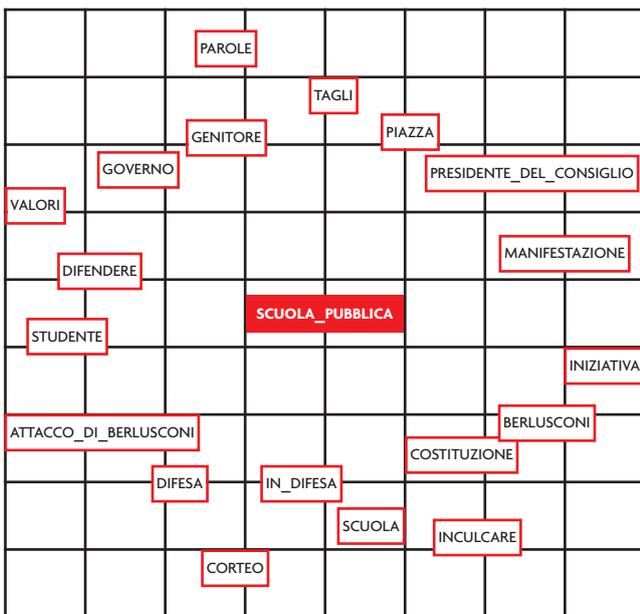


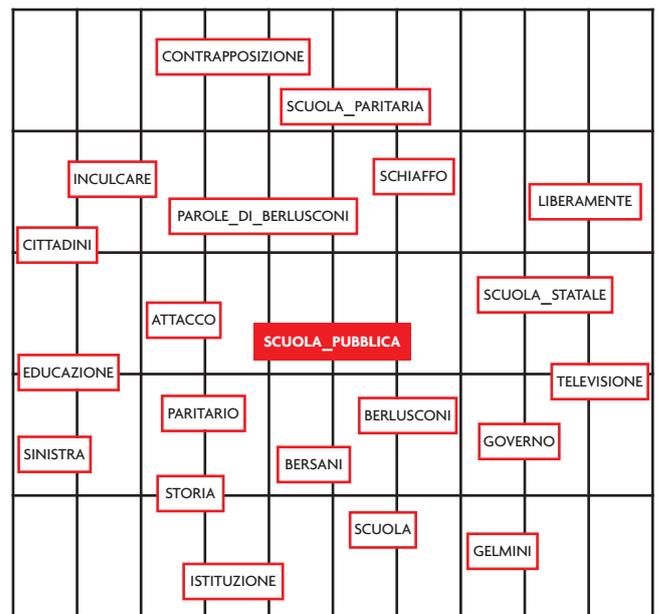
Scuola pubblica... al testoscopio (III)

di SILVIA BENVENUTI

Nella prima parte di quest'articolo (XlaTangente n. 26, pp. 13-16), abbiamo analizzato, con l'aiuto di un software per l'analisi testuale, tutti gli articoli contenenti la stringa "scuola pubblica", usciti su la Repubblica e il Giornale nei giorni immediatamente successivi alle esternazioni del Presidente del Consiglio sul tema. L'ultima delle analisi descritte in quelle pagine produce, per ogni termine selezionato dall'utente, un grafico in cui il termine stesso è evidenziato in rosso e gli altri lemmi, in nero, sono distribuiti attorno, a distanza inversamente proporzionale al grado di associazione di ognuno con quello selezionato. Ebbene: in modo del tutto controintuitivo (e anche un po' inquietante) i grafici centrati sulla stringa "scuola_pubblica" risultano molto simili, nel primo e nel secondo quotidiano. Possiamo allora concluderne che il Giornale e la Repubblica ce la raccontano, sulla scuola pubblica, sostanzialmente allo stesso modo?



Associazioni della stringa scuola_pubblica nel primo quotidiano



Associazioni della stringa scuola_pubblica nel secondo quotidiano

Eccoli qui, i due grafici incriminati. Se dovete scommettere, qual è quello che corrisponde al quotidiano *la Repubblica*? E quale a *il Giornale*? Se da una parte potrebbero entrambi appartenere al quotidiano diretto da Mauro, difficilmente ci sentiremmo di attribuirne uno a quello diretto da Sallusti. Come vedete, infatti, in entrambi i grafici compaiono le parole "attacco", "inculcare", "parole", "Berlusconi" e "governo". Uno dei due, è vero, parla anche di "corteo", "piazza", "manifestazione" e "difesa", mentre l'altro non vi fa cenno; il secondo, però, sembra rincarare la dose con il termine "schiaffo", che suona come un giudizio di merito sulle parole rivolte dal premier alla scuola. Che il programma abbia fatto qualche errore? Ebbene, lettori e direttori, la risposta è no, nessuno sbaglio. E il mistero è presto risolto se teniamo presente che, in

certe analisi, un po' di cautela è d'obbligo: un buon testoscopio infatti, proprio come un buon microscopio, mostra cose che possono essere "viste" solo da chi ha il tempo, la pazienza e le conoscenze necessarie per interpretarle correttamente. Ed è quello che proviamo a fare qui di seguito. Innanzi tutto osserviamo che il termine scuola_pubblica, nel quotidiano *il Giornale*, è sotto-rappresentato (40 occorrenze contro 315): il grafico corrispondente sarà quindi meno rappresentativo, rispetto a quello relativo all'altro quotidiano, che ha scuola_pubblica come specificità tipica. Inoltre, come possiamo facilmente accertare mediante *click* sulle varie caselle, il termine scuola_pubblica, nel quotidiano corrispondente al secondo grafico, compare prevalentemente nei virgolettati che riportano le affermazioni di esponenti dell'opposizione: «La **scuo-**

la pubblica è nel cuore degli italiani – tuona Bersani – da Berlusconi arriva uno **schiaffo** inaccettabile» – ed ecco spiegata la presenza della parola “schiaffo” nel grafico stesso.

A fugare ogni ulteriore dubbio controlliamo, mediante un altro semplice *click* sul secondo grafico, i contesti elementari in cui la parola “attacco” compare nel quotidiano corrispondente:

DATE: 6/5/2011 - 11:32:45 AM

Subject : LEMMA ASSOCIATIONS < SCUOLA_PUBBLICA > AND < ATTACCO >

**** *QUOT_2

perciò si fonda su un presupposto sbagliato perché da questo governo non c'è stato nessun **attacco** alla **scuola pubblica**. Anzi, ha osservato Gelmini, molti di quelli che sono scesi in piazza in difesa della scuola pubblica, mandano i figli a quella che loro chiamano scuola privata e lo trovo un po' incongruente.

**** *QUOT_2

Chi legge le parole di Berlusconi come un **attacco** alla **scuola pubblica** lo fa perché ancora legato alla erronea contrapposizione tra scuola statale e scuola paritaria, prosegue il ministro. Contrapposizione che per il Pdl non esiste perché, conclude la Gelmini, la scuola può essere sia statale sia paritaria perché è sempre un'istituzione pubblica al servizio dei cittadini.

**** *QUOT_2

Ieri il quotidiano diretto da Concita De Gregorio allegava un inserto contenente le 120mila firme raccolte in favore della **scuola pubblica** dopo l'**attacco** del presidente del consiglio agli insegnanti. Un inserto fortemente sconsigliato dagli oculisti. Infatti la maggioranza dei nomi è stampata in corpo tipografico uno. Così minuscolo da essere quasi impercettibile a occhio nudo.

Su quale sia il secondo quotidiano, a questo punto, potremmo davvero scommettere. L'analogia tra i due grafici, in conclusione, non implica affatto che il tema in esame sia trattato allo stesso modo. Se, infatti, invece di fermarci al grafico puro e semplice, sfruttiamo adeguatamente tutte le informazioni fornite dal programma, possiamo far emergere le differenze, con buona pace dei rispettivi direttori, che possono continuare a dormire sonni tranquilli!

Noi, intanto, ci dedichiamo a un'altra funzione di T-Lab, *l'analisi tematica dei contesti elementari*. Come dicevamo nell'articolo precedente, per T-lab il *corpus* da analizzare è costituito da tanti contesti elementari (frammenti di testo di lunghezze comparabili), ognuno dei quali è caratterizzato dalle parole che contiene. La funzione in esame consente di estrarre dal testo un certo numero di *cluster*, ciascuno dei quali è costituito dall'unione di più contesti elementari caratterizzati da parole omogenee, e si può quindi descrivere mediante le parole stesse.

In altri termini, ogni *cluster* rappresenta il filo di un discorso (che l'utente deve interpretare!) all'interno della trama complessiva del *corpus*, e l'insieme dei *cluster* individuati da T-Lab costituisce il riassunto delle tematiche trattate nel testo. Nel sottoinsieme costituito dagli articoli de *la Repubblica*, per esempio, T-Lab individua quattro *cluster*, per ognuno dei quali fornisce l'elenco delle parole caratterizzanti.

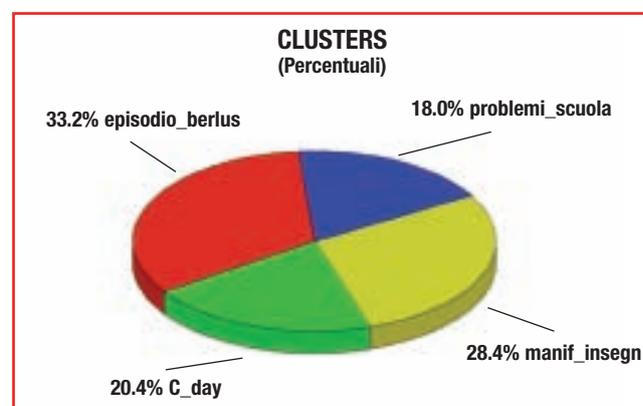
Un'analisi accurata di tali elenchi consente di “dare un nome” a ciascuno dei *cluster*, ovvero a ciascuno dei fili attorno ai quali ruota il discorso portato avanti dal quotidiano.

Il primo *cluster*, ad esempio, è caratterizzato dalle parole *scuola pubblica*, *insegnanti*, *presidente del Consiglio*, *inculcare*, ecc.: pare dunque lecito identificarlo come l'episodio delle esternazioni di Berlusconi.

Il secondo, caratterizzato dalle parole *italiani*, *Costituzione*, *festeggiamento*, *risorgimento*, ecc. merita il nome di *C_day*, a indicare che si riferisce alla manifestazione in difesa della Costituzione.

Il terzo, individuato dai termini *mobilitazione*, *palco*, *a difesa*, ecc... sarà denominato “manifestazione degli insegnanti” e l'ultimo, caratterizzato dalle parole *tagli*, *dirigenti scolastici*, *difficoltà*, *finanziamento*, *investire*, ecc. merita il nome di “problemi della scuola”.

Il grafico a torta che vedete in figura consente di apprezzare il peso che ogni *cluster* riveste nel complesso degli articoli del quotidiano selezionato.



Di più, T-lab fornisce un file html (vedi sotto) in cui all'interno del *corpus* in esame vengono selezionati i contesti elementari rilevanti, evidenziandoli con un colore diverso per ogni *cluster*. Selezionando quindi un colore per volta, otteniamo una sorta di “riassunto” per tematica.

LEGENDA : CLUSTER N. 1 CLUSTER N. 2 CLUSTER N. 3 CLUSTER N. 4

*QUOT_REPUBBLICA

[...]

La giornata “a difesa della Costituzione” è organizzata dalle tre associazioni Articolo 21, Valigia blu e Libertà e Giustizia: prudenti sui numeri, puntano a riempire piazza del Popolo a Roma, dove parlerà Roberto Vecchioni, e largo Cairoli a Milano, con Dario Fo e Salvatore Borsellino sul palco.

[...] e luoghi istituzionali nelle metropoli europee di Parigi, Londra ed Edimburgo, Madrid, Siviglia, Praga. A Firenze gli organizzatori chiedono che si lascino a casa le bandiere di partito [...] In quel momento ho pensato di non aver abbastanza inculcato nella mente di Noemi valori diversi. [...]

Costituzione day, domani iniziative in cento città. Centoquattro città, dicono gli ultimi aggiornamenti, ospiteranno cortei e happening in difesa della Costituzione e della scuola pubblica.

Costituzione, l'Italia delle 100 piazze. L'Italia è una Repubblica democratica fondata sul lavoro... Si comincia con l'articolo

1 (e con un minuto di silenzio per le vittime giapponesi) ma subito dopo, sul palco di piazza del Popolo a Roma, sarà l'articolo 111 a essere recitato [...]

Il merito deve trionfare, ma i punti di partenza devono essere uguali per tutti. È la democrazia. Nel Duemila si è congelato. Con che pensione? Millesettecento euro al mese, dopo 37 anni di insegnamento.

La Gelmini non è riuscita a mettere nelle sue riforme le parole uguaglianza e serenità. [...]

Vogliamo tornare sulle frasi di Berlusconi, i professori che inculcano agli studenti pensieri diversi da quelli delle famiglie? Berlusconi è impreparato, nonostante sbandieri la sua conoscenza del greco. Inculcare è un verbo violento, d'obbligo. Il problema è che non gli drena la testa, quel che gli arriva dice: l'istinto è quello. E dice cose fortemente repressive.

È un valore la difesa della scuola pubblica? Essenziale. La scuola pubblica non è facoltativa, non può dividersi con quella privata. Per questioni morali, disciplinari, civili, politiche. Garantisce l'uguaglianza ed è la cosa da sovvenzionare di più. La scuola pubblica regala cittadini liberi.

Ma passiamo all'ultima delle analisi, che costituisce un ulteriore passo avanti nell'individuazione di somiglianze e differenze tra sottoinsiemi del corpus. Si tratta dell'analisi delle corrispondenze, il cui prodotto più affascinante è una serie di grafici bidimensionali (e, nella versione più recente, anche 3D) in cui i diversi testi compaiono come pallini, disposti nel piano con una geometria governata dalla loro composizione rispetto ad assi fattoriali individuati dal programma (e interpretabili dall'utilizzatore). Secondo la teoria alla base di queste tecniche, dette di *analisi fattoriale*, due testi vicini (risp. lontani) relativamente a un asse del piano fattoriale comunicano messaggi simili (risp. diversi) relativamente al fattore corri-

spondente. In altri termini: la posizione dei punti nel piano fattoriale è la traduzione grafica della distribuzione delle parole in ciascun testo e, secondo la teoria sopra esposta, le "distanze ideologiche" tra i messaggi trasmessi da due testi sono direttamente proporzionali alle distanze metriche tra le loro rappresentazioni nel piano fattoriale.

Confusi? Vediamo un esempio: dal momento che l'analisi delle corrispondenze si può effettuare solo se il corpus è costituito da almeno tre sottoinsiemi, ed è tanto più significativa quanto più sono i sottoinsiemi analizzati, abbandoniamo per un istante la nostra comparazione *la Repubblica vs il Giornale* per dedicarci a un argomento di più stretta attualità, le elezioni del sindaco di Milano, conclusesi il 30 maggio scorso con la vittoria di Giuliano Pisapia. Il corpus in analisi è quindi costituito dai programmi elettorali dei nove candidati a sindaco, ciascuno dei quali costituisce un sottoinsieme a parte. T-Lab individua otto fattori, e quindi ben $7 + 6 + 5 + 4 + 3 + 2 + 1$ diversi grafici bidimensionali in cui piazzare i simboli corrispondenti ai nove programmi. Vediamo, in particolare, un paio di quelli che hanno in ascissa il fattore 3 (Figura 1 e 2).

Osservandoli, notiamo subito che, rispetto al fattore in ascissa, i candidati Mantovani e Montuori hanno posizioni estreme. Ora: sapendo che Mantovani è il candidato di Forza Nuova e Montuori è quello del Partito Comunista dei Lavoratori, non ci stupisce più di tanto che ci sia una qualche contrapposizione ideologica nei loro programmi. Un

po' più sorprendente è che la individui T-Lab, questa polarità; T-lab che non sa niente di politica e che trae le sue conclusioni solo analizzando i lemmi usati nei testi e il modo di associarli – ma senza conoscerne il senso!

All'utente, a questo punto, resta il compito interpretativo. In altri termini: T-Lab indi-

Il numero di parole conosciute e usate è direttamente proporzionale al grado di sviluppo della democrazia e dell'uguaglianza delle possibilità. Poche parole e poche idee, poche possibilità e poca democrazia: più sono le parole che si conoscono, più ricca è la discussione politica e, con essa, la vita democratica"

Gustavo Zagrebelsky

vidua degli assi fattoriali, rispetto ai quali colloca la posizione dei vari sottoinsiemi del corpus. L'utente, per poter interpretare le mappe così ottenute, deve "interpretare gli

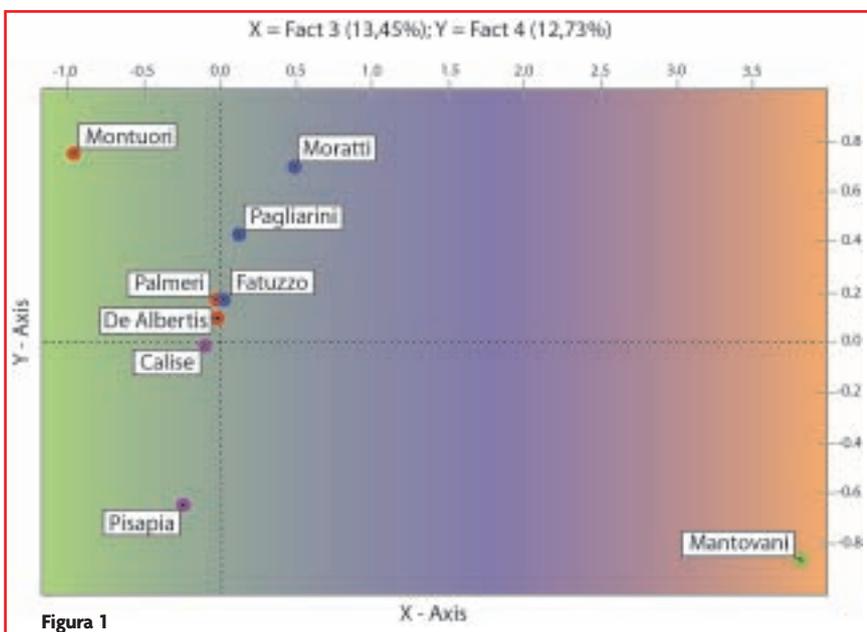


Figura 1

«Interpretare un asse fattoriale significa trovare ciò che vi è di analogo, da una parte tra tutto ciò che sta alla destra dell'origine, dall'altra tra tutto ciò che sta a sinistra, ed esprimere poi con concisione ed esattezza l'opposizione tra i due insiemi». Operazione senz'altro più facile a dirsi che a farsi, e che comunque T-Lab rende ragionevolmente agevole fornendo le tabelle dei lemmi in base ai quali ha individuato i singoli fattori.

Tornando dunque ai due grafici che abbiamo riportato, la prima cosa da fare è "interpretare il fattore 3", verificando i termini corrispondenti. I due grafici, inoltre, enfatizzano anche una con-

trapposizione, decodificata dal fattore 4, tra il programma dell'ex sindaco Moratti e quello dell'attuale sindaco Pisapia, e una contrapposizione uno contro tutti formalizzata nel fattore 5, che possiamo definire "il fattore Calise". Con 28 grafici a disposizione, capite bene che c'è di che divertirsi!

Ma concludiamo con qualcosa di più semplice, che non richiede software complicati e che d'altra parte non può mancare, in un articolo sull'analisi automatica del testo. Parliamo della leggibilità, valutabile tramite un indice, il *Gulpease*¹, che misura la scorrevolezza del testo oggetto di analisi. Si tratta di un numero, funzione di parametri quantitativi quali il numero di lettere per parola e il numero di parole per frase, calcolato secondo la formula seguente:

$$89 + \frac{300 \cdot (\text{numero delle frasi}) - 10 \cdot (\text{numero delle lettere})}{(\text{numero delle parole})}$$

Più alto è il *Gulpease*, maggiore è la scorrevolezza del testo: i lettori che hanno un'istruzione elementare trovano difficili i testi che presentano un indice inferiore a 80, quelli con un'istruzione media leggono con fatica i testi con indice inferiore a 60 e i lettori con un diploma superiore comprendono senza difficoltà testi con indice maggiore di 40.

Vediamo quindi qual è il piazzamento dei diversi testi di cui abbiamo parlato: *Giornale e Repubblica* quasi pareggiano, 48 il primo e 49 il secondo. Risultato discreto, se pensate che, sebbene la *Gazzetta dello Sport* si piazzino mediamente a 54 e i *Promessi Sposi* si piazzino a 52, i documenti scaricabili dal sito del governo italiano totalizzano mediamente 45, mentre i bugiardini di molti medicinali non superano il 40.

Meno bene se la cava la quasi totalità degli aspiranti/uscanti/attuali sindaci, dal 34 di Pisapia al 46 di Pagliarini. Su tutti svetta il candidato Marco Mantovani, che con un più che onorevole 59 sbaraglia nettamente la concorrenza.

Chiunque abbia letto il suo programma (Forza Nuova Milano. Chi siamo. Giovani soprattutto. Padri di famiglia. Madri e figli. Imprenditori, operai. Studenti. [...]) troverà questa vittoria prevedibile. Altrettanto prevedibile è il 49 conseguito da Silvio Berlusconi con la sua "lettera ai milanesi" a sostegno della candidatura del sindaco uscente. Una lettera del tutto fedele al suo stile linguistico, che ha indubbiamente rappresentato una svolta nel linguaggio politico italiano. Tanto da meritare uno studio quantitativo², a

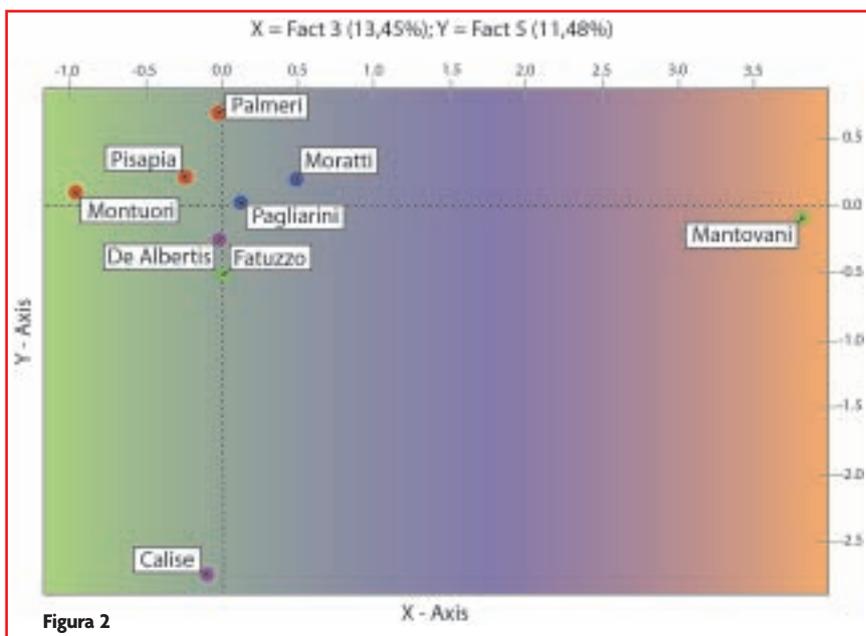


Figura 2

base di rigorosi strumenti statistici, che ne misura inequivocabilmente novità e semplicità (oltre a molte altre cose). Bravo. Ammesso che ci sembri davvero un affare il passaggio dalla retorica al linguaggio di impresa, dai periodi agli slogan, dai pensieri ai pensierini.

D'altra parte, leggibilità non è affatto sinonimo di comprensibilità: se guardate la formula, infatti, vedrete come l'indice *Gulpease* si disinteressa completamente dei vocaboli usati. Paradossalmente, quindi, può risultare altamente leggibile un testo fatto di parole inventate di sana pianta, e quindi del tutto incomprensibile, a patto che parole e frasi siano sufficientemente brevi. Per questo, nel valutare la comprensibilità di un testo, è bene affiancare all'indice di leggibilità una valutazione accurata delle parole che lo compongono. Aiuta in questo senso il ricorso al *Vocabolario di base* di Tullio De Mauro: si tratta di un elenco di parole, ottenuto grazie a criteri prevalentemente statistici, che rappresentano la porzione della lingua italiana usata e compresa dalla maggior parte di coloro che la parlano (e leggono). Meno di 7000 vocaboli, contro i 150.000 del Devoto-Oli: il che forse chiude un po' il cerchio ... Invece di straparlare di quali principi inculchino, i nostri insegnanti, perché non dar loro qualche strumento in più per far bene il loro lavoro?

Note

1. L'indice *Gulpease* è stato formulato attorno al 1987 da un gruppo di linguisti dell'Università La Sapienza di Roma, riunitisi attorno a Tullio de Mauro a costituire il Gruppo Universitario Linguistico Pedagogico (GULP).
2. S. Bolasco, N. Galli de' Paratesi, L. Giuliano, *Parole in libertà*, Manifestolibri, 2006. Si veda anche G. Carofiglio, *La manomissione delle Parole*, Rizzoli, 2010.

"Le parole sono fatte, prima che per essere dette, per essere capite: proprio per questo, diceva un filosofo, gli dei ci hanno dato una lingua e due orecchie. Chi non si fa capire viola la libertà di parola dei suoi ascoltatori. È un maleducato, se parla in privato e da privato. È qualcosa di peggio se è un giornalista, un insegnante, un dipendente pubblico, un eletto dal popolo. Chi è al servizio di un pubblico ha il dovere costituzionale di farsi capire."

Tullio De Mauro