

I termini specifici

Mauro La Torre

Nelle scienze dell'educazione, e nel parlare e scrivere di fatti concernenti la scuola, coesistono linguaggi di diversa origine e diverse finalità, che utilizzano spesso le stesse parole con sensi diversi, a volte contrapposti.

Basta confrontare un testo ministeriale, con un saggio accademico, o con un documento didattico prodotto da un Collegio di docenti; ma è anche istruttivo comparare come, rispetto ad un certo problema disciplinare, supponiamo di apprendimento del linguaggio, si esprimono diversi professionisti: un glottodidatta, uno psicolinguista, un filosofo del linguaggio, un insegnante di materie letterarie, un semiologo, un docente di linguistica generale.

Ora, se già un certo grado di "polisemia" - se non di vera e propria "confusione" - è constatabile all'interno di una società come la nostra, con un sistema scolastico e universitario relativamente unitario (?!), quale sarà il livello di intercomprensione effettiva negli organismi internazionali che si occupano di istruzione, come quelli dell'Unione Europea, o nei gruppi di lavoro di un convegno dell'Unesco, dove alla pluralità delle teorie scientifiche e delle norme amministrative si sovrappone la pluralità delle tradizioni culturali nazionali?¹

Naturalmente, chi si occupa di questi problemi sa bene che non si tratta di questioni meramente linguistiche: molto spesso l'incomprensione è legata alla non condivisione delle concezioni. Se, ad esempio in un determinato testo si parla di scuola "pubblica", di università "libera", di "professione" o "professionale", il significato delle parole cambia, perfino all'interno della stessa lingua, a seconda che ci si riferisca alla Germania orientale o occidentale, alla Gran Bretagna o agli USA o all'India.

1

Non a caso questi organismi interstatali spendono gran parte delle proprie risorse in lavori di traduzione, interpretazione e in servizi terminologici. Vedi, per esempio, Selten R. (red.), *I costi della (non) comunicazione linguistica*, Roma, ERA, 1997

Come abbiamo visto in nostre analisi² e progetti³ precedenti, un'attività nel campo della terminologia della formazione, come la costruzione di lessici settoriali, glossari, tesauri relazionali, o di "semplici" dizionari bilingui, combina sempre, in un modo che può essere più o meno esplicito, un intento descrittivo con un intento normativo. In altre parole si osserva l'uso effettivo dei termini, per proporre, se necessario, una standardizzazione delle forme verbali e delle loro diverse accezioni.

Anche in questa stessa rubrica abbiamo avviato un'indagine lessicostatistica della rivista *Cadmo*⁴, rivolgendo una particolare attenzione all'uso - peraltro molto limitato - di forestierismi. Si tratta di una ricerca che merita di essere estesa all'intero *corpus* di articoli, che con questo numero completano il primo lustro (superando i 5 milioni di grafemi), con l'opportunità di evidenziare, per rubrica per autore e per annata, i temi e i termini maggiormente ricorrenti. L'applicazione di nuovi metodi di ricerca contestuale consentirà poi anche l'esplorazione delle cooccorrenze, come indici di approfondimento sottotematico o di incrocio transdisciplinare.

Ma l'interesse per la terminologia settoriale e per la sua *normalizzazione*⁵ travalica ormai i confini nazionali, ispirando progetti e realizzazioni di servizi con dimensioni transculturali. In ambito europeo si comincia a disporre di tesauri multilingui, come quello del Consiglio d'Europa sulla formazione, che nel 1991 comprendeva già 9 lingue per 42 microtesauri⁶. La stessa Unione Europea cura il dizionario terminologico *Eurodicatom*, pubblicamente accessibile sulla rete Internet⁷.

Appena nello scorso luglio è stato fondato il CEL (Consiglio Europeo per le Lingue), che coordina le iniziative di decine di istituzioni (tra cui la nostra

2

La Torre M. (1993b), "Un lessico per la tecnologia dell'istruzione", in *Cadmo*, n. 1, Tecnodid

3

La Torre M. (1992e), "Per una banca testuale e terminologica", in *IAD - Istruzione a distanza*, n. 6, Tecnodid

4

La Torre M. (1996), "Chi è senza barbarismo ... lanci il primo sasso", in *Cadmo*, n. 12

5

Usiamo il termine *normalizzazione* senza nessuna connotazione ideologica, come semplice derivato di *norma*, sinonimo di *standardizzazione*.

6

CEE/Consiglio d'Europa (1991), *Thesaurus europeo dell'educazione* (versione in lingua italiana), Lussemburgo, Ufficio delle Pubblicazioni Ufficiali delle Comunità Europee

⁷ tramite l'indirizzo: <<http://www2.echo.lu/edic/>>

Università di Roma Tre) nel campo della formazione linguistica e della politica interculturale, con tutto ciò che consegue: sistemi di prove diagnostiche, dizionari-macchina, istruzione a distanza, ecc.⁸

Pochi giorni dopo, ad Adelaide in Australia, si svolgeva un Colloquio internazionale sul ruolo della lingua esperanto nel programma educativo *Linguapax* dell'Unesco. In tale occasione è stata presentata una nuova ricerca empirica sugli usi terminologici in campo didattico ed educativo da parte della comunità esperantofona⁹.

Concretamente si tratta dello spoglio di alcune annate del periodico IPR (*Internacia Pedagogia Revuo*), svolto con mezzi informatici. La regolarità della lingua ha contribuito notevolmente nel facilitare l'analisi lessicostatistica, in quanto ha consentito la costruzione di programmi computazionali completamente automatici, sia per quanto riguarda la lemmatizzazione delle varianti morfologiche, sia per l'analisi a livello dei singoli morfemi. Per avere un'idea delle potenzialità di questo tipo di ricerca, iniziata su un primo campione di testi, conviene dare uno sguardo ad alcuni dati.

Della rivista IPR, che viene pubblicata più o meno regolarmente dal 1970, sono usciti più di 80 numeri: il campione analizzato è costituito dagli ultimi 10 quaderni, contenenti testi per una lunghezza di 930.000 caratteri di stampa. La segmentazione del corpus ha fornito le seguenti unità lessicali: 140.000 occorrenze di parole, 25.000 forme, 15.000 lemmi. Ciò implica una ripetizione media di circa 9 volte per lemma, e una varietà morfologica di 1,66 forme per lemma. Il computo mirava primariamente alla compilazione di formari e lemmari di frequenza. Nella figura 1 vediamo infatti - per pura curiosità - il segmento iniziale del formario alfabetico generale.

8

Vedi, ad esempio, il sito Internet dello stesso *CEL*, presso la Libera Università di Berlino: <http://userpage.fu-berlin.de/~elc>

9

La Torre M. (1997), "Desuprismo kaj desubismo en terminara laboro", UK/IKU, Adelaide (Australia)

Figura 1 - Inizio del formario alfabetico (rivista IPR, 1995-97), in cui compaiono le forme flesse (VOF), la categoria morfologica (GR), la frequenza assoluta (OFT), il lemma (LEM)

<i>VOF</i>	<i>GR</i>	<i>OFT</i>	<i>LEM</i>
abatej o	o	1	abatej o
abc- ordo	o	1	abc- ordo
abdi ki s	i	4	abdi ki
abel uj on	o	1	abel uj o
abi o	o	1	abi o
abi turi enta	a	1	abi turi enta
abi turi entaj	a	1	abi turi enta
abi turi entan	a	1	abi turi enta
abocon	o	1	aboco
abomen'	r?	1	abomen
abomeni	i	1	abomeni
abomeni s	i	1	abomeni
.....			

Più interessante è invece - nella figura 2 - la testa della lista di frequenza dei lemmi sostantivali

Figura 2 - Inizio del lemmario sostantivale in ordine di frequenza (rivista IPR, 1995-97)

<i>GR</i>	<i>OFT</i>	<i>LEM</i>	<i>FRE</i>	<i>FRE2</i>	<i>FREDF</i>	<i>FREKV</i>
o	866	esperanto	0,006219	0,007112	-0,000893	0,874
o	780	lingvo	0,005601	0,003594	0,002007	1,558
o	395	kurso	0,002837	0,000979	0,001858	2,898
o	297	lernanto	0,002133	0,000228	0,001905	9,355
o	284	jaro	0,002039	0,002373	-0,000334	0,859
o	254	instruisto	0,001824	0,000448	0,001376	4,071
o	233	lernejo	0,001673	0,000422	0,001251	3,964
o	228	instruado	0,001637	0,000324	0,001313	5,052
o	219	infano	0,001573	0,000531	0,001042	2,962
o	208	lando	0,001494	0,001702	-0,000208	0,878
o	206	vorto	0,001479	0,000795	0,000684	1,860
o	181	informo	0,001300	0,001622	-0,000322	0,801
o	171	verbo	0,001228	0,000024	0,001204	51,167
o	162	numero	0,001163	0,001015	0,000148	1,146
o	162	ekzemplo	0,001163	0,000652	0,000511	1,784
o	156	revuo	0,001120	0,000703	0,000417	1,593
o	139	artikolo	0,000998	0,000966	0,000032	1,033
o	138	grupo	0,000991	0,000700	0,000291	1,416
o	132	teno	0,000948	0,000811	0,000137	1,169
o	130	leciono	0,000934	0,000117	0,000817	7,983
o	127	metodo	0,000912	0,000202	0,000710	4,515

Emergono termini più o meno pertinenti alla materia in questione (in italiano:

lingua, corso, allievo, anno, insegnante, scuola, insegnamento, bambino, Paese, parola, informazione, ...). Anzi, se si confronta la lista della figura 2 con quella estratta dalla prima annata di Cadmo¹⁰, si vede che la maggior parte di queste parole (quelle sottolineate) appartengono come significato anche alla testa dell'altra lista.

Naturalmente si può obiettare che con un computo di questo tipo si raccolgono anche parole che sono molto frequenti nella lingua comune, o genericamente scientifica, ma non specifiche del settore di ricerca indagato (nel nostro caso l'istruzione e la formazione). Ma è in qualche modo possibile evidenziare i termini specifici con un sistema automatico? Sì, se con termini *specifici* intendiamo quelli statisticamente caratteristici, ovvero quelle voci che in un corpus di testi settoriali come il nostro compaiono con frequenza *significativamente superiore* alla frequenza che essi presentano in un campione di testi, che sia rappresentativo della lingua comune o che comunque rappresenti uno spettro tematico molto ampio¹¹. La difficoltà pratica, in tali indagini, sta spesso nella disponibilità di dati affidabili e comparabili sulla lingua comune e sul lessico "generico".

Nel caso della ricerca menzionata, si è costruito - per utilizzarlo come termine di paragone - un campione relativamente corposo di un altro periodico nella stessa Lingua internazionale, costituito da 89 numeri consecutivi del notiziario *Eventoj*, con una lunghezza testuale di quasi 5 milioni di grafemi, 700.000 occorrenze, 74.000 forme, e 45.000 lemmi. Questo secondo insieme di testi è stato trattato con gli stessi metodi lessicostatistici, applicati alla rivista pedagogica IPR, in modo da acquisire dati quantitativi sulla lingua "generica", perfettamente confrontabili con quelli ottenuti per il linguaggio settoriale. I risultati del confronto tra le rispettive frequenze sono leggibili nella figura 2, dove accanto ad ogni lemma (LEM) compaiono le frequenze relative, rispettivamente nella rivista IPR (FRE) e nel periodico *Eventoj* (FRE2), seguite dalla loro differenza algebrica (FREDF) e dal loro rapporto aritmetico (quoziente) FREKV. Leggendo la colonna delle differenze FREDF, si notano termini con differenza positiva (più o meno grande) - come *lingvo*, *kurso*, *lernanto*, *instruisto*, *lernejo*, *instruado*, *infano* - che sono infatti caratteristici del settore studiato, ma anche voci a differenza negativa (come *esperanto*, *aro*, *lando*, *informo*, ...), che sono sì frequenti nel corpus IPR, ma in misura tutto sommato minore che in un periodico non specializzato.

Le analisi lessicostatistiche più elementari, quelle che producono i vocabolari di frequenza, possono, come si è visto, contribuire alla individuazione del lessico settoriale; tuttavia se ci si limita a computare le forme linguistiche, senza tenere conto dei diversi sensi che le stesse forme possono avere, si rischia di confondere termini che vanno invece tenuti distinti. Per esempio, in italiano, 'disciplina'/'disciplinare' designano in certi contesti le materie di studio e in altri

¹⁰ La Torre (1996), op. cit.

¹¹

Trascuriamo qui di distinguere tra le parole più frequenti per ragioni tematiche, e quelle emergenti per ragioni stilistiche. Per questa discriminazione, peraltro importante, vedi anche:

Cossette A. (1994), *La richesse lexicale et sa mesure*, Parigi, Champion

il corretto comportamento degli allievi; perciò la presenza in una lista lessicale dell'aggettivo 'disciplinare' è scarsamente significativa, anzi è ambigua, in quanto - pur restando all'interno del linguaggio relativo alla formazione - può appartenere di fatto a due *termini* diversi. Infatti in un repertorio terminologico dovremo avere:

- disciplina 1 (materia)
- disciplina 2 (comportamento).

Esistono numerosi altri casi di forme, la cui ambiguità è anche più complessa e più varia, soprattutto se tali voci vengono considerate in una prospettiva interculturale ('educazione', 'formazione', 'obiettivo', 'valutazione', ecc.).

La disponibilità di "banche" di testi settoriali in forma computabile permette di ottenere facilmente, per una voce ambigua, una lista di contesti in cui essa compare, che può essere utilizzata come base d'appoggio per un'analisi delle accezioni reali, e quindi dell'eventuale polisemia (o semplicemente omografia). Un esempio di questo tipo, per la ricerca su menzionata, è dato dalla parola *pedagogo*, la cui univocità d'uso in ambito internazionale può essere facilmente messa in dubbio.

Figura 3 - Lista di contesti in cui compaiono forme del lemma *pedagogo* (rivista IPR)

pedagogo	
(1 P951/7545:)	metodologo kaj pedagogo kia Claude Gacond
(2 P952/2444:)	Al pedagogoj kiuj instruas Esperanton ankaŭ propedeŭtikcele
(3 P952/11943:)	da nepagipovaj pedagogoj
(4 P953/10590:)	la grandan pedagogon estintaj forte influataj de liaj
(5 P953/10811:)	nia elstara pedagogo La memoradon helpas interesaj artikoloj
(6 P953/11952:)	la lokaj pedagogoj institutoj verkemuloj ktp
(7 P954/6274:)	Vi karaj pedagogoj eksciuj kiujn nacilingvajjn poemojn ŝatas
(8 P961/11828:)	kunveno de pedagogoj okazos inter la 24-a kaj
(9 P961/13238:)	Kial ĝuste pedagogoj ofte montras nesufiĉan scion pri
(10 P962/805:)	plej gravaj pedagogoj
(11 P962/856:)	la analizitaj pedagogoj pere de mia verko
(12 P962/11774:)	nur al pedagogoj Eble por ili taŭgus pli
(13 P962/12861:)	du samtempaj pedagogoj Baghy kaj Freinet
(14 P963/1432:)	de progresaj pedagogoj kaj uzis iliajn rezultojn Lia
(15 P963/1639:)	pensojn de pedagogo laŭ kiu la homo estas
(16 P963/12541:)	malbone iuj pedagogoj regas la gramatikon de la
(17 P964/3837:)	KCE svisa pedagogo kaj verkisto naskiĝis antaŭ 250
(18 P964/4518:)	tiu altranga pedagogo Pestalozzi
(19 P971/12319:)	diplomitaj Internaciaj pedagogoj de E-o Estus interese scii
(20 P972/3800:)	inter E-parolantaj pedagogoj ktp
(21 P972/4445:)	Mi estas pedagogo li estas lingvisto

Nella figura 3 si può leggere un insieme di contesti, con le rispettive collocazioni, tratto dal suddetto campione della rivista IPR, in cui tale voce compare. Dalla lettura di tali frasi o da ulteriori espansioni, si può verificare la monosemia della voce, o, in caso di

deviazioni dal senso principale, si possono circoscriverle, spiegarle ed eventualmente "correggerle" (come in ogni lingua ancora "viva" è consentito!).

Vogliamo infine esemplificare quanto accennato all'inizio circa le facilitazioni, che l'analisi di testi scritti in una lingua regolare come l'esperanto consente anche a livello morfemico. Supponiamo di voler condurre un'indagine su un certo campo semantico, per esempio intorno al significato dell'imparare. In italiano potremmo cominciare ad inseguire le tracce - più o meno camuffate dalla storia - di lessemi come "apprend-, impar-, acquis-, stud-, disc-, sc[u]ol-, ...". Le difficoltà di automatizzazione in questo genere di esplorazione stanno nel fatto che, da un lato è difficile evitare che la macchina prenda parole estranee ('*discente*' sì, anche '*discepolo*', ma non '*discobolo*'), dall'altro si rischia di lasciare fuori voci poco imparentate come '*allievo*', '*alunno*', ecc.. In una lingua "unimorfica", in cui - cioè - i morfemi coincidono con i morfi¹², la "macchina" si muove più facilmente e con minore rischio di errore. Vedi a questo proposito, nella figura 4, la varietà di termini dell'imparare, trovati dal calcolatore nel corpus IPR in base alla semplice condizione di contenere il morfema *lern-* (imparare, studiare) in prima posizione.

12

In italiano, ad esempio, il primo morfema che compone la parola '*ven-ire*' si presenta, nella coniugazione e nella derivazione, sotto numerose altre forme (dette "morfi"): '*veng-o*, '*vien-i*, '*verr-ò*, '*venn-i*, ...'; perciò in italiano, come in molte altre lingue, un morfema è costituito da molti diversi morfi.

Figura 4 - Sottinsieme del lemmario IPR costituito dalle parole inizianti con il morfema lern.

LEM	GR	OFT	FRE	LEM	GR	OFT	FRE
lern	r?	1	0,000007	lernigo	o	1	0,000007
Lern-	p?	1	0,000007	lernilo	o	18	0,000129
lern-facilig	o	1	0,000007	lernini	i	1	0,000007
lerna	a	1	0,000007	lerninta	a	2	0,000014
lernadi	i	1	0,000007	lerninte	e	1	0,000007
lernado	o	98	0,000704	lerninto	o	5	0,000036
lerna^o	o	1	0,000007	lernita	a	12	0,000086
lernanta	a	4	0,000029	lernkapablo	o	1	0,000007
lernantaro	o	4	0,000029	lernmaterialo	o	3	0,000022
lernante	e	6	0,000043	lernmetodo	o	1	0,000007
lernantino	o	3	0,000022	lerno	o	16	0,000115
lernanto	o	297	0,002133	lerno-intereso	o	1	0,000007
lernata	a	1	0,000007	lernobjekto	o	2	0,000014
lernata^o	o	1	0,000007	lernoĵanbro	o	1	0,000007
lernebla	a	9	0,000065	lernogrup	o	2	0,000014
lernebleco	o	4	0,000029	lernohoro	o	2	0,000014
lerneblo	o	1	0,000007	lernojaro	o	22	0,000158
lernej-ĵanbro	o	1	0,000007	lernokurso	o	2	0,000014
lernej-estro	o	1	0,000007	lernolibra	a	3	0,000022
lernej-vesto	o	1	0,000007	lernolibro	o	77	0,000553
lerneja	a	54	0,000388	lernomaterialo	o	4	0,000029
lernejano	o	17	0,000122	lernonerkato	o	1	0,000007
lerneje	e	2	0,000014	lernonotodo	o	1	0,000007
lernejestro	o	9	0,000065	lernonta	a	1	0,000007
lernejo	o	233	0,001673	lernonto	o	2	0,000014
lernejo-fonda^o	o	1	0,000007	lernoobjekto	o	2	0,000014
lernejpolitiko	o	1	0,000007	lernookupo	o	1	0,000007
lernelenento	o	1	0,000007	lernoperiodo	o	1	0,000007
lernena	a	1	0,000007	lernoplano	o	2	0,000014
lerneno	o	1	0,000007	lernoprocezo	o	1	0,000007
lernemlo	o	3	0,000022	lernoprograno	o	1	0,000007
lernenda	a	4	0,000029	lernotago	o	2	0,000014
lernenda^o	o	1	0,000007	lernotenpo	o	1	0,000007
lernfacileco	o	1	0,000007	lernplifacilig	o	3	0,000022
lernfaciliga	a	1	0,000007	lernprogramo	o	1	0,000007
lernfacilig	o	2	0,000014	lernsituacio	o	1	0,000007
lernformo	o	1	0,000007	lerntempo	o	3	0,000022
lerngrupo	o	1	0,000007	lernulo	o	1	0,000007
lerni	i	174	0,001250	lernvojo	o	1	0,000007
lernigi	i	5	0,000036	lernvolo	o	1	0,000007