneaux & Barbotin (1990; Marie Galante) for Guadeloupean Creole; Valdman (2010) for Louisiana French; Michaelis & Rosalie (2009), Bollé (1977), Corne (1977); Beniamino (1996) for Reunion Creole; Pfänder (2000) for French Guiana Creole.

Lexical repositories such as dictionaries are undoubtedly a first profitable tool to inspect the presence of the agglutinated nouns in the lexicon of each language, but

⁴ Different labels have been proposed in previous studies. The most traditional one – which is also used in this paper – is “agglutination”. Bonami & Henri (2012) prefer “article incorporation”, clarifying that the term “agglutination” could cause ambiguity with morphological processes. The most recent denomination proposed is “article fusion”, by Lam (2019).

⁵ For reasons of space, I will not summarize the content of past research here, but I will only list here the related main bibliographic references: Holm (1988); Parkvall (2000); Baker (1992) and (1994); Grant (1995); Bonami & Henry (2012).

insufficient to state its real distribution in quantitative term, and unfortunately a corpus for these creoles is not always available.

The *An Crúbadán Project - Building a Corpus for Minority Languages*⁶ has lately included Seychelles Creoles, Haitian Creole, and Reunion Creole among the number of its languages. Using web crawlers, this tool can allow access to a broad number of lexical types retrieved from online texts, providing a better picture of the incidence of the article agglutination in creoles.

2. Phono-morphology of article agglutination: from French to creole

The most productive article agglutinations in French creoles are the following:

1. The French feminine singular form of the definite article *la*, in creole nouns like *lavi* ‘life’ (< Fr. *la vie* ‘the life’) or *lari* ‘street’ (< Fr. *la rue* ‘the street’).⁷

In addition, there is the high rate of French nouns beginning (phonetically) with a vowel which receive an agglutination in creole, consequently to the already active phonetic initial vowel elision in French, reducing the definite article to the sole initial consonant *l*. This French phonetic feature is inherited in the phonology of creole words, like *léglis* ‘church’ (< Fr. *l’église* ‘the church’) or *lenfè* ‘hell’ (< Fr. *l’enfèr* ‘the hell’).

½. Some nouns exhibit a form that seems to be derived from the French masculine singular article *le*, like *lili* ‘bed’ (< Fr. *le lit* ‘the bed’(?)) or *lelu* ‘wolf’ (< Fr. *le loup* ‘the wolf’). For most items it is difficult to discern whether *li-* or *le-* are outputs of the French singular article *le* or whether they derive from the plural article *les*, since the phonological results in creole are the same for both French forms; the only exception is when the article is etymologically positioned before a vowel-initial noun (i.e., *le, li* with a rise, depending on the variety and probably vowel harmony). We may postulate that for nouns referring to unique referents like *lekrist* ‘Jesus Christ’ or *lepap* ‘the pope’ the article must logically have been a singular (e.g. allegedly, there exist no more than one Jesus Christ or Pope in every speaker’s mind), but for other nouns like *lisyien* ‘dog’, we cannot be certain about the etymological singular or plural number of the French noun phrase that originated the agglutinated article in creoles (many nouns for animals belong to class 2 and derive from French plurals).

For this reason, we must consider this category as an “hybrid”, e.g. an ‘overlapping’ group between 1 and 2 (hence, we have decided to number it as ½), remarking that the substantial ambiguity concerns not only the phonological aspects but also its semantic basis.

2. The definite plural form *les*, like in *lezòt* ‘others’ < Fr. *les autres* ‘the others’, always in the case in which the French noun had a phonetic initial vowel. In fact, the

⁶ Available at <http://crubadan.org/>. Developed by Kevin Scannel of the Department of Mathematics and Computer Science, Saint Louis University, Saint Louis, Missouri, USA, and officially presented in 2007, the project is an attempt to build large bodies of corpora for minority languages with small numbers of speakers and/or limited computational resources through the use of web-crawlers, e.g. exploiting the quantity of texts freely available online in order to be able to work on statistical data.

⁷ Where not specified otherwise, the words in “creole” in the examples are reported from the Mauritian Creole variety, hence, some sounds and/or graphic rendering that may change for other creoles are not reported alongside. This choice is purely formal and depends on the tradition of studies for this creole, which is predominant if compared to others.

occurring of a plural definite article *les* is witnessed by the resistant *-s/* sound in the creole word, correspondent to the *-z/* element in French, needed for reasons of *liaison*, that in some cases is the only element that survives in the agglutination (like in *zòt*, a grammaticalized variant of *lezòt*, the direct object personal pronoun for the second person plural ‘you’). Note that, if the French noun does not start with a vocalic sound, it is practically impossible to estimate if the article agglutinated corresponds to a definite singular or plural form; according to the processes of phonological vowel reductions and change from the French vocalic system to the creole one, the agglutinated syllable *le* or *li* (as found in *lili* or *lelu*, cfr. above) could in fact also derive from French plurals *les lits* and *les loups*, while the outcome in creole would not change phonologically nor morphologically.⁸

3. The partitive morpheme *de*, in its articulated masculine singular form *du*, like in *diri* ‘rice’ (see example 1 above), or in conjunction with the definite article when a French noun begins with a vowel, *dilo/delo/dlo* ‘water’ < Fr. *de l’eau* ‘(some) water’, *dilwil* ‘oil’ < Fr. *de l’huile* ‘(some) oil’.

4. Indefinite article *un/une*, although rare, in lexemes like *nam* ‘soul’ < Fr. *une âme* ‘a soul’, and/or *nom* ‘man’ < Fr. *un homme* ‘a man’.

5. Strandquist (2005: 24) reports a sole agglutinated lexeme in Mauritian Creole, *dizef* ‘egg’, which would have derived from the plural partitive French *des* agglutinated to the French plural noun *oeufs*.

Table 1. sums up functions and forms of the elements involved in article agglutination from French to creoles.

3. Alternations

If we look at the article-agglutinated nouns’ lemmas in dictionaries or lexicons, we can frequently observe that a non-agglutinated version of the same word is co-cited along the agglutinated one (or vice versa), like for the lemmas reported in (3), from the dictionary of Guadeloupean Creole (Marie Galante) – French by Tourneux & Barbotin (1990):

(3) a. *louwa* (cfr. *wa*)
misyé Louwa • monsieur le Roi (personnage des contes)

b. *lo*
a lo • à l'eau

c. *listwa* (cfr. *istwa*)
Histoire, science historique

⁸ Cfr. Strandquist (2003: 8-9) on the vocalic reductions and changes in Mauritian Creole of French vocalic system.

French				creoles			
Functions	Forms		Examples (art.+noun)	Gloss	Forms	Examples (noun)	Gloss
Definite article	sg.	f. <i>la</i>	<i>la rue</i>	'the street'	<i>la-</i>	<i>lari</i>	'street'
		m./f. <i>l' + V-</i>	f. <i>l'église</i> m. <i>l'éléphant</i>	'the church'; 'the elephant'	<i>l-</i>	<i>léglis;</i> <i>lelefan</i> (<i>lefan</i>)	'church'; 'elephant'
	1/2	m. <i>le(s)(?)</i>	<i>le(s) lit(s),</i> <i>le(s) loup</i>	'the bed(s);' 'the wolf(-ves)'	<i>li-, le-</i>	<i>lili; lelu</i>	'bed'; 'wolf'
Definite article	pl.	<i>les + V-</i>	<i>les autres</i>	'the others'; 'you'	<i>lez-, les-, le-</i> (?), <i>liz-, z-, s-</i>	<i>lezòt; zòt</i>	'other' 'others', 'you'
Partitive article	m.	<i>du</i>	<i>du riz</i>	'some rice'	<i>di-</i>	<i>diri</i>	'rice'
	f./m.	<i>de l' + V-</i>	f. <i>de l'eau</i> m. <i>de l'huile</i>	'some water'; 'some oil'	<i>di-, de-, d-</i>	<i>dilo; dibwil</i>	'water'; 'oil'
Singular indefinite article	m.	<i>un</i>	<i>un homme</i>	'a man'	<i>n-</i>	<i>nom</i>	'man'
	f.	<i>une</i>	<i>une âme</i>	'a soul'		<i>nam</i>	'soul'
Plural indefinite article		<i>des</i>	<i>des oeufs</i>	'some eggs'	<i>diz-</i>	<i>dizèf</i>	'egg, testicle'

Table 1. Forms of the elements involved in article agglutination.

The article agglutination is not the only form in which (some) article-agglutinated creole nouns may be featured. They can also appear as non-agglutinated, and this formal *alternation* (Zribi-Hertz & Glaude 2014) usually triggers a change of meaning.

Here is an example of the distribution of agglutinated and non-agglutinated version of the same noun in Mauritian Creole:

- (4) a. Mauritian Creole (Bonami & Henry 2012: 11)
Donn mwa enn liv pwason.
 Give 1SG INDF pound fish
 'Give me a pound of fish.'

- b. Mauritian Creole (Bonami & Henry 2012: 11)
Komie ou dir laliv?
 How.much 2SG say pound
 'How much do you say the pound?'

Such flexibility of forms concerns almost exclusively nouns that belong to the feminine definite article *la* or the consonantal *l* agglutination class (class 1 from Table 4), but the process is very productive for this class and applies to neologisms or loanwords (cfr. Grant 1995 defines it an "open class"). Indeed, the alternation is a morphological device which functions particularly well for nouns of class 1.

4. The lexical semantic aspect

Baker (1984: 111-112) first noticed that nouns that underwent article agglutination in Mauritian Creole can be ascribed to certain types. In his view, these are nouns of places that allegedly were single entities in the culture of Mauritian slaves, such as *lavil* ‘the city’ with reference to the one main city, *latab* ‘the table’ with reference to the only table available for slaves to eat, or nouns of items which occurred in pairs or groups, like furniture, cutlery, body parts, and mass nouns.

Shortly afterwards, a study of Ndayiragije (1989) pointed out that some alternating forms of Haitian Creole showed a correspondence with prefix-alternating nouns in Fon, a language of the Kwa family, claimed to be, with other West African languages, substrate to several Caribbean creoles (Lefebvre 1986 and 1998) – the languages share anyway numerous structural similar features (cfr. Aboh 2006; Aboh, Smith & Zribi Hertz 2011).

In Fon some nouns can acquire a syllabic or vocalic prefix, a classifier, which transforms their meaning, allowing a shift from denoting something concrete or specific to something abstract or generic (row a in Table 2), or vice versa (row b). The Fon nouns that can undergo such alternations have often equivalents in Haitian (and other) Creole(s), as Table 2 shows.

	Fon	Haitian Creole	glosses
a.	tàn	istoua	‘story, legend’
	itàn	listoua	‘history’
b.	kú	lanmò	‘death’
	mèkúkú	mò	‘human corpse’

Table 2. Fon/Haitian Creole (Ndayiragije 1989: 315).

In a more recent analysis by Zribi-Hertz & Jean-Louis (2014), the authors account for the stable/unstable (alternating) singular definite article agglutinated nouns in Martinican Creole, suggesting that *l-*, *la-* and *le-/li-* are attached to “semantically definite DPs denoting individual terms in the manner of definite proper names” (Zribi-Hertz & Jean-Louis 2014: 269). As a framework, they take up Löbner (2011)’s semantic theory of definiteness, according to which ‘sortal’ and ‘individual’ nouns correspond *grosso modo* to Carlson (1977)’s ‘kind-referring’ and ‘object-referring’ nouns, determined respectively by a generic or specific reference domain. The authors explain that in Martinican Creole, whenever a noun is accompanied by the French agglutinated definite article, it corresponds to an individual reading of the noun, *contra* the sortal reading that it has when it appears bare. They also notice that the article agglutination in quality of marker of definiteness is usually present on nouns denoting abstract concepts (5a), instrument markers (5b), and times and places (5c):

- (5) a. Martinican Creole (Zribi-Hertz & Jean-Louis 2014: 282-283)
 *(La-)doulè pran’y.
 La-pain take.3SG
 ‘(S)he was suddenly in pain.’ (lit. ‘pain took hold of her/him’)

- (5) b. Martinican Creole (Zribi-Hertz & Jean-Louis 2014: 282-283)
*Jan ka travèsé lanmè-a a- *(la-){ranm/ *(la-)vwèl}.*
 John IMPFV cross sea-DET INS la-oar/la-sail
 ‘John is {rowing/sailing} across the sea.’ (lit. ‘John is crossing the sea à-la {oar/sail}’)
- (5) c. Martinican Creole (Zribi-Hertz & Jean-Louis 2014: 282-283)
Mari la-plaj.
 Mary la-beach
 ‘Mary is at the beach.’ [the type of place called ‘beach’]

The analysis carried out by Zribi-Hertz and Jean-Louis outlines the semantic mechanism underneath the grammatical alternations in creole, e.g. that of the generic and specific meaning of a noun, or – with Carlson 1977, Krifka et al. 1995 – of its reference to a kind or to an object. The article agglutination on a noun occurs when the latter is introduced in a sentential or phrasal context that do not need it to identify an object, but instead it need it to identify a kind. On the contrary, an object referring noun appears in its bare form.

5. Kind-referring and object-referring NPs

(Semantic) *referentiality*⁹ is a property of the meaning of a noun to denote an entity identifiable in the real world (Givón 2001).

In this paper, we take on board the assumption that a nominal lexeme or lexical basis denotes primarily a *kind*, in line with the fundamental semantic theory on genericity elaborated by Carlson (1977) and Krifka (2004), and incorporating the cognitive-semantic notion of the noun *schema* (Langacker 2008)¹⁰ according to which the content of a nominal lexical stem would *per se* be associated in the first place to the concept of an ideal referent, of a category (a *kind*)-prototype, characterized by a stable set of common properties. Only when this content is brought into the universe of discourse, can its denotation be pinned down to identify one specific object.¹¹

5.1 Noun referentiality (and non-referentiality)

In the functional discourse grammar tradition, nouns denoting at the kind-reading level are called *non-referentials* (Hengeveld 1992), while those denoting at an object-reading level are called *referentials*;¹² but, although any noun can potentially make reference to a

⁹ Distinct from the *referential function* (pragmatic referentiality) of a noun or of any other linguistic element.

¹⁰ Such an assessment was already found under the form of the concept of *sense* of a word in Frege (*Sinn*; Frege 1982), according to whom a word is endowed with one lexical-semantic sense, e.g. the cognitive or epistemic concept associated to it, depicting a universal category of referents, while only when it is inserted in the discourse it will denote a single specific referent (takes a reference or denotation function; *Bedeutung*).

¹¹ The so-called *relation of instantiation*, cfr. Ojeda (1993).

¹² Cfr. Aguilar-Guevara et al. (2014) about all the elements that can play a role in non-referentiality, apart from the lexical semantics of a noun, and about the linguistic bibliography on referentiality.

kind and to an object in a sentence, from a lexical-semantic point of view, any noun contains some information on the reference domain it instantiates as correlated to the inherent concepts it represents (Wierzbicka 1988; Ojeda 1993), which is not always associated to an *object* prototype.

As already argued by Quine (1960),¹³ every word possesses build-in modes of dividing reference, even if their denotation identifies the same entity in the real world. Entities that instantiate individual reference are called *individuals* in the philosophical tradition (Strawson 1959): “consider ‘shoe’, ‘pairs of shoes’ and ‘footwear’: all three range over the same [...] stuff, and differ from one another solely in that two of them divide their reference differently, and the third one not at all” (Quine 1960: 91).

If referentiality is measured according to the possibility of identifying individuals, and a noun or an NP prototypically denotes an *object* like *shoe* – quintessentially an individual –, the more the conceptual structure of a noun represents in individual terms, the higher its reference force will be.

Also, the “reference” argument is at the bottom of the countable/uncountable distinction. A countable word stem denotes an atomic and individual structure of the entity and provides a semantic criterion for individual identification. On the contrary, uncountable stems do not suggest an individual identification, or at least an atomically instantiated reference. According to the chosen definition of referentiality, all the categories that we are considering as non-referential are uncountable nouns, being their structure more like that of a kind than that of an object.

5.2 Prototypical and non-prototypical NPs

From a denotational point of view, the category prototype for nouns (and NPs) typically consists in the conception of a physical object (e.g. object) (Langacker 2008: 103), holding the properties of unity, identity, functional cohesion, perceptual boundedness, and instantiation (Jackendoff 1991; Acquaviva 2008: 90-118). In addition, a prototypical NP has a nominal head, denotes by lexical means, and is used referentially (Hengeveld 1992; Hengeveld & MacKenzie 2008).

When an NP denotes another entity than object, it is usually marked with dedicated morphological or syntactic means in natural languages, diverging from a default nominal encoding. According to Molmann (2013), besides kind terms, a typology of grammatically marked nouns includes: 1. noun phrases with nominalizations or functional nouns as heads; 2. expressions in subject or complement positions that are not NPs (for instance, completive clauses); 3. bare plurals and mass nouns; 4. special quantifiers and pronouns.

According to Hengeveld & Mackenzie (2008: 131), in terms of denotation, the contents of non-prototypical nouns are: (some) individuals such as *chair*, properties such as *colour*, state of affairs such as *meeting*, location such as *country*, time such as *week*. As Hengeveld (1992) explains, natural languages might allow reference to these elements on a different conceptual basis than that required by a semantic prototypical object: for instance, although the referent for *chair* is an individual, it is not always conceptualized (and consequently categorized in languages) as an object, but rather as an element of a larger group of the same kinds of objects with – in this case – the same function (*chairs*),

¹³ Cited in Ojeda (1993: 116).

e.g. a set or a plurality (*set-objects*) to which “we can impute a sufficiently regular behaviour” (Chierchia 1998: 348) or characteristics, shifting the focus from the *object* to the *kind*. Likewise, fruits like apples or pears, or tools like spikes or knives and forks can be viewed as set-objects, their nouns exhibiting degrees of markedness in natural languages¹⁴ (ibid.: 511–512; Rijkhoff 2002; Langacker 2008).

Other entities that are treated differently than prototypical nouns in languages that systematically reflect the ontological difference between an object and a non-object are masses (*water, sand*), aggregates and cohesive aggregates (*bones, fingers*; or dual concepts like *scissors, trousers*), collectives (*furniture, police*) (Wierzbicka 1988; Jackendoff 1991; Ojeda 1993; Chierchia 1998; Borer 2005; Langacker 2008; Acquaviva 2008; Harbour 2011; Borik & Espinal 2012).¹⁵ Also, properties, events, state of affairs and propositions – traditionally the ‘abstract objects’ of metaphysics – are typically encoded in languages as marked nouns and NPs, sometimes by dedicated grammatical strategies (Strawson 1959; Lyons 1977; Asher 1993; Moltmann 2013).

Interestingly (as it will be showed below), creole nouns marked by article-agglutination mainly identify those concepts that make good candidates for non-prototypical classes of nouns,¹⁶ such as properties, state of affairs and propositional acts; masses and aggregates (including set-objects); locations and times (Wierzbicka 1988; Goddard & Wierzbicka 2002; Furukawa 2010; Corblin 2011), and *weak-definite* referents in general, usually objects or people that have a special social and cultural function or characteristics (it is the case of the much debated word *newspaper*: Hawkins 1978; Carlson et al. 2006; Aguilar-Guevara et al. 2014; Schwarz 2014).

5.3 The morpho-syntax of NPs denoting kinds

As said, every noun can refer either to an object or to a kind, and it is usually the morpho-syntactic context of the NP and the meaning of the other elements of the sentence that attributes the interpretation. Beside the genericity of the NP, the second type of genericity enounced by Krifka et al. (1995) is the one related to the predication in a sentence, when it reports about a generic event or property rather than about a specific and isolated episodic event.

In English, a *bare* noun (Chierchia 1998; Longobardi 2001) is usually read as a generic NP; in a sentence like *roses are red*, the noun *roses* is the subject of what is known as *characterising sentence*,¹⁷ whose predication *are red* predicates a universal and characterising property of the flower *rose*, so that the subject NP cannot but be interpreted as kind-referring.

Differently, in French every noun must be preceded by an article (with few exceptions), and bare nouns are seldom tolerated, also when a kind term is subject of a characterizing sentence: the English sentence *roses are red* turns in fact in French in *Les*

¹⁴ Although not all languages do; cfr. Wierzbicka (1988: 499-555).

¹⁵ There is still a vivid debate going on about the internal structure and conceptualizations of these entities, cross-linguistically encoded with various strategies; we have here followed the denominations used in Acquaviva (2008); on the topic and for examples from the world’s languages, cfr. also Ojeda (1998), Wierzbicka (1998), Chierchia (1998).

¹⁶ For reasons of space, I cannot give here in the text a full coverage of the theories about each category. I will just enter the most prominent bibliographic references for each of them.

¹⁷ For more on the definition of characterising sentences, cfr. Carlson (1977), Chierchia (1998), Krifka et al. (1995); for syntactic tests of genericity, cfr. Zamparelli (2002).

roses sont rouges, with the obligatory presence of a definite article, just as in any episodic sentences whose subject is a definite and object-referring NP (*Les filles sont arrivées hier soir* ‘The girls arrived yesterday evening’).

Nevertheless, in certain syntactic contexts, such as that of indefinite specific existential sentences (6), we can find another article in front of the same noun, the partitive *des*:

(6) French (Moltmann 2013: 10)

J’ai vu des roses.
 1SG.SBJ=have.1SG.PRS.IND seen some rose.PL
 ‘I have seen (some) roses.’

Creole languages allow much less variety than French in marking both pragmatic and semantic definiteness by pronominal articles, as Table 3, adapted from Guillemin (2007: 1), shows.

	Features	French	Mauritian C.	translation
Count nouns	Singular [+ definite]	la table	latab	‘the table’
	Singular [- definite]	une table		‘a table’
	Plural [+ definite]	les tables		‘the tables’
	Plural [- definite]	des tables		‘tables’
Mass nouns	Singular [+ definite]	l’eau	dlo	‘the water’
	Singular [- definite]	de l’eau		‘water’

Table 3. Count nouns and mass nouns in French and Mauritian Creole.

The opposition count/mass exemplified in Table 3 represents the French system to have developed mark the different values that a noun expresses in terms of the semantic properties of its referents, and this is reflected in the language by an asymmetry in the morphological number system of nouns (mass nouns select by default the singular number) and whether it is definite or indefinite. French-based creoles do not rely on pronominal articles nor on the inflection of morphemes to convey such information, as it is clear from the two lexemes *latab* and *dlo*, highlighted in the table, which never change their form.

The postnominal definite article system in French-based creoles is widely described and discussed in a variety of papers, for instance Zribi-Hertz & Glaude (2006); Déprez (2003 and 2008). The system includes markers for pragmatic definiteness (usually *-la* or *-a*, lexically derived from the French locative adverb *là*), as well as morpho-syntactically related markers that mark plurality and/or demonstrativity, e.g. features that encode or reflect semantic definiteness, and which can transform and enable interpretations of nouns as kind- or object-referring. For instance, in the examples in (7) it is shown that in Haitian Creole the adding of the plural marker *yo* after the noun for ‘letter’, changes the interpretation from kind-referring (7a, *lèt*, translated in English with mail refers generically to the type of documents that Paul is writing, not to the single letters that he is writing) to object-referring (7b, the presence of *yo* already implies the definiteness of the noun, and in addition it adds specificity as a demonstrative adjective would do):

- (7) a. Haitian Creole (Zribi Hertz & Glaude 2006: 259-260)

Pòl ekri lèt.

Paul write letter

‘Paul is writing mail.’

- b. Haitian Creole (Zribi Hertz & Glaude 2006: 259-260)

Pòl ekri lèt **yo**.

Paul write letter **PL**

‘Paul is writing *the(se)* letters.’ ≠ ‘Paul is writing letters.’

Similarly, the article agglutination appears as a marker on the nominal lexeme which gives information on the lexical semantics of the referents, e.g. whether their reference domain structure corresponds to a representation of a kind or of an object. Nouns with agglutinated articles fall exactly in those categories that natural languages can treat differently than standard nouns, because their referents drift away from the nominal semantic core, e.g. a prototypical object.

As said, in those creoles that allow alternation, the agglutinated morpheme is to be viewed as a morpho-syntactic device which marks the different use of the noun as kind- or object-referring (just as the postnominal articles do), e.g. it depends on the meaning of the noun given by the pragmatic context; but what about those agglutinated nouns that do not allow alternations and are only featured as agglutinated, or about creoles where alternations are not possible at all and nouns only come in agglutinated forms?

Our argument is that in both cases the article agglutination corresponds for form and function to a classifier, responsible for a formal classification of nouns in various classes and whose motivation is primarily semantic and acts according to the types of denotation showed in § 7.1, e.g. marking nouns whose semantics is not prototypical. The semantic parameter is eventually that of the referentiality of the noun, and only when alternations are possible it actively takes the role of a grammatical marker.

In some African languages which exhibit a complex noun-classes system, two noun-classes can actually be dedicated to the pragmatic-semantic specific/generic interpretation of the meaning of a noun, and, interestingly, the types of nouns concerned usually overlap with the semantic types encoding non-prototypical entities and concepts, even partially. For instance, in Bantu languages classifiers from noun class 16 and 17 can appear on the same noun of location to denote typically a specific and generic place. In Kikuyu (or Gikuyu, Bantu; Englebretson 2015), the two classifier marking classes 16 and 17 are applied on nouns denoting in the first case a specific location, of which the speaker is aware of (he/she knows where the place is): *handu* ‘(the) place’; while in the second case, the noun denotes a non-specific location, but rather the roundabouts of a place: *kundu* ‘place’. In Ndali (Aikhenvald 2000: 64), the same classes 16 and 17 host nouns denoting respectively places specific and proximal (class 16, classifier *pa*) «motion to/from, situation; proximity to someone or something near the speaker», and places generic and distal (class 17, classifier *ku*) «motion to/from, situation; proximity to someone or something far from the speaker» (ibidem), confirming the relationship in languages between the semantic specificity/pragmatic definiteness and the dimension of spatial proximity. Similarly, in Guadeloupean Creole, some nouns for objects can take the agglutination of the article and produce a new meaning by the generic/distal semantic effect. The word *pot* means the object ‘door’, while the noun *lapot* takes a figurative

locative meaning, denoting or the space in front of the door (*veranda*), or even all the space around the house, the roundabouts of the property (8):

- (8) Guadeloupean Creole (Tourneaux & Barbotin 1990)
S' ou pasé douvan lapòt, an mwen ou vwè pòt-la femè,
 If 2SG pass in.front door of 1SG 2SG see door-DET close
an pa la.
 1SG NEG there
 'If you pass in front of my door and you see the door close, I am not there.'

6. An attempt of classification: formally and semantically related agglutinated nouns in French creoles

Comparing data from various lexical sources (see § 1), we can summarize the match between the semantic classes of nouns and the forms of agglutinated articles.

6.1 Main matches

Table 4 sums up the most typical matches found in creoles.

	Agglutinated article	Semantic classes of nouns
1.	<i>li-, le-; la-; l-</i>	a. Properties, events, state of affairs, facts, propositions b. (Mostly granular) masses (e.g. <i>rain; sand; sea</i>) c. Unique referents (e.g. <i>the king, the sun</i> ; including proper names and names of countries e.g. <i>France</i>) d. Locations (e.g. <i>hospital</i> ; including natural geographical entities e.g. <i>pond</i>) e. Time reference (e.g. <i>day, week</i>) f. Collectives (e.g. <i>family, population</i>) g. Single body parts (e.g. <i>beard, head</i>) h. Tools (instruments, furniture, and vehicles, e.g. <i>key, lamp, car</i>)
¹ / ₂ .	<i>li (?), le-(?),</i>	a. Dual or plural body parts (e.g. <i>kidneys; fingers; arms</i>) b. Animals (e.g. <i>rat; dog</i>)
2.	<i>lez-, les-, liz-, z-, s-</i>	a. Dual or plural body parts (e.g. <i>feet; bones</i>) b. Animals (e.g. <i>bird; monkey</i>) c. Plants and fruits (e.g. <i>almond; bush</i>) d. People with specific social function (<i>child; officer</i>) e. Properties, events, state of affairs, facts, propositions (often hypernyms, e.g. <i>thing; affair</i>)
3.	<i>di-, de-, d-</i>	a. (Mostly homogeneous) masses (e.g. <i>oil; honey</i>) b. Properties events, state of affairs, facts, propositions

Table 4. Semantic and formal correspondences.

6.2 Class 1

Class 1 is formally characterized by the agglutination of the French singular definite article, and, as said, it is the most numerous one. Among the number of nouns in this class, some semantic consistencies can be found in the types of entities they denote.

Class 1a. The French definite singular article is frequently agglutinated to nouns expressing properties, events, state of affairs¹⁸, facts and propositions, which are usually expressed in French by lexical nominalizations, like *beauté* ‘beauty’, *patience* ‘patience’ or *pêche* ‘(the) fishing’, resulting in creole nouns like *laboté*, *lapasians* or *lapèch*.

Class 1b. Mass nouns that take the singular definite French article in creole are mainly those expressing granular mass concepts, like *lapli* ‘rain’, *lajan* ‘money’, *lapusyè* ‘dust’, although also homogeneous masses can be part of this category, like *lakòl* ‘glue’.

Class 1c. contains nouns for *unique referents* (Russell 1905), like *lerwà* ‘the king’ or *labib* ‘the Bible’, which almost always in creoles receive article agglutination without allowing alternations.

Nevertheless, the cognitive and linguistic principle which is responsible for *uniqueness* of reference is opposite to that of genericity of reference, since unique referents are intrinsically specific and definite. Indeed, according to the theory of definiteness (Frege 1892; Strawson 1950; Christophersen 1939; Searle 1969; Hawkins 1976), uniqueness is when a referent is not associable to a kind, but it only identifies an entity that is unique in the real world, such as the satellite *moon* and the star *sun*, literally ‘one of a kind’. Linguistically, this condition implies that the nouns like *moon* and *sun* are endowed with maximal values of referentiality and pragmatic definiteness, too, since no prior context need to be provided for the speakers to identify the referent or to acquire knowledge of it. Such definiteness is also the reason for these nouns to be almost always accompanied by a definite article, like English *the moon*, *the sun*.

However, as Hawkins (1967) points out, uniqueness of reference can also be featured in special situations where the unique reference is produced as a pragmatic implicature. This might be the case of two speakers having cognition of some entities in a restricted-quantification domain, so that these entities are considered unique, and the reference is exclusive (Hawkins 1967: 161). For instance, two people that live in the same small village where there is only one bar, one church and one bakery, presumably will think and refer of those locations as to *the bar*, *the church* or *the bakery*, without the need of specifying furthermore which of the all bars, churches or bakeries in the world they are talking about. Similarly, when we refer to the king of France or to the Prime Minister, we are presupposing that the person we are speaking to has in mind the individual who, at the time of utterance, impersonates the role of king or Prime Minister, although they are surely not the only human being that could potentially fill that role, as many others have had or will in history. In other words, the kind ‘king of France’ or ‘Prime Minister’ do exist, but these nouns are typically used as definite because they are usually used to identify an individual person, so that their occurrence with an indefinite article corresponds to a marked use.

It is thus plausible that nouns denoting unique referents in creole could be originated by the systematic collocation in French of the article before the noun. Some article agglutinations on names of countries, for instance, can be interpreted as marking this unique referent function (ex. *Lasiri* ‘Syria’), taking after the same marking in French. The

¹⁸ For the difference between events and state of affairs, facts and propositions cfr. Vendler (1967).

same happens with names of cities or island, such as *Lapwènt* ‘Point-à-Pitre’ in Guadeloupean Creole, or *Ladig* ‘La Digue’ in Seychelles Creole. In addition, nouns denoting some historic events are marked with the ‘uniqueness’ definite article/marker. Grant (1995) mentions that the noun *lakatastrof* ‘catastrophe’ in Martinican Creole refers to the particular episode of the massive eruption of the Vulcan Mount Pelée in May 1902.

Furthermore, Hawkins interestingly adds that in some cases the pragmatic presupposition which allows a unique reference reading implies also an inclusiveness of reference. In sentences like *Fred brought the wickets in after the game of cricket* (from Hawkins 1967: 160), the reference to *the wickets* includes all of seven wickets that are needed in the game of crickets, e.g. to the entire set, but it is also unique, because those wickets are the only ones that Fred used.

In a wider but similar perspective, masses can generally be viewed as having an inclusive profile: when we refer to ‘the sand’ we are including all the items that can possibly be described as *sand* in the domain of quantification, but indeed this argument could be extended to all *kinds*. In fact, according to Hawkins, also when we refer to the Prime Minister or the King of France presupposing that there are no other Prime Ministers or Kings of France, the reference is indeed inclusive, even if the domain of quantification includes only one entity. This inclusive concept is for Hawkins the one really motivating the presence of the definite article before the noun in English, because either a kind-referring noun, a mass noun, a unique referent noun and a quantificationally-restricted noun of aggregates require the definite article as default marking. In other words, Hawkins’ considerations can be seen as a connection piece between the structure and reference of unique entities and concepts and kinds or categories, since, as explained in the introduction, the conceptualization of a kind or of other entities that are construed in a similar way (masses, properties, indefinite events) happens through the same pattern: these entities are viewed as whole (*kinds*) disregarding of their internal substructure. In other words, every noun marked with a definite article expressing one of these entities (the sand, the wisdom, the swimming) would subsume an inclusive profile, without the need of pragmatically restricting the context of utterance.

However, Hawkins’ vision has not been further developed in cognitive or semantic studies, and that unique referents are usually considered at the opposite pole of indefinites.

Class 1d. contains nouns for entities that can represent locations and places.

As explained in § 5.2, although in the right syntactic and pragmatic context nouns like *house*, *hospital*, *beach* can refer to individual objects, their *kind* denotation is likely to be associated with a locative concept (a place), as well as nouns denoting natural entities like *river*, *pond*, *hill*, which are often conceptualized as geographical references and locations (Bromhead 2017). Consequently, these nouns have in a sentence a high chance to fill the *Location* theta role and to be syntactically encoded in prepositional phrases, such as *at/in/by the hospital*, *house*, *beach*, *river*, etc.

Class 1e. hosts nouns for times reference, such as names of the days of the week or months, nouns for parts of the day and hours.

Class 1f. contains nouns for collective entities, such as *family*, *population*, but also *police* or *band*.

Class 1g. is for nouns denoting single body parts. In the perspective of the conceptually and pragmatically restricted domain of human body, parts such as *beard*, *head* or *mouth* represent indeed unique referents, since they are logically the one and only beard, head

or mouth in a body.¹⁹ In a language such as French they are always marked as definite when used in weak-definiteness contexts, e.g. in constructions such as collocations or idiomatic fixed expressions (ex. Fr. *se prendre la tête* ‘to fight’, *se faire la barbe* ‘to shave’, *fermer la bouche* ‘to be quiet’). These factors must have favoured the agglutination of the article in creole.

Finally, class 1h. contains nouns for tools, such as *key* or *wheel*, or music instruments, such *arp* or *organ*.

Table 5 reports examples for each subclass from eight creoles (pure lemmas are given).

	Mauritian	Seychelles	Reunion	Haitian	Lesser Antillean	Louisiana	Tayo
a. Properties, events, state of affairs, facts, propositions	<i>lamor</i> ‘death’ <i>lafreyer</i> ‘fear’	<i>lafanmin</i> ‘hunger’ <i>lidé</i> ‘idea’	<i>lapasians</i> ‘patience’ <i>loner</i> ‘honour’	<i>lapèch</i> ‘fishing’ <i>lajenès</i> ‘youth’	<i>lanmou</i> ‘love’ <i>lanvi</i> ‘desire’	<i>lapè</i> ‘peace’ <i>labitid</i> ‘habit’	<i>lafèt</i> ‘feast, party’
b. (mostly granular) masses	<i>lapo</i> ‘skin’ <i>larecolte</i> ‘harvest’	<i>lafarin</i> ‘flour’ <i>lasoup</i> ‘soup’	<i>lapli</i> ‘rain’ <i>lo</i> ‘water’	<i>lajan</i> ‘money’ <i>lapousyè</i> ‘dust’	<i>lasid</i> ‘acid’ <i>lété</i> ‘ether’	<i>lasòs</i> ‘sauce’ <i>lwil</i> ‘oil’	<i>lapli</i> ‘rain’ <i>larcha</i> ‘money’
c. unique referents	<i>labib</i> ‘the Bible’	<i>lerwà</i> ‘king’ <i>lekrist</i> ‘Christ’	<i>lesegnèr</i> ‘the Lord’ <i>levanzil</i> ‘gospel’	<i>lepep</i> ‘pope’ <i>letazini</i> ‘the US’	<i>lakatastrof</i> ‘catastrophe’ (Grant 1995)	<i>lafrik</i> ‘Africa’	<i>lerwa</i> ‘king’
d. locations	<i>ladress</i> ‘adress’ <i>lakot</i> ‘the coast’	<i>lotèl</i> ‘hotel’ <i>labutik</i> ‘store’	<i>letan</i> ‘pond’ <i>lekol</i> ‘school’	<i>lakay</i> ‘home, house’ <i>lapos</i> ‘post office’	<i>lekòl</i> ‘school’ <i>lapòt</i> ‘patio’	<i>lekol</i> ‘school’ <i>leglis</i> ‘church’	<i>lakay</i> ‘house’ <i>larut</i> ‘road’
e. time reference		<i>lèr</i> ‘hour’ <i>letàn</i> ‘time’	<i>lerla</i> ‘then’ <i>lezour</i> ‘day’	<i>lavèy</i> ‘eve’ <i>lanwit</i> ‘night(-time)’	<i>lè</i> ‘hour’ <i>lapremidi</i> ‘afternoon’	<i>lanwi</i> ‘night’	
f. collectives		<i>limanité</i> ‘humanity’ <i>lorkès</i> ‘orchestra’	<i>lansan</i> ‘set, ensemble’	<i>laband</i> ‘band’ <i>lapolis</i> ‘police’	<i>larmé</i> ‘army’ <i>laronn</i> ‘patrol’	<i>lodyens</i> ‘audience’	<i>latribe</i> ‘tribe’
g. single body parts	<i>labarb</i> ‘beard’	<i>labouche</i> ‘mouth’ <i>lefwa</i> ‘liver’	<i>latet</i> ‘head’	<i>lalang</i> ‘tongue’ <i>lapat</i> ‘paw’	<i>lapendisit</i> ‘appendicitis’	<i>latèt</i> ‘head’	
h. tools		<i>lasel</i> ‘saddle’ <i>lasyet</i> ‘dish’	<i>loto</i> ‘car’	<i>laplanche</i> ‘board’ <i>labanyè</i> ‘flag’	<i>letabli</i> ‘table, shelf’	<i>lalarm</i> ‘clock’ <i>lavyon</i> ‘airplane’	<i>lakle</i> ‘key’

Table 5. Examples for class 1.

In recognizing the presented classes, we acknowledge some ambiguity about the semantic types to which certain nouns can be assigned, which usually depends on the degree of abstractness of their referents.²⁰ It is typically the case for nouns denoting imaginary,

¹⁹ For a study on these unique referents in the human body, therefor unique in a defined conceptual context, cfr. Guéron (1983); Vergnaud & Zubizarreta (1992); Ojeda (1993).

²⁰ The concept of abstractness is extremely wide and responds with different definitions to different traditions of philosophical and linguistic studies (Rosen 2011). In this paper, we have therefore decided to

fictitious concepts such as *lènfè* ‘hell’ or *lekwatèr* ‘the Equator’, which are likely to be conceptualized as “places”, but which indeed represent conventional entities. In actual facts, they also represent unique referents, since allegedly there is only one *hell* and one *Equator*. Hence, the uncertainty about which semantic type hold the respective agglutinated nouns.²¹

In addition, natural elements or phenomena that are acquired through sensorial experience but that are not physically bounded, such as *lè* ‘air’ or *larkansièl* ‘rainbow’ are more difficult to associate with the given classes, although some association can be made with masses (e.g. *air*) and/or state of affairs/events (e.g. *rainbow*).

In some creoles, some nouns that are listed in class 2. belong to class 1, or to both, being able to agglutinate both the *l-/la/le* and the *lez-/z-* article, generating no difference in meaning. For instance, in Seychelles Creole, nouns that encode individuals with a particular characteristic or a peculiar social role typically belong to class 2., but in a few cases these referents are encoded by nouns starting with *l-/la/le* (ex. *letourdi* ‘stupid, fool’, *letranzé* ‘stranger’).

6.3 Class ½

This intermediate class is here added as an acknowledgement of uncertainty. The uncertainty given by the fact that the non-linearity between the resulting phonological form of and French articles in creoles and their etymology. In fact, if different (French) original forms have the same phonological output because of phonological change and reduction in creole, it is difficult to decide which singular or plural forms of French article was the selected one at the moment of agglutination. In particular, a few masculine French nouns that start with a consonant and undergo article-agglutination show this feature.

We include in this category nouns for animals, such as *lera* ‘rat’, *lisyen* ‘dog’, *lelu* ‘wolf’; nouns for body parts, such as *lebra* ‘arm’, *ledwa* ‘finger’, *leren* ‘kidney’, *lipyé* ‘foot’, *ledan* ‘tooth’. Of course, it would make sense to reckon that nouns with the latter referents are intrinsically plural, since logically they denote dual objects (ex. body parts *foot/feet*) or plural concepts or aggregates (ex. *tooth/teeth*), as well as that the article derives from the plural French form *les* (we have evidence from overt markers that it usually does). Nevertheless, from a purely linguistic point of view we cannot affirm that *le-* or *li-* is the phonological reanalysis of a plural form instead of a singular one.

On the other hand, nouns that usually agglutinate the plural forms of the article – and it shows because it is phonologically overt – sometimes simply agglutinate the singular form, or both, resulting in two different lexical types in the same language. It is often the case of nouns denoting individuals (people) with peculiar social or personal characteristics, such as *letourdi* ‘the stupid, dumb’, *letranzé* ‘the stranger, foreigner’, *landomì* ‘the slow-minded’, *lanonym* ‘anonymous’, *lapost* ‘apostle’.

avoid the label ‘abstract’ as much as possible in order to avoid vagueness in definition, although many linguists would have compacted many of the here singled out semantic types of nouns as ‘abstract nouns’.

²¹ It is fair to say that in these cases the decision cannot be but arbitrary. We have for instance decided to list ‘hell’ among unique referents as well as nouns that are names of countries, that we intend as proper names denoting one unique referent, such as *lafrik* ‘Africa’.

6.4. Class 2

Class 2 is formally characterised by the agglutination of the original French plural article to French nouns starting with a vowel. These nouns' referents are mostly entities that might be conceptualized as aggregates, and here are the most common types:

Class 2a. hosts nouns for dual or plural body parts, such as (dual) *zyè* 'eye(s)', *zorey* 'ear(s)', or (plural) *lezo* 'bone(s)'. Nouns for animal body parts are counted in this class, too; such is the case for nouns *zèl* 'wings', or *zécal* (< Fr. *les écals*) 'crustacean shell(s)'. Finally, we have considered in this class the noun for offal and giblets, *zaba*, e.g. a noun denoting internal organs as food (in the food semantic area), from French *les abats*, although etymologically the French noun is not the name of any body part.

Class 2b. includes nouns for animals, such as *zwazo* 'bird', *zéléfan* 'elephant', *zarigné* 'spider', including the noun *zanimò* 'animal' as well.

Class 2c. contains nouns' referents are plants and fruits. *Zamann* 'almond', *zabriko* 'Antillean apricot', *zepina* 'spinach', *zaboka* 'avocado' are just few examples of the nouns converging in this class.

Class 2d. consists of nouns denoting peculiar people with a specific or relevant social function, which can be identified by a specific age range, ex. *zanzan* 'child', by a specific job ex. *zofisié* 'policeman', *zelèv* 'student', by a physical peculiarity, ex. *zaveg* 'blind people', by a peculiar relationship, ex. *zanmi* 'friend'²². Here, we have also included the pronoun 'other(s), *lezòt*, that in Lesser Antillean creoles has grammaticalized into the second person plural pronoun 'you (2pl.)'.

We also find nouns for properties, events, state of affairs, facts, propositions with an agglutinated plural article, such as *zafèr* 'business' and others in class 2e.

Class 2f. hosts nouns for tools, among which the noun itself for 'tool', *zouti*.

Finally, we have reserved class 2g. for nouns denoting entities that have to do with nature, such as natural phenomena or formations that usually manifest themselves in plural instances, such as *zèklè* 'thunder' or *zil* 'island'.

Table 6. provides for some examples for each listed type in the creole considered.

	Mauritian	Seychelles	Reunion	Haitian	Lesser Antillean	Louisiana	Tayo
a. Dual or plural body parts	<i>lezo</i> 'bone'	<i>zorey</i> 'ear'	<i>lezo</i> 'bone'	<i>zécal</i> 'crustacean shell'	<i>zaba</i> 'offal, giblets'	<i>zortei</i> 'toe'	<i>son</i> 'nail'
b. animals	<i>zanimò</i> 'animal, beast'	<i>zwazo</i> 'bird'	<i>zako</i> 'monkey'		<i>zagouti</i> 'agouti'	<i>zozo</i> 'bird'	
c. plants and fruits	<i>zamann</i> 'almond'	<i>zèpis</i> 'spice'	<i>zananas</i> 'pineapple'	<i>zepina</i> 'spinach'	<i>zaboka</i> 'avocado'	<i>zanana</i> 'pine-apple'	
d. people with specific social function		<i>zofisié</i> 'policeman'	<i>zanzan</i> 'child'	<i>lezòt</i> 'other'	<i>zanzan</i> 'child'	<i>zami</i> 'friend'	<i>lesot</i> 'other'
e. Properties, events, state of affairs, facts, propositions		<i>zistwar</i> 'story'	<i>zafèr</i> 'business, thing'		<i>zeprèv</i> 'challenge'		
f. Tools		<i>zouti</i> 'tool'	<i>zouti</i> 'tool'	<i>zéguiy</i> 'pin'	<i>zannò</i> 'earring'	<i>zouti</i> 'tool'	
g. Natural feature/phenomenon		<i>zil</i> 'island'	<i>zetwal</i> 'star'	<i>zespri</i> 'ghost, presence'	<i>zèklè</i> 'thunder'		

Table 6. Examples for class 2.

²² Cfr. Wierzbicka (1988) on connotations of nouns denoting social functions and properties as kinds, such as 'fools' or 'hunchbacks', and on shape and age as crucial properties for the conceptualization of kinds (Wierzbicka 1988: 476-480).

6.5 Class 3

Class 3a. reunites nouns mainly expressing mass concepts. In particular, these entities seem to correspond to homogeneous masses, but of course some exceptions are present, for instance nouns for granular entities such as ‘rice’.

In a few creoles, for instance Réunion Creole, a noun like water can enable alternation: *dolo/lo*, very unlikely to be found in other creoles. While *dolo* clearly enables a generic or kind-referring reading of the referent (‘water’), *lo* is used to refer to an object, specifically identifiable in the Universe of Discourse, a meaning that in French (or in English) would be probably encoded by a plural form strategy (Fr. *eaux*):

(9) a. Réunion Creole

Di pa «Fontène zamé ma boire ton lo».
 Say NEG fountain never 1SG drink 2SG.POSS water
 ‘Don’t say «Fountain, I will never drink your water(s).’

b. Réunion Creole

Dolo i pé dormi; mé zamé mon zénemi.
 Water 3SG PROG sleep but never 1SG.POSS enemy
 ‘The water sleeps, but my enemy(-ies) never do(es).’

Nouns indicating any of the instance and bounded parts of materials like *dibwa* ‘wood’ or food, like *dilar* ‘bacon’, come from French collective nouns, that have a perfectly parallel internal structure to mass nouns (Acquaviva 2003: 224); they indeed select the same marker.

In a few cases, class 3b. holds nouns for properties, events, state of affairs, facts and proposition, such as *ditò* ‘tort, wrongdoing’ or *dibwi* ‘noise’. Table 7. shows the examples for class 3.

	Mauritian	Seychelles	Reunion	Haitian	Lesser Antillean (St.Lucien)	Louisiana	Tayo
a. (mostly) homogeneous masses	<i>dibwa</i> ‘wood’ <i>dilar</i> ‘bacon’	<i>dilo</i> ‘water’	<i>debuà</i> ‘wood’ <i>dolo</i> ‘water’	<i>ditè</i> ‘tea’ <i>diven</i> ‘wine’	<i>diwi</i> ‘rice’ <i>diven</i> ‘wine’	<i>diri</i> ‘rice’	<i>dlo</i> ‘water’
b. Properties, events, state of affairs, facts, propositions					<i>ditò</i> ‘wrongdoing’ <i>dibwi</i> ‘noise’		

Table 7. Examples for class 3.

7. Data from An Crubàdàn corpus

Table 8. shows the numeric data for the occurrence of agglutinated nominals in Reunion

Creole, Seychelles Creole, and Haitian Creole²³, as taken from the An Crubàdàn corpus' lexical search.

An Crubadàn Project – Building a Corpus for Minority Languages is a project that builds small corpus for minority languages through web crawlers, hence it allows access to a broad number of lexical types retrieved from online texts, providing the possibility to work with quantificational data (although still in limited terms). Originally designed and restricted to Celtic languages, it was later applied to various minority languages, among which some French creoles, and in particular Seychelles Creoles, Haitian Creole, and Reunion Creole, which represent a good enough variational sample for our research, since they are illustrative of the two main linguistic areas that productively report the phenomenon of article agglutination, e.g. the Indian Ocean creoles (Seychelles and Reunion) and Atlantic creoles (Haitian).

The total number of lexical *tokens* crawled for Reunion Creole amounting to 35.302, for Seychelles Creole to 206.404, and for Haitian Creole to 5.676.584. In the table, I will be reporting for each language the number of nominal *types* whose semantics is ascribable to the established subclasses, to give evidence of the phenomenon in each language²⁴ and overall.

Agglutinated article	Semantic categories of nouns	Reunion Creole (n. types)	Seychelles Creole (n. types)	Haitian Creole (n. types)
<i>li-, le-; la-; l-</i>	Properties, events, state of affairs, facts, propositions	82	379	843
	(Mostly granular) masses	11	33	62
	Unique referents	25	27	134
	Locations	13	43	44
	Times	7	17	13
	Collectives	4	16	29
	Single body parts		14	10
	Tools		20	4
<i>li(?), le-(?)</i>	Dual or plural body parts		4	26
	Animals		2	30
<i>lez-, les-, le-(?), liz-, z-, s-</i>	Aggregates	21		45
	Dual or plural body parts		6	11
	Animals		4	10
	Plants and fruits		5	2
	People with specific social function		11	16
	properties, events, state of affairs, facts, propositions	11	8	18
<i>di-, de-, d-</i>	(Mostly homogeneous) masses	6	12	51
	Other (properties, events, state of affairs, facts, etc.)	1	3	12

Table 8. Data from the *An Crubadàn* corpus.

²³ Available at: Reunion Creole: <http://crubadan.org/languages/rcf>; Seychelles Creole: <http://crubadan.org/languages/crs>; Haitian Creole: <http://crubadan.org/languages/ht>.

The choice of selecting these languages is methodological: at least one enough-well documented creole for the major groups (Indian Ocean and Caribbean) of French Creole had to be picked. Reunion Creole was added to check how a variety considered 'a creoloid' would have performed with respect to the given phenomenon. For a definition of 'creoloid', cfr. Platt (1977).

²⁴ Graphic variations due to graphic transcription uncertainties or phonetic variations have not been considered. For instance, entries like *la-lwa*, *lalwà* or *laloï* have all been reconducted to the same type. Data related to *tokens* can be found consulting the corpus itself at the link given above.

8. Conclusion: a noun-class principle

Most of the mentioned creoles have an instrument to encode a referential and a non-referential reading of a noun, the alternation. In parallel with a plural morpheme on a mass noun in English, which can turn its meaning to denote specific instances of that mass substance (cfr. 8. RC: *dolo/lo*; Engl: *water/waters*), alternating agglutinated articles can be viewed as morphological markers for the category of semantic referentiality.

However, it is difficult to decide which of the lexical variants of an alternating noun is the marked one in creole. In a typological sense, it should be the one actually marked with the morpheme, but, looking at the semantic classes of nouns that do not allow alternations, it really seems that the article agglutination marks those categories of nouns whose referents are not conceptualized as objects, but rather as kinds.

Our statement is that in French-based creoles there are lexical classes of nouns that are marked with specific morphemes – that hence behave like classifiers – according to their semantic content, may that be a mass concept, an aggregate concept, a property or event concept, and (almost) each of these referent classes receive their dedicated agglutinated article.

Indeed, creole article agglutination tick many of the boxes required from the list proposed from Aikhenvald (2000) for the theoretical identification of classifiers: their choice is based on the semantic principles; a classifier is systematically present on those noun and encoded as part of its lexical-semantic information, to express inherent semantic characteristics or properties of a noun's referent; they are a type of non-agreeing noun categorization device, so they do not match any inflectional properties of nouns with other concorded words in the sentence, being their scope the sole NP²⁵; one noun can be used with different classifiers, enabling a change of meaning.

Interestingly, this last point resorts a good explanation to account for the presence of alternations in these languages: it is possible that the opposition 'reference to object/reference to kind' was finally grammaticalized in French creoles as a \emptyset /agglutinated-article opposition. Alternatively, alternations could be viewed as discourse features which are sorted out based on the type of reference that the speaker wants to set, allowing to infer a referential/non-referential interpretation of the noun in a sentence, alongside articles and other inflectional morphemes in Indo-European languages.

However, regardless of the perspective we decide to take, we conclude that, due to their semantic contents, agglutinated nouns are marked in language as less referential than prototypical nouns, and there is no doubt that the semantic classes of nouns which enable article-agglutinations (cfr. Table 8) are the same that cross-linguistically face a peculiar treatment.

9. Bibliography

- Acquaviva, Paolo. 2008. *Lexical plurals*. Oxford: Oxford University Press.
Acquaviva, Paolo. 2018. Categorization as noun construction. Gender, number, and entity types. In Mathieu, Éric & Dali, Myriam & Zareikar, Gita (a cura di). *Gender and noun classification*, 41–66. Oxford: Oxford University Press.

²⁵ Differently than what happens with gender morphemes, cfr. Corbett (1991).

- Aguilar-Guevara, Ana & Le Bruyn, Bert & Zwarts, Joost. 2014. *Advances in weak referentiality*. Amsterdam/Philadelphia: John Benjamin.
- Aikhenvald, Alexandra. 2000. *Classifiers. A typology of Noun Categorization Devices*. Oxford: Oxford University press.
- Asher, Nicholas. 1993. *Reference to Abstract Objects*. Dordrecht: Kluwer.
- Baker, Philip. 1984. Agglutinated nominals in Creole French: their evolutionary significance. *Te Reo* 27. 89–129.
- Baker, Philip. 1972. *Kreyol: a description of Mauritian Creole*. London: Hurst.
- Baker, Philip & Hookoomsing, Vinesh Y. 1987. *Morisyen-English-Français. Diksyoner Kreol Morisien. Dictionary of Mauritian Creole. Dictionnaire Créole Mauricien*. Paris: L'Harmattan.
- Bakker, Peter. 1997. *A language on our own. The Genesis of Michif, the Mixed Cree-French Language of Canadian Métis*. New York-Oxford: Oxford University Press.
- Beniamino, Michel. 1996. *Le Français de la Réunion. Inventaire des particularités lexicales*. Paris: Edicef.
- Bernabé, Jean. 1983. *Fondal Natal. Grammaire basilectale approchée des créoles guadeloupéen et martiniquais*, vol. 2–3. Paris: L'Harmattan.
- Bollée, Annegret. 1977. *Le créole français des Seychelles. Esquisse d'une grammaire-textes-vocabulaire*. Tübingen: Niemeyer Verlag.
- Bonami, Olivier & Henri, Fabiola. 2012. Predicting article agglutination in Mauritian. (Paper presented at «*Formal Approaches to Creole Studies*» 3, Universidade de Lisboa, November 15–16.)
- Borer, Hagit. 2005. *In Name Only*. Oxford: Oxford University Press.
- Borik, Olga & Espinal, Maria Teresa. 2012. On definite kinds. In Mari, Alda (a cura di), *New perspectives on genericity at the interfaces (Recherches Linguistiques de Vincennes 41)*, 123–146. Vincennes: Presses universitaires de Vincennes.
- Carlson, Gregory N. 1977. *Reference to Kinds in English*. University of Massachusetts at Amherst (PhD Dissertation.)
- Caudill, Herb & Diallo, Ousmane Besseko. 2000. *Learner's Guide to Pular (Fuuta Jallon)*. Guinea: Peace Corps.
- Chaudenson, Robert. 1981. *Textes créoles anciens: La Réunion et Ile Maurice. Comparaison et essai d'analyse*. Hamburg: Buske.
- Chaudenson, Robert. 2003. *La créolisation: théorie, applications, implications*. Paris: L'Harmattan.
- Chierchia, Gennaro. 1998. *Plurality of Mass Nouns and the Notion of Semantic Parameter*. In Rothstein, Susan (a cura di). *Events and Grammar*, 53–103. Dordrecht: Kluwer.
- Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Munksgaard.
- Corblin, Francis. 2011. Des définis para-intensionnels: être à l'hôpital, aller à l'école. *Revue langue Française* 171. 55–75.
- Corne, Chris. 1977. *Seychelles Creole Grammar. Elements for Indian Ocean Proto-Creole Reconstruction*. Tübingen: TBL Verlag Gunter Narr.
- Danny, Peter J. & Creider, Chat A. 1976. The semantics of noun classes in Proto-Bantu. *Studies in African Linguistics* 7. 1–30.
- Ehrhart, Sabine. 1992. The development of the preverbal markers in St-Louis-Creole: The formation of a TAM system? *Language Sciences* 14. 233–247.

- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100(1). 25–50.
- Furukawa Naoyo. 2010. Article défini, son emploi ‘intensionnel’ et énoncé tautologique. *Bulletin d'études françaises* 41. 51–71.
- Jennings, Williams & Pfänder, Stephan. 2015. French Guianese Creole. Its emergence from contact. *Journal of Language Contact* 8. 36–69.
- Goddard, Cliff & Wierzbicka, Anna. 2002. *Meaning and Universal Grammar – Theory and Empirical Findings*, vol. I-II. Amsterdam: John Benjamins.
- Grant, Anthony. P. 1995. Article agglutination in Creole French: a wider perspective. In Baker (ed.). *From contact to creole and beyond*, 149–176. London: University of Westminster Press.
- Guéron, Jacqueline 1983. L'emploi ‘possessif’ de l'article défini en français. *Langue Française* 58. 23–35.
- Guillemin, Diane. 2011. *The syntax and the semantics of a Determiner System. A case study of Mauritian Creole*. Amsterdam-Philadelphia: John Benjamins.
- Harbour, Daniel. 2011. Descriptive and explanatory markedness. *Morphology* 21. 223–240.
- Hawkins, John. 1978. *Definiteness and Indefiniteness. A study of reference and grammaticality prediction*. London: Croom Helm.
- Hengeveld, Kies. 1992. Non-verbal predication: theory, typology, diachrony. Functional Grammar Series 15. Berlin-New York, Mouton de Gruyter.
- Holm, John. 1988. *Pidgins and creoles*, vol. 1. Cambridge: Cambridge University Press.
- Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41. 9–45.
- Kihm, Alain. 1995. Tayo. The strange bird from New Caledonia. Determiners and Tense-Aspect in Tayo and their implications for creolization theories. *Journal of Pidgin and Creole Languages* 10. 225–252.
- Ladhams, John. 2012. Article agglutination and the African contribution to the Portuguese-based Creoles. In Bartens, Angela & Baker, Philip (a cura di), *Black through White*, 31–50. London: Battlebridge Publications.
- Lam, Chi Dat. 2019. Determiner-noun fusion in Haitian Creole: a statistical learning approach. (Paper presented at the *Annual Society for Pidgin and Creole Linguistics Summer Meeting*, Universidade de Lisboa, June 17–19.)
- Langacker, Ronald. 2008. *Cognitive Grammar: A basic introduction*. Oxford, Oxford University Press.
- Löbner, Sebastian. 2002. *Understanding Semantics*. London: Arnold Publishers.
- Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.
- Longobardi, Giuseppe. 1994. Reference and Proper Names: a Theory of N-Movement in Syntax and Logical Form. *Linguistic Inquiry* 25. 609–665.
- Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. (Cambridge Studies in Linguistics, 88). Cambridge-New York: Cambridge University Press.
- Li, Jessica. 2015. Nouns and Noun phrases. In Englebretson, Robert (a cura di), *A Basic Sketch Grammar of Gikūyū* (Rice Working Papers in Linguistics 4), 18–51, Special Issue. Available online at: <http://www.ruf.rice.edu/~reng/kik/sketch.pdf>
- Lyons, John. 1977. *Semantics*, voll. 1-2. Cambridge: Cambridge University Press.
- McWhorter, John & Parkvall, Mikael. 2002. Pas tout à fait du français : une étude créole. *Études créoles* 25 (1). 179–231.

- Michaelis, Susanne Marie & Rosalie, Marcel. 2009. Loanwords in Seychelles Creole. In Haspelmath, Martin & Tadmor, Uri (a cura di), *Loanwords in the World's Languages: A Comparative Handbook*, 215–229. Berlin: Mouton De Gruyter.
- Ndayiragije, Juvénal. 1989. La source du déterminant agglutiné en créole haïtien. *Canadian Journal of Linguistics* 34. 313–317.
- Ojeda, Almerindo. 1993. *Linguistic Individuals* (CSLI Lecture Notes 31). Stanford: CSLI Publications.
- Parkvall, Mikael. 2000. *Out of Africa. African influences in Atlantic Creoles*. London: Battlebridge Publications.
- Platt, John T. 1977. The creoloid as a special type of interlanguage. *Interlanguage Studies Bulletin* 2(3). 22–38.
- Russell, Bertrand. 1905. On Denoting. *Mind* (New Series) 14. 479–493.
- Searle, John R. 1969. *Speech Acts: an essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Strandquist, Rachel E. 2003. *Article Incorporation in Mauritian Creole*. University of Victoria (PhD Dissertation.)
- Strawson, Paul Francis. 1959. *Individuals. An Essay in Descriptive Metaphysics*. London: Routledge.
- Strawson, Paul Francis. 1950. On referring. *Mind* (New Series) 59. 320–344.
- Targète, Jean & Urciolo, Raphael G. 1993. *Haitian Creole-English Dictionary*. Kensington (USA): Dunwoody Press.
- Tourneau, Henry & Barbotin, Maurice. 1990. *Dictionnaire pratique du Créole de Guadeloupe (Marie-Galante)*, Paris : Karthala.
- Valdman, Albert. 2010. *A dictionary of Louisiana French. As spoken in Cajun, Creole and American Indian Communities*. Jackson: University Press of Mississippi.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*, Ithaca (NY): CLC Publications.
- Vergnaud, Jean-Roger & Zubizarreta, Maria Luisa. 1992. The definite determiner and the inalienable constructions in French and English. *Linguistic Inquiry* 23. 595–652.
- Wierzbicka, Anna. 1988. *The Semantics of Grammar*. Amsterdam-Philadelphia: John Benjamins.
- Zribi-Hertz, Anne & Glaude, Herby. 2006. Bare NPs and deficient DPs in Haitian and French: from morphosyntax to reference construal. In Baptista, Marlyse & Gueron, Jacqueline (a cura di), *Noun phrases in Creole languages*, 265–300. Amsterdam: Benjamins.
- Zribi-Hertz, Anne & Jean-Louis, Loic. 2014. From Noun to Name: definiteness marking in Modern Martinikè. In Cabredo Hofherr, Patricia & Zribi-Hertz, Anna (a cura di) *Crosslinguistic Studies on Noun Phrase Structure and Reference*, 269–315. Leiden-Boston: Brill.

Club Working Papers in Linguistics

Sezione

“Leggere per comprendere
a partire dai dati INVALSI”

Insegnare a leggere e capire

Mario Ambel

insegnare – Rivista CIDI

insegnare.ambel@gmail.com

Abstract

Il contributo espone le condizioni essenziali che consentono di predisporre contesti e attività finalizzate a incrementare le competenze di comprensione dei testi, in particolare espositivi.

Dopo aver ricordato il debito che la didattica della comprensione deve alle “Dieci Tesi per l’educazione linguistica democratica” (1975) e all’innovazione didattica degli anni Ottanta del secolo scorso, nella prima parte del contributo vengono sommariamente presentati alcuni errori di politica scolastica, sia complessiva che disciplinare, compiuti nel frattempo e che hanno inciso non poco sulle pratiche di classe e sulla loro efficacia.

Nella seconda parte viene presentato e discusso un approccio strategico alla comprensione del testo, che in qualche misura rappresenta la sistemazione di esperienze che si sono via via consolidate nel tempo, soprattutto dove e quando l’attività didattica non è stata condizionata da pratiche routinarie e valutative, bensì è stata orientata alla ricerca e alla sperimentazione di strategie adeguate di insegnamento-apprendimento dei processi di comprensione del testo.

In Appendice, infine, viene proposto un esempio di attività didattica, elaborata e sperimentata in contesti di formazione in servizio, secondo le metodiche della ricerca azione.

1. Una precisazione sul titolo e una premessa (apparentemente) datata

Il titolo del contributo, così come compariva nel programma del Convegno, focalizzava l’attenzione su *Imparare a leggere e capire*, anziché su “insegnare” a leggere e capire. Ma è proprio questo secondo aspetto che ho affrontato nella relazione e di cui tratterò anche qui: ovvero di che cosa si potrebbe e dovrebbe fare (ma anche di ciò che non si è fatto o non si dovrebbe più fare) affinché gli allievi imparino a leggere e capire oppure affinché migliorino le loro competenze di lettura e comprensione del testo. Come si diceva negli anni Ottanta del secolo scorso, prima che attorno al concetto di “competenza” si scatenasse un tornado concettuale e istituzionale dai contorni quanto mai contraddittori e dagli esiti non sempre confortanti.

In quegli anni, infatti, era usuale, fra chi si occupava di questi problemi, guardare alla comprensione del testo come a una competenza strategica articolata in sottocompetenze fra loro intrecciate, che avrebbero dovuto consentire al lettore di ricavare dal testo quanto il testo conteneva, magari cercando di essere consapevole dei rischi insiti nel far dire

al testo cose che non diceva o della parzialità, talvolta però funzionale, di ricavare dal testo solo quanto più utile ai propri scopi e interessi:

Il modo di leggere varia, oltre che in funzione dello scopo, anche in base al tipo di testo che si legge. [...]

Ora, poiché entrambi gli elementi, lo scopo e il tipo di testo, sono importanti nel determinare strategie efficaci, non si potrà non tenerne conto in ambito didattico dove insegnare a leggere vorrà dire, da un lato, insegnare a riconoscere eventualmente classificare le diverse forme testuali e a coglierne le caratteristiche distintive; dall'altro, a simulare o dare scopi di lettura. Per far sì che si legga in maniera finalizzata e dunque efficace ed economica.¹

Ed era anche usuale chiedersi che cosa davvero significasse capire o non capire un testo, magari per scoprire che non era (e non è tuttora) così semplice trovare risposte soddisfacenti:

Chiunque abbia una certa esperienza di aggiornamento nella scuola media inferiore o anche superiore sa che una delle lamentele più frequenti formulate dagli insegnanti (e spesso non soltanto da quelli di lettere) è «che gli allievi non sanno *nep-pure* leggere»; inoltre, se si va ad approfondire un'altra delle più frequenti accuse mosse agli allievi, «non sanno ragionare», si scopre che spesso c'è ancora dietro, almeno in parte, il «non sanno leggere», perché in realtà si rimprovera agli allievi di non saper interpretare, collegare e rielaborare concetti per lo più veicolati attraverso il testo scritto.

Il problema è generalmente formulato dagli insegnanti in termini troppo generici per essere anche soltanto verificabile, non parliamo poi di «risolvibile»: se si chiede agli insegnanti che cosa intendono con il loro «non sanno leggere», si ottiene una pluralità di risposte che, a seconda anche della fascia di scolarità a cui si riferiscono, vanno dal «non sanno leggere con espressione» al «non capiscono perché questo racconto è bello», dal «non sanno fare il riassunto» a «non capiscono il significato delle parole».²

Val forse la pena annotare come questa attenzione al processo di lettura-comprensione, alle sue dinamiche, al modo di orientarlo o rinforzarlo, o anche dalla pluralità e complessità del “non capire”, che era al centro della ricerca didattica, compiuta soprattutto dal basso e dagli insegnanti, dentro le scuole e presso le associazioni professionali, si sia di fatto arrestata o sia stata deviata verso altro. Che siano di fatto inesistenti o assai carenti, nel nostro sistema scolastico, luoghi e tradizioni di adeguata formazione alla professione docente e di reale ricerca didattica sperimentale è certamente una delle cause di una condizione sostanzialmente improvvisata e spesso aleatoria delle pratiche didattiche.

¹ Cfr. Pozzo (1982a: 9); queste parole di Graziella Pozzo sono tratte dalla *Introduzione* al volume *Insegnare la lingua. La comprensione del testo*, che raccoglieva gli atti di una delle tre giornate dell'omonimo Convegno organizzato da CIDI, GISCEL e LEND, a Ivrea il 5-6-7 marzo 1982. Le altre due giornate, e i rispettivi volumi, riguardano *Parlare e scrivere* e *Quale grammatica?*

² Cfr. Bertocchi (1982: 15); nella giornata del Convegno del 1982 dedicata alla “comprensione del testo” la relazione introduttiva venne tenuta da Daniela Bertocchi, un nome importante nella storia della didattica delle abilità linguistiche, *in primis* della lettura, che mi è caro ricordare qui, anche per il ruolo che ebbe nella qualificazione delle Prove INVALSI, che fanno da controcanto di questi ragionamenti.

2. Qualche errore del passato

Anche se apparentemente estranei al tema dell'intervento, è forse opportuno annotare qualche errore di politica scolastica, che è stato compiuto negli anni che sono intercorsi da quella stagione di reale rinnovamento della didattica dell'educazione linguistica e della comprensione. Si tratta di circostanze che hanno inciso non poco sulle pratiche didattiche e sulla loro efficacia. Le elenco per sommi capi, scusandomi per non avere lo spazio per argomentarle con maggior compiutezza.

In generale, a partire da quei presupposti, avrebbero dovuto consolidarsi pratiche didattiche curricolari che progressivamente dotassero l'allievo di strategie adeguate di lettura e comprensione dei testi. In secondo luogo, ma essenziale, si sarebbero dovuti sperimentare e tradurre in procedure didattiche efficaci i rapporti fra comprensione, riflessione linguistica e scrittura. Che cosa è accaduto, invece, ad impedire questa evoluzione, restaurando di fatto modalità contro cui si era mosso il rinnovamento promosso dalla "ricerca e dall'intervento sul campo dell'educazione linguistica" o, all'opposto, orientando le pratiche didattiche verso altro e altre modalità di "innovazione" spesso presunta?

Il discorso si farebbe lungo e impegnativo. Mi limito a indicare alcuni filoni di eventi che hanno indotto, più o meno consapevolmente o volutamente, questo arretramento e questa inversione di tendenza:

a) la persistenza di quella che nelle "Dieci Tesi" viene definita la "pedagogia linguistica tradizionale";

b) una maldestra gestione della dialettica (spesso inopportuna gestita come opposizione) fra piacere (della lettura) e competenza (di comprensione);

c) l'impatto complessivamente negativo delle prove INVALSI e delle ricadute anche indirette sulle metodologie didattiche;

d) più recentemente, la reazione di molti docenti sia alla didattica per competenze³ che al *teaching to test*⁴ (spesso in area umanistica erroneamente assimilati), che ha portato, per reazione, non solo a privilegiare un massiccio e spesso conservatore ritorno alla letteratura, ai classici, ai testi impegnativi e fortemente significativi, ma persino alla storia della letteratura e alle "poetiche" nella scuola di base.

Come se non bastasse, hanno pesato alcuni fattori estranei allo specifico dell'insegnamento-apprendimento delle strategie linguistiche, ma che – data la loro portata generale – hanno avuto ripercussioni anche su questo ambito:

e) la gestione ministeriale improvvida e pervasiva delle politiche di valutazione (sia degli apprendimenti individuali che di sistema) e il ritorno, nella scuola di base, della valutazione decimale attraverso i "voti", che, approfittando dell'ipocrisia o dell'ignoranza delle normative, ha orientato i docenti ad utilizzare non solo la finalità premiale della

³ Di competenze si è molto discusso in questi anni, anche se raramente si sono davvero perseguite metodologie di coinvolgimento attivo e consapevole degli allievi nei processi di apprendimento, che sarebbero il presupposto fondamentale di una effettiva didattica per competenze.

⁴ Letteralmente sta a indicare un insegnamento che prepara a rispondere a quesiti, spesso a scelta multipla, la cui diffusione, nonostante le stesse indicazioni ministeriali, è stata anche in Italia favorita dall'incidenza istituzionale delle prove standardizzate e di molta editoria scolastica di tipo addestrativo.

valutazione in luogo di quella formativa, ma anche ad estenderla dalle procedure di valutazione di lungo periodo previste dalla legge a quelle quotidiane;⁵

f) il caos e spesso l'improvvisazione che regnano nell'applicazione di molte reali o presunte innovazioni metodologiche e strumentali, spesso tentate e realizzate fuori o a scapito di una ragionevole e attenta analisi del rapporto fra metodologie didattiche coerenti ed epistemologie disciplinari o eventualmente interdisciplinari e spesso con una sostanziale inversione di priorità fra mezzi e fini;

g) il perdurare, accresciuto in questi ultimi anni, della pressione di contenuti di attualità sempre più "sensibili" e cogenti, che hanno fortemente accentuato la dimensione tematica (e talvolta valoriale) dell'approccio ai testi, a scapito di quelle retorico-formali e procedurali.

E infine, come conseguenza di questi fattori di disorientamento:

h) la sostanziale fragilità delle pratiche messe in campo per fronteggiare le trasformazioni linguistiche, testuali e cognitive degli universi della comunicazione contemporanea, prima fra tutte, ovviamente, la massiccia digitalizzazione dei mezzi e delle procedure.

Uno degli aspetti più rilevanti di questi eventi è stato il progressivo spostamento da quella attenzione al processo di comprensione e alle sue implicazioni di cui si diceva ad un eccesso di attenzione al prodotto della comprensione e alla sua verifica. In virtù delle tematiche e delle ragioni stesse del convegno Giscel dedicato alla lettura e alla comprensione è opportuna una precisazione relativa, in tal senso, al ruolo dell'INVALSI. Questo spostamento di attenzione dalla ricerca sui processi di comprensione alla verifica anche ossessiva del prodotto (spesso presunto) della avvenuta comprensione è stato uno degli effetti collaterali più gravi dell'introduzione delle prove INVALSI, del loro uso come forma di valutazione degli apprendimenti individuali, della strumentalizzazione degli esiti che ne hanno spesso fatto i media nel discorso pubblico sulla scuola. Ciò è avvenuto in parte malgrado le intenzioni stesse dell'Istituto o nonostante le raccomandazioni delle Indicazioni nazionali,⁶ ma complice una editoria improvvida (e i docenti che l'hanno assecondata), che ha trasformato le attività funzionali all'acquisizione di strategie di comprensione in una stanca ripetizione di domande desunte, spesso anche maldestramente, dalle metodologie di indagine adottate dall'INVALSI per verificare la comprensione, al di là di quanto fossero davvero efficaci e adeguate a ottenere lo scopo.

È certamente giunto il momento di interrogarsi sull'impatto complessivamente negativo delle ricadute anche indirette sulle metodologie didattiche del profluvio di "prove" che ha invaso la scuola negli ultimi vent'anni, dovuto sostanzialmente al pessimo rapporto che si è generato fra scuole e "prove" istituzionali. Ne sono scaturiti due esiti altrettanto negativi e complementari: l'addestramento pedissequo alle prove oppure il rifiuto pregiudiziale e contrastivo.

⁵ Fra il tempo della relazione orale (dicembre 2019) e quello della stesura della relazione scritta (aprile 2021) è intervenuta la soppressione dei voti decimali (Ordinanza Ministeriale del 4.12.2020), almeno nella scuola primaria, che potrebbe riaprire una stagione di attenzione qualitativa e formativa ai processi di apprendimento a scapito dell'attenzione ponderale e classificatoria ai risultati, anche se proprio la limitazione alla sola primaria, grava il provvedimento di motivazioni parziali e controproducenti.

⁶ Cfr. MIUR (2012: 19): "L'Istituto nazionale di valutazione rileva e misura gli apprendimenti con riferimento ai traguardi e agli obiettivi previsti dalle Indicazioni, promuovendo, altresì, una cultura della valutazione che scoraggi qualunque forma di addestramento finalizzata all'esclusivo superamento delle prove".

3. La differenza fra attività (per la comprensione) e prove di verifica (della comprensione)

Anche per i motivi fin qui esposti il primo e ineludibile presupposto di una didattica funzionale alla comprensione – che restituisca centralità ai processi di apprendimento e circoscriva le metodiche di verifica sommativa agli ambiti, ai tempi e ai fini che compongono loro – è stabilire una netta distinzione fra attività finalizzate e funzionali all’acquisizione e al consolidamento di strategie di comprensione e prove finalizzate alla verifica ed eventualmente alla valutazione dell’avvenuta comprensione, previa delimitazione seria e coerente di che cosa si debba intendere per comprensione.

Tra queste due dimensioni, ovvero le attività per l’apprendimento e le prove per la verifica, vi sono e vanno mantenute differenze sostanziali di varia natura e importanza: di fine, di tempo, di modalità di attuazione, di pratiche didattiche conseguenti. Si tratta infatti di due modalità di approccio al testo che rispondono a finalità profondamente diverse:

- accompagnare e rinforzare il processo di comprensione;
- ricostruire e verificare il prodotto della comprensione.

L’attività di apprendimento della comprensione (talvolta anche di parti o di aspetti specifici della testualità) comporta da parte dell’allievo-lettore un atteggiamento attivo, problematico, analitico, mentre la verifica induce atteggiamenti di dimostrazione, risposta, sintesi. Potremmo dire che chi sta eseguendo una attività di apprendimento si pone domande e le pone al testo, mentre chi svolge una prova di verifica (e viene egli stesso per questo sottoposto a domande) cerca risposte sul testo e in sé. E quale delle due procedure sia quella che può generare apprendimento è evidente. O meglio dovrebbe essere evidente da molto tempo.⁷

Inoltre un’attività di apprendimento spesso seleziona, focalizza uno o più aspetti del testo o una particolare procedura di relazione fra testo e lettore, procede per sottrazione e non per accumulo, mentre la verifica inevitabilmente deve o pretende di assumere un atteggiamento esaustivo, anche se spesso lo persegue parcellizzando la comprensione in una miriade di domande che randomizzano le procedure di comprensione invece di orientarle, andando alla ricerca di frammenti del processo di significazione, anziché puntare al cuore dei problemi posti dal testo.

Inoltre un’attività di apprendimento dovrebbe seguire, orientare, facilitare i processi linguistico-cognitivi compiuti dal lettore prima, durante e dopo la lettura, mentre una prova di verifica inevitabilmente segue a calco una qualche teoria del prodotto della comprensione. Un esempio di protocollo interessante al riguardo è dato dal “Quadro di Riferimento” INVALSI, che espone una coerente idea di che cosa si debba intendere per prodotto della comprensione, ma che non è assolutamente detto che coincida con il processo attraverso il quale il lettore raggiunge quel risultato.

Senza contare, infine, che una strategia di apprendimento mira a seguire e rinforzare il processo di comprensione e quindi le attività che propongono tendono a ricostruire la linearità o la consapevole selettività dell’approccio con il testo, mentre una verifica ri-

⁷ Sull’ampia bibliografia dedicata al rapporto dinamico e interrogativo fra lettore e testo, si segnalano, tra gli altri Pozzo (1982c e 1982b); Bertocchi (1982 e 1983); Della Casa (1987); Colombo (2002); Ambel (2006, 2010, 2012, 2015); Lumbelli (2009).

schia spesso di alterare volutamente il rapporto fra sequenzialità della lettura e luoghi del testo su cui ogni singolo quesito tende a porre l'attenzione. Quando addirittura l'alterazione della *ratio* delle domande non è dovuta a fattori del tutto estranei alla serietà procedurale, come evitare che gli allievi copino fra loro le risposte!

Per tutti questi motivi è assai pericoloso ridurre l'attività didattica di comprensione a una successione di messe in atto o di simulazioni di prove di verifica, cosa che invece troppo spesso accade, o nei fascicoletti predisposti all'uso, da una editoria più sollecitata alle sirene del mercato che alle esigenze pedagogiche, oppure nella standardizzazione delle attività proposte dopo i singoli testi, che non sempre rispondono a una chiara metodologia di approccio all'apprendimento e al rinforzo dei processi di comprensione.

Non a caso, negli anni Ottanta del secolo scorso, in quella stagione di rinnovamento di cui si sono inizialmente ricordate le motivazioni di fondo, si andarono via via consolidando pratiche di analisi e interazione con il testo che ruotavano a quelle che, semplificando, possiamo definire le "giuste domande", anche se non si trattava solo di quesiti aperti o chiusi in senso stretto, ma di una pluralità di azioni linguistiche e cognitive sul testo e a partire dal testo: sottolineature, titolazioni, sintesi parziali, raccolta di informazioni, discussioni sul senso del testo o di sue parti, rappresentazioni schematiche del contenuto, analisi delle caratteristiche formali, ecc.

Si maturò allora la consapevolezza – che si tradusse in pratiche didattiche e anche in alcuni prodotti editoriali, e che dovrebbe guidarci ancora oggi – della necessità che le attività didattiche proposte coniughino e portino a sintesi alcune variabili:

- la natura del testo;
- lo scopo per cui si legge;
- il livello di conoscenze e competenze di chi legge;
- il fine didattico che si persegue.

Mentre ora, troppo spesso, le attività proposte si limitano a quella imitazione di cui si è detto della *ratio* (neppure sempre ben capita e padroneggiata) delle prove INVALSI. L'effetto che ne scaturisce è duplice e davvero perverso. Da un lato, infatti, non si impara a comprendere i testi, perché ripetere ossessivamente prove può forse allenare ad affrontarle, ma non rinforza certo le procedure linguistiche e cognitive della comprensione con effetti a lungo termine. Dall'altro, a lungo andare, si induce (o si è già indotta?) la scuola a sottovalutare e dimenticare come sia fatta una reale didattica della comprensione, per affidarsi alla sola routine di addestrare al superamento di prove. È uno dei tanti modi con cui la didattica, in particolare degli ultimi vent'anni, si è allontanata da finalità e pratiche di reale emancipazione, a vantaggio di pratiche di addestramento passivizzante.

C'è infatti un ultimo corollario che non va dimenticato e sottovalutato; una buona didattica della comprensione tende a conciliarne due dimensioni altrettanto importanti: quella individuale di lettura silenziosa, comprensione e studio e quella cooperativa, collettiva, sociale, mentre il *teaching to test* esalta e al contempo deprime la dimensione individualistica e competitiva dell'approccio al testo. Nel primo caso ci si muove in una logica processuale di *problem solving*, nel secondo di dimostrazione produttivistica del risultato ottenuto.

4. Per un approccio strategico alla comprensione del testo

L'approccio alla comprensione del testo che viene qui proposto ha poco o nulla di innovativo o di almeno in parte non visto o già sperimentato. È piuttosto la sistemazione delle esperienze che in questi anni si sono via via consolidate quando l'attività didattica non è stata condizionata da pratiche routinarie e valutative, bensì è stata orientata alla ricerca e alla sperimentazione di strategie adeguate di insegnamento-apprendimento dei processi di comprensione del testo.

4.1. Impianto metodologico

Tale approccio si fonda anzitutto sulla distinzione fra attività da compiere in momenti diversi della lettura-comprensione, ovvero da sollecitare e guidare:

- PRIMA
- DURANTE
- DOPO

la lettura del testo.

Le attività da proporre agli allievi in questi diversi momenti si differenziano per natura, modalità di attuazione, ma soprattutto per il fine che si pongono e le strategie linguistiche e cognitive che sollecitano in rapporto al testo, alle sue parti, ai suoi vari aspetti. Si tratta infatti di strategie da adottare e far progressivamente acquisire, tali da orientare/facilitare/accompagnare/esercitare/verificare/usare il processo della comprensione attraverso attività e domande da svolgere quindi in fasi diverse, ciascuna di queste fasi ha le sue peculiarità e le sue modalità di attuazione.

Indico pertanto, per sommi capi, come si dovrebbero caratterizzare le strategie di comprensione da attivare...

- PRIMA della lettura: orientate al lettore e preparatorie del processo di lettura/comprendimento, sono finalizzate a mettere a fuoco il tema, richiamare alla mente conoscenze, suscitare aspettative, indirizzare il processo di lettura e comprensione e si esercitano attraverso attività e domande di
 - ANTICIPAZIONE: formulazione di ipotesi su contenuti e scopi; richiamo di conoscenze;
 - ORIENTAMENTO: sollecitazione aspettative; orientamento del processo di lettura.
- DURANTE la lettura: orientate e funzionali al processo di lettura/comprendimento, sono finalizzate a facilitare e guidare il processo di lettura, sciogliere dubbi e superare difficoltà e ostacoli, per disinnescare inciampi o fraintendimenti e si esercitano attraverso attività e domande di
 - ACCOMPAGNAMENTO: ricostruzione e accompagnamento del processo di lettura, rinforzo delle procedure funzionali alla comprensione;
 - FACILITAZIONE: facilitazione del processo di lettura, approccio alle questioni lessicali, interventi sui nessi coesivi, ecc.
- DOPO una o più letture: orientate al testo e funzionali alla sistematizzazione della comprensione, sono finalizzate a rinforzare – con modalità esercitative ed euristiche – la comprensione, la consapevolezza e il controllo del testo o di alcune sue parti e si esercitano attraverso attività e domande di

- RICOSTRUZIONE: ricostruzione della coerenza testuale e della struttura tematica, delle informazioni e dei contenuti del testo;
- RIELABORAZIONE: rielaborazione del testo o di sue parti attraverso modalità diverse di sintesi, riscrittura, risposta, riuso.

4.2. Un esempio: contesto e modalità di attuazione

È importante rilevare che l'esempio che viene proposto in Appendice è stato elaborato e sperimentato in contesti di formazione in servizio, realizzati secondo le metodiche della ricerca-azione: prima, in una versione più ampia con docenti di scuola secondaria di I grado di Piosasco (TO) e, successivamente, nella versione qui allegata, con docenti di scuola primaria e secondaria di I grado di Cadeo (PC).⁸

È bene fare questa precisazione poiché la progettazione delle attività, soprattutto nella prima stesura più ampia e qui non riproposta, risente delle particolari condizioni in cui è stata realizzata. Scopo del lavoro era infatti realizzare un'attività da proporre ad una classe prima per verificarne e discuterne l'efficacia.

L'individuazione e la costruzione delle attività tenevano quindi conto delle variabili di contesto (ordine di scuola e classe, tipologia degli allievi, collocazione nel progetto curricolare), ma anche della necessità di sottoporre a verifica sperimentale scelte metodologiche e procedurali, oltre alle ipotesi relative all'impianto epistemologico disciplinare. Da questo punto di vista, tra l'altro, è bene osservare la natura dei testi scelti in queste sperimentazioni: quasi sempre testi divulgativi che affrontano tematiche proprie di "altre" discipline (storia, geografia, scienze, tecnologia, diritto, ecc.). In tal senso, nei gruppi di ricerca-azione, si è a lungo discusso di quale fosse la collocazione ideale del lavoro attorno a testi che possono essere oggetto di approcci disciplinari (o interdisciplinari) diversi. Qui si è optato per finalità e procedure legate all'incremento delle competenze di comprensione (e quindi in un contesto di "italiano"). Va da sé che le attività proposte dovrebbero essere in buona misura diverse se il testo fosse utilizzato in attività finalizzate, per esempio, all'incremento delle abilità di studio, oppure in una unità didattica disciplinare, in questo caso di scienze sulla migrazione degli uccelli.

Molte delle attività qui esemplificate pongono questo argomento al centro del lavoro sul testo, ma perché e in quanto costituisce il focus delle motivazioni e dei significati del testo stesso e non perché sia oggetto di uno specifico approccio disciplinare.

In generale va detto che i complessi problemi posti dai testi espositivi, infatti, dovrebbero essere oggetto di attenzioni molto più sollecite e accurate di quanto non accada, nella scuola, spesso a vantaggio dei testi letterari oppure di quelli argomentativi.

Per esempio – durante la progettazione, l'attuazione in classe e poi la discussione professionale sugli esiti della sperimentazione – sono state di estremo interesse le notazioni fatte in risposta alla domanda "Ma perché gli uccelli migrano?", posta da alcune attività e che ha coinvolto in un processo di progressiva scoperta le stesse insegnanti di

⁸ Negli ultimi anni questi spunti metodologici sono stati sperimentati in alcuni gruppi volontari di ricerca-azione, attivi presso alcuni Cidi territoriali (a Pordenone, Brescia, Pescara, Cosenza), in alcune attività istituzionali di formazione in servizio e sperimentazione sul campo (a Travagliato-BS, Piosasco-TO, Cadeo-PC), in un progetto di ricerca sui rapporti fra comprensione e riflessione grammaticale promosso da Indire (a Bolzano); esprimo qui il mio ringraziamento a tutte le colleghe e i colleghi che hanno variamente partecipato e stanno tuttora lavorando a queste modalità di incremento delle competenze professionali.

“italiano” protagoniste dei gruppi di ricerca. Se da un lato si è così potuto verificare quanto l’enciclopedia del lettore e le sue specifiche conoscenze siano fondamentali nella comprensione (come nella non comprensione) di questo tipo di testi, dall’altro, in quanto, docenti di lingua, abbiamo spesso sottoposto questi testi ad indagini critiche sulla reale consistenza della loro testualità, spesso non priva di incoerenze, di approssimazioni o improprietà coesive e persino sintattiche. Aspetti che hanno generato anche interessanti discussioni sulla legittimità dell’impiego didattico di testi che, se da un lato non sono certo esemplari, dall’altro costituiscono un importante terreno di esercizio delle competenze di comprensione negli universi attuali della comunicazione sociale. Nonostante o forse proprio in virtù delle loro approssimazioni o carenze formali e talvolta, anche, delle evidenti manipolazioni del contenuto.

5. Qualche implicazione vincolante

Per collocare queste proposte in un’ottica curricolare e disciplinare coerente e completa, sono indispensabili altre due precisazioni conclusive.

Ritengo anzitutto che per consolidare e realizzare adeguate strategie di insegnamento-apprendimento della comprensione del testo, oltre a predisporre attività coerenti ed efficaci, sia necessario affrontare contestualmente alcune questioni non del tutto risolte:

- Studiare specificità e differenze che esistono fra apprendimento strategico e verifiche sistematiche, per mettere in atto rapporti virtuosi e non condizionanti a ritroso fra modalità di apprendimento e momenti di verifica sommativa dei risultati raggiunti.
- Consolidare e tradurre in coerenti progettazioni curricolari i rapporti fra differenti strategie di lettura-comprensione (la fruizione, la comprensione, l’analisi, l’interpretazione) e diversi di scopi di riscrittura ad esse funzionali e complementari (parafrasi, sintesi, recensione, commento, imitazione, citazione, confutazione responsiva), su una gamma ampia di tipologie testuali organizzate in modo coerente e conseguente.
- Approfondire, consolidare e tradurre in pratiche didattiche i rapporti teorici, metodologici e procedurali fra comprensione dei testi – riflessione sulla lingua e grammatica – riuso, ri/scritture e scritture.

Infine, ma solo su queste basi, sarà necessario affrontare la più attuale delle questioni irrisolte:

- Studiare e consolidare le strategie di comprensione e riuso dei testi realizzati e veicolati su supporti digitali, sia che essi siano fruiti che prodotti.

Sarà però importante affrontare quest’ultima questione, dopo aver adeguatamente risolto e portato a compimento le tre precedenti, perché il farlo senza quei presupposti significherebbe avventurarsi in uno scenario ancor più impegnativo e talvolta ingannevole, senza un solido bagaglio epistemologico e metodologico alle spalle. E potrebbe rivelarsi molto pericoloso.

E forse è già un po' tardi! Troppi usano testi su supporti cartacei, digitali o misti, spesso anche relativi a tematiche assai complesse, e vi si accostano con eccessiva superficialità e approssimazione. I rischi sono già evidenti: non solo quello di non accrescere i meccanismi di comprensione, ma soprattutto quello di sostituire, fra gli esiti parziali o negativi della comprensione non avvenuta, il fraintendimento con la presunzione di comprensione e di conoscenza, che è assai peggio, a livello individuale ma soprattutto sociale. Oggi, infatti, nell'affrontare le strategie comunicative, dobbiamo essere operativamente consapevoli di una finalità di estremo rilievo, soprattutto in ottica di cittadinanza consapevole: combattere la presunzione di comprensione, alimento fondamentale di quella presunzione di conoscenza, che sta inquinando la vita democratica, certamente nel dibattito pubblico e "social", ma non raramente anche negli ambiti decisionali.

6. Cornici culturali di contesto

In una prospettiva curricolare più ampia, che ponga in rapporto di interdipendenza fra loro e in prospettiva di emancipazione individuale e sociale le diverse "abilità" e competenze interessate dall'educazione linguistica, così come le *Dieci tesi* e Tullio De Mauro ci hanno insegnato a fare, vorrei concludere indicando tre cornici culturali programmatiche, nelle quali tutto l'impianto metodologico di insegnamento-apprendimento delle strategie di comprensione e riuso dei testi andrebbe collocato.

Per quanto guardino inevitabilmente anche al futuro, tali cornici hanno radici in buona misura datate.

La prima riguarda quella che alla fine degli anni Settanta veniva indicata come la "trasversalità" dell'educazione linguistica: proclamata nella settima delle *Dieci tesi*, caldamente raccomandata nei "Programmi per la scuola media" del 1979 ed eternamente disattesa in questi ormai più di quarant'anni da tutti gli ordini di scuola e dalla stragrande maggioranza delle pratiche didattiche:

VII. Limiti della pedagogia linguistica tradizionale

A) La pedagogia linguistica tradizionale pretende di operare settorialmente, nell'ora detta «di Italiano». Essa ignora la portata generale dei processi di maturazione linguistica (tesi I) e quindi la necessità di coinvolgere nei fini dello sviluppo delle capacità linguistiche non una, ma tutte le materie, non uno, ma tutti gli insegnanti (Educazione fisica, che è fondamentale, se è fatta sul serio, compresa). La pedagogia linguistica tradizionale bada soltanto alle capacità produttive, e per giunta scarse e scarsamente motivate da necessità reali. Le capacità linguistiche ricettive sono ignorate, e con ciò è ignorata non tanto e solo la metà del linguaggio fatta di capacità di capire le parole lette e scritte, ma proprio quella metà che è condizione necessaria (anche se non sufficiente) per il funzionamento dell'altra metà: come il bambino impara prima a individuare le frasi, ad ascoltare e capire, e poi impara a produrre parole e frasi, così da adulti prima dobbiamo leggere e rileggere e udire e capire una parola, poi ci avventuriamo ad usarla. Ma la pedagogia linguistica tradizionale non fa alcun conto di ciò. Anzi, fa peggio. Molto spesso il bambino (e così l'adulto) controlla la bontà della ricezione col collaudo. Vi sono insegnanti che non si rendono conto di ciò e condannano le sperimentazio-

ni con cui l'allievo controlla sue ricezioni parziali o sue ipotesi provvisorie sulla funzione e il valore di un elemento linguistico appena appreso.⁹

La scuola italiana amplia, talvolta a dismisura, i terreni tematici e le ambizioni culturali delle problematiche che affronta, spesso anche nella totale disattenzione delle enciclopedie di riferimento degli allievi e della complessità delle problematiche pluridisciplinari messe in campo, ma continua spesso a pretendere di farlo nell'ora detta «di Italiano», come affermava quella Tesi. Ora, in verità, con la reintroduzione della “educazione civica”, sembra delinarsi uno spazio che viene definito “trasversale”, nel quale trovano spazio quelle tematiche. E quindi, probabilmente l'approccio ai testi che le affrontano. Ciò che va assolutamente salvaguardato e curato con estrema attenzione è che l'approccio a quei testi e a quelle problematiche, mentre doverosamente e inevitabilmente esce dalla sola ora “di italiano”, venga affrontato con strategie conoscitive e procedurali adeguate, nelle quali si riconoscano e separino oppure si intreccino consapevolmente, ma comunque trovino concreta e graduale attuazione didattica, in ottica di apprendimento e non solo di uso irriflesso, sia le strategie linguistico-testuali che il progressivo controllo/acquisizione delle conoscenze e delle enciclopedie di riferimento. E non è facile.

La seconda cornice riguarda invece la prospettiva del plurilinguismo, che deve inevitabilmente far da sfondo problematico all'insegnamento-apprendimento delle abilità e delle strategie linguistiche. Oggi il concetto di plurilinguismo si è andato ampliando in una società che ha assorbito pluralità di idiomi all'interno degli spazi più ampi di relazione istituzionale, economica e sociale, ma anche all'interno dei confini nazionali e delle stesse territorialità circoscritte, in cui si compiono la vita quotidiana e le relazioni che essa implica.¹⁰

Ma anche in questo caso, non è male tornare ad un concetto e a un insegnamento antico in tema di “plurilinguismo”, o quanto meno di varietà delle forme testuali oggetto dei processi di insegnamento:

con plurilinguismo intendiamo qui la compresenza sia di linguaggi di tipo diverso (verbale, gestuale, iconico, ecc), cioè di diversi tipi di semiosi, sia di idiomi diversi, sia di diverse norme di realizzazione d'un medesimo idioma. Esso pare una condizione permanente della specie umana e, quindi, di ogni società umana. L'età contemporanea, attraverso i grandi moti migratori, l'eccezionale sviluppo dei linguaggi artificiali e di quelli formalizzanti o formali delle scienze, l'introduzione e diffusione di tecnologie che consentono la riproduzione e diffusione mondiale e di massa delle produzioni segniche d'ogni tipo, esalta oltre ogni limite noto la capacità plurilingue delle società umane.¹¹

Infine la terza cornice di riferimento riguarda la necessità di muoversi in una prospettiva multimediale e soprattutto multimodale. Anche in questo caso non si tratta di una novità assoluta, poiché da molti decenni ormai l'insegnamento-apprendimento della lingua dovrebbe aver abbandonato la centralità del linguaggio formale scritto, a vantaggio di

⁹ Cfr. Gisel (1975); si legge nella presentazione del testo ora in rete sul sito del GISCEL: “Le Dieci tesi in questa redazione sono un testo collettivo preparato dai soci del GISCEL nell'inverno e primavera del 1975 e definitivamente approvato in una riunione tenutasi alla Casa della Cultura di Roma il 26 aprile 1975”. Per i “Programmi” del 1979 vedi De Mauro & Lombardo Radice (1979).

¹⁰ Si vedano Minardi (2020), Cavalli (2020), Paone (2019).

¹¹ Cfr. De Mauro (1977: 144).

una miscela assai polivalente e varia di media, supporti e contesti di esercizio della comunicazione formale e informale.

Oggi questo scenario è reso certamente più problematico dall'impatto con la lettura e la scrittura su supporti e media informatici, e ormai anche il parlato e l'ascolto, che tra l'altro l'emergenza pandemica ha amplificato ben oltre i limiti sui quali solo un anno fa eravamo tutti pronti ad attestarci.

Ma è indubbio che anche il rapporto con gli strumenti della comunicazione in senso lato debba compiere un salto di qualità ed essere assunto in un'ottica più complessa, che si occupi dell'intero retroterra personale e sociale, culturale, ideologico e politico che si cela dietro l'uso di media diversi o della comunicazione che si realizza attraverso una pluralità di linguaggi, strumenti e contesti.¹² La salvaguardia e il rinforzo delle competenze di scrittura e lettura del testo scritto si inquadra oggi in uno scenario di maggior complessità degli universi della comunicazione e delle strategie necessarie per viverci oggi in modo consapevole e critico. Per farlo da protagonisti e non da "sudditi" cognitivi, linguistici e culturali sono necessarie strategie adeguate e di non facile acquisizione e controllo. E forse le strategie di comprensione e riuso consapevole del testo scritto su supporto cartaceo continuano a costituire la base anche per l'acquisizione di quelle funzionali a comunicazioni miste e veicolate su supporti diversi. E in ogni caso, i compiti che la pedagogia della comprensione dei testi nella società contemporanea ha di fronte sono tali che sarebbero indispensabili l'innalzamento dell'obbligo scolastico e una maggior uniformità fra le finalità dei vari indirizzi dei livelli scolastici di II grado. Ma questo è un altro discorso. O forse no.

¹² Cfr. Paone (2019).

Appendice

PRIMA PARTE.

Guardare, ascoltare e comprendere un video

Gli uccelli migratori

- Guardate e ascoltate questo servizio del Tg Regionale del Trentino
<https://www.rainews.it/tgr/trento/video/2018/10/tnt-Fauna-Uccelli-Muse-Caset-Ledro-b44e92ec-e1c3-405f-b918-d2b8985614a4.html>
- Ora rispondete a coppie per scritto a questa domanda:
Perché i ricercatori dell'osservatorio mettono le reti e catturano gli uccelli?
- Il titolo del servizio è:
Uccelli migratori: sempre meno quelli in rotta verso sud
Secondo voi, il servizio spiega perché? Confrontate e discutete le vostre opinioni.
- Ora riguardate il servizio per verificare le vostre impressioni.
- Questo è un fermo immagine del servizio:



Scrivete che cosa stanno facendo le due persone ritratte nell'immagine.

SECONDA PARTE.

Prima di leggere

Leggendo i diversi testi (anche quello sulla rondine), rispondendo alle domande, discutendo fra voi vi siete certamente fatti un'idea del perché gli uccelli migrano.

Scrivi la tua opinione.

- Ora leggerai un breve testo intitolato
Ma perché gli uccelli migrano?
che dovrebbe quindi rispondere a questo interrogativo!

Durante la lettura

- Durante una prima lettura (individuale e silenziosa) sottolinea le frasi che rispondono a quella domanda.
- Durante una seconda lettura (collettiva e fatta ad alta voce da alcuni di voi) chiedetevi e verificate che cosa intende il testo con “risorse vitali”

Testo

Gli uccelli migratori si spostano principalmente per nidificare e cercare cibo sufficiente e più adatto per allevare i loro piccoli. La migrazione è un evento cruciale nella vita di migliaia di specie di uccelli, è necessaria alla loro stessa sopravvivenza.

Se molte specie di uccelli non migrassero, aumenterebbe notevolmente la competizione per le risorse vitali all'interno di uno stesso ambiente. Per questo motivo, gli uccelli hanno sviluppato diversi schemi di migrazione in diversi periodi dell'anno, seguendo rotte diverse per garantire a se stessi e ai propri giovani maggiori possibilità di sopravvivenza.

Non tutti gli uccelli però migrano. Alcune specie si sono adattate bene al proprio ambiente, imparando a trarre vantaggio a livello alimentare nelle diverse stagioni dell'anno. Questo gli ha permesso di diventare stanziali e non dover viaggiare tutto l'anno. Altri uccelli ancora si sono abituati ai climi freddi cominciando a immagazzinare riserve di grasso e cambiando il proprio piumaggio per sopravvivere alle stagioni più rigide. In ogni caso ben oltre la metà delle specie di uccelli nel mondo ha bisogno di migrare.

Dopo la lettura

- Confrontate e commentate le frasi che avete individuato durante la lettura
- Ora rispondete individualmente per scritto:

Tutti gli uccelli migrano? Tutti migrano nello stesso periodo?

TERZA PARTE.

Riscrittura

- Sulle cause delle migrazioni degli uccelli, e in particolare delle rondini, ti proponiamo altri due brevi testi tra loro in qualche modo collegati: il primo è tratto dal compito di un bambino come voi (di qualche anno fa); il secondo da un articolo su un sito scientifico.

Le rondini preferiscono fare il nido in stalle e costruzioni in piena campagna: rimesse, garage e altri edifici rurali. Questo perché la specie è molto legata alla presenza di grandi mammiferi; in Africa possono essere bufali, gazzelle o mucche domestiche, che camminando nella savana spaventano e fanno alzare i piccoli insetti di cui si cibano le rondini. Anche se da noi si devono accontentare di qualche mucca o cavallo al pascolo.

<https://www.focus.it/ambiente/animali/perche-la-rondine-non-si-posa-mai-per-terra>

La migrazione degli uccelli
Il freddo rappresenta un pericolo per la vita di molti animali.
Gli insetti all'arrivo del freddo muoiono; allora gli uccelli che si cibano d'insetti, come la rondine, non hanno più nulla da mangiare e per sopravvivere migrano nei luoghi dove fa caldo e dove trovano gli insetti.
Gli uccelli che mangiano solo insetti si chiamano insettivori.

- Questi due testi aggiungono qualcosa a quello che già sapevi sulla migrazione delle rondini?
- Ora, usando le informazioni che hai raccolto durante le diverse attività, scrivi un breve testo espositivo (circa 10,12 righe) per spiegare perché le rondini, in autunno, migrano verso paesi più caldi.

Bibliografia

Ambel, Mario. 2006. *Quel che ho capito*. Roma: Carocci.

Ambel, Mario. 2012. *Le "prove Invalsi" di comprensione del testo: analisi riflessiva e proposte didattiche*, Indire – Valore scuola Risorse. (http://www.scuolavalore.indire.it/nuove_risorse/le-prove-invalsi-di-comprensione-del-testo-analisi-riflessiva-e-proposte-didattiche/) (Consultato il 03.08.2021.)

- Ambel, Mario. 2015. *Le strategie di lettura e uso dei testi e le relative "comprensioni"*, Indire – Valore scuola Risorse http://www.scuolavalore.indire.it/nuove_risorse/le-strategie-di-lettura-e-uso-dei-testi-e-le-relative-comprensioni/ (Consultato il 03.08.2021.)
- Ambel, Mario (a cura di). 2020. *Una scuola per la Cittadinanza*. Varazze: PM edizioni.
- Bertocchi, Daniela. 1982. Curricolo di lettura, lettura nel curricolo. In Pozzo (1982c), 15–37.
- Bertocchi, Daniela. 1983. *La lettura*. Lecce: Milella.
- Cavalli, Marisa, 2020, PLURILINGUISMO - Per una cittadinanza rispettosa dei diritti linguistici, in Ambel (2020), vol. 2: 206–212.
- Colombo, Adriano 2002. *Leggere. Capire e non capire*, Bologna: Zanichelli.
- Bruni, Emilio Emma Maria. 2019. *Una pedagogia possibile per l'intercultura*. Milano: Franco Angeli.
- Della Casa, Maurizio. 1987. *La comprensione dei testi*. Collana a cura del Cidi. Milano: Franco Angeli.
- Lumbelli, Lucia. 2009. *La comprensione come problema*, con "Prefazione" di Tullio De Mauro. Bari: Laterza.
- De Mauro, Tullio. 1977. Il plurilinguismo nella società e nella scuola italiana. In De Mauro, Tullio, *Scuola e linguaggio*, Roma: Editori Riuniti; già in AA.VV. *La radio nella scuola di oggi*, Torino: Stampatori, 1975. In De Mauro (1977), 138-151 è reperibile anche una stesura delle "Dieci tesi per l'educazione linguistica democratica".
- De Mauro, Tullio & Lombardo Radice, Lucio, (a cura di). 1979. *I nuovi programmi della scuola media*, Roma: Editori Riuniti.
- GISCEL. 1975. *Dieci tesi per l'educazione linguistica democratica*. <https://giscel.it/dieci-tesi-per-leducazione-linguistica-democratica/> (Consultato il 03.08.2021.)
- Invalsi. 2013. *Quadro di Riferimento della prova di italiano*. https://invalsi-areaprove.cineca.it/docs/file/QdR_ITALIANO.pdf (Consultato il 03.08.2021.)
- Minardi, Silvia. 2020. L'educazione alla cittadinanza democratica ha bisogno di più lingue, in Ambel (2020), vol.1, 172–178.
- MIUR. 2012. *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione*. Roma: Annali della Pubblica Istruzione.
- Paone, Fiorella. 2018. Sulla strada delle 'Dieci tesi' in prospettiva multimodale. *insegnare*, 19.10.2018. <http://www.insegnareonline.com/rivista/cultura-ricerca-didattica/strada-tesi-prospettiva-multimodale> (Consultato il 03.08.2021.)
- Paone, Fiorella. 2019, Diritti linguistici a scuola: itinerari e prospettive del plurilinguismo. In Bruni (2019), 102–117.
- Pozzo, Graziella. 1982a. Introduzione. In Pozzo (1982c), 8–14.
- Pozzo, Graziella. 1982b. Sulle tracce del testo. In Pozzo (1982c), 75–88.
- Pozzo, Graziella (a cura di). 1982c. *Insegnare la lingua. La comprensione del testo. Atti del Convegno Insegnare la lingua, CIDI, GISCEL e LEND, Ivrea 5-7, 1982*. Milano: B. Mondadori.

Text comprehension: the role of individual differences

Lucia Mason

Università degli Studi di Padova

lucia.mason@unipd.it

Angelica Ronconi

Università degli Studi di Padova

angelica.ronconi@studenti.unipd.it

Abstract

This contribution of this paper is to present a narrative review of current issues of psychological research about text comprehension. The process of meaning construction is described according to a well-known model which indicates that comprehension occurs at different levels and can be evaluated with reference to them. The importance of inferential activity for establishing coherence during the comprehension process is highlighted. Readers' self-monitoring of the ongoing process of comprehension and judgments of their own comprehension are also introduced as fundamental. As for recent advances in technology, reading is not only from paper but also from screen, text comprehension is examined in two main research areas of digital reading: one regards the selection, evaluation, and comprehension of information in multiple, often controversial documents on the Internet; the other concerns the analysis of text comprehension in relation to the reading medium, paper or the screen of a digital device. The role of readers' individual differences in text comprehension is then reviewed. The contribution of various linguistic, cognitive, motivational, and emotional factors is highlighted to explain how the complex activity of meaning construction from text can be sustained or hindered by readers' characteristics.

1. Text comprehension: the process

Successful text reading is at the very basis of learning in school from the last years of primary school onwards. Students at all educational levels are asked to read a variety of texts to acquire disciplinary knowledge. Reading comprehension is therefore a primary learning activity and it is not surprising that it correlates with academic achievement in general (cfr. Savolainen et al. 2008) and in specific domains. For example, in the domain of science learning, reading comprehension skills showed to be highly associated with proficiency as measured by the Programme of International Student Assessment (PISA) (cfr. Cromley 2009).

Text comprehension is a complex cognitive activity of meaning construction that starting from the processing of letters, words, and sentences, then relies on the selection

of relevant information and the construction of semantic representations. This occurs in a process in which textual information is connected with background knowledge and used appropriately. Various models of text comprehension have been proposed in psychological research and despite some divergences, all they share the idea that meaning construction from text implies an active role of the reader who must be involved in both bottom-up (e.g., decoding) and top-down processes (use of background knowledge) that operate flexibly and iteratively (cfr. Sinatra & Broughton 2011). The final outcome of text comprehension is a mental representation that includes both information explicitly provided in the text and information derived from prior knowledge.

To exemplify this cognitively complex activity, we describe the most well-known model of text comprehension, Kintsch's (1998) Construction-Integration (CI) model. It posits that two main stages are involved in discourse comprehension. During the first (*construction*), readers construct a network of propositions based on information provided in the text. As they try to construct a coherent and fully connected network, in many cases they need to make associations and inferences (see 1.1.) based on their relevant background knowledge. What they already know about the text content is retrieved from long-term memory and activate in working memory¹. Consider, for instance, the short text: *The doorbell rang. Toni jumped up spilling his coffee on the table. It was early morning.* The text explicitly contains four propositions; however, the propositions are not sufficient to construct a coherent representation of the text. Prior knowledge should be used to comprehend why Toni Jumped up (because the doorbell rang), what he was doing (having his morning coffee), where he was (in his kitchen) and who can be at the door (a visitor?). That is, readers should comprehend that Toni is surprised because normally we do not expect people to visit in the morning. Thus, it must be a person who delivers mail or some other professional delivery person. During the second stage (*integration*), a continuous spread of activation within the propositional network occurs and only the concepts that are connected to many others are maintained in it. As a consequence, for example, unimportant propositions are omitted and contradicting pieces of information are connected.

According to the CI model, multiple levels of memory representations are formed during the comprehension process. The first is the verbal or linguistic level which includes the exact words and propositions of the text. The second level of representation during the comprehension process is the *textbase* which involves extracting semantic textual information at both local (microstructure) and global (macrostructure) levels, and encoding it into a coherent network of propositions. The third and deeper level of representation is the *situation model* which is built when information provided by the text are integrated with the reader's prior knowledge. Learning from text, therefore, is more than forming a good textbase. It implies a situation model as a mental representation of

¹ Working memory is a complex multi-componential structure that includes an attentional control system, termed the central executive, and two short-term storage systems. The system for the storage of verbal-acoustic material is the phonological loop, while the system for the storage of visual material is the visuo-spatial sketchpad. The complex structure is also supplemented by the episodic buffer that holds multi-dimensional episodes or chunks of visual and auditory information. This buffer provides a temporary store in which various types of information pertaining to the two short-term storage systems interfaces with information from long-term memory. The working memory system has a brief limited capacity, thus only a limited number of information chunks can be processed and held temporarily. In contrast, long-term memory is a long-lasting with unlimited capacity, which represents the final stage of information processing where encoded and processed information enters and from which it can then be retrieved (cfr. Baddeley 1990).

the text topic linked to the reader's knowledge base. It is such integration that allows to use and apply the information learned from a text to another context, that is, to transfer the newly acquired information in order to provide an answer and solve a problem in a different situation. Without integration between the new and old information, no meaningful learning occurs (cfr. Butterfuss & Kendeou 2018).

1.1. Inferential activity

A text does not give all of the information required for its comprehension, thus the reader has to fill in the information gaps by inferential activity. Inferences are information that are not explicitly presented in the text but are generated during the comprehension process. If, in some cases, inferential processes are rather easy, in others that may require demanding reasoning processes (cfr. Lumbelli, 2009). Cognitive research has indicated that readers' low comprehension is related to low inferential skills (cfr. Cain & Oakhill, 1999).

Two types of inference have been mainly examined: *coherence inferences* and *elaborative inferences*. Coherence inferences are necessary to establish connections between propositions. They are local when they connect separate propositions usually signaled by a synonym, pronoun, or category exemplar, for instance, "He finished the orange juice quickly. The drink was very refreshing" (cfr. Currie & Cain 2015: 58). They are global coherence inferences when they connect to the text topic propositions that are distributed throughout the text; for example, readers can infer the seaside setting of a story via links to semantically related concepts such as "building sandcastles", "paddling in the water", and the presence of a pier (Currie & Cain 2015: 58). Elaborative inferences enrich the mental representation of the text and are often based on the integration of textual information with prior knowledge. They are not indispensable for text comprehension but crucial when reading complex expository texts that presented concepts to be learned.

Research with readers at different ages has shown that vocabulary knowledge sustain inferential skills both by providing knowledge of word meanings that are needed to generate inferences, and by contributing to memory processes. A rich semantic network with strong links between the meanings of words associated with the text topic is essential for easy and correct inference generation (cfr. Currie & Cain 2015). Moreover, given the multicultural backgrounds of students in today's classrooms, it is important to take into account that inferencing skills are also supported by sharing the same cultural background a text assumes. Struggling readers must be supported in developing inferences skills (cfr. Lumbelli 2009). Teachers' modeling of inferencing is recommended with this respect. They can read aloud to their students and "think aloud" their thoughts while processing a text. In this way teachers make the thinking process explicit while drawing an inference and asking themselves questions to show how they monitor their own comprehension (cfr. Kispal 2008).

As text comprehension occurs at different levels, it is also measured at various levels. Representation of the meaning at the textbase level is captured using a recall task where readers can freely express what they have read, or using questions whose answers require information presented in the text. Answers to questions that involve local and global coherence inferences are useful to measure comprehension at both textbase and situational model levels, depending on the complexity of inferencing. The deepest level of text comprehension is learning from text which occurs when a concept explained in a

text, for example, is applied to a new situation or context (cfr. Mason et al. 2015). Problem-solving and transfer questions are appropriate to reveal meaning construction at this more demanding level.

1.2. Metacomprehension

The process of successful text comprehension also requires that readers be active on metacognitive level. This is the level of cognition about cognition, that is a second-order cognition, in which students actively monitor and regulate their processes, in this case reading comprehension (cfr. Mason 2019). Readers are simultaneously engaged in constructing meaning from text and self-monitoring the ongoing process of comprehension. Readers' judgments of their own comprehension during or at the end of the comprehension process are very important to self-regulate. If they are aware, in fact, that their comprehension is not enough, they should not give up reading prematurely, but regulate their reading behavior accordingly.

The correspondence between students' judgments of their comprehension and actual comprehension performance as measured by a given task, is termed "calibration" (cfr. Thiede et al. 2002). To exemplify, a case of good calibration is when a student evaluates that she has comprehended a text passage well, and her actual performance in a comprehension task is good. In contrast, a case of poor calibration is when another student estimates that he has studied sufficiently to do well on a history test, but the actual score he receives is low. Both over-confidence and under-confidence are signs of poor calibration (cfr. Alexander 2013). As students' monitoring accuracy is typically poor, instructions should be provided to sustain effective monitoring and use of regulatory strategies to facilitate comprehension performance.

Research has documented that poor calibration may be due to students' reliance on superficial information ("cues") during monitoring, instead of cues that are truly diagnostic of their understanding, for example whether they can explain causal relations introduced in a text. Instructional interventions to support students' monitoring and self-regulation may consist on directly telling students what to do to promote this critical metacognitive activity. In this regard, they may be asked to think how well they would be able to explain the text content to a peer (cfr. Stoeger et al. 2014). Less direct interventions help readers generate valid diagnostic cues for monitoring, for example when they are asked to transform a text into visual representations using diagrams, concept maps, or drawings (cfr. van de Pol et al. 2020).

2. Digital reading

With the spread of digital technology, reading does not take place only on paper but also, and more and more frequently, on the screen of a digital device, such as computer, ipad, e-reader, and smartphone. On-screen activities are dominating the lives of our children at an increasingly early age and home literacy for preschoolers now includes interactions with electronic storybooks. Learning to read in primary school may also be supported by digital-based activities. When considering digital reading, two main research areas are identifiable in the educational psychology literature, which are briefly described in the next sections.

2.1. Reading on the Internet: source reliability matters!

From the last years of primary school onward, when children have learned to read and start reading to learn, Internet becomes “the” source of information to know more on a topic. Although digital inequalities still exist even amongst young people – as distance education is revealing during the health emergency caused by Covid-19 pandemic – in the digital era it is not necessary to possess books or to visit libraries to know more on unfamiliar topics. Immersed in technology, students are almost always connected through a smartphone, tablet, or laptop. They are the ‘Google generation’ who search the Web for information for a wide variety of purposes. Even though this generation of students has been raised in the Internet era, their information literacy has not improved as a result of the wide access to technology. Even older and ‘Internet savvy’ students have difficulties identifying fake news when it came to civic online reasoning (cfr. Wineburg & McGrew, 2016). It is, therefore, vital that seekers of online information know how to distinguish between reliable and unreliable sources.

The huge array of texts that they can access by a simple click pose new challenges to them. Even if they are able to locate and select relevant webpages – which may not always be the case – students must then be able to evaluate their credibility or authoritativeness according to appropriate criteria. The term “sourcing” is used to refer to readers’ use of information about the author, genre, and date of publication (cfr. Wennås Brante & Strømsø, 2017). The importance of source evaluation skills does not emerge solely in relation to the Internet, as it also regards information on paper, so it is not a novel issue. However, the Internet is unparalleled in this respect because for any given topic students are able to access multiple documents which, especially for controversial issues, offer conflicting information.

Digital literacy necessarily includes sourcing skills. What do we have to believe in? Why? How can this information be reconciled with the others and to what extent? These are crucial questions to deal with when reading multiple online pages on the same content, not only to be able to discern the sources that deserve to be followed up, but also to comprehend source content better (cfr. Salmerón et al. 2018). Sourcing is associated with deep multiple-document comprehension, which is necessary to construct meaning from a set of texts on the same issue or question.

With respect to Kintsch’s (1998) construction-integration model that describes superficial and deep levels of single-text comprehension (surface code, textbase, and situation model), Perfetti et al.’s (1999) model includes the additional layer of *intertext model*. This regards comprehension of the relationships between sources and text contents (source-content link), as well as the relationships among sources (source-source links) through rhetorical predicates (e.g., confirms, contradicts, etc.). During the comprehension process, readers achieve overall coherence through the intertext model despite inconsistent or conflicting information. However, according to the document model, readers may not want to integrate uncertain or inconsistent information, especially if the source is considered biased and unreliable. Nevertheless, they may read and consider the information, marking it as closely related to a particular source (cfr. Britt & Rouet, 2012).

Of note is that students who are more able to process a webpage for both its content and “metadata” regarding the authoritativeness of the source, are also those who are more able to comprehend conflicting information (cfr. Bråten et al. 2014). It is therefore

important to teach how to evaluate Web sources and the contents provided as there is evidence that at all educational levels students may not take into account source authority. If it is not surprising that a child of primary school believes that everything read on the Internet is true, it is worrying that high school (or even university) students consider more the form and appearance of a site than its expertise and credibility, or use naïve criteria, such as popularity, to evaluate information sources (cfr. Mason et al. 2011).

2.2. Reading on paper or screen?

As students are increasingly using digital devices to read, research has investigated whether reading comprehension on a screen is the same as reading comprehension on paper. In recent years, some meta-analyses have been conducted to determine the influence of the reading medium on reading comprehension. In what follows we refer to the most recent ones as a synthesis of research data collected in the two last decades involving readers at all educational levels, but with a great prevalence of university students.

Kong et al. (2018) analysed 17 studies conducted between 2000 and early 2016, focusing on the comparison between reading on paper and on screen for reading performance and speed. Results of this meta-analysis showed that reading comprehension was better when reading on paper, while no significant difference in reading speed was observed. The researchers suggested that the advantage of reading from paper was linked to readers' greater experience with printed texts, which might have led to the adoption of more effective reading strategies. Furthermore, the use of digital devices for reading might have determined higher cognitive load while performing the task, with negative effects on reading comprehension. However, the increased cognitive load did not influence the time used to read the texts, which did not differ across mediums. Nevertheless, a better performance while reading on paper for a similar investment of time may suggest a greater effectiveness of printed texts. Comparing differences for reading comprehension in studies conducted before and after 2013, Kong et al. (2018) found a decreasing trajectory of paper superiority over time which, although not significant, was potentially connected to technology improvement and greater familiarity with digital media.

In their meta-analytic study Delgado et al. (2018) focused on 54 studies conducted between 2000 and 2017, which were analysed separately according to the adopted experimental design, *between* or *within subjects*². The results of both meta-analyses revealed lower reading comprehension on screen than on paper, especially when the reading task was performed under limited time frames. This outcome supported the *screen inferiority effect* on text comprehension. It may be related to a decrease of the cognitive resources available for the construction of a coherent a complete situation model (cfr. Kintsch 1988) when complex written material is read on a digital device. Furthermore, Delgado and colleagues (2018) highlighted that the advantage of paper reading increased rather than decreased over the years, thus contradicting Kong et al.'s (2018) results. Overall, these data supported the hypothesis that the daily use of digital

² In a between-subjects design study each reader participant is assigned to a condition of reading from paper or a condition of reading from screen, thus the reader interacts with only one reading medium. In contrast, in a within-subjects design study each reader participant reads one or more texts both from paper and screen, thus the reader interacts with both reading mediums.

devices, characterized by frequent multitasking and rapid interactions (cfr. Kirschner & De Bruyckere 2017), may promote a more superficial and incomplete processing of information (*shallowing hypothesis*; cfr. Annisette & Lafreniere 2017). Such processing negatively affects reading comprehension as this cognitive activity requires focused attention for an extended period of time.

Similar results of screen inferiority effects were also found by Clinton (2019) in a meta-analysis that focused not only on differences in reading comprehension but also in reading time and calibration. The author included 33 studies conducted between 2008 and 2019 and considered age group (child or adult) as a moderating variable. However, it resulted not to be significant, suggesting that screen negative effects on reading performance affected both children and adults. Furthermore, Clinton (2019) observed a great variability in reading times, as some studies reported longer reading times in the paper condition, while others reported longer times in the screen condition. The lack of significant differences in reading time related to the medium may suggest greater effectiveness of printed text when the time used to read is similar. The calibration results were more homogeneous, indicating that the use of digital devices for reading generates a greater degree of inaccuracy in self-evaluating one's reading performance. Excessive overconfidence, for example, may lead the reader to end the activity before an adequate mental representation of the text content is constructed, negatively affecting the overall comprehension of the text. Besides the shallow hypothesis, the mind wandering hypothesis can also explain poorer comprehension and calibration from reading on screen. As focusing on text content on a screen is reported as more difficult by readers, distraction and task-unrelated thoughts may be more associated with reading digital than printed texts.

3. Individual differences in text comprehension

As an essential cognitive activity in many contexts besides the educational one, text comprehension has been widely investigated by psychological research in terms of individual factors – linguistic, cognitive, motivational, and emotional – that may sustain or hinder meaning construction from written materials.

3.1. Linguistic factors

Decoding, or reading of written words accurately and fluently, is at the very basis of text comprehension as it allows the processing of the linguistic information at the word-level. Faster and more efficient decoders do not spend cognitive resource in applying the rules to convert graphic stimuli into their linguistic referents. Thus, they can devote greater resources to the processes involved in meaning construction from text (cfr. Florit & Cain 2011). Word reading skill is particularly important in young readers; its relative contribution to reading comprehension then decreases during the primary school years as word reading fluency increases. Of note is that children take longer to become accurate and fluent in word reading when learning to read in a more opaque (deep) orthography (e.g., English and French: highly irregular relationship between phoneme and grapheme)

compared to a relatively more transparent (shallow) one (e.g., Italian and Finnish: highly regular relationship between phoneme and grapheme).

Readers' *vocabulary* knowledge is another essential language-related factor to be considered as involved in text comprehension. Vocabulary breadth is the number of words known, whereas vocabulary depth regards what an individual knows about those words and their semantic associations inferences. It's quite evident that being able to define words and understand the interrelations between words is important (cfr. Oakhill & Cain 2012). Research has documented that vocabulary knowledge is related to inference making, in particular global coherence inferences. When readers have a rich and precise semantic representation of words, they are facilitated in making thematically-related inferences to establish coherence during the meaning construction process. As such inferences are not triggered by "cohesive ties" like anaphors, they depend on readers' activation processes (cfr. Cain & Oakhill 2014).

3.2. Cognitive factors

Executive functions are among the cognitive factors that contribute to text comprehension. They are fundamental mechanisms underlying complex cognitive tasks that require planning, impulse control, flexible use of strategies, and goal-directed behavior. Three are the main executive functions: *updating*, *shifting*, and *inhibition* (cfr. Miyake et al. 2000). Updating is the skill to update task-relevant information by replacing the old information in working memory and maintaining it. Shifting is the skill to switch between task goals or sets of tasks. Inhibition is the skill to suppress dominant and automatic but inappropriate responses and produce task appropriate responses. The contribution of executive functions to text comprehension has been explained in relation to the aforementioned Kintsch's (1998) model, as they are essential in the two main phases of construction and integration (cfr. Butterfuss & Kendeou 2018). Specifically, updating allows readers to select and maintain the activation of relevant information in working memory during reading (cfr. Pelegrina et al. 2015; Zaccoletti et al. 2020b); shifting is involved in paying attention to phonological and semantic text features and switching between reading strategies in a flexible way (cfr. Guajardo & Cartwright 2016); inhibition is required for the exclusion of irrelevant information and prevention of its intrusion from memory (cfr. Borella et al. 2010).

Prior knowledge is also a cognitive factor that largely contributes to text comprehension. As described above, a text does not give all the information required for its comprehension, thus readers have to fill in the information gaps by inferential activity, that is to use their knowledge of the text topic and structure (cfr. Tarchi 2015). The formation of a situation model in deep text comprehension requires the reader to have adequate knowledge of a topic, especially when a text has low coherence. Coherence reflects the degree to which the sentences and parts of a text are connected, for example by coreference and causal links, and facilitate an accurate text representation. If a text is characterized by low coherence, readers' prior knowledge becomes even more important to bridge text gaps by inferential activity. Research has documented that readers with high prior knowledge showed greater deep text comprehension after reading a low-coherence text. This occurs because a less coherent text stimulates active processing in a reader who is equipped with high-level of knowledge. In contrast, readers with little prior knowledge about the text topic comprehended better a globally coherent version of the same text (cfr. Boscolo & Mason 2003; McNamara & Kinsch 1996). The explanation for this outcome

is that a well coherent text is an easy one, thus it is not demanding as it reduces the need for active processing.

3.3. Motivational factors

Interest is a motivational variable that is considered in relation to the useful distinction between individual and situational interest (cfr. Hidi 2001). Individual interest is a relatively stable evaluative orientation toward certain objects, ideas, or events, which reflects in a relatively enduring predisposition to re-engage with them over time. Situational interest is generated by certain conditions and/or environmental stimuli, such as novelty and intensity. For example, particular textual characteristics, like readability and suspense, can trigger situational interest. Topic interest can be considered both individual and situational as it is at the same time an expression of individual interest for a subject matter (e.g., history), but also an expression of situational interest related to a specific content (e.g., the Roman Empire). Most topics, therefore, have both individual and situational sources (cfr. Renninger 2000). Research has indicated that topic interest is positively associated with text comprehension as it sustains deeper cognitive information processing and affective investment during the reading process (cfr. Linnenbrink-Garcia et al. 2012; Mason et al. 2008). Topic interest can also compensate for low prior knowledge of the content to be learned (cfr. Boscolo & Mason 2003).

Another motivational factor that sustains text comprehension is *self-efficacy*. It concerns a person's beliefs about his or her capabilities to learn or perform at a designated level, having control on the outcome (cfr. Bandura, 1986). Self-efficacy beliefs are nurtured by four sources: personal attainment, vicarious or observational experience, social persuasion, and physiological reaction. Self-efficacy clearly increases through success. Students' self-efficacy is also formed by comparing their performance with those of others and by observing similar peers. When peers perform a task well, other students may believe that they are also able of executing it well. Their self-efficacy is also promoted by encouraging teachers, parents, and peers ("you can do it"). Moreover, sweating and heart rate, for example, are physiological indicators that may sign lack of competence, whereas experiencing decreased anxiety may increase self-efficacy (cfr. Schunk 2003). The literature shows rich evidence that self-efficacy is positively correlated with performance in a domain or task, including reading (e.g., Guthrie et al. 2013). This positive association is explained in terms of increased effort and persistence in difficult tasks, and resilience in disadvantageous situations (cfr. Schunk & Pajares 2005).

A further motivational factor that can explain motivation to and performance in reading comprehension regards *achievement goals*. They refer to the reasons why students' engage in a task or activity (cfr. Ames & Archer 1988). Three types of achievement goal are commonly examined in research about school learning: mastery, performance-approach, and performance-avoidance. Mastery-oriented students are focused on learning and understanding, and acquiring new skills. In contrast, performance-oriented students are focused on getting a good grade, reward, or performing better than others, whereas performance-avoidance oriented students are focused on avoiding social judgments that they are less able than their peers (cfr. Elliot & Covington 2001). Research has provided clear evidence that mastery and, relatively more recently, also performance-approach goals are positive predictors of students' deeper processing strategies, persistence and effort expenditure. In contrast, there is also clear evidence that

performance-avoidance goals are always associated with negative outcomes in terms of low engagement and achievement when reading comprehension performance is examined (cfr. Liem et al. 2008).

3.4. Emotional factors

Research on the role of affect in learning processes and outcomes is relatively recent as for long time emotions have been considered as a disturbing factor, given that only rationality was thought to be involved in academic performance and achievement (cfr. Pekrun 2006). However, test anxiety has a long tradition of research in academic contexts (cfr. Hembree 1988; Zeidner 2007). Emotions can be described according to two orthogonal dimensions (cfr. Feldman Barrett & Russell 1998): valence in terms of pleasantness (e.g., joy) and unpleasantness (e.g., shame), and activation as physiologically activating (e.g., anger) or deactivating (e.g., boredom).

Emotions in school are termed *achievement emotions* as they are related to activities and outcomes that are judged against competence-based standards of quality (cfr. Pekrun et al. 2006). Enjoyment or boredom, for example, are experienced during learning and teaching, hope and pride in relation to success, whereas anxiety, anger, and shame are associated with failure. Achievement emotions may be linked to a specific subject matter (e.g., mathematics) but may also depend on a specific task, like reading.

According to the control-value theory of achievement emotions (cfr. Pekrun 2006), these have two motivational antecedents: the control students perceive in a learning situation and the importance they assigned to an activity and to performing well in that activity. When considering reading comprehension, this means that the emotions generally associated with this crucial learning activity depend on both students' control appraisal and value appraisal. If students believe that they will be able to successfully comprehend a material to read, and that to comprehend well is very important to them, their reading-related emotions are positive (e.g., hope and joy). In contrast, if they believe that they are not able to achieve a good comprehension result and/or reading comprehension is not valued with respect to what they think it is important to do well, negative emotions are associated with their reading activity. Research has indicated that learning-related emotions contribute to achievement: positive activating emotions, such as enjoyment, are positively related to achievement, while negative deactivating emotions, such as boredom, are usually negatively related to achievement (cfr. Pekrun & Perry, 2014). Why?

According to the resource allocation model (cfr. Ellis & Ashbrook 1998), intense affective states (e.g., anxiety) drain attentional resources that are necessary to execute a cognitive task, since they generate extra-task processing (thinking about own emotional state, mind wandering) and irrelevant task processing. In reading comprehension tasks, emotions may reduce the cognitive resources for the information processing required in the construction and integration phases of Kintsch's (1998) model. As a consequence, emotions may hinder an accurate text representation. However, there is also evidence that positive affective states may sustain students' willingness to engage and comprehend a text deeply. In contrast, negative affective states may hinder an accurate text representation (Scrimin & Mason 2015).

Emotions have been recently investigated theoretically (cfr. Bohn-Gettler 2019) and empirically (cfr. Mason et al. 2017) in relation to reading activity. For example, enjoyment of reading, as feeling of pleasure and happiness, has been the focus of a study with a very large sample of 15-year-old students in 65 countries involved in PISA 2009 (cfr. Cheema 2018). Enjoyment revealed to be a positive predictor of reading achievement at country level, meaning that the positive emotion sustains reading comprehension. However, differences also emerged at cross-country level, after controlling for gender and socio-economic status. The positive association of enjoyment of reading with reading achievement was identified in countries characterized by high academic performance, whereas a negative association of the two variables appeared in low-academic performance countries. One explanation for the last puzzling outcome is that students who enjoy reading not necessarily are very proficient readers. It is also possible that the negative association in those countries is driven by a variable that had not been taken into account. Moreover, another speculative interpretation is that the countries in which the negative association occurs between reading enjoyment and reading achievement are developing nations, and because of their poor economic resources, the poor quantity and quality of the reading materials in schools may be at the basis of low enjoyment of reading (cfr. Cheema 2018).

On a smaller scale, Zaccoletti et al. (2020) investigated the contribution of three more frequently experienced emotions at school – enjoyment, anxiety, and boredom – to reading comprehension in fifth graders. The findings showed that the motivational antecedent of control in reading negatively predicted anxiety which, in turn, negatively predicted students' text comprehension. These outcomes mean that the less the students' perceive control of reading activity, the greater their anxiety which partly mediates the link between the motivational factor (control) and the cognitive performance (text comprehension), after controlling for a number of potentially interfering variables, such as gender, vocabulary knowledge, reading speed, and non-verbal skills.

Reading-related emotions have also been investigated in relation to the three main executive functions described above. Updating of working memory emerged as positively related to text comprehension in eighth graders, as well as activating negative emotions such as anxiety, anger, and shame, and deactivating negative emotions such as boredom and hopelessness. In addition, an interaction between the activating negative emotions and updating emerged: when anxiety, anger, and shame combine with low or moderate levels of updating ability, students' text comprehension gets worse (cfr. Zaccoletti et al. 2020). These findings reveal the negative contribution of students' negative emotions to their reading comprehension performance, especially the activating emotions in interaction with low ability to continuously refresh information in working memory during reading.

A particular type of emotion is termed *epistemic emotions* as they are elicited by the characteristics and the processing of the information to be learned (cfr. Pekrun et al. 2016). Curiosity, frustration or confusion may be experienced while reading about a disciplinary topic. For example, confusion may be triggered when readers perceive a conflict between their knowledge and the new information, but this conflict is not immediately regulated and resolved. They are therefore in a state of cognitive disequilibrium that implies a consequential disruption in the information processing. To overcome an impasse related to confusion, readers need to deepen their engagement in the activity. With this respect, it has been demonstrated that confusion is not only a negative emotion when comprehending complex texts, but also beneficial as it may drive

conceptual learning (cfr. D’Mello et al. 2014). For this to occur, however, confusion should be experienced in an optimal “dose”, neither too little nor too much, and appropriately regulated.

The importance of emotional factors during reading has also been revealed by research on the comprehension of multiple conflicting texts on the Internet. Emotions were measured not only using self-report questionnaires but also at physiological level. Empirical data revealed that high physiological arousal (activation) acted as an impairment for the complex cognitive task of constructing a coherent and integrated representation of a set of four texts with some emotional content. In contrast, physiological emotional regulation positively contributed to multiple-text comprehension. In other words, the less the students were emotionally reactive to the content read from real websites – the controversial issue of the health risks associated with the continuous use of mobile phone – and the more they were able to self-regulate while reading the four texts, the greater their multiple-text comprehension as reflected in an argumentative short essay (cfr. Mason et al. 2017, 2018).

Conclusion

Reading comprehension is a fundamental activity not only for students’ success at different educational levels, but also for any citizens’ active participation in our information society. Either on paper or a screen of a digital device, and for enjoying a novel, studying a textbook, scrutinizing the news, or searching for more information on an unfamiliar topic, reading must lead to comprehending what has been read. The comprehension process may be very demanding, especially when texts require readers to be cognitively active in inferential activity based on relevant and accurate prior knowledge. Successful comprehension also requires readers to be metacognitively involved in monitoring their ongoing meaning construction. Such complex process is sustained by individual differences of various nature, from linguistic to motivational, from cognitive to emotional which can all be promoted by appropriate educational interventions.

References

- Alexander, Patricia A. 2013. Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction* 24. 1–3.
- Ames, Carole & Archer, Jennifer. 1988. Achievement goals in the classroom: Students’ learning strategies and motivation processes. *Journal of Educational Psychology* 80(3). 260–267.
- Annisette, Logan E. & Lafreniere, Kathryn D. 2017. Social media, texting, and personality: A test of the shallowing hypothesis. *Personality and Individual Differences* 115. 154–158.
- Baddeley, Alan D. 1990. *Human memory [La memoria umana]*. Hillsdale, NJ: Erlbaum [trad. it. Bologna: Il Mulino, 1995].
- Bandura, Albert. 1986. *Social foundation of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

- Bohn-Gettler, Catherine M. 2019. Getting a grip: The PET framework for studying how reader emotions influence comprehension. *Discourse Processes* 56. 386–401.
- Borella, Erika & Carretti, Barbara & Pelegrina, Santiago L. 2010. The specific role of inhibitory efficacy in good and poor comprehenders. *Journal of Learning Disabilities* 43. 541–552
- Boscolo, Pietro & Mason, Lucia. 2003. Prior knowledge, text coherence, and interest: How they interact in learning from instructional texts. *Journal of Experimental Education* 71(2). 126–148.
- Bråten, Ivar & Ferguson, Leila E. & Strømsø, Helge I. & Anmarkrud, Øistein. 2014. Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology* 84. 58–85.
- Britt, M. Anne & Rouet, Jean-Francois. 2012. Learning with multiple documents: Component skills and their acquisition. In John R. Kirby & Michael J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes*, 276–314. New York: Cambridge University Press.
- Butterfuss, Reese & Kendeou, Panayiota. 2018. The role of executive functions in reading comprehension. *Educational Psychology Review* 30. 801–826.
- Cain, Kate & Oakhill, Jane. 1999. Inference making ability and its relation to comprehension failure in young children. *Reading and Writing* 11. 489–503.
- Cheema, Jehanzeb R. 2008. Adolescents' enjoyment of reading as a predictor of reading achievement: new evidence from a cross-country survey. *Journal of Research in Reading* 41S1. S149–S162.
- Clinton, Virginia. 2019. Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading* 42(2). 288–325.
- Cromley, Jennifer G. 2009. Reading achievement and science proficiency: International comparisons from the Programme on International Student Assessment. *Reading Psychology* 30(2). 89–118.
- Currie, Nicola K. & Cain, Kate. 2015. Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology* 137. 57–75.
- Delgado, Pablo & Vargas, Cristina & Ackerman, Rakefet & Salmerón, Ladislao. 2018. Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review* 25. 23–38.
- D'Mello, Sidney & Lehman, Blair, & Pekrun, Reinhard, & Graesser, Art. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29. 153–170.
- Elliot, Andrew J & Convington, Martin V. 2001. Approach and avoidance motivation. *Motivation and Emotion* 30. 111–116.
- Ellis, Henry C. & Ashbrook, P W. 1988. Resource allocation model of the effects of depressed mood states on memory. In Klaus Fiedler & Joseph D. Forgas (Eds.), *Affect, cognition, and social behavior: new evidence and integrative attempts*, 25–43. Toronto: Hogrefe.
- Feldman Barrett, Lisa & Russell, James A. 1998. Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology* 74. 967–984.
- Florit, Elena & Cain, Kate. 2011. The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review* 23(4). 553–576.
- Guajardo, Nicole R. & Cartwright, Kelly B. 2016. The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers' language

- comprehension and later reading awareness and comprehension in elementary school. *Journal of Experimental Child Psychology* 144. 27–45.
- Guthrie, John T. & Klauda, Susan L. & Ho, Amy N. 2013. Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly* 48. 9–26.
- Hembree, Ray. 1988. Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research* 58(1). 47–77.
- Hidi, Suzanne. 2001. Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review* 13. 191–209.
- Kintsch, Walter. 1998. *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kirschner, Paul A. & De Bruyckere, Pedro. 2017. The myths of the digital native and the multitasker. *Teaching and Teacher Education* 67. 135–142.
- Kispaal, Anne. 2008. *Effective teaching of inference skills for reading*. National Foundation for Educational Research: Research Report DCSF-RR031.
- Kong, Yiren & Seo, Young Sik & Zhai, Ling. 2018. Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education* 123. 138–149.
- Liem, Arief D. & Lau, Shun & Nie, Youyian. 2008. The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology* 33. 486–512.
- Linnenbrink-Garcia, Lis & Middleton, Michael J. & Ciani, Keith D. & Easter, Matthew A. & O’Keefe, Paul A. & Zusho, Akane. 2012. The strength of the relation between performance-approach and performance-avoidance goal orientations: Theoretical, methodological, and instructional implications. *Educational Psychologist* 47. 281–301.
- Lumbelli, Lucia (2009). *La comprensione come problema* [Comprehension as a problem]. Rome: Laterza.
- Mason, Lucia 2019. *Psicologia dell’apprendimento e dell’istruzione* [Psychology of learning and instruction]. 3 ed. Bologna: Il Mulino.
- Mason, Lucia & Ariasi, Nicola & Boldrin, Angela. 2011. Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction* 21(1). 137–151.
- Mason, Lucia & Pluchino, Patrik & Tornatora, Maria Caterina. 2015. Eye-movement modeling of text and picture integration during reading: effects on processing and learning. *Contemporary Educational Psychology* 41. 172–187.
- Mason, Lucia & Scrimin, Sara & Tornatora, Maria Caterina & Zaccoletti, Sonia. 2017. Emotional reactivity and comprehension of multiple online texts. *Learning and Individual Differences* 58. 10–21.
- Mason, Lucia & Scrimin, Sara & Tornatora, Maria Caterina & Zaccoletti, Sonia & Goetz, Thomas. 2018. Webpage reading: Psychophysiological correlates of emotional arousal and regulation predict multiple-text comprehension. *Computers in Human Behavior* 87. 217–326.
- McNamara, Danielle S. & Kintsch, Walter. 1996. Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes* 22. 247–288.
- Miyake, Akira & Friedman, Naomi P. & Emerson, Michael J. & Witzki, Alexander H. & Howerter, Amy & Wager, Tor D. (2000). The unity and diversity of executive

- functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology* 41. 49–100.
- Oakhill, Jane & Cain, Kate. 2012. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading* 16(2). 91–121.
- Pekrun, Reinhard. 2006. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review* 18. 315–341.
- Pekrun, Reinhard & Perry, Raymond P. (2014). Control-value theory of achievement emotions. In Reinhard Pekrun & Lisa Linnenbrink-Garcia (Eds.), *International handbook of emotions in education*, 130–151. New York: Routledge.
- Pekrun, Reinhard & Vogl, Elisabeth & Muis, Krista R & Sinatra, Gale M. (2016). Measuring emotions during epistemic activities: the Epistemically Related Emotion Scales. *Cognition and Emotion* 31(6). 1268–1276.
- Pelegrina, Santiago & Capodiecì, Adele & Carretti, Barbara & Cornoldi, Cesare. 2015. Magnitude representation and working memory updating in children with arithmetic and reading comprehension disabilities. *Journal of Learning Disabilities* 48(6). 658–668.
- Perfetti, Charles A. & Rouet, Jean-Francois & Britt, M. Anne. 1999. Towards a theory of documents representation. In Hans van Oostendorp & Susan R. Goldman (Eds.), *The construction of mental representations during reading*, 99–122. Hillsdale, NJ: Erlbaum.
- Renninger, K. Anne. 2000. Individual interest and its implications for understanding intrinsic motivation. In Carol. Sansone & Judith M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance*, 373–404. San Diego, CA: Academic Press.
- Salmerón, Ladislao & Gil, Laura, & Bråten, Ivar. 2018b. Effects of reading real versus print-out versions of multiple documents on students’ sourcing and integrated understanding. *Contemporary Educational Psychology* 52. 25–35.
- Savolainen, Hanny & Ahonen, Timo & Aro, Mikko & Tolvanen, Asko & Holopainen, Leena. 2008. Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning and Instruction* 18(2). 201–210.
- Schunk, Dale H. 2003. Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly* 19. 159–172.
- Schunk, Dale A., & Pajares, Frank (2005). Competence perceptions. In Andrew J. Elliot & Carol S. Dweck (Eds.), *Handbook of competence and motivation*, 85–104. New York: Guilford Press.
- Scrimin, Sara & Mason, Lucia. 2015. Does mood influence text processing and comprehension? Evidence from an eye-movement study. *British Journal of Educational Psychology* 85(3). 387–406.
- Sinatra, Gale M. & Broughton, Suzanne H. 2011. Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly* 46. 374–393.
- Stoeger, Heidrun & Sontag, Christine & Ziegler, Albert. 2014. Impact of a teacher-led intervention on preference for self-regulated learning, finding main ideas in expository texts, and reading comprehension. *Journal of Educational Psychology* 106. 799–814.

- Tarchi, Christian. 2015. Fostering reading comprehension of expository text through the activation of readers' prior knowledge and inference-making skills. *International Journal of Educational Research* 72. 80–88.
- Thiede, Keith W. & Anderson, Marie C. M. & Therriault, David. 2003. Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology* 95(1), 66–73.
- van de Pol, Janneke & van Loon, Mariëtte & van Gog, Tamara & Braumann, Sophia & de Bruin, Anique. 2020. Mapping and drawing to improve students' and teachers' monitoring and regulation of students' learning from text: Current findings and future direction. *Educational Psychology Review*.
- Wennås Brante, Eva & Bråten, Helge I. 2017. Sourcing in text comprehension: a review of interventions targeting sourcing skills. *Educational Psychology Review* 30. 773–799.
- Wineburg, Sam & McGrew, Sarah. 2016. Why students can't Google their way to the truth. *Education Week* 36. 22–28.
- Zaccoletti, Sonia & Altoè, Gianmarco & Mason, Lucia. 2020a. Enjoyment, anxiety and boredom, and their control-value antecedents as predictors of reading comprehension. *Learning and Individual Differences* 79. e101869.
- Zaccoletti, Sonia & Altoè, Gianmarco & Mason, Lucia. 2020b. The interplay of reading-related emotions and updating in reading comprehension performance. *British Journal of Educational Psychology* 90(3), 663–682.
- Zeidner, Moshe. 2007. Test anxiety in educational contexts: Concepts, findings, and future directions. In Paul A. Schutz & Reinhard Pekrun (Eds.), *Emotions in education*, 165–184. Amsterdam: Elsevier-Academic Press.

Qualche osservazione sociolinguistica sulle prove INVALSI

Emanuele Miola

Università di Bologna

emanuele.miola@unibo.it

Abstract

Il contributo mette a tema le INVALSI, indagando, dalla specola di osservazione del sociolinguista, specialmente le prove di italiano proposte nel torno di anni scolastici che va dal 2016/17 al 2018/19. Si discute in prima battuta quale spazio sia dedicato alla sociolinguistica nelle Indicazioni nazionali delle INVALSI, ritrovando richiami alla dimensione della variazione della lingua, in particolare per quanto riguarda il livello di analisi del lessico. Si prendono quindi in esame le prove effettivamente somministrate a scolare e scolari delle scuole primarie e secondarie di primo grado: si dà qualche saggio della ‘lingua delle INVALSI’, si esaminano alcuni tratti notevoli di questa varietà e, infine, analizzando alcune domande di stampo sociolinguistico che si incontrano nei test, si osserva in che modo la variazione è presente e in che modo è verificata nelle prove.

1. Introduzione

Lo scopo che si prefigge questo contributo è guardare, almeno cursoriamente, ai testi dell’Istituto Nazionale per la Valutazione del Sistema dell’Istruzione (cioè delle INVALSI) attraverso la lente della sociolinguistica. Ci appelleremo alla sociolinguistica principalmente per provare a vedere che varietà ‘parlano’ le prove che vengono somministrate agli studenti nel passaggio da un ordine di scuola all’altro o da una classe all’altra. In buona sostanza, dunque, questa indagine potrebbe valere anche da banco di prova per valutare quale sia la varietà di lingua a cui i nostri studenti siano abituati come lingua di prestigio, insegnata dalle grammatiche e nelle aule di scuola, non sanzionata e anzi usata nei test di verifica: insomma, forse un po’ per approssimazione, la varietà standard della nostra lingua (Ammon 1986; Berruto 2007; Berruto 2012: 70–73).¹

È già così, perciò, evidente che la specola di osservazione della sociolinguistica può essere utile, nell’economia di una discussione sulle INVALSI, per almeno per due motivi.

¹ Primi sondaggi volti a saggiare il rapporto tra norma prescrittiva e norma interiorizzata per quanto riguarda l’italiano degli studenti e dei parlanti nativi, quindi volti anche a saggiare in certa misura la norma ‘trasmessa’ dalla scuola dell’obbligo, si possono leggere in Grandi (2018) e in Ondelli & Romanini (2018).

Tanto per cominciare, perché la scuola ha il compito, tra gli altri o forse *primus inter pares*, di fornire ai ragazzi gli strumenti che permettano loro di orientarsi nella vita della loro società. Per questo sarebbe necessaria una tensione, per dirla con Tullio De Mauro e con le sue *Dieci tesi*,² allo “sviluppo delle capacità linguistiche nel loro insieme [...] rispettoso ma non succubo” della varietà linguistica, ovvero delle diverse lingue e varietà di lingua, che possono essere oggi nelle orecchie degli studenti, oppure di fronte ai loro occhi. In secondo luogo, poi, perché, come accennato poco fa, un’indagine condotta con l’occhio del sociolinguista ci può aiutare a raffinare le nostre conoscenze su quale sia la varietà di lingua che mediamente gli italiani, o almeno quelli che frequentano le scuole, oggi vedono, leggono e, quindi, riconoscono come standard, ovvero come varietà verso cui tendere nella comunicazione di alta o media formalità.

Insomma, per fare scuola oggi non si deve soltanto avere ben presente quali lingue sono disponibili agli studenti e quindi quali ha senso trattare in classe o considerare quando si fa lezione;³ ma anche aver presente e insegnare, se necessario, a comprendere e a decrittare la varietà che adopera nelle circolari, nei testi scolastici e, infine, nelle prove di ammissione alle classi superiori. In definitiva, poi, l’insegnante deve avere ben chiaro quanto questa o queste varietà si discostino da quella usuale, cioè usata più di frequente nella vita quotidiana, per gli studenti, per i parlanti in genere.

2. Lo spazio della sociolinguistica nelle *Indicazioni nazionali*

Prima di verificare se e in che modo i testi delle INVALSI tengano conto della sociolinguistica e dei problemi ad essa connessi, e se e in che modo interrogino le competenze sociolinguistiche degli scolari, mi sembra utile richiamare quale spazio sia destinato alla sociolinguistica secondo le direttive curriculari per la scuola. Si prendano dunque a esempio le *Indicazioni nazionali per il curriculum della scuola dell’infanzia e del primo ciclo d’istruzione* (da adesso in poi, e in bibliografia, solo *Indicazioni*). È chiaro che per questi ordini scolastici le competenze in uscita non potranno essere estremamente raffinate, ma tuttavia le formulazioni del Ministero, quantomeno tra le righe, cioè pur senza nominarli direttamente, toccano aspetti che richiamano la diversità linguistica e i quattro canonici assi di variazione sociolinguistica.

Pare un chiaro riferimento alla *varietas linguarum* e alla dimensione diafasica quanto si legge a p. 36 delle *Indicazioni*:

Nel nostro paese l’apprendimento della lingua avviene oggi in uno spazio antropologico caratterizzato da una varietà di elementi: la persistenza, anche se quanto mai ineguale e diversificata, della dialettologia; la ricchezza e la varietà delle lingue minoritarie; la compresenza di più lingue di tutto il mondo; la presenza infine dell’italiano parlato e scritto con livelli assai diversi di padronanza e con marcate varianti regionali. Tutto questo comporta che nell’esperienza di molti bambini e ragazzi l’italiano rappresenti una seconda lingua.

Io sarei addirittura più categorico: se l’occorrenza di *italiano* nell’ultima frase della citazione si riferisce all’*italiano standard*, allora per ogni bambino o ragazzo – salvo rare eccezioni – questo rappresenta almeno in parte una seconda lingua (oppure una varietà

² Se ne veda ora l’edizione con commento a cura di Loiero & Lugarini (2019).

³ Il lettore interessato a una panoramica in questo merito potrà consultare Fiorentini et al. (2020).

non nativa della propria lingua). Nessuno infatti nasce parlante della varietà standard di una lingua: nel caso italiano, “in pratica tutti gli italofoeni [hanno] imparato l’italiano standard propriamente detto (la varietà scolastica a base ‘fiorentina’ emendata) [...] attraverso l’educazione scolastica, e non nella socializzazione primaria” (Berruto 2003: 4). Ciò che al giorno d’oggi i bambini e i ragazzi apprendono nell’interazione con i genitori e con i loro pari è un italiano, parlato, che ha piccole o grandi differenze rispetto allo standard scolastico, dal punto di vista della pronuncia e del lessico, sì, ma talvolta anche della morfosintassi. L’insegnante deve essere cosciente che nelle consegne degli esercizi e sui libri di testo in genere viene adoperata quella che per tutti i ragazzi potrebbe essere in buona sostanza assimilabile a una L2.

La variazione diafasica, nelle *Indicazioni*, è anche occhieggiata in formulazioni quali “l’esperienza dei diversi usi della lingua”; “un’ampia gamma di testi appartenenti ai vari tipi e forme testuali” che si consiglia di leggere; “le potenzialità espressive della lingua italiana” che gli studenti devono esperire.

L’accento sociolinguistico sembra posto tuttavia specialmente sul lessico, come si legge a p. 38:

Data la grande importanza della comprensione e dell’uso attivo del lessico, il primo compito dell’insegnante è proprio quello di rendersi conto, attraverso attività anche ludiche e creative, della consistenza e tipologia (varietà) del patrimonio lessicale di ognuno. [...] Infatti l’uso del lessico, a seconda delle discipline, dei destinatari, delle situazioni comunicative e dei mezzi utilizzati per l’espressione orale e quella scritta richiede lo sviluppo di conoscenze, capacità di selezione e adeguatezza ai contesti. [...] Va, in questo senso, tenuta in considerazione la ricchezza delle espressioni locali, «di strada», gergali e dei molti modi di dire legati alle esperienze, che spesso racchiudono un senso identitario e capacità narrative e che rappresentano un bagaglio attraverso il quale ampliare l’espressione anche in italiano corretto.

Ritorna qui, oltre a *varietà* (anche se usato in un senso che non corrisponde a quello del tecnicismo della sociolinguistica, che ricorre però subito dopo, a p. 39), l’idea che sia importante padroneggiare adeguatamente situazioni e contesti comunicativi, così come l’espressione orale e quella scritta, e non mi pare assente neppure un cenno alla dimensione diastratica, nelle ultime righe del passo citato sopra. Tutto questo bagaglio di conoscenze serve principalmente ad orientare l’espressione in italiano *corretto*, termine qui adoperato come sinonimo di standard, che si legge poco oltre.

Come traguardo sociolinguistico le *Indicazioni* propongono la possibilità che lo scolaro “riconosc[a] che le diverse scelte linguistiche sono correlate alla varietà di situazioni comunicative” e che sia “consapevole che nella comunicazione sono usate varietà diverse di lingua e lingue differenti (plurilinguismo)” (p. 40). Per quanto riguarda gli obiettivi, ha sapore sociolinguistico quello della terza classe della scuola primaria: “Confrontare testi per coglierne alcune caratteristiche specifiche (ad es. maggiore o minore efficacia comunicativa, differenze tra testo orale e testo scritto, ecc.)” (p. 41), cui si aggiunge in quinta la capacità di adattare il lessico “ai destinatari, alle situazioni” e “alla forma testuale scelta” (p. 42).

Per gettare uno sguardo oltre la scuola elementare, sarà richiesto poco altro di specificamente sociolinguistico in terza media: possiamo citare i traguardi relativi alla conoscenza dei registri informale e formale e il riconoscimento, all’interno del repertorio linguistico, del “rapporto tra varietà linguistiche/lingue diverse (plurilinguismo) e il loro uso nello spazio geografico, sociale e comunicativo” (p. 43). Come obiettivo per l’ultima

classe della secondaria di primo grado, oltre al “[r]ealizzare scelte lessicali adeguate in base alla situazione comunicativa, agli interlocutori e al tipo di testo”, ancora improntato solo sul lessico, si trova anche quello di “stabilire relazioni tra situazioni di comunicazione, interlocutori e registri linguistici; tra campi di discorso, forme di testo, lessico specialistico” (p. 45).

Siamo di fronte a obiettivi e traguardi sociolinguisticamente ambiziosi, a giudicare dalla competenza sociolinguistica dimostrata poi dagli studenti universitari, per esempio dalle matricole con cui sono, per ragioni di lavoro, a contatto. Questa mia personale e impressionistica opinione sembra corroborata da ricerche specifiche anche recenti, come Cacchione & Rossi (2016).

Che cosa si propongono di saggiare, dal punto di vista sociolinguistico, le INVALSI? Iniziamo dando uno sguardo al *Quadro di Riferimento delle prove INVALSI di Italiano* (da adesso in avanti, e in bibliografia, QdR). Le conoscenze sociolinguistiche vengono qui addirittura menzionate direttamente, quando si parla delle competenze di letto-scrittura e in particolare della competenza lessicale esplicita (p. 5, corsivo mio):

la competenza lessicale esplicita comprende la capacità di motivare l’appropriatezza e la correttezza delle scelte lessicali e di esplicitare il significato di una parola o di un’espressione, o di una sua particolare accezione in rapporto ad altre, anche quando queste sono slegate da un contesto, e di analizzare le scelte lessicali facendo appello alle proprie conoscenze linguistiche (ad esempio morfologiche, etimologiche, *sociolinguistiche*, ecc.).

È sempre sul lessico, mi pare, che il quadro di riferimento insiste, ed è quindi ancora nel solo lessico che il QdR rischia di relegare la sociolinguistica: così anche a p. 3, quando si fa appello a due concetti centrali della diafasia come quello di linguaggio specialistico e di registro, le formulazioni “comprendere il significato di parole afferenti a lessici specialistici (per esempio scientifico, economico, artistico, ecc.)” e “riconoscere l’appartenenza di parole a determinati registri” trattano pur sempre soltanto di “lessico” e di “parole”. Tuttavia, per quanto fugace, qualche richiamo al registro e allo stile c’è anche quando si parla di dimensione pragmatica, e quando si approfondisce la dimensione grammaticale viene fatto un cenno alla comprensione del “significato dell’ordine ‘marcato’ – cioè differente dalla convenzione abituale (soggetto, verbo, complementi) – delle parole nella frase” (p. 4), un significato che, oltre ad avere risvolti evidenti nell’organizzazione pragmatico-informativa, non può essere del tutto dissociato dalla dimensione sociale dell’italiano.

La sociolinguistica, in un certo senso, esula dai livelli di analisi tradizionali, perché li può toccare e inglobare facilmente tutti, quindi non è facile, almeno per me, dire come possa essere adeguatamente inserita nei programmi scolastici e verificata proficuamente da prove del tipo delle INVALSI.⁴ Né mi è possibile ora suggerire una linea di progressione lungo i vari gradi d’istruzione che possa valere da percorso utile per stimolare l’attenzione dei ragazzi su fatti linguistici che si legano a fatti sociali come le diverse varietà geografiche di italiano, l’ambito formale o informale, il plurilinguismo.

⁴ Ma sul perché sia proficuo inserirla e sui limiti dei percorsi di educazione linguistica che la trascurino ha scritto Berruto (1976), (1977a), (1977b) e (1979), pur in un contesto sociale e linguistico profondamente diverso da quello odierno. Per alcune proposte didattiche atte a insegnare la variabilità sociolinguistica, cfr. Berruto et al. (1976) e, in anni più recenti, Andorno et al. (2003); Baratto et al. (2017) e Favata & Tronci (2019). Le risposte aperte fornite alle INVALSI dagli studenti delle terze medie del 2017/18 sono studiate in ottica variazionistica da Leonetti & Viale (2020).

Posso, più modestamente, provare ad osservare che cosa effettivamente di sociolinguistico venga interrogato dalle prove INVALSI, e ‘che lingua fa’ nelle prove INVALSI, riprendendo una formula che è tornata in voga tra letterati e linguisti qualche anno fa, e che carsicamente riaffiora sulle riviste, nei libri e nelle altre pubblicazioni a tema.⁵

3. Che lingua fa nelle INVALSI?

Il campione dell’indagine è piuttosto sommario. Ho infatti preso in esame solamente i testi delle prove INVALSI di italiano somministrati nell’arco di tempo che va dall’anno scolastico 2016/17 al 2018/19 nelle scuole primarie e nella secondaria di primo grado. È un campione certo non amplissimo, dal quale tuttavia spero possa scaturire qualche considerazione non inutile.

3.1 L’italiano delle INVALSI

Quanto alla varietà di italiano che si legge nelle prove, si dirà, innanzitutto, che le variazioni di registro dovute ai diversi tipi di testo somministrato possono essere avvertite, probabilmente, anche dagli stessi studenti, che vi riconosceranno testi narrativi, da una parte, e testi più espositivi, dall’altra. In questo modo i ragazzi sono messi fin da subito, o almeno fin dalla preparazione alle prove che si fa a scuola, di fronte alla multiformità della norma e alla variabilità della lingua.

La marcatezza diatopica dei testi narrativi è spesso evidente. Nella prova per la quinta elementare del 2018/19, si incontra, per esempio, il regionalismo *babbo*. Non soltanto questa parola è presente nel brano tolto da *L’uomo nero e la bicicletta blu* di Baldini, ma compare anche in una domanda del test e pur trovandosi sempre vicino a “nonno”, con cui condivide il campo semantico, non è oggetto di nessuna spiegazione: non si fa cenno, insomma, al fatto che *babbo* è di uso comune, come voce per indicare il genitore, solo in area toscana, mentre nel resto della Penisola, tanto al nord quanto al sud e specialmente nei grandi centri, il suo impiego si è fortemente ridotto di fronte all’avanzata del sinonimo *papà*.⁶

Tra i materiali di accompagnamento della stessa prova, invece, viene discusso il termine *cocomero*:

⁵ Qualche esempio recente di riemersione si legge nei titoli di Antonelli (2016) e Berruto (2013). Il conio pare essere di Pasolini, del marzo 1965.

⁶ Sacche di resistenza di *babbo*, che nel gergo giovanile ha preso a significare anche ‘stupido’, ‘sciocco’ (cfr. Ambrogio & Casalegno 2004: s.v. babbo), si trovano p. es. in Sardegna, nel Mantovano, nella Tuscia. Sulla questione si trovano accenni *passim* nella raccolta curata da Nesi (2013).

In questo testo si parla di cocomeri. Siccome tu potresti essere abituato a chiamare il cocomero con un altro nome, sotto trovi un'immagine del cocomero del racconto e altri due nomi con cui può essere chiamato.

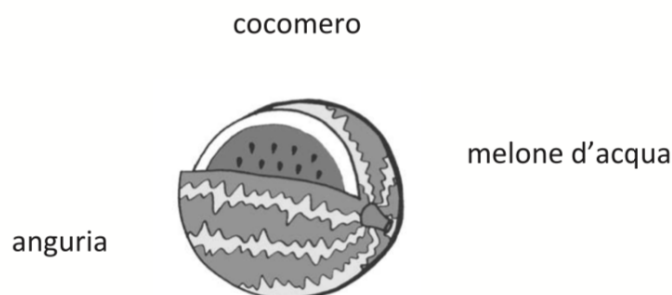


Figura 1. Da 2019, 05, materiali di accompagnamento al testo A.⁷

Si sa, quindi, la capacità del bambino e del ragazzo di comprendere un caso di geosinonimia, peraltro abbastanza noto. Il momento del test diventa quindi anche un momento in cui si scopre (o si riscopre nel migliore dei casi) la diversità dell'italiano nelle sue varietà macroregionali. Il materiale di accompagnamento serve ai quei parlanti che non siano toscani o non siano dell'Italia centrale, secondo quanto ci rivelano i dati della prima inchiesta dell'ALIQUOT.

C'è da sperare che i ragazzi abbiano già capito come gestire situazioni simili, in quanto potrebbero anche essere disorientati, almeno di primo acchito, da un brano in cui il frutto, che riveste un ruolo centrale, in una delle scene porta un nome che non è immediatamente perspicuo a chi non sia nelle aree in cui questo termine è in uso.

Come *babbo* e *cocomero* rimandano a un italiano regionale toscano o centrale, è notevole anche l'impiego fisso del passato remoto, che forse tanti scolari delle regioni del nord scoprono e adoperano solo nella scuola e certamente per tradizione scolastica.

Per passare a fatti che non correlano più con la diatopia, ma con caratterizzazioni sociolinguistiche diverse, negli stessi testi narrativi menzionati poc'anzi sono frequenti i casi di dislocazione e di frasi sintatticamente marcate, senza che sia fatto cenno alcuno alle particolarità informative di queste strutture. Ancora nella prova per la quinta classe elementare 2018/19 si incontrano casi come quelli mostrati negli esempi (1)-(3):

- (1) 2019, 05, testo A
 - *Cosa è successo?*
 - *Stanotte ci hanno rubato dei polli.*
- (2) 2019, 05, A1
 - Il racconto si apre con la scena in cui il protagonista e Francesco vedono la macchina dei carabinieri (da riga 1 a riga 11).*

⁷ Da qui in avanti, quando necessario le prove INVALSI verranno citate con l'abbreviazione consueta, che indica in successione l'anno di somministrazione, il livello (quelli che ci riguardano sono 05, per la quinta primaria, e 08, per la terza secondaria di primo grado), la domanda o la sezione in questione.

*In base al racconto, **che cosa** può aver pensato il protagonista alla vista della macchina dei carabinieri?*

- (3) 2019, 05, A12
“Cosa è successo?” (riga 21) chiede il protagonista alla mamma. Con questa domanda si inizia a raccontare ciò che è accaduto a casa del protagonista mentre lui non c’era.

Vi si apprezza la disinvolta alternanza delle forme del pronome interrogativo “neutro” che sono però marche sociolinguistiche forti: *cosa*, da solo, è neo-standard, di contro allo standard (in specie, scolastico) *che cosa*. Generalmente nelle consegne e nelle domande della prova si impiega la seconda forma, ma il quesito posto nell’esempio (3) interroga proprio il “Cosa è successo?” presente nel testo.⁸

Nel test per la quinta elementare dell’anno precedente possiamo trovare casi di glossatura o spiegazione del lessico più ricercato tramite il trattino (-) seguito da un altro nome, che del primo può dirsi sinonimo: come in *situazione-problema* (2018, 05, A1), *espediente-trucco* (2018, 05, A17), ecc.;⁹ e casi di superfetazione di aggettivi possessivi nelle domande o nelle consegne degli esercizi, o negli esempi (4)-(6):

- (4) 2018, 05, A1
*Alessandro ha paura del **su**o maestro.*
- (5) 2018, 05, A2
*Ad esempio un cambiamento è: Alessandro riesce a suonare bene il **su**o violino.*
- (6) 2018, 05, A10
*Mostrare che Alessandro rifiuta di fare i compiti che gli dà il **su**o maestro.*

Si noti che in nessun caso il *maestro* o il *violino* di Alessandro possono essere confusi con altri maestri o altri strumenti menzionati nel discorso, quindi specificare con l’aggettivo possessivo non è necessario. Sembra che si faccia sentire qui interferenza dell’inglese, nel cui standard si avrebbe per esempio *Alessandro plays his violin well*. Se non è, perciò, solo ed esclusivamente un tratto del doppiaggese che sta entrando anche nello standard italiano, il *su*o degli esempi (4)-(6) è forse dovuto alla volontà di essere tanto chiari, fino a risultare ridondanti e quasi didascalici, nelle domande INVALSI, con l’effetto di elevare di registro, fino a quelli più sorvegliati, un tratto ancora non-standard, come questo.

Sempre a motivi di didascalismo può essere ascritto l’uso abnorme dei dispositivi di strutture marcate dal punto di vista informativo, come nel caso dell’esercizio in Figura 2:

⁸ E si noti in questo stesso quesito l’impiego di *ciò che*, tratto standard letterario rispetto al neo-standard *quello che* (v. Sabatini 1985: 158).

⁹ Sul piano teorico, questi si potrebbero anche ritenere esempi di composti NN in cui il primo termine è immediatamente giustapposto al secondo, che funziona da aggettivo, e in cui i due termini sono legati da un legame implicito equivalente a una relativa appositiva (molto frequenti nel neo-standard, cfr. Berruto 2012: 90). Si tratta tuttavia di un tipo di grafia ben poco attestata in italiano. *Espediente-trucco* potrebbe essere trattato come un binomio contestualizzato (Masini 2006), che si presenta però, rispetto ai casi prototipici di questa categoria, collegato senza congiunzione. La resa grafica con il trattino è tuttavia ben poco attestata in italiano (dove prevale semmai l’impiego del segno interpuntivo di barra obliqua, cfr. Fiorentini & Miola 2020).

RAGIONI	ESIGENZE		
	VUOLE ESERCITARSI	È <u>DAL CINESE</u> CHE VUOLE ESERCITARSI	È <u>IN SEGRETO</u> CHE VUOLE ESERCITARSI
a) Non vuole far sapere a nessuno dei suoi trucchi. È UNA RAGIONE PER CUI...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Il vecchio cinese lo ha aiutato fin dall'inizio. È UNA RAGIONE PER CUI...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Vuole farcela da solo. È UNA RAGIONE PER CUI...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figura 2. Quesito 2018, 05, A5.

Si compiti il sottoquesito b), la cui risposta corretta è da porre nella colonna di mezzo: “Il vecchio cinese lo ha aiutato fin dall’inizio. È una ragione per cui è dal cinese che vuole esercitarsi”. Ne risulta un costrutto sintatticamente molto marcato. Sarebbe bastato, a ben vedere, “è una ragione per cui vuole esercitarsi dal cinese”, con l’aggiunta, rispetto alla *esigenza* che si legge nella prima colonna, del solo sintagma dal *cinese*, che poteva essere eventualmente segnalato con una ulteriore sottolineatura.

Dunque la lingua usata nelle prove INVALSI, anche a considerare soltanto quella adoperata nelle consegne e negli esercizi dei test rivela un’articolazione tutt’altro che banale e inequivocabile: il quadro non sembra completamente omogeneo con ciò che viene detto dalle grammatiche più tradizionali, né con l’italiano che si parla, più spontaneamente, nelle nostre aule scolastiche e fuori da esse. Soprattutto, sembra che le consegne presentino anche tratti e costruzioni usati talvolta in modo abnorme rispetto allo standard scolastico o al neo-standard, cosa che rischia, se non proprio di confondere gli scolari che devono compilare il test, almeno di condizionare ciò che essi ritengono essere standard o neo-standard. Non è escluso che anche questo contribuisca a far assegnare, dagli scolari, alcuni di quei tratti a registri formali e alti e come tali a (tentare di) riprodurli nei gradi e ordini di scuola successivi.

3.2 Domande sociolinguistiche nelle INVALSI

Qualche parola, infine, su quali conoscenze sociolinguistiche sono testate nelle INVALSI, e come.

In prima battuta – lo abbiamo già detto – i test delle INVALSI mettono gli scolari di fronte alla variazione diafasica e diamesica, presentando sempre due tipi differenti di testi: uno di genere narrativo e l’altro di genere più espositivo o scientifico. È questo un primissimo grado di conoscenza sociolinguistica, che mette sì in luce l’esistenza di diversi registri, ma che si situa ancora a un livello molto intuitivo e per nulla specifico. Molti sono i tratti geograficamente marcati che s’incontrano nei testi narrativi, come visto al paragrafo 3.1 per il caso del *cocomero*. Infatti, anche nella prova somministrata nel 2016/17 nelle scuole secondarie di primo grado si poteva apprezzare almeno un po’ di attenzione per la variazione, e in particolare sempre alla diatopia. Nella lettura proposta

da *Mai devi domandarmi* di Natalia Ginzburg, il plurilinguismo coinvolge una lingua morta, il latino che l'io narrante fa trapelare dalle parole della maestra con il cappello di feltro: *Te Deum* viene immediatamente storpiato dalla giovane Natalia in *Tedem*, appioppato come soprannome alla maestra. Poi:

(7) 2017, 08, testo A

Comunque grazie alla maestra Tedem fui promossa agli esami di licenza elementare. Mia madre m'informò che ora m'aveva iscritto «al ginasio»: pronunciava questa parola con una enne sola. Il ginasio era il luogo dove avevo fatto gli esami.

Così il brano, in ragione della pronuncia marcatamente settentrionale che Ginzburg vuole riprodurre nel testo, propria della madre della protagonista. Tuttavia, anche in questo caso la dimensione diatopica, sebbene presente, non viene esplicitamente interrogata nella prova, benché vi si trovi una domanda che chiede proprio conto di cosa sia il *ginasio*:

A4. Il "ginasio" era

- A. la scuola preferita dalla mamma della protagonista
- B. la scuola dove la protagonista aveva frequentato le elementari
- C. la scuola dove la protagonista aveva conseguito la licenza elementare
- D. la scuola che la maestra Tedem aveva consigliato alla protagonista

Figura 3. Quesito 2017, 08, A4.

Come si vede, piuttosto che la dimensione sociolinguistica (che poteva essere sondata con una domanda del genere di "Che tipo di pronuncia è registrata da Ginzburg nella parola 'difficile' e sconosciuta alla protagonista detta dalla mamma?") viene richiesto un esercizio di comprensione del testo, che non mette effettivamente, o non mette del tutto, in gioco le competenze maturate dagli scolari in merito al variare della pronuncia dell'italiano al variare dello spazio geografico: si dà in un certo senso per scontato che nella conoscenza enciclopedica degli scolari lo scempiamento delle geminate (o almeno la pronuncia di *ginasio* "con una enne sola") sia un fatto tipico dell'italiano delle regioni settentrionali, ma non lo si verifica direttamente.

Un altro caso nel quale la competenza sociolinguistica mi sembra entrare almeno tangenzialmente è il quesito A3 della già citata prova per la quinta elementare del 2017/18:

A3. “Era che da una settimana la sognavo tutte le notti, ‘sta cosa” (righe 9-10); “ ‘sta cosa” si riferisce a una parte di testo che abbiamo sintetizzato in una parola. Quale parola sintetizza questa parte di testo?

- A. Arresto
- B. Sgridata
- C. Furto
- D. Litigata

Figura 4. Quesito 2018, 05, A3.

In questo quesito viene verificata la capacità di riconoscere a che cosa si riferisce *'sta cosa*. Non soltanto l'incapsulatore stesso interrogato è in una varietà di italiano colloquiale, e certamente distante dall'italiano standard che a scuola ci si aspetta di trovare, ma anche l'intera porzione di testo in cui *'sta cosa* si trova presenta un ordine degli elementi marcato rispetto a quello atteso in italiano: in “era che da stamattina la sognavo tutte le notti, 'sta cosa” abbiamo, infatti, il dispositivo sintattico della scissa della polarità, o costruzione inferenziale – ancora proprio della lingua parlata piuttosto che scritta. Anche in questo caso osserviamo una discesa verso il registro che definirei del parlato-parlato, che è proprio di molti testi narrativi, ma che può creare disorientamento nell'allievo se non è stato precedentemente addestrato a riconoscerlo e a trattarlo con i giusti strumenti linguistico-grammaticali.

Questo addestramento è ancor più necessario per l'esercizio C4 delle INVALSI per la classe terza della secondaria di primo grado 2016/17:

C4. Qual è il soggetto delle frasi che seguono? Scrivilo vicino ad ognuna.

Attenzione: scrivi il soggetto anche quando è sottinteso.

- a) L'hai avuto l'invito?
- b) A lei non piace la verdura.
- c) Dove l'avete messa la mia cartella?
- d) Il mio libro l'hai preso tu?
- e) Vi interessa questo spettacolo?
- f) Correavano tutti verso la piazza.

Figura 5. Quesito 2017, 08, C4.

Il *Rapporto tecnico delle rilevazioni degli apprendimenti* (INVALSI 2017) riporta che:

Il quesito più difficile è C4, è una domanda a risposta aperta univoca afferente all'ambito della sintassi, che richiede di rintracciare per ognuna delle sei frasi

proposte il soggetto anche quando non è in posizione preverbale e/o non è un soggetto agente.

Un adeguato trattamento in classe delle strutture a dislocazione avrebbe potuto forse dare qualche chance in più su almeno due dei quesiti posti in questa domanda: il (d) e forse anche (a) e (c). Ma, come si è cercato di far vedere, queste competenze che pure sono ritenute necessarie per il buon superamento delle INVALSI, solo di rado sono approfondite nelle grammatiche scolastiche e vengono, più sovente che no, delegate esclusivamente all'insegnante di buona volontà e lungimirante, per quanto, forse, uno studente possa essere esposto, e perciò abituato, a questi tratti di parlato-parlato pan-italiani più che ai regionalismi visti all'inizio del § 3.1. Infatti, stando al campione di grammatiche scolastiche studiato da Bachis (2019), sono pochi i testi pubblicati, anche dopo il 1980, che si occupano dei vari fenomeni di enfasi e dislocazione; di questi, poi, alcuni, pur non rifiutandolo categoricamente, specificano che il loro uso va accolto con cautela. Si noti peraltro che le frasi con dislocazione non sono per nulla estranee all'italiano comune, nemmeno nei secoli in cui la norma andava fissandosi e confluendo così nelle grammatiche che possiamo chiamare tradizionali. Si tratta dunque a buon diritto di un costrutto che dovrebbe essere incluso, dai grammaticografi, nello standard o almeno nel cosiddetto neo-standard. La forma di messa in rilievo più trattata dalle grammatiche recenti, comunque, rimane senz'altro la posposizione del soggetto rispetto al predicato, di cui si occupa la quasi totalità dei testi rivolti alla scuola media e a quella superiore.¹⁰

Effettivamente, come si è notato nei paragrafi precedenti, è sulla differenza lessicale al variare del registro che si osserva la maggiore attenzione sociolinguistica delle prove INVALSI, specialmente da quando, nel 2017/18, è stata inserita una parte atta a verificare più da vicino il lessico nella secondaria di primo grado, in ragione del fatto che – per tornare ancora una volta alle *Indicazioni* – “l'uso del lessico, a seconda delle discipline, dei destinatari, delle situazioni comunicative e dei mezzi utilizzati per l'espressione orale e quella scritta richiede lo sviluppo di conoscenze, capacità di selezione e adeguatezza ai contesti” (p. 38) e in ragione del fatto che i ragazzi debbano essere guidati a essere consapevoli della variabilità sociolinguistica della lingua (p. 43).

Dunque, mentre la lingua, per dir così, ‘parlata’ dalle INVALSI, almeno in qualche caso, scarta dall'italiano standard letterario e scolastico, mostrando tendenze che talvolta vanno addirittura oltre l'italiano parlato comune, ovvero oltre ciò che siamo soliti chiamare neo-standard, le competenze sociolinguistiche saggiate dalle INVALSI non sono poi molte, almeno per gli ordini scolastici precedenti alla scuola secondaria di secondo grado.

Si può dire, a mo' di concisa conclusione, che nelle *Indicazioni* per le INVALSI la sociolinguistica non è affatto negletta e anzi vi sono chiari richiami a questa dimensione della lingua. Tuttavia, nei test essa fa capolino quasi solo quando si tratta del lessico, della buccia della lingua, e non certo della polpa. Ma la sociolinguistica, adeguatamente presentata e insegnata, potrebbe certamente dare di più agli studenti: essere cioè quantomeno ancillare agli altri livelli tradizionali del lessico, morfologia, sintassi, insomma alla grammatica tout court, e venire in qualche caso in aiuto durante i test di verifica della preparazione.

¹⁰ Infatti queste domande sono frequentissime nei test per la secondaria di primo grado, come mostra Lo Duca (2018).

Bibliografia

- ALQUOT = Tosques, Fabio & Castellarin, Michele. 2013–2017. *Atlante della Lingua Italiana QUOTidiana* (<https://www.atlante-aliquot.de>) (Consultato il 14.07.2021).
- Ambrogio, Renzo & Casalegno, Giovanni. *Scrostati gaggio! Dizionario storico dei linguaggi giovanili*. Torino: UTET.
- Ammon, Ulrich. 1986. Explikation der Begriffe ‘Standardvarietät’ und ‘Standardsprache’ auf normtheoretischer Grundlage. In Holtus, Günter & Radtke, Edgar (eds), *Sprachlicher Substandard*. Vol. 1: 1–62. Tübingen: Niemeyer.
- Andorno, Cecilia & Bosc, Francia & Ribotta, Paola. 2003. *Grammatica. Insegnarla e impararla*. Perugia: Guerra.
- Antonelli, Giuseppe (a cura di). 2016. *Che lingua fa?* Numero monografico di *Nuovi argomenti* 73. Milano: Mondadori.
- Bachis, Dalila. 2019. *La lingua delle grammatiche scolastiche italiane edite dal 1919 al 2018*. Firenze: Accademia della Crusca.
- Baratto, Gabriele & Duberti, Nicola & Sordella, Silvia. 2014. Registri di classe. Note su un’esperienza didattica di educazione alla variabilità diafasica. In Cerruti, Massimo & Corino, Elisa & Onesti, Cristina (a cura di), *Lingue in contesto. Studi di linguistica e glottodidattica sulla variazione diafasica*, 137–162. Alessandria: Dell’Orso.
- Berruto, Gaetano. 1976. Recenti pubblicazioni sulla didattica dell’italiano. *Orientamenti pedagogici* 23. 1040–1059.
- Berruto, Gaetano. 1977a. Dodicesima lezione: L’insegnamento della lingua e la sociolinguistica. In Berruto, Gaetano & Berretta, Monica (a cura di), *Lezioni di sociolinguistica e linguistica applicata*, 139–149. Napoli: Liguori.
- Berruto, Gaetano. 1977b. Sociolinguistica e educazione linguistica. In Berruto, Gaetano (a cura di), *Scienze del linguaggio ed educazione linguistica*, 120–168. Torino: Stampatori.
- Berruto, Gaetano. 1979. Sociolinguistica e grammatiche scolastiche. In Albano Leoni, Federico & Pigliasco, Maria Rosaria (a cura di), *La grammatica: aspetti teorici e didattici. Atti del IX Congresso Internazionale della Società di Linguistica Italiana (SLI), Roma 31 maggio-2 giugno 1975*, 119–136. Roma: Bulzoni.
- Berruto, Gaetano. 2003. Sul parlante nativo (di italiano). In Radatz, Hans-Ingo & Schlösser, Rainer (eds), *Donum grammaticorum. Festschrift für Harro Stammerjohann*, 1–14. Tübingen: Niemeyer.
- Berruto, Gaetano. 2007. Miserie e grandezze dello standard. Considerazioni sulla nozione di standard in linguistica e sociolinguistica. In Molinelli, Piera (a cura di), *Standard e non standard tra scelta e norma. Atti del XXX Convegno della Società Italiana di Glottologia, Bergamo, 20-22 ottobre 2005*, 13–41. Roma: il Calamo.
- Berruto, Gaetano. 2012. *Sociolinguistica dell’italiano contemporaneo*. Roma: Carocci [1a ed. Roma: La Nuova Italia Scientifica, 1987].
- Berruto, Gaetano. 2013. Ma che lingua fa? Alcune riflessioni sullo standard e sul neo-standard. In Bombi, Raffaella (a cura di), *Manuale di comunicazione istituzionale e internazionale*. 45–67. Roma: il Calamo.
- Berruto, Gaetano & Berretta, Monica & Calleri, Daniela & Canobbio, Sabina. 1976. *Il libro di grammatica*. Torino: Marietti.
- Cacchione, Annamaria & Rossi, Luca. 2016. La lingua troppo (poco) variabile: monolinguisimo e mistilinguisimo in testi funzionali di matricole universitarie. In Ruffino, Giovanni & Castiglione, Marina C. (a cura di), *La lingua variabile nei testi*

- letterari, artistici e funzionali contemporanei: analisi, interpretazione, traduzione. Atti del XIII Congresso SILFI (Palermo, 22-24 settembre 2014)*, 457–489. Firenze: Franco Cesati Editore.
- Favata, Giovanni & Tronci, Liana. 2019. Fare sociolinguistica attraverso la letteratura: una proposta didattica per studiare le varietà dell'italiano. *Italiano a scuola* 1. 25–46.
- Fiorentini, Ilaria & Gianollo, Chiara & Grandi, Nicola (a cura di). 2020. *La classe plurilingue*. Bologna: Bononia University Press.
- Fiorentini, Ilaria & Miola, Emanuele. 2020. Disjunctive/conjunctive/whatever: the development of Italian barra ('slash') as a non-exhaustive connective. *Language Sciences* 81. 1–11.
- Grandi, Nicola. 2018. Sulla penetrazione di tratti neo-standard nell'italiano degli studenti universitari. Primi risultati di un'indagine empirica. *Griseldaonline* 17(1). 1–24 (<https://griseldaonline.unibo.it/article/view/9021>) (Consultato il 14.07.2021).
- Indicazioni = INVALSI. 2012. Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. *Annali della Pubblica Istruzione* LXXXVIII. Firenze: Le Monnier.
- INVALSI. 2017. Rilevazioni Nazionali degli Apprendimenti 2016-2017. Rapporto tecnico. (http://www.invalsi.it/invalsi/doc_eventi/2017/Rapporto_tecnico_SNV_2017.pdf) (Consultato il 14.07.2021).
- Leonetti, Eliana & Viale, Matteo. 2020. La variabilità linguistica delle risposte a domande aperte di comprensione del testo nelle prove INVALSI *computer based*. *Italiano a Scuola* 2(1). 103–130.
- Lo Duca, Maria Grazia. 2018. Le prove di grammatica dell'INVALSI e la progressione dei contenuti grammaticali: il caso del soggetto. In Calaresu, Emilia & Dal Negro, Silvia (a cura di), *Attorno al soggetto. Percorsi di riflessione tra prassi didattiche, libri di testo e teorie*. 123–138. Milano: Studi AItLA.
- Loiero, Silvana & Lugarini, Edoardo (a cura di). 2019. *Tullio De Mauro: Dieci tesi per una scuola democratica*. Firenze: Franco Cesati Editore.
- Masini, Francesca. 2006. Binomi coordinati in italiano. In Cresti, Emanuela (a cura di), *Prospettive nello studio del lessico. Atti del IX Congresso SILFI (Firenze, 14-17 giugno 2006)*, 563–571. Firenze: FUP.
- Nesi, Annalisa (a cura di). 2013. *La lingua delle città. Raccolta di studi*. Firenze: Franco Cesati Editore.
- Ondelli, Stefano & Romanini, Fabio 2018. Norma interiorizzata e uso: un'indagine preliminare su parlanti italiani. *Italica Wratislaviensia* 9(1). 185–207.
- QdR = INVALSI. 2018. *Quadro di Riferimento per le prove INVALSI di Italiano*. (https://invalsi-areaprove.cineca.it/docs/file/QdR_ITALIANO.pdf) (Consultato il 14.07.2021).
- Sabatini, Francesco. 1985. 'L'italiano dell'uso medio': Una realtà tra le varietà linguistiche italiane. In Holtus, Günter & Radtke, Edgar (eds), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 154–184. Tübingen: Gunter Narr Verlag.

Quattro caratteristiche dell'analisi logica. Considerazioni a margine delle prove INVALSI del grado 8

Diego Pescarini

CNRS, Université Côte d'Azur, BCL

diego.pescarini@univ-cotedazur.fr

Abstract

Questo contributo discute i quattro principî che, a mio avviso, dovrebbero guidare l'insegnamento dell'*analisi logica*. Intendo qui abbozzare alcune idee che spero possano risultare utili per valutare criticamente gli strumenti didattici in commercio, per riflettere su quanto richiesto dalle prove INVALSI e per aiutare gli insegnanti nella progettazione delle proprie attività didattiche.

1. Introduzione

L'analisi logica rimane un cardine della formazione linguistica, sebbene l'impianto della disciplina e gli strumenti adottati – a partire dai manuali – siano stati e continuano ad essere oggetto di critica da parte di esperti di varia estrazione.^{1,2} Questo contributo offre alcune brevi riflessioni sul ruolo e sull'*identità* dell'analisi logica nella scuola secondaria prendendo spunto da alcune prove INVALSI per il grado 8 (ovvero, il terzo anno della scuola media).

Ho scelto di illustrare le mie idee usando gli esercizi dell'INVALSI perché si tratta di strumenti per la valutazione molto più avanzati e ben calibrati di quanto si trovi mediamente nei manuali per la scuola. Tali esercizi sono inoltre un modello per la didattica della grammatica dei prossimi anni: data la natura e le proporzioni delle

¹ Ho cercato di includere in questo lavoro osservazioni e commenti ricevuti nel corso della giornata di studio *Leggere per comprendere a partire dai dati INVALSI* (Bologna, 6 Dicembre 2019) e del LIII Congresso della Società di Linguistica Italiana (*Insegnare linguistica: basi epistemologiche, metodi, applicazioni*), Como, 19-21 settembre 2019). Ringrazio gli organizzatori ed i partecipanti ad entrambi gli eventi. Ringrazio inoltre due revisori anonimi che mi hanno fornito preziosi suggerimenti e correzioni per migliorare questo testo.

² Da qui in poi userò il termine "analisi logica" per riferirmi all'analisi sintattica in generale, inclusa quindi l'analisi della frase complessa, tradizionalmente indicata come "analisi del periodo".

rilevazioni INVALSI, è inevitabile che il contenuto dei testi scolastici e, di conseguenza, delle lezioni di sintassi saranno condizionate dall'orientamento dell'INVALSI. Ciò pone un duplice fardello sulle spalle degli esperti INVALSI, che da un lato devono confezionare degli strumenti efficaci per la valutazione del presente, ma dall'altro devono anche riflettere su quali saranno gli effetti di lungo periodo. Realisticamente, non possiamo pensare che un tipo di rilevazione come l'INVALSI sia uno strumento di indagine neutrale rispetto al proprio oggetto di valutazione.

Riflettere *criticamente* sulle prove INVALSI diventa quindi un esercizio utile per cercare di capire la direzione verso cui stiamo andando. Studiare come sono strutturate le prove INVALSI dovrebbe inoltre essere un esercizio abituale fra gli insegnanti, per riflettere sulle ragioni, gli obiettivi, la “tecnica” delle prove, per comparare questi aspetti con le attività didattiche proposte quotidianamente in classe e per valutare la bontà di quanto si trova nei manuali adottati. Le prove INVALSI – o gli esempi delle prove – si trovano facilmente online³ e rappresentano un ottimo punto di partenza che gli insegnanti dovrebbero sfruttare per costruire i propri esercizi, senza ricorrere a strumenti didattici supplementari.

Ciò che segue è una serie di considerazioni condotte a margine della lettura di alcuni quesiti INVALSI. Si tratta di osservazioni che scaturiscono dalla mia formazione e dalla mia esperienza di ricerca quotidiana; quindi, non tutto quello che dirò sarà immediatamente valido o trasponibile nella prassi didattica. Spero però che questo breve scritto possa essere di aiuto a quanti, soprattutto fra i docenti, sono alla ricerca di qualche linea di indirizzo per navigare nel mare della riflessione grammaticale.

L'articolo è strutturato come segue: nel §2 cerco di individuare alcune motivazioni della “crisi d'identità” che sembra investire l'analisi logica nella scuola di oggi; nei §§3-6 cerco di discutere quattro idee-chiave attorno alle quali, a mio avviso, va ridefinita la disciplina. La prima idea è che l'analisi logica sia prima di tutto un'attività *metacognitiva* che mira a ricostruire l'insieme di istruzioni logiche che ci consentono di costruire il significato di una frase a partire dai significati delle parole e degli elementi funzionali che la compongono. La seconda idea è che l'analisi logica debba essere il più possibile *autonoma* rispetto ad altre attività di riflessione linguistica e comprensione testuale. La terza idea è che l'analisi logica debba essere *razionale*, debba cioè basarsi su principi logicamente consequenziali ed empiricamente verificabili. La quarta idea è che l'analisi logica debba essere il più possibile *generale* e *generalizzabile* a tutte le lingue studiate, compreso l'inglese o le lingue classiche.

2. Alla ricerca di una nuova identità?

L'insegnamento della grammatica poggiava in origine su un solido contrafforte: il latino. Attribuire un ruolo autonomo alla grammatica – ed in particolar modo alla pratica dell'analisi logica – è una questione su cui si dibatte sin dal periodo postunitario (Demuru 2018) ed è rimasto un nodo non del tutto risolto in seguito all'abbandono del latino nella scuola media (fine anni '70) e con il progressivo ridimensionamento del ruolo delle lingue classiche nel secondo grado della secondaria. L'impressione è che nel corso del tempo si sia elusa la domanda principale: perché continuare ad investire sforzi ed energie di insegnanti e studenti in una pratica – l'analisi sintattica – che non ha alcun fine pratico

³ Come nota un revisore anonimo, questo era vero prima dell'introduzione delle prove computerizzate.

immediato?

I tentativi di rispondere a questa domanda non mancano e le indicazioni – incluse quelle ministeriali – sono tutto sommato chiare. Tuttavia, osservando i libri di testo e gli esercizi proposti in classe, mi sembra che nel complesso si navighi ancora a vista. Alcuni fra insegnanti ed intellettuali sostengono che sarebbe meglio tornare al passato: ripristinare lo studio del latino come panacea di tutti i mali della grammatica o, almeno, riproporre strumenti didattici obsoleti e latino-centrici come l'*Analisi logica* di Vittorio Tantucci (1955), ancor oggi adottato data “la considerazione della immutata validità della parte teorica” (cito dall'introduzione all'edizione del 1982). L'adozione di un volume come quello di Tantucci è, a mio avviso, una spia, un indicatore dell'assenza di idee chiare su come valutare uno strumento didattico in relazione agli obiettivi della disciplina. Torniamo quindi al problema centrale da cui eravamo partiti: l'assenza di un'identità della disciplina e la necessità di fornirgliene una.

Nella tradizione italiana, il termine *analisi logica* si riferisce allo studio della sintassi della frase semplice che, per ragioni pedagogiche (ma non necessariamente linguistiche), è solitamente mantenuto distinto dallo studio della sintassi della frase complessa (o analisi del periodo). La scelta del termine “analisi logica” (anziché il più trasparente “sintassi”) ha radici molto lontane. Logica e grammatica si trovano spesso appaiate nella visione dei filosofi, ma non nella pedagogia delle lingue, dove l'imitazione, la ripetizione e la memorizzazione hanno tradizionalmente giocato un ruolo primario sin dall'antichità. Nel XVII secolo, tuttavia, si registra un vero e proprio cambio di paradigma, che gli storici della linguistica fanno solitamente coincidere con le innovazioni pedagogiche adottate dai monaci benedettini del monastero di Port-Royal. Il metodo di Port-Royal rappresentava un'innovazione rispetto alla didattica tradizionale, che, come accennato, si basava ancora ampiamente sulla memorizzazione e sullo studio della retorica. Questa prassi, a Port-Royal, venne scalzata da un metodo di apprendimento basato sul ragionamento, la comparazione fra strutture linguistiche affini, l'osservazione dei dati empirici. I principî ispiratori della pedagogia di Port-Royal furono poi condensati nella *Grammaire générale et raisonnée* (1660; il sottotitolo recita *contenant les fondements de l'art de parler, expliqués d'une manière claire et naturelle*). La grammatica contiene un solo capitolo sulla sintassi – il ventiquattresimo e ultimo – che consiste in poche note di sintassi minima tese ad illustrare alcuni principî *naturali*, ovvero generalissimi, seguiti da alcune annotazioni sulle figure di discorso, che sono quindi concepite come deviazioni dalla sintassi naturale. I monaci di Port-Royal produssero anche un manuale di logica (*La logique ou l'art de penser*, 1662). La logica e la grammatica di Port-Royal furono evidentemente concepite come due opere parallele, ma senza alcun nesso esplicito fra *l'art de parler* e *l'art de penser*. Solamente con la riflessione grammaticale degli enciclopedisti si afferma una concezione della sintassi come *analisi logica*, ovvero come trait d'union fra l'analisi del discorso (meglio: delle parti del discorso) e l'analisi del pensiero (Chevalier 1979). Ai primi dell'Ottocento il termine *analisi logica* si è già imposto come sinonimo di analisi sintattica (contrapposta all'analisi morfologica o grammaticale), ad esempio nel titolo dei due fascicoli pubblicati nel 1827 da Noël e Chapsal: *Leçons d'analyse grammaticale* e *Leçons d'analyse logique* (Chevalier 1979: 31).

Per quanto riguarda la ricezione del termine in Italia nel periodo preunitario, non sono a conoscenza di ricerche sul tema, ma il termine *analisi logica* è già dato come corrente nei primi documenti postunitari (Demuru 2018, De Roberto 2018), sebbene di “logico” non fosse rimasto che il nome. Pedagogisti, linguisti, letterati e politici erano più o meno

tutti concordi nel sostenere l'inutilità dell'analisi logica, ritenuta una prassi propedeutica all'insegnamento del latino, ma priva di alcun valore intrinseco. Questo ostracismo nei confronti dell'analisi rifletteva le reali pratiche d'insegnamento e le condizioni sociali dell'epoca. La maggior parte degli studenti aveva ben altri bisogni educativi⁴ e l'insegnamento era ancora ampiamente basato sull'apprendimento mnemonico.

Da allora, tuttavia, il contesto sociale e pedagogico è radicalmente mutato e, nel corso del Novecento, abbiamo assistito ad un rinnovato interesse per la disciplina (almeno a parole...). Oggi gli apprendenti sono per lo più italo-foni (inclusi coloro che parlano italiano accanto a una lingua familiare straniera), il nozionismo è – almeno in linea di principio – del tutto superato, gli esercizi hanno un peso preponderante rispetto alla teoria, è scomparsa praticamente del tutto – almeno nella scuola primaria e nel primo ciclo della secondaria – una visione dell'analisi orientata verso il latino (sebbene un certo latino-centrismo si possa ancora intravedere in certe prassi didattiche più volte criticate come l'ipertrofica classificazione dei complementi). Ancora, la linguistica ha dato nuova linfa agli studi di grammatica, richiamandosi anche ai principi razionalistici della scuola Port-Royal. Dato questo contesto, credo sia quindi finalmente possibile ripensare l'insegnamento della sintassi e, a questo proposito, vorrei qui discutere quattro le linee guida che, a mio avviso, dovrebbero guidare la riflessione di insegnanti e studenti (ognuna sarà discussa più in dettaglio nei paragrafi successivi):

1. L'analisi logica è prima di tutto un'attività metacognitiva, ovvero un procedimento attraverso il quale gli studenti mettono a nudo dei meccanismi cognitivi. Sarebbe quindi riduttivo pensare che l'analisi logica serva ad 'esprimersi correttamente' o ad 'evitare di commettere errori nello scritto'.
2. L'analisi logica dovrebbe essere il più possibile indipendente da altri processi inferenziali che scaturiscono da conoscenze enciclopediche (ovvero da ragionamenti che dipendono dalla conoscenza del mondo).
3. L'analisi logica è un metodo scientifico, basato sulla verifica delle ipotesi, sulla replicabilità delle osservazioni, sulla generalizzabilità delle definizioni. La costruzione di un apparato teorico pienamente razionale è un obiettivo di lungo periodo, da raggiungere progressivamente mediante approssimazioni successive, in base alle capacità degli studenti e, possibilmente, in parallelo con l'acquisizione delle lingue straniere. Se quindi è accettabile un certo grado di approssimazione nei materiali per gli studenti, gli insegnanti non dovrebbero perdere però di vista l'obiettivo finale dell'analisi, quello appunto di costruire un procedimento logico basato su inferenze e deduzioni.
4. L'analisi logica non è l'analisi dell'italiano. Concetti come soggetto, predicato, complemento, valenza, accordo, frase, sintagma, ecc. non si applicano solo all'italiano, ma sono strumenti descrittivi per comprendere la struttura di tutte le lingue. La storia del nostro sistema educativo ha tradizionalmente assegnato al docente di italiano il compito di insegnare l'analisi logica, ma, in linea di principio, non si capisce perché questa incombenza non possa essere condivisa con gli insegnanti delle lingue straniere.

⁴ Ricordiamo che l'insegnamento della grammatica italiana si rivolgeva ad una platea di studenti in larga parte dialettofoni. A questo proposito, è interessante sottolineare la proposta di Ascoli, secondo cui "[l]'insegnamento grammaticale può essere lasciato alle classi per cui i programmi attualmente lo prescrivono, ma purché sia fatto proporzionalmente al dialetto locale" (Ascoli 1875, cit. in Demuru 2018: 340).

3. Un'analisi metacognitiva

L'analisi logica è lo studio delle inferenze logiche che hanno luogo quando più parole si combinano formando una frase. L'analisi logica è quindi un esercizio di metacognizione attraverso il quale i parlanti ricostruiscono alcuni dei processi, largamente impliciti, che ci consentono di dare un significato ad una frase. L'interpretazione di una frase non è la somma dei significati delle singole parole, ma è il risultato di una serie complessa di operazioni astratte che, a partire da una stringa di suoni (o caratteri), ricostruisce una *Forma Logica*, ovvero un insieme di condizioni che ci consentono di discriminare in quali contesti la frase è vera o falsa.

Tali inferenze logiche dipendono strettamente dalla costituenza sintattica: ad esempio, le condizioni di verità variano enormemente se, a parità di elementi, l'ordine dei costituenti è invertito senza alcun processo di dislocazione o focalizzazione:

- (1) a. *Un cane ha morso un bambino.*
b. *Un bambino ha morso un cane.*

Uno dei primi compiti dell'analisi logica è mostrare perché i due sintagmi nominali (SN) *un cane* e *un bambino* in (1), pur essendo entrambi argomenti dello stesso verbo, hanno interpretazioni diverse – indipendentemente dalle proprietà lessicali dell'individuo/entità cui il SN si riferisce – e soggiacciono a regole sintattiche diverse. Si dovrebbe quindi porre particolare attenzione alla correlazione sistematica fra aspetti interpretativi (chi è l'agente, chi è il paziente) e proprietà grammaticali: quale costituente controlla l'accordo del verbo flesso, quale può essere passivizzato come in (2), quale può essere sostituito da un pronome clitico come in (3), ecc.:

- (2) a. *Un bambino è stato morso da un cane.* \approx (1a), \neq (1b)
b. *Un cane è stato morso da un bambino.* \approx (1b), \neq (1a)
- (3) a. *Un cane l'ha morso.* \approx (1a), \neq (1b)
b. *Un bambino l'ha morso.* \approx (1b), \neq (1a)

Osservare l'interpretazione dei costituenti senza tenere in considerazione le proprietà morfosintattiche (o viceversa) ci porta ad effettuare un ragionamento circolare, che non ci dice molto sulla struttura logico-sintattica della frase.

Per mostrare quanto forti siano le inferenze logiche attivate dalla struttura sintattica potrebbe essere utile anche lavorare su frasi prive di senso o su frasi costruite con pseudoparole, come (4a) e (4b):

- (4) a. *Un armadio ha inondato un rospo.*
b. *Un artiso ha perato un vammo.*

Pur non riuscendo a comprendere il senso di (4a) e ancor meno quello di (4b), tutti noi (ed i nostri studenti) siamo in grado di comprendere che qualcuno o qualcosa (*un armadio* o *un artiso*) ha compiuto un'azione (*inondare* o *perare*) che ha avuto delle conseguenze su un'altra entità (*un rospo* o *un vammo*). Allo stesso modo, sappiamo anche che i costituenti *un rospo* e *un vammo* possono essere passivizzati o cliticizzati come in (2) e

(3). Esercizi di questo tipo dovrebbero aiutare lo studente ad osservare la struttura sintattica facendo astrazione dalle parole che la compongono, arrivando quindi a cogliere come tutte le frasi transitive in (1) e (4) abbiano la medesima architettura, più o meno quella in (5):⁵

(5) [S [V O]]

Lavorando su frasi, comparandole, manipolandole, sostituendone dei pezzi, trasformandole, usando delle pseudoparole, l'analisi logica dovrebbe quindi consentire di isolare i passaggi fondamentali attraverso cui si costruisce la Forma Logica (significato) della frase. Individuare gli elementi nucleari della frase è un passaggio-chiave di un procedimento affascinante – ed in parte ancora misterioso – che ci consente di rappresentare nella nostra mente eventi, azioni, stati a partire da una sequenza di suoni. In quest'ottica, credo sia possibile cercare di dare una luce diversa ai vari compiti di identificazione del soggetto, che sono molto comuni anche nelle prove INVALSI come (6):

(6) es. C8, a.s. 2014-5:

Nelle seguenti frasi individua il soggetto e scrivilo nella riga sotto ciascuna frase.

- 1) Oggi finalmente è venuto fuori il perché del tuo nervosismo.
- 2) Il caffè oggi te lo offro io.
- 3) La settimana scorsa è arrivata la nuova professoressa di italiano.
- 4) Ogni giorno sui giornali si trovano le previsioni del tempo.

Un esercizio come (6), ad esempio, potrebbe essere riproposto in classe come se fosse un esercizio di *coding*, immaginando che l'obiettivo non sia individuare il soggetto (cui prodest? – potrebbe obiettare qualche studente particolarmente intraprendente), ma progettare un algoritmo che individui la struttura della frase e la traduca in una lingua straniera che presenta una sintassi completamente diversa dall'italiano. Per fare questo, il nostro algoritmo deve prima di tutto osservare alcune proprietà sintattiche, ad esempio:

1. Individuare i costituenti nucleari;
2. Assegnare ad ogni costituente nucleare un ruolo/attante (es. agente, paziente, esperiente, strumento, luogo, ecc.);
3. Individuare i costituenti nucleari che non sono introdotti da preposizioni;
4. Osservare il primo costituente nucleare privo di preposizione:
 - a. Controlla l'accordo del verbo flesso? Sì → è il soggetto; No → proseguire
 - b. Controllare se il costituente può essere passivizzato o sostituito da un pronome clitico come *lo, la, li, le*? Sì → è l'oggetto.
5. Procedere con il secondo costituente nucleare, ecc.

La struttura sintattica non gioca un ruolo evidente solamente nel caso dei costituenti nucleari (che sono il principale oggetto di analisi dei modelli *valenziali*), ma anche nella sintassi e nell'interpretazione degli elementi extranucleari, come illustrato dai ben noti casi di ambiguità sintattica come (7) e (8):

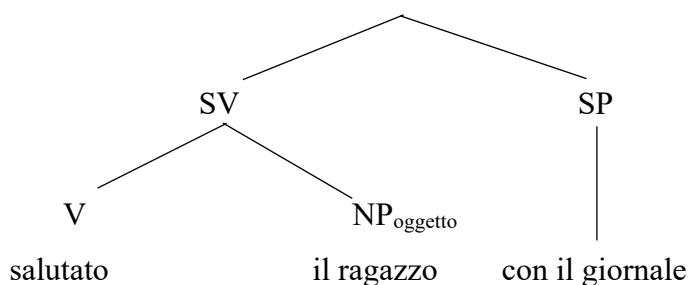
⁵ La struttura in (5) – od ogni sua possibile variante notazionale – indica che le funzioni di S(oggetto) e O(gggetto) non sono simmetriche e, quindi, non si possa modificare l'ordine di S e O (al netto di dislocazioni e/o focalizzazioni) senza modificare il senso della frase. Vari test di costituenza sintattica ci portano a pensare che, nella struttura della frase, S occupi una posizione strutturale più “esterna” rispetto al nucleo [V O], ma non è questa la sede per approfondire la questione.

- (7) a. *[[ho salutato il ragazzo] con il giornale]*
 ‘Ho salutato il ragazzo muovendo il giornale’
 b. *[ho salutato [il ragazzo con il giornale]]*
 ‘Ho salutato il ragazzo che aveva il giornale’
- (8) a. *[[mi sono sposato e abito] a Verona]*
 ‘Mi sono sposato a Verona e vivo nella stessa città’
 b. *[mi sono sposato] e [abito a Verona]*
 ‘Dopo essermi sposato, sono andato a vivere a Verona’

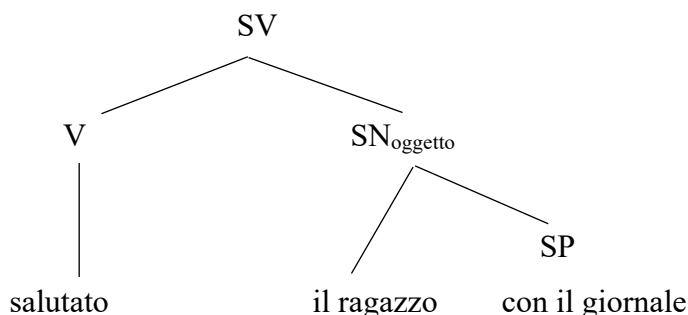
Casi come (7) e (8) illustrano come la nostra mente operi una segmentazione e una gerarchizzazione dei costituenti sintattici tale per cui, a parità di sequenza lineare, la frase in (7) e (8) ottengono due interpretazioni possibili (e, quindi, due intonazioni differenti). In (7a) e (8a) il sintagma preposizionale (SP) ha portata ampia (esso cioè modifica tutta la frase), mentre in (7b) e (8b) l’SP modifica solamente il costituente limitrofo.

Esempi come (7) e (8) sono forse la dimostrazione più chiara del significato del termine *analisi logica*. La struttura logica di queste frasi cambia infatti se la modificazione apportata dal SP extranucleare si applica all’intera struttura sintattica o ad un livello gerarchico inferiore. In questo senso, credo che imparare a rappresentare questo tipo di fenomeni mediante diagrammi ad albero come quelli in (9) offra agli studenti uno strumento importantissimo per visualizzare questo tipo di relazioni logiche:

- (9) a.



- b.



La pratica di rappresentare le frasi mediante diagrammi ad albero è molto utile per mettere alla prova le capacità analitiche degli studenti (accettando comunque una certa flessibilità nel modo di rappresentare i costituenti sintattici, che deve essere commisurato al livello

raggiunto dagli studenti). Lo scopo di un diagramma come (9), o di una sua possibile variante più o meno raffinata, è quello di mettere nero su bianco le relazioni gerarchiche fra costituenti, dando quindi una veste grafica all'ordine dei processi logico-interpretativi.

Questo tipo di esercizi orienta verso un tipo di analisi metacognitiva che si concentra sugli aspetti strutturali della lingua e su come questi diano forma al processo logico attraverso cui i parlanti attribuiscono un senso a espressioni complesse. Lavorando in questo modo, lo studente dovrebbe infine apprendere a ragionare in modo più astratto, comparando frasi che sono apparentemente diverse, ma strutturalmente identiche (come (1) e (4)) o, viceversa, comparando frasi che sono apparentemente identiche, ma strutturalmente diverse (come (7) e (8)). Questo tipo di esercizi dovrebbe quindi abituare gli studenti a vedere le frasi come oggetti a più strati, in cui la dimensione orizzontale – la sequenza di parole, ognuna con il proprio significato – si combina con la dimensione verticale/gerarchica che stabilisce l'ordine con cui tali significati interagiscono per dare luogo al senso complessivo della frase.

4. Un'analisi indipendente

L'analisi logica ha come oggetto di studio una minima sottoparte delle inferenze logiche necessarie alla costruzione del significato delle espressioni linguistiche. Molto di quello che capiamo di una frase non dipende infatti dalla nostra capacità di computare la struttura sintattica, ma da ragionamenti deduttivi o induttivi che sono legati a conoscenze *enciclopediche* più ampie o da principi di comunicazione più generali. Basta prendere come esempio una semplice frase interrogativa come *sai l'ora?* per rendersi conto che l'analisi logica rende conto di una piccola porzione del senso di una frase – una porzione significativa, ma non sempre sufficiente.

Il limite fra inferenze linguistiche ed inferenze enciclopediche è talvolta labile⁶ o difficile da definire, ma credo sia importante sottolineare questo punto perché, specialmente nel quadro delle verifiche, sono spesso assegnati compiti in cui la soluzione può essere data esclusivamente sulla base di conoscenze extralinguistiche. In pratica, in un test che misura la capacità di analizzare la grammatica si finisce con il misurare un altro tipo di abilità che con la grammatica (e talvolta con il linguaggio tout court) ha poco o nulla a che vedere. Per illustrare questo punto, prendiamo ad esempio l'interpretazione del pronome *lo* in una frase come (10):

(10) *Gianni stava giocando con il cane e per sbaglio l'ha morso.*

La teoria sintattica ci insegna che *lo* si comporta come un pronome *libero*, ovvero come un pronome che può prendere come antecedente qualsiasi referente attivo nel contesto discorsivo. Poiché (10) non ha un vero contesto, è lecito assumere che *lo* avrà solo due antecedenti possibili: *Gianni* o *il cane*. In italiano, però, il soggetto sottinteso della frase sceglie di solito come antecedente il topic che lo precede (Calabrese 1986), lasciando quindi come referente più probabile del clitico solo *il cane*. Eppure, l'interpretazione secondo cui *lo* = *Gianni* non sarà esclusa del tutto, in quanto si tratta di una lettura più plausibile dell'altra: la nostra conoscenza del mondo ci dice che è più probabile che un

⁶ Come notato da un revisore anonimo, il confine è particolarmente sottile nel caso di inferenze legate alla pragmatica (implicature).

cane morda un uomo che il contrario. Nell'analisi logica, tuttavia, dobbiamo abituarci (e dobbiamo abituare i nostri studenti) a fare astrazione da tali inferenze enciclopediche. Ad esempio, possiamo sostituire le parole in (10) con pseudoparole, come in (11), in modo che le nostre conoscenze enciclopediche diventino inutili ed i legami di coreferenza che si basano unicamente su fattori sintattici appaiano più chiari:

(11) *Il pocco stava giovendo con il marro e per sbaglio l'ha rudato.*

Come detto in precedenza, sarebbe opportuno riflettere maggiormente su esempi in cui la riflessione sulle strutture della lingua sia chiaramente scindibile da altre forme di ragionamento logico, come nel caso in (10), in cui sintassi e conoscenze del mondo portano attivate letture opposte. In questo modo, sarà possibile indicare agli studenti quali indizi linguistici seguire nella ricostruzione della struttura logico-grammaticale della frase e quali invece ignorare.

A questo proposito, vorrei fare qualche considerazione in merito ad alcuni quesiti INVALSI. Nelle attuali prove, la quantità di test inerenti la riflessione sulla grammatica è inferiore – giustamente, beninteso – rispetto allo spazio dedicato alla comprensione del testo. La comprensione del testo è il frutto di inferenze logiche di varia natura, mentre – idealmente – l'analisi logica dovrebbe essere il luogo in cui lo studente ha la possibilità di concentrarsi su un sottoinsieme di inferenze linguistiche. Come già accennato commentando (10)-(11), scindere questi due aspetti non è né intuitivo, né immediato.

Si prenda come illustrazione l'esercizio 42 della simulazione INVALSI 2017-8, riprodotto in (12):

(12) es. 42, a.s. 2017-8

Nella frase “Secondo una prima ricostruzione, il ladro sarebbe entrato da una finestra sul retro”, il verbo indica

- un'azione certa
- un'azione futura
- un'azione improbabile
- un'azione possibile

In tale esercizio si chiede di indicare quali inferenze logiche possiamo trarre a partire dal modo e dal tempo verbale. Si tratta di un esercizio di per sé utile, che consente allo studente di ricostruire i valori di verità dell'enunciato a partire dall'analisi grammaticale della forma verbale. Peccato però che, nella frase in (12), tale sforzo sia probabilmente annullato dalla presenza di un'informazione ridondante (“secondo una prima ricostruzione”), che di fatto restringe a prescindere il tipo di risposta possibile. In altre parole, penso che molti studenti abbiano potuto indicare la risposta (d) anche se il verbo fosse stato all'indicativo presente (“secondo una prima ricostruzione, il ladro è entrato da una finestra sul retro”).

Concludendo, l'autonomia della sintassi rispetto ad altre componenti cognitive è oggetto di dibattito fra gli studiosi da più di mezzo secolo. Non spetta a noi, in questa sede, intervenire o approfondire la questione, ma, da un punto di vista pedagogico, credo sia opportuno cercare di individuare per quanto possibile delle abilità specifiche, che vanno coltivate e monitorate per verificarne lo sviluppo. In quest'ottica, credo che gli insegnanti debbano porre particolare attenzione nel canalizzare l'attenzione degli studenti verso gli aspetti strutturali delle frasi, escludendo per quanto possibile l'interferenza di

altri indizi linguistici o extralinguistici che potrebbero aiutare (o in alcuni casi sviare) nella ricostruzione della forma logica delle frasi. Chiaramente, nella vita di tutti i giorni tutti gli aspetti linguistici ed extralinguistici sono fittamente intrecciati. Tuttavia, ridurre i fenomeni complessi in parti elementari credo sia un procedimento didattico assai comune, che la riflessione sulla grammatica dovrebbe adottare in maniera sempre più rigorosa.

5. Un'analisi (progressivamente) scientifica

Uno dei fondamenti della grammatica di Port-Royal, a cui ho brevemente accennato nel §2, è che l'analisi linguistica debba essere *ragionata*. Creare un impianto rigoroso e fornire delle regole chiare, facilmente applicabili, sono aspetti in parte coincidenti, in parte inconciliabili. O meglio: i due aspetti della grammatica – la teoria e la pratica – si combineranno in un quadro unitario nel corso del tempo, ma nell'insegnare i rudimenti dell'analisi sintattica sarà necessario procedere per approssimazioni successive, sacrificando in principio la coerenza teorica per favorire l'apprendimento delle nozioni più elementari. L'importante, però, è che nella mente del docente sia chiaro sin da subito il percorso da intraprendere.

Idealmente, quello che si vorrebbe insegnare attraverso l'analisi logica è l'applicazione di un metodo scientifico all'analisi delle frasi. Si tratta quindi di invitare gli studenti a (1) fare delle osservazioni generali e (2) controllare la veridicità delle proprie osservazioni mediante il confronto o la manipolazione di esempi per tornare, eventualmente, al punto (1) e riformulare le proprie osservazioni in modo più raffinato. Esiste un'ampia letteratura che illustra tale procedimento, ad esempio in merito all'analisi dei soggetti. Si parte da una prima definizione di tipo *semantico-nozionale* (Lo Duca 2010), che è quella solitamente proposta dai manuali (il soggetto è “colui che compie l'azione”). Come detto in precedenza, se tale tipo di definizione può essere accettata agli esordi della riflessione grammaticale, è inaccettabile che studenti del grado 8 arrivino ad affrontare le prove INVALSI senza che quell'apparato definitorio sia stato riformulato di pari passo con lo sviluppo intellettuale degli studenti. In questo senso, mi sembra molto sensato l'esercizio 39 nella prova INVALSI 2018-9, in cui, attraverso una risposta a scelta multipla, si richiede di individuare le proprietà di un soggetto non agentivo:

(13) es. 39, a.s. 2018-9

Leggi la seguente frase.

“Maria ha ricevuto assieme a me il premio dell'amicizia.”

Sono una delle seguenti affermazioni è vera. Quale?

- il soggetto si trova alla fine della frase
- il soggetto non compie l'azione
- il soggetto è sottinteso
- il soggetto è un pronome personale

Anziché proporre il classico esercizio in cui si chiede di individuare il soggetto di una lista di frasi (come richiesto in (6)), un esercizio come quello in (13) costringe gli studenti a concentrare la propria attenzione sulle proprietà. Combinando più esercizi di questo tipo, proprietà diverse (e applicandole alle diverse classi di verbi e a diversi tipi di frasi)

gli studenti dovrebbero essere in grado di costruire una matrice di proprietà, verificando ad esempio quali sono quelle più affidabili (l'accordo, tipicamente).

Insomma, non dovrebbero esistere, nella scuola secondaria, definizioni che non siano state accuratamente verificate mediante prove empiriche (un po' come in geometria non dovrebbero esistere teoremi lasciati indimostrati). In questa prospettiva, trovo che una delle più grandi lacune nell'insegnamento dell'analisi logica riguardi le cosiddette *prove di costituenza*, che sono invece essenziali per almeno due motivi: in primo luogo, le prove di costituenza consentono di portare alla luce i confini dei sintagmi e di individuare le loro relazioni gerarchiche, cfr. (9); in secondo luogo, le prove di costituenza aiutano gli studenti ad apprendere come *manipolare* le frasi per ottenere coppie minime. Riprendiamo ad esempio l'ambiguità sintattica delle frasi in (7) e (8). L'idea, di per sé intuitiva, che a diverse strutture di costituenti corrispondano diverse forme logiche può essere dimostrata attraverso un semplice test di sostituzione come in (14) o provando a trasformare la frase, ad esempio utilizzando una struttura scissa come in (15). La sostituzione in (14) mostra come solamente nel secondo caso il sintagma preposizionale sia un modificatore dell'oggetto, mentre (15) mostra che due frasi coordinate possano co-occorrere nella medesima frase scissa solo entrambe sono modificate dal costituente estratto (nel nostro caso il SP *a Verona*):

- (14) a. *Chi [[hai salutato ~~il ragazzo~~] con il giornale]] ?*
 b. *Chi [hai salutato [~~il ragazzo con il giornale~~]] ?*
- (15) a. *[[mi sono sposato e abito] a Verona]*
 = *è a Verona che [[mi sono sposato e abito] ~~a Verona~~]]*
 b. *[mi sono sposato] e [abito a Verona]*
 ≠ *è a Verona che [mi sono sposato] e [abito ~~a Verona~~]*
 = *[mi sono sposato] ed è a Verona che [abito ~~a Verona~~]*

Piccoli esercizi come quelli in (14) e (15), se dovutamente spiegati e illustrati, dovrebbero portare gli studenti ad individuare le relazioni strutturali fra costituenti e, da lì, a ricostruire quali sono i rapporti logici fra i significati di tali costituenti, da rappresentare infine attraverso schemi ad albero come quelli in (9).

Un altro esercizio interessante per portare alla luce le proprietà 'nascoste' delle strutture sintattiche seguendo il metodo scientifico è ragionare sugli aspetti distribuzionali, ovvero su come le varie categorie della frase si combinano fra loro. Di solito, l'analisi logica poggia sull'analisi grammaticale, intesa come abilità nel riconoscere le parti del discorso (N, V, P, ecc.). Anche per questo aspetto, le nozioni operative su cui gli studenti basano le proprie analisi sono soprattutto di tipo semantico-nozionale (V denota un'azione, N un individuo o una cosa, ecc.). Si presta invece poca attenzione alla definizione delle categorie sulla base delle relazioni sintagmatiche con gli altri elementi della frase (ad esempio, sul fatto che i nomi sono spesso introdotti da un determinante, che i verbi possono ospitare dei pronomi clitici, ecc.). Questo tipo di osservazioni si prestano ad essere verificate (o falsificate), discusse, riformulate, aggiustate secondo quel meccanismo circolare, cui si accennava all'inizio del paragrafo, che conduce dalla teoria alla pratica e viceversa. Quello è il tipo di meccanismo che occorre rafforzare.

Riporto qui un esempio di discussione particolarmente complesso, che mi è stato suggerito dall'osservazione del quesito 44 dell'esempio di prova 2017-8. L'esercizio, che

qui non riporterò nella sua interezza, riguarda l'identificazione di strutture passive. La mia attenzione, tuttavia, è stata attirata da un altro aspetto, che non è direttamente rilevante nella prova, ma che, volendo, potrebbe essere oggetto di approfondimento, discussione e confronto in classe. Mi riferisco alla distinzione fra il participio passato di una costruzione passiva (per es. *la porta viene chiusa alle 9*) e l'aggettivo di una costruzione predicativa (per es. *la porta è chiusa dalle 9*). Uno dei criteri per distinguere aggettivi e participi è di solito la presenza di un complemento di agente, però talvolta l'agente non è espresso oppure è molto difficile distinguere l'agente da altri tipi di complemento che indicano una causa/mezzo (per es. *la porta è chiusa dal chiavistello*). Le specificazioni di tempo ci danno ulteriori indicazioni utili per discriminare fra le due interpretazioni, visto che i passivi solitamente denotano azioni dotate di una propria struttura temporale, mentre le costruzioni predicative sono stative. Tuttavia, con verbi che denotano azioni durative, la distinzione non è affatto chiara. Ci sono però altri test a nostra disposizione, che ora proverò ad applicare a due delle frasi presenti nella batteria di esempi del quesito 44 (per questioni di spazio, mi concentro solamente sulle frasi (a) e (d) e tralascio gli altri due esempi riportati nell'esercizio, che non sono qui pertinenti):

(16) es. 44, a.s. 2017-8

- a) I nuovi arrivati furono ospitati nella casa di un amico
- d) Nelle ore di punta le vie della città sono intasate dal traffico

La domanda che io mi porrei, confrontando (a) e (d), è se entrambe queste frasi siano passive o se una delle due sia una frase predicativa che contiene un aggettivo predicativo. Proviamo allora ad applicare qualche ulteriore test per cercare di capire se *ospitati* in (a) e *intasate* in (d) siano entrambi participi o no.

In primo luogo, sappiamo che il passivo si può costruire anche con l'ausiliare *venire*, mentre una costruzione predicativa richiede sempre un verbo di tipo copulativo:

- (17) a. *I nuovi arrivati furono/vennero ospitati nella casa di un amico.*
- d. *Nelle ore di punta le vie della città sono/^ovengono intasate dal traffico.*

Ragionando sul significato di (17d), credo che gli studenti possano avvertire una forzatura nel significato quando si impiega l'ausiliare *venire*. Non che la frase divenga agrammaticale (nel qual caso l'avrei segnata con un asterisco), ma è chiaro che l'ausiliare *venire* dona alla frase una sfumatura agentiva che è invece assente nella frase con *essere*. Ciò si collega con il fatto che le strutture passive in genere implicano un agente (magari implicito) e che quindi possono essere modificate da avverbi che presuppongono la presenza di un'agente. Come sopra, la presenza di tali avverbi in un contesto come (18d) appare invece piuttosto strana:

- (18) a. *I nuovi arrivati furono ospitati volutamente/di proposito nella casa di un amico.*
- d. *Nelle ore di punta le vie della città sono intasate ^ovolutamente/^odi proposito dal traffico.*

Inoltre, sappiamo che gli aggettivi e participi possono combinarsi con modificatori di tipo diverso (degli aggettivi, inoltre, possono modificare il grado):

- (19) a. **I nuovi arrivati furono molto ospitati/ospitatissimi nella casa di un amico.*
 d. *Nelle ore di punta le vie della città sono molto intasate/intasatissime dal traffico.*

Un'ultima proprietà che distingue i participi dagli aggettivi è che i primi, essendo verbi, si possono combinare con i pronomi clitici, mentre gli aggettivi non lo possono fare. Quindi, se si provasse a manipolare le due frasi in questione in modo da ottenere due subordinate implicite, ci si attende che la pronominalizzazione abbia luogo solamente nel contesto in cui il predicato corrisponde ad un participio passato:

- (20) a. *Ospitati i nuovi arrivati, il mio amico si dileguò nel nulla.*
 → ospitati-*li*, il mio amico si dileguò nel nulla.
 d. *Intasate le strade della città, il traffico si spostò sulla tangenziale.*
 → ^{??}intasate-*le*, il traffico si spostò sulla tangenziale.

Le prove in (17)-(20) mostrano che la frase (d) tenda ad avere un'interpretazione stativa, piuttosto che passiva. Questo piccolo esempio di analisi mostra come, attraverso il confronto e l'elaborazione, sia possibile pervenire ad una comprensione sempre più sottile e raffinata delle strutture linguistiche. Probabilmente il tipo di raffronto in (16)-(20) è troppo complesso per essere sottoposto all'analisi di studenti del grado 8, tuttavia ritengo che gli esercizi del grado 8 siano un buon punto di partenza per gli insegnanti del secondo ciclo. Inoltre, quanto discusso in (16)-(20) intende soprattutto illustrare un metodo per costruire un'argomentazione razionale che, attraverso la comparazione di coppie di esempi, discuta la natura di due categorie linguistiche: participi vs aggettivi.

Si obietterà che, oltre un certo livello di dettaglio, l'analisi logica può diventare un esercizio barocco, ma, in verità c'è un ulteriore fattore – il quarto – che merita di essere preso in considerazione: non è detto che ciò che può sembrare irrilevante o eccessivamente minuzioso per l'italiano lo sia anche per le altre lingue. In un'ottica di apprendimento multilingue, quindi, può valere la pena spingere verso l'analisi più approfondita di alcuni fenomeni linguistici laddove questo determini un effettivo vantaggio nell'acquisizione delle lingue straniere.

6. Un'analisi generalizzabile

Lo scopo di questo paragrafo è riflettere attorno al concetto di *generalizzabilità*. Le idee, le definizioni, i metodi, le prove impiegate nell'ambito dell'analisi logica non dovrebbero avere valore solamente per l'italiano, ma dovrebbero essere facilmente trasponibili nello studio e nell'analisi delle altre lingue, sia quelle classiche che quelle moderne.

Anche in questo caso, si tratta di un principio nato attorno alla grammatica di Port-Royal e poi sviluppato dalle grammatiche del periodo illuminista, come testimoniato dal seguente passo tratto dalla *Grammaire générale, ou exposition raisonnée des éléments nécessaires du langage* di N. Beauzée (1767), che trovo particolarmente illuminante:

La grammatica generale è la scienza ragionata dei principi immutabili e generali del linguaggio parlato o scritto, in qualunque lingua. Una grammatica particolare è l'arte di applicare ai principi immutabili e generali del linguaggio parlato o scritto le istituzioni arbitrarie e usuali di una lingua particolare. [...] La scienza grammaticale

è anteriore a tutte le lingue, giacché i suoi principi non suppongono che la possibilità delle lingue e sono gli stessi di quelli che dirigono la ragione umana nelle sue operazioni intellettuali; in una parola, sono veri di una verità eterna. (cit. in Verga 1970: 8)

Questo aspetto della generalizzabilità dei principi dell'analisi logica lo si trova espresso, con parole più o meno simili, anche nelle grammatiche moderne, per esempio nella già citata (e criticata) *Analisi logica* di Tantucci (p. 6):

Si è detto in precedenza che i vari popoli parlano lingue diverse e che la grammatica, e quindi l'analisi grammaticale di ogni singola lingua, è lo strumento indispensabile per apprenderle. Gli uomini però, al di là delle differenze di evoluzione storica, di cultura o di razza [*sic*], pensano ovunque in modo analogo; se possedessero la facoltà di comunicare per telepatia, un cinese e un italiano potrebbero ad esempio comprendersi senza eccessiva difficoltà. Da ciò possiamo capire che anche l'analisi dell'organizzazione del pensiero è fondamentalmente la stessa, indipendentemente dalla lingua alla cui sintassi essa è applicata. Lo studio dell'analisi logica riveste pertanto un'importanza fondamentale per l'apprendimento del latino come del tedesco e di qualsiasi altra lingua, antica o moderna che sia.

Quindi, anche se le vicende storiche del sistema educativo italiano hanno fatto sì che l'analisi logica entrasse nel dominio dell'insegnamento etichettato come "italiano", l'analisi logica dovrebbe essere uno strumento trasversale all'acquisizione e allo studio di tutte le lingue.

Questo tema della generalizzabilità dell'analisi è molto delicato, anche perché tocca aspetti di organizzazione della didattica sui quali non interverrò, mancandomi lo spazio, ma soprattutto la competenza. L'unico aspetto cui mi premeva accennare qui è che un'analisi logica approfondita della lingua prima è il fulcro per l'insegnamento delle altre lingue, ammesso che gli insegnanti condividano un bagaglio disciplinare condiviso, di cui la sintassi dovrebbe essere parte.

In un'ottica contrastiva – l'ho già accennato – aspetti grammaticali apparentemente minuziosi possono rivelarsi utili nella traduzione da una lingua all'altra. Per esemplificare, torno per un momento alla distinzione sottile fra participi e aggettivi discussa alla fine del §5: nel caso della frase *le vie della città sono intasate dal traffico*, abbiamo visto che in italiano ci sono indizi, forse non conclusivi, per sostenere che *intasate* non sia un participio, ma piuttosto un aggettivo e che, di conseguenza, *dal traffico* non sia un complemento d'agente. Questo tipo di indicazioni, sicuramente di dettaglio, si rivelano però centrali nella traduzione verso la lingua straniera, che, proprio come l'italiano, negli stessi contesti sembra adottare la stessa struttura predicativa dell'italiano, per es. *jammed with/*by traffic*.

7. Conclusioni

In questa nota ho cercato di mettere in luce quattro principî-cardine che dovrebbero guidare gli insegnanti nel progettare le proprie attività didattiche relative alla cosiddetta *analisi logica*. Questi quattro principî – di cui non ripeto la definizione, si veda il §2 – sono la centralità degli aspetti metacognitivi, la specificità e, quindi, l'indipendenza dell'analisi sintattica rispetto ad altri processi di comprensione, la scientificità del

metodo, che si deve basare sulla verifica continua e replicabile di concetti chiari e non ambigui, e la generalizzabilità dell'analisi ad altre lingue e contesti.

Seguendo questi principi, spero che gli insegnanti possano orientare le proprie attività senza dover ricorrere a strumenti didattici spesso effimeri, ma costruire dei percorsi ragionati a partire dalla riflessione critica sugli strumenti già a disposizione. In questo senso, una rilettura critica delle prove INVALSI è un ottimo punto di partenza su cui costruire un modello didattico solido.

Bibliografia

- Ascoli, Graziadio Isaia. 1875. S.t. In *Atti del IX Congresso Pedagogico Italiano e della V esposizione scolastica*. Bologna: Regia Tipografica.
- Calabrese, Andrea. 1986. Some properties of the Italian pronominal system: An analysis based on the notion of thema as subject of predication. In Stammerjohann, Harro (a cura di), *Tema-Rema in Italiano*, 25–36. Tübingen: Gunter Narr Verlag.
- Chevalier, Jean-Claude. 1979. Analyse grammaticale et analyse logique : esquisse de la naissance d'un dispositif scolaire. *Langue Française* 41. 20–34.
- De Roberto, Elisa. 2018. La frase semplice. In Antonelli, Giuseppe & Motolese, Matteo & Tomasin, Lorenzo (a cura di), *Storia dell'italiano scritto. IV. Grammatiche*, 357–399. Roma: Carocci.
- Demuru, Cecilia. 2018. «Un guazzabuglio di cose diverse». La classificazione dei complementi nelle grammatiche postunitarie. *Italiano Lingua Due* 1. 337–354.
- Lo Duca, Maria Giuseppa. 2010. Analisi Logica. *Enciclopedia dell'Italiano* (https://www.treccani.it/enciclopedia/analisi-logica_%28Enciclopedia-dell%27Italiano%29/) (Consultato il 13.10.2021.)
- Tantucci, Vittorio. 1955. *Analisi Logica*. Bologna: Poseidonia.
- Verga, Leonardo. 1970. La teoria del linguaggio di Port-Royal. *Rivista di Filosofia Neo-Scolastica* 62(1/2). 1–100.

Heri dicebamus. Competenze linguistiche prima e dopo la pandemia

Alberto A. Sobrero

Università del Salento

alberto.sobrero@gmail.com

Abstract

Si analizzano i dati ottenuti dall'INVALSI nei rilevamenti censuari degli ultimi anni relativamente a quattro parametri: divario fra macroregioni settentrionali, centrali e meridionali, livello socioeconomico e culturale delle famiglie, immigrati e nativi, variabilità fra scuole e classi. Le prospettive di analisi sono due: identificazione delle linee di tendenza che caratterizzano le variazioni delle competenze in lingua italiana relativamente a questi fattori (confronto fra i dati del 2018 e quelli del 2019): analisi dei dati e discussione sulla direzione, sul senso e sulle motivazioni dei cambiamenti – quasi sempre significativamente negativi – che il lockdown e l'esperienza DAD ha apportato alle competenze linguistiche negli ultimi due anni scolastici (confronto fra i dati del 2019 e quelli del 2021). Si coglieranno, in conclusione, spunti per un ri-orientamento della didattica, che tragga spunto dalle enormi difficoltà e dai risultati negativi di questi due anni, nella prospettiva dell'adeguamento di strumenti e metodi alle nuove esigenze didattiche e formative.

1. Introduzione

Il coronavirus ha rivoluzionato le nostre abitudini, ha cancellato eventi di tradizione secolare. Chissà se quest'anno ci risparmierà anche una radicata tradizione, quella delle polemiche che da ormai un decennio accompagnano puntualmente le prove INVALSI della tarda primavera, su quotidiani e riviste più o meno mainstream. Indimenticabile la campagna orchestrata due anni fa: una mitragliata di oltre 200 articoli, fra luglio e settembre, affidati a esperti (filologi bizantinisti, scrittori, psicanalisti ecc.) tutti accomunati dalla stessa caratteristica: non avere conoscenze serie né di scuola né di indagini statistiche. Gli articoli seri e documentati non furono neppure una dozzina. Venne ben presto allo scoperto l'obiettivo strategico reale di questa massiccia campagna: utilizzare il pretesto INVALSI per una battaglia tutta politica, che aveva messo nel mirino non le prove standardizzate ma gli orientamenti innovativi – egualitari e democratici – nel campo della pedagogia e della didattica, per affermare antiche e sempre ricorrenti istanze restauratrici.

La strumentalizzazione dell'INVALSI non era casuale: l'attività dell'Istituto sin dai primi anni di attività è stata oggetto di critiche anche sostanziali, che hanno dato luogo a manifestazioni organizzate anche da forze politiche e sindacali di primo piano, a scioperi, a contestazioni. Contro questo bersaglio era facile trovare alleanze anche corpose. E così si cercò di fare. E si fece.

Lasciando da parte gli esiti di quella campagna, qui preme sottolineare che in quella occasione, come in tante altre, dell'utilità, delle funzioni e del funzionamento, dei pregi reali e dei difetti reali dei rilevamenti INVALSI poco o nulla è stato detto.

Nonostante il gran parlare, c'è ancora una domanda in buona parte inevasa, almeno per la maggior parte degli insegnanti, ed è invece una domanda fondamentale: a che cosa servono i dati INVALSI? Per quali fini possono essere utilizzati i copiosissimi dati che ogni anno sono raccolti in tutte le scuole, presentati e commentati?

Le possibili risposte a questa domanda si muovono su due direttrici: dare consapevolezza dei traguardi effettivamente raggiunti dagli allievi nelle diverse tappe del loro percorso educativo, misurando le loro prestazioni con strumenti il più possibile vicini alla precisione e all'oggettività; fornire agli insegnanti strumenti utili e affidabili, in quanto tarati sulle Indicazioni Nazionali e sulle Linee Guida ministeriali, sul Common European Framework of Reference e sul Quadro di Riferimento INVALSI. Per la prima di queste due direttrici è sufficiente – almeno in prima istanza – la lettura attenta dei dati e dei commenti forniti a conclusione dei rilevamenti, nei Rapporti pubblicati ogni anno sul sito INVALSI.

2. Livelli di apprendimento e variabili di contesto

Una delle obiezioni ricorrenti è quella dell'inutilità: tanto spreco di risorse e di energie, tanto tempo rubato al lavoro in classe per la preparazione all'INVALSI, per ottenere che cosa? La conferma, di anno in anno, che va meglio a scuola chi ha alle spalle una famiglia di buon livello economico e culturale, che le scuole del Nord ottengono risultati migliori di quelle del Sud e delle isole, che i figli degli immigrati hanno gravi problemi con la lingua italiana? Ma lo sappiamo già, è inutile rilevarlo e confermarlo tutti gli anni.

È vero. I report annuali confermano e quantificano, anno per anno, caratteri che sembrano costitutivi della nostra popolazione scolastica. Ma dicono anche qualcosa di più, qualcosa di cui non si parla mai abbastanza. Prendiamo i risultati delle somministrazioni della primavera 2019, e leggiamo con attenzione il relativo rapporto. Troviamo quattro parametri importanti, i cui dati sono rilevati sistematicamente, anno dopo anno.

2.1 Divario tra i risultati delle femmine e dei maschi

Nelle prove di italiano risultano costantemente avvantaggiate le femmine rispetto ai maschi: nello stesso 2019 le percentuali di ragazzi che non raggiungono il livello 3 superano quelle delle ragazze del 3% in seconda elementare, ma poi il dislivello aumenta: arriva al 9% al termine della scuola primaria e ancora al 9% all'altezza della licenza di scuola media. La differenza è ancora significativa nelle superiori, in particolare nei licei e negli istituti professionali.

2.2 Divario fra macro regioni settentrionali, centrali e meridionali

Si veda la tabella 1, che riporta le percentuali degli studenti che nel 2019 non raggiungono il livello 3 (su 5):¹

Grado ²	Nord-ovest	Nord-est	Centro	Sud	Sud e isole
08	30	28	32	40	46
10	21	20	29	40	44
13	22	23	34	46	50

Tabella 1. Percentuali di studenti che non raggiungono il livello 3, per macroregioni e per grado.

La forbice è molto ampia. Alla fine delle scuole secondarie di primo grado il divario fra le regioni meridionali e le isole maggiori da una parte e le regioni settentrionali dall'altra è molto alto: il Sud ha circa il 50% in più di livelli inferiori a 3. E alla fine delle superiori aumenta ancora: la differenza registra una percentuale doppia, perché rispetto ai gradi inferiori la situazione è migliore al Nord ma peggiore al Sud.

I divari sono significativi: anche se i dati non sono stati raccolti con rilevamenti longitudinali, ed è dunque scorretto trasformare il tempo apparente in tempo reale, una distribuzione così netta fa pensare alla persistenza dello storico, secolare divario Nord-Sud, con prospettive di aggravamento.

2.3 Livello socioeconomico e culturale della famiglia

Per identificare il livello socioeconomico della famiglia di appartenenza di ogni ragazzo si usano vari indicatori (ricavati dal questionario che accompagna le prove): status occupazionale dei genitori, livello di istruzione, possesso di beni materiali propri di un contesto economico e culturale favorevole all'apprendimento. Tutti confluiscono nell'indice ESCS (Economic Social Cultural Status Index). Consideriamo il quadro dei discostamenti dal punteggio medio 200 di ragazzi appartenenti, rispettivamente, ai livelli più alti e ai livelli più bassi dell'ESCS:³

¹ Ricordiamo che il livello è la descrizione di quello che lo studente sa fare in un certo ambito, e della sua capacità di compiere specifiche operazioni cognitive. Il livello minimo è 1, il massimo 5. Il livello 3 è considerato il livello minimo per il raggiungimento degli obiettivi indicati dalla Indicazioni Nazionali per ogni classe.

² I rilevamenti INVALSI riguardano i gradi: 02 = seconda classe della scuola primaria; 05 = quinta classe della scuola primaria; 08 = terza classe della secondaria inferiore; 10 = seconda classe della secondaria superiore; 13 = ultima classe delle scuole secondarie superiori.

³ Nella tabella il valore di riferimento non è 200 ma 400 perché in ognuna delle due righe si sommano i valori delle due classi ESCS qui considerate insieme (rispettivamente: indice alto + indice medio-alto; indice basso + indice medio-basso). I valori sono arrotondati all'unità.

Indice ESCS	Grado 05	Grado 08	Grado 10
Alto e medio-alto	417	423	430
Basso e medio-basso	384	379	396
Forbice	33	44	34

Tabella 2. Punteggio medio, per indice ESCS e per grado.

Come ci aspettavamo, il livello socioeconomico della famiglia incide tuttora in modo decisamente significativo sulle prestazioni relative all'italiano, e in misura particolarmente rilevante nella scuola dell'obbligo.

2.4 Immigrati e nativi

Dà i risultati attesi – e persino ovvi – anche il quadro dei risultati ottenuti dai giovani immigrati rispetto ai nativi: considerando 200 il punteggio medio di riferimento, ecco di quanto se ne discostano le prestazioni di giovani immigrati e nativi:

Grado	Nativi	Immigrati
02	+ 3	- 24
05	+ 3	- 28
08	+ 2	- 31
10	+ 6	- 18
13	+ 2	- 16

Tabella 3. Discostamento dal punteggio medio di riferimento, per condizione di nativo / immigrato e per grado.

La differenza è notevole, paragonabile a quella che riconduciamo al fattore 'contesto familiare'. Si nota un'attenuazione delle differenze nel passaggio dalle medie alle superiori, probabilmente da attribuire alla selezione avvenuta nel passaggio dalla scuola secondaria inferiore alla superiore con l'autoesclusione dalle iscrizioni, per i motivi facilmente ipotizzabili.

Questi sono i dati evidenziati nel rapporto INVALSI 2019. Ma ce ne sono anche altri, forse più interessanti.

Dopo i rilevamenti del 2019, grazie a innovazioni sostanziali nella metodologia di scelta e somministrazione dei quesiti, è possibile – ed è scientificamente corretto – confrontare i dati di due anni diversi: è possibile, in altre parole, un'analisi longitudinale, che consente di stabilire progressi e regressi significativi nella competenza linguistica degli alunni, anno dopo anno.

Se proviamo a confrontare i dati del 2019 con quelli del 2018 troviamo importanti conferme.

2.5 Confronto dei divari tra i risultati delle femmine e dei maschi

La forbice tra femmine e maschi si ripropone con la stessa ampiezza nei due anni consecutivi: una differenza minima all'inizio della scuola primaria (+3); un incremento significativo (+9) al termine della primaria e alla fine della scuola secondaria inferiore.

Il fenomeno non è solo costante (si ripete puntualmente di anno in anno) ma è anche internazionale: si registra praticamente in tutte le rilevazioni internazionali. Le spiegazioni che in genere se ne danno sono sia 'interne', rispetto alle prove (il formato dei quesiti: i test a scelta multipla favoriscono i maschi, quelli a risposta aperta le femmine; la somministrazione via computer o su carta: la prima modalità favorisce i ragazzi, la seconda le ragazze) che 'esterne', cioè generali: in particolare la pervasività degli stereotipi culturali, trasmessi attraverso condizionamenti familiari e sociali radicati nel tempo.

Su questo fronte, dunque, non ci possiamo aspettare una riduzione del divario in tempi brevi. È opportuno monitorare i cambiamenti, ma la complessità e la forza dei fattori di contesto inducono ad orientarci piuttosto verso la 'lunga durata'.

Altre variabili danno però risultati più interessanti, in parte inattesi, e di sicuro interesse per insegnanti e dirigenti scolastici, perché orientano su tempi più vicini a noi, anzi – volendo – imminenti.

2.6 Confronto dei divari fra macroregioni settentrionali, centrali e meridionali

Al termine della terza media la percentuale di alunni che raggiungono o superano il livello 3 fra il 2018 e il 2019 diminuisce di 2,4 punti nel Nord-ovest e di 2,9 nel Centro, mentre nel Sud aumenta di 4,6 punti percentuali (nelle altre aree non ci sono variazioni significative). Anche nella seconda superiore si registra una tendenza analoga: + 2,3 nel Nord-ovest, + 4 nel Centro, + 2,5 nel Sud e + 4,1 nel Sud e nelle isole maggiori. Due linee di tendenza che contrastano con la "narrazione" corrente: nel passare dal 2018 al 2019 la competenza linguistica non diminuisce ma aumenta, e aumenta in misura maggiore proprio nelle aree storicamente svantaggiate: le regioni del Mezzogiorno, Sicilia e Sardegna.

Troviamo anche un andamento simile se consideriamo il variare della distribuzione per livelli nelle prove di inglese del 2019 rispetto al 2018. Ecco il prospetto di quanto raggiungono il livello previsto (A1 per il grado 05; A2 per il livello 08):

	Nord-ovest	Nord-est	Centro	Sud	Sud e isole
grado 05 prove di ascolto	+ 6,1	+ 2,3	+3,9	+ 8,2	+ 4,3
grado 08 prove di lettura	+ 1,9	+ 3,2	+ 2,0	+ 8,2	+ 2,9

Tabella 4. Inglese. Differenze 2019-2018 nelle percentuali di studenti che raggiungono il livello previsto, per macroregioni e per grado.

2.7 Confronto dei divari legati al livello socioeconomico e culturale delle famiglie

Fra il 2018 e il 2019 si registra anche un'altra tendenza: non forte, tutta da verificare ma – ci sembra – significativa, se non altro per il suo occorrere in tutti e tre i gradi considerati. È l'ampiezza della forbice tra i punteggi medi dei ragazzi con indice ESCS alto e medio-alto e quelli dei ragazzi a indice medio-basso.

	05: diff. 2018	05: diff. 2019	08: diff. 2018	08: diff. 2019	10: diff. 2018	10: diff. 2019
Alto e medio-alto	419	417	428	423	422	430
Basso e medio-basso	381	384	375	379	386	396
Forbice	38	33	53	44	36	34

Tabella 5. Italiano. Punteggi medi degli studenti, per indice ESCS e per grado, nei rilevamenti 2018 e 2019.

Come si vede, le due medie si sono avvicinate, in tutti gli ordini e gradi: la distanza fra i due valori, rispetto all'anno precedente, è diminuita di 5 punti in quinta elementare, di 9 punti in terza media e di 2 punti in seconda superiore. Nella scuola dell'obbligo, in particolare, si è registrato sia un incremento nei livelli più bassi che un decremento dei livelli più alti: non è ancora un livellamento (obiettivo al momento utopistico) ma un avvicinamento abbastanza significativo.

Una tendenza anche più marcata si rileva in inglese: tanto nelle prove di ascolto quanto nelle prove di lettura le distanze fra i due blocchi (alto e medio-alto; basso e medio-basso) si riducono in modo drastico: al grado 5 e al grado 8 si dimezzano.

2.8 Confronto diacronico tra i dati di immigrati e nativi

Nello stesso 2019 si può constatare un altro cambiamento interessante – e non banale –, che riguarda gli immigrati di seconda generazione: la differenza tra questi e i nativi italofoni diminuisce con l'incremento della scolarità.

Fra le prestazioni medie della prima e della seconda generazione di immigrati c'è sempre differenza: i ragazzi immigrati da poco hanno punteggi più lontani dai punteggi dei nativi, in ogni ordine e grado di scuola. Questo in ragione della progressiva esposizione a contatti linguistici (attivi e passivi) con produzioni linguistiche in italiano, che aumenta col passar del tempo, assicurando un progressivo incremento delle acquisizioni linguistiche. Ma lo scostamento fra le prime e le seconde generazioni rispetto ai nativi nel 2019 è inferiore all'anno precedente, in ogni grado. Si può pensare, con un certo fondamento, a un 'assestamento' dei processi di integrazione via via che passano le generazioni, anche grazie all'effetto scuola: è a scuola che si moltiplicano gli effetti dell'esposizione a

produzioni linguistiche in italiano; e questi dati suggeriscono che forse la scuola, almeno in questo ordine di problemi, non funziona così male come spesso siamo portati a pensare.

Grado	Immigrati di prima generazione rispetto ai nativi	Immigrati di seconda generazione rispetto ai nativi	Differenza fra seconda e prima generazione
02	- 27	- 21	6
05	- 31	- 21	10
08	- 33	- 18	15
10	- 24	- 12	12

Tabella 6. 2019: differenza di punteggio medio fra immigrati di prima e seconda generazione e nativi, per grado.

Grado	Immigrati di prima generazione rispetto ai nativi	Immigrati di seconda generazione rispetto ai nativi	Differenza fra seconda e prima generazione
02	- 25	- 17	8
05	- 32	- 16	16
08	- 38	- 17	21
10	- 25	- 13	12

Tabella 7. 2018: differenza di punteggio medio fra immigrati di prima e seconda generazione e nativi, per grado.

3. Quando una classe è troppo omogenea

I rilevamenti INVALSI consentono di misurare anche la variabilità, cioè di quanto ogni risultato di ogni alunno si scosta dagli altri alunni e dalla media. Raggruppando opportunamente i vari indici di scostamento si può stabilire quanta variabilità è dovuta a differenze fra le scuole, fra le classi di ogni scuola, o fra gli alunni della stessa classe. Sono dati importanti.

La variabilità fra gli alunni della stessa classe rimanda a differenze individuali (a loro volta legate a background sociali e culturali diversi, a fattori psicoevolutivi individuali ecc.); ma i decisori del processo educativo sono particolarmente interessati alle altre due dimensioni della variabilità, che sono indicatori di diseguaglianze nel processo di formazione: una variabilità significativa vuol dire che non a tutti la scuola ha dato le stesse opportunità educative. Se è alta la differenza fra scuole vuol dire che le caratterizzazioni socioeconomiche e culturali della società si riflettono senza mediazioni sulla differenziazioni fra questo e quell'Istituto scolastico; come se ogni scuola avesse come

obiettivo non l'assicurazione di pari opportunità per tutti ma la perpetuazione delle disuguaglianze oggi presenti.

Ancora più importante è la variabilità fra le classi di uno stesso istituto scolastico perché chiama in causa direttamente l'organizzazione interna di una scuola: dipende infatti dai criteri con cui sono formate le classi. La ricerca pedagogica ha rivelato che se i ragazzi più capaci e/o quelli socioeconomicamente e culturalmente avvantaggiati sono raggruppati o sono comunque numericamente prevalenti in una certa classe, quella classe si avvantaggerà dell' 'effetto di contesto', cioè dell'influenza reciproca, della tendenza di ciascuno ad adeguarsi al livello delle prestazioni degli altri per un naturale impulso di integrazione nel gruppo-classe; mentre su una classe che raggruppasse ragazzi meno capaci o svantaggiati l'effetto di contesto avrà risultati opposti, poiché la tendenza all'integrazione nel gruppo-classe porterà a livelli di prestazione uniformemente bassi.

Nei dati INVALSI la variabilità fra le classi è presente fin dall'inizio delle elementari ed è ancora rilevata – con intensità addirittura crescente – al termine delle elementari, e poi delle medie. Ma in modo geograficamente ben differenziato: la variabilità interessa soprattutto le aree del sud e delle isole, dove ha un'intensità anche doppia e tripla rispetto al Nord-ovest e al Nord-est. La deduzione è facile, e preoccupante: nel Mezzogiorno e nelle isole prevale la formazione di classi omogenee, di antica tradizione classista, pre- e antidemocratica. In certe classi si mettono insieme i bambini delle classi alte e medio-alte, in altre tutti gli altri, col risultato di indurre apprendimenti polarizzati: verso l'alto nella prima, verso il basso nella seconda. È l'esatto contrario dello spirito e della lettera dell'obbligo statuito dall'articolo 3 della Costituzione.

Possiamo dunque dare una prima risposta, sia pure parziale e provvisoria, alla domanda che ci siamo posti all'inizio. I copiosi dati INVALSI, non solo danno consapevolezza dei traguardi effettivamente raggiunti dagli studenti, ma danno importanti notizie sui fattori scolastici ed extrascolastici che influiscono – a volte anche in modo determinante – sui risultati raggiunti, e sul rispettivo 'peso' condizionante. Ma non solo. Offrono alla nostra riflessione i dati di un processo in corso, del quale ben pochi indizi circolavano – e circolano – nel pur vasto e variopinto campo dell'informazione scritta e parlata: diversamente dalla *communis opinio* il divario fra i risultati di classi tradizionalmente, storicamente lontane (per banalizzare: Nord e Sud, immigrati e nativi, ricchi e poveri) non sta aumentando, ma al contrario subisce piccoli ma generali, diffusi (e spesso progressivi) processi di riduzione. Tra il 2018 e il 2019 tende a ridursi la distanza tra macroregioni del Nord e macroregioni di Sud e Isole, fra ragazzi con background socioeconomico e culturale alto e medio-alto e ragazzi con background basso e medio-basso, fra immigrati di seconda e di prima generazione. Avviene con le prove di italiano ma anche con quelle di inglese e – sia pure con minore rilevanza – con quelle di matematica.

Se è vero che – come gli economisti e i sociologi costantemente rilevano non solo nella società italiana ma un po' in tutto l'Occidente – negli anni più recenti tende ad aumentare la divaricazione fra classi sociali, Nord e Sud ecc., determinando strutture sociali sempre più polarizzate sulle ali estreme, i movimenti segnalati dai questionari e dalle prove raccolte nelle scuole, in significativa controtendenza, fanno pensare che proprio nella scuola si registrino i primi segnali di un'inversione di rotta. Se sarà confermata, vorrà dire che la scuola ha qualche merito importante: se non nel contrastare quanto meno nel non assecondare affatto il preoccupante processo di polarizzazione in corso.

Per contro, gli stessi dati ci dicono che in molte scuole – almeno a livello decisionale – resistono comportamenti ispirati alla riproduzione delle disparità anche profonde che caratterizzano il contesto sociale e culturale nel quale esse si trovano: la formazione di classi tendenzialmente omogenee è il riflesso di questi atteggiamenti e a sua volta si riflette nei risultati dei test INVALSI, quando si considera la variabilità fra classi dello stesso Istituto. Il percorso di una scuola rinnovata sembra avviato, ma è accidentato e impervio.

Due facce della stessa medaglia: quella di una scuola che prova ancora una volta a cambiare in meglio, ma che porta ancora in sé antiche contraddizioni.

4. Postilla

Quando le bozze erano già impaginate l'INVALSI ha presentato un compendio dei risultati delle prove standardizzate del 2020-2021, le prime dopo lunghi periodi di sospensione delle lezioni e di Didattica a distanza a causa della pandemia. I dati sono interessantissimi, e in più sedi sono stati commentati. Qui ne richiamo solo alcuni, che consentono di riprendere e aggiornare le osservazioni condotte sin qui.

1. *La scuola primaria* tiene. Ottiene risultati in linea con quelli degli anni precedenti, anzi leggermente superiori: nel 2019 nella scuola primaria si collocava sui livelli uguali o superiori a 3 (su 5) il 75% dei bambini; nel 2021 addirittura l'83%. La scuola secondaria di primo grado ha risultati decisamente inferiori: nel 2018 e nel 2019 gli studenti che non raggiungevano il livello 3 erano il 34%, quest'anno sono il 39%. La situazione è ancora peggiore nella scuola secondaria di secondo grado, dove il punteggio medio diminuisce del 5% e la percentuale degli studenti che non raggiungono risultati adeguati (il livello 3) passa dal 35% del 2019 al 44%.
2. Il *contesto socio-economico-culturale* influisce sui risultati di tutte le materie in modo ancor più pesante di prima: fra gli studenti più svantaggiati non solo si abbassano i punteggi medi, ma diminuisce anche la quota di quelli che forniscono prestazioni di livello elevato. Questa tendenza arriva a livelli particolarmente alti alla fine delle superiori, quando la percentuale di studenti in difficoltà aumenta dell'8% fra gli alunni con ESCS alto e addirittura del 15% fra gli alunni con ESCS basso.
3. *Differenza dei risultati fra scuole e fra classi*: si riscontra già nelle scuole elementari, soprattutto nelle aree meridionali, e si accresce ulteriormente nelle scuole secondarie di primo e secondo grado.
4. *Divari territoriali*: sono aumentati. In molte regioni del Mezzogiorno (Campania, Calabria, Puglia, Sicilia, Sardegna, Abruzzo) più della metà degli studenti 'maturandi' non raggiunge la soglia minima di competenze in italiano, mentre al Nord e al Centro tutte le regioni si collocano al disotto del 40%.

Niente di nuovo sotto il sole, insomma, con o senza la pandemia? Sì, ma qualche osservazione interessante si può fare, ad esempio a proposito delle differenze territoriali.

Consideriamo i risultati del grado 8 e del grado 13 (ultimo anno, rispettivamente, delle scuole medie e delle superiori). Le differenze fra regioni del Nord e regioni del Sud sono

enormi: ai due estremi troviamo il Nord-ovest, in cui la percentuale di coloro che non raggiungono il livello di accettabilità è del 28% e le regioni del Sud, in cui la percentuale è del 61%. Eppure qualcosa sta cambiando. Se consideriamo, per ogni area, la differenza tra il 2019 e il 2021, l'incremento delle 'insufficienze' si distribuisce nelle diverse aree geografiche come qui si riporta alla Tabella 8.

Grado	Nord-ovest	Nord-est	Centro	Sud	Sud e isole	Italia
8	20	17,9	12,5	12,5	8,7	14,7
13	22,7	43,5	26,5	32,6	14	25,7

Tabella 8. Incremento delle 'insufficienze' nelle prove INVALSI tra il 2019 e il 2021.

In tutta Italia sono aumentati gli 'insufficienti', ma più nelle aree settentrionali (si veda in particolare il Nord-est per il grado 13) che in quelle meridionali; all'interno di queste, più nel mezzogiorno peninsulare che nelle isole maggiori. Si noti che, se limitiamo la nostra attenzione ai dati 2021, proprio Sicilia e Sardegna sono in testa a questa specie di 'lista nera'.

In definitiva, questi dati aggiungono informazioni importanti al quadro che si delineava sulla base del confronto fra i dati dei due ultimi anni pre-pandemia (§§ 2 e 3).

- In italiano si registra, dopo la pandemia, un calo generalizzato, con una forte riduzione del punteggio medio e un allarmante incremento delle percentuali di studenti che non raggiungono il livello minimo di accettabilità.
- Persiste, e anzi si aggrava lo storico, secolare divario fra regioni settentrionali e regioni meridionali-insulari, così come si accresce l'influenza dei fattori socio-economico-culturali di contesto sui processi di acquisizione delle competenze, determinando uno squilibrio ancor più forte fra classi superiori e classi inferiori del tessuto sociale. Sulle perdite di competenza legate a questi fattori il tempo lungo della didattica a distanza ha prodotto danni evidenti. Si conferma anche la variabilità dovuta a differenze fra scuole e fra classi: ma nell'eccezionalità dell'anno scolastico pandemico questo dato è di difficile lettura.
- Le tendenze generali all'abbassamento dei livelli di prestazione nell'intervallo pre-postpandemia hanno tuttavia interessanti differenziazioni areali: sono meno accentuate nelle aree meridionali, alle quali possiamo dunque attribuire una sia pur minima resilienza, da collegare al miglioramento delle prestazioni che si registrava già nel passaggio dal 2018 al 2019 (§ 2.6): insieme, queste due tendenze suggeriscono l'idea di un controcanto, o meglio di una contro-narrazione rispetto alle geremiadi correnti sui divari incolmabili del Mezzogiorno, che travalica il tempo della pandemia.
- I rilevamenti INVALSI del 2021, infine, dicono una parola forse definitiva sulla forzata sperimentazione DAD: il fatto che i risultati migliori siano stati registrati nelle scuole primarie, nelle quali la didattica a distanza è stata adottata solo in parte (spesso in minima parte) può solo essere letto come un *caveat* per la ripetizione (o, peggio

ancora, la generalizzazione) della didattica via computer. Con un'ovvia specificazione: i dati si riferiscono alla DAD così com'è stata realizzata nel 2020 e 2021: con tanta buona volontà e tanto impegno, ma con mezzi spesso insufficienti, in condizioni di contesto ambientale spesso difficili (se non impossibili) e soprattutto in assenza dei processi di formazione dei docenti. Intervenire in questi settori pare oggi urgente, anzi indifferibile, se si vuole non condannare sic et simpliciter la Didattica a distanza – irrinunciabile strumento didattico per la scuola del XXI secolo – ma integrarla nel processo educativo dei nostri ragazzi, in una scuola moderna, democratica e inclusiva.

CLUB Working Papers in Linguistics

A cura di Emanuele Miola e Rosa Pugliese

Volume 5, 2021

ISBN: 9788854970649

Contributi di

Mario Ambel

Cecilia Andorno

Anna Boario

Massimo Cerruti

Francesca Gallina

Lucia Mason

Giulia Minori

Emanuele Miola

Diego Pescarini

Angelica Ronconi

Alberto A. Sobrero

Laura Tramutoli