
Maurizio Lana
Università del Piemonte Orientale “A. Avogadro”.
20 Marzo 2004

Maurizio Lana, Dipartimento di Studi Umanistici, Università del Piemonte Orientale “A. Avogadro”. (m.lana@lett.unipmn.it).

Sono laureato in lettere classiche e fin dagli anni 80 mi sono interessato di filologia computazionale (partendo dagli scritti e dall’opera di dom Froger, dom Quentin, G. P. Zarri) e di studio di testi con il computer (incominciando con J. Abercrombie, *Computer programs for the analysis of texts*). I miei interessi hanno riguardato in passato la stilometria (analisi dei testi allo scopo di individuarne caratteristiche formali e di contenuto tali da permetterne l’attribuzione, cioè il riconoscimento dell’autore), l’attribuzione di testi apocrifi o anonimi, l’analisi statistica dei testi in senso lato. Attualmente i miei interessi di ricerca, pur senza lasciare completamente da parte quelli precedenti, si orientano soprattutto verso l’analisi del contenuto come evoluzione dell’analisi dei testi e l’analisi delle corrispondenze lessicali su testi contemporanei.

I miei interessi riguardano soprattutto i testi, in senso stretto. Ho lavorato su testi antichi (la Costituzione degli Ateniesi attribuita a Senofonte, le Storie di Tucidide, l’opera di Senofonte), e lavoro ora su testi contemporanei letterari (Collodi, Le avventure di Pinocchio), e no (gli articoli pubblicati dai principali quotidiani italiani dal 12 al 29 luglio 2001 sul G8 di Genova).

In particolare, in relazione agli articoli dei quotidiani sul G8 mi interessa mettere in evidenza, attraverso lo studio del lessico utilizzato, l’evoluzione temporale del linguaggio usato, dal 12 al 29 luglio in relazione agli avvenimenti che si verificarono; mi interessa altresì caratterizzare, sempre in base al lessico utilizzato, i giornalisti e le testate, così da individuarne *distanze* e *vicinanze*. La ricerca è in corso.

Sto anche (marzo-aprile 2004) lavorando per prepararmi all’*ad-hoc attribution contest* bandito da P. Juola della Duquesne University di Pittsburgh. Il *contest* consisterà nel riconoscere quali tra una serie di campioni di testo anonimi siano di un determinato autore del quale pure viene fornito un campione di testo. Anche questa ricerca è in corso. I risultati saranno resi noti da P. Juola in occasione della ALLC-ACH Conference di giugno 2004 a Göteborg.

Nell’*ad-hoc attribution contest* sto utilizzando T-LAB per vedere se riesco ad ottenere classificazioni corrette dei testi basate sull’analisi delle corrispondenze dei segmenti ripetuti. Nel lavoro sul G8 lo utilizzerò per mettere a fuoco aree semantiche e nuclei concettuali specifici dei vari quotidiani e dei loro giornalisti.

Dal punto di vista dello studio dei testi, mi aspetto di trovare conferma alla mia convinzione che i metodi statistici (analisi delle corrispondenze) conducano a due tipi di esiti:

- da una parte dare corpo ed evidenza fattuale (quantitativa, numerica!) a intuizioni che nella lettura tradizionale, sequenziale del testo rimangono inafferrabili (ricerca di conferme ad ipotesi preesistenti);
- dall’altro, con una logica di tipo *text mining*, estrarre e porre in evidenza una serie di fenomeni linguistici e di contenuto (es.: negli articoli sul G8 del quotidiano A il nome del ministro Scajola è associato ai termini *a*, *b*, *c*; mentre nel quotidiano B è associato ai termini *r*, *s*, *t*) tra i quali il ricercatore può talora individuare quelli che alla sua analisi e interpretazione appaiono essere fatti rilevanti (esplorare materiali testuali troppo estesi per poter essere afferrati nel loro senso complessivo).

Le ricerche descritte sono ancora in corso e le relative pubblicazioni, non appena definite, potranno essere indicate a chi mi scriverà in posta elettronica.