

Tra lessicografia e sociolinguistica: la frequenza dei regionalismi nell'italiano parlato nel Canton Ticino

Elena Maria Pandolfi

Casa S. Michele, 6947 CH – Vaglio

Osservatorio Linguistico della Svizzera Italiana (Bellinzona)

Abstract

The variety of Italian in Canton Ticino (Italian speaking Switzerland) shows lexical and partially morphosyntactic features that are well known in literature. Studies on the real presence of these features in spoken discourse are however lacking. This paper aims at applying quantitative methods in studying the Swiss variety of Italian providing frequency lists of lexical regionalisms on a basis of a corpus consisting of 213.007 words (approximately about 30 recording hours). The corpus includes different varieties of spoken language: spontaneous face-to face conversation, telephone calls, radio-tv entertainment and news. From this corpus 787 tokens (corresponding to 229 types) have been identified as lexical regionalisms, meaning the presence of one lexical regionalism every 2 minutes and 17 seconds or respectively every 271 words. The distribution of the regionalisms is quite widespread even if they appear with a greater frequency in informal conversational speech.

1 Introduzione: obbiettivi della ricerca

Se si esclude il LIP (De Mauro *et alii* 1993), le caratteristiche lessicali delle varietà regionali di italiano non hanno sinora mai costituito l'oggetto di lavori svolti in prospettiva di linguistica dei *corpora* e di lessicografia computazionale. E lo stesso LIP, i cui *corpora* parziali sono stati com'è noto raccolti in quattro diverse città, non ha tuttavia dato luogo ad analisi specifiche dei regionalismi lessicali (sui lavori di carattere quantitativo basati su corpora cfr. comunque ora Cresti – Moneglia 2005 e De Mauro – Chiari 2005). A quindici anni di distanza dall'ultima importante indagine sociolinguistica sulle particolarità lessicali dell'italiano regionale ticinese (Petralli 1990) si è voluto indagare, con un approccio quantitativo, l'effettiva presenza nell'italiano parlato nel Canton Ticino di particolarità lessicali dell'Italiano Regionale Ticinese, IRT (sulla situazione generale dell'italiano in Svizzera cfr., tra gli altri, Bianconi 2001). L'indagine di cui qui si presentano i risultati è parte di una ricerca più ampia svolta presso l'Osservatorio linguistico della Svizzera italiana (OLSI) su "Regionalismi e forestierismi nell'italiano parlato nel Canton Ticino. Un'indagine quantitativa"¹ che concerneva l'allestimento e l'analisi di liste di frequenza sia dei regionalismi dell'italiano ticinese sia

¹ Attualmente in corso di stampa nella collana dell'OLSI presso Dadò, Locarno.

anche dei forestierismi. L'indagine è consistita nell'allestimento di un opportuno *corpus*, per un totale complessivo di 29 ore e 53 minuti di parlato trascritto e di 213.007 parole grafiche, da cui sono state ricavate liste di frequenza di regionalismi ritenuti regionalmente marcati, che sono state anche messe in correlazione con le variabili sociolinguistiche classiche, situazione comunicativa, fascia di età, istruzione e sesso dei parlanti. I regionalismi sono inoltre stati classificati secondo categorie formali e secondo parametri sociolinguistici.

2 Metodologia dell'indagine

I materiali raccolti specificamente per questa indagine costituiscono un *corpus* di circa 145.000 parole per 20 ore e 33 minuti di registrazione. Il materiale raccolto è stato trascritto in grafia convenzionale secondo criteri CHAT (MacWhinney 1997) molto semplificati. Le analisi statistiche sono state effettuate in Excel. A questi sono stati aggiunti per l'analisi i materiali registrati e trascritti da Franca Taddei Gheiler (2004 e 2005) che costituiscono un *corpus* di altre 68.000 parole per circa 10 ore di registrazione. Dal punto di vista della classificazione secondo il mezzo di comunicazione, i testi appartengono a due macro-categorie: parlato radiofonico e televisivo da un lato (quello che Sabatini 1997 definisce 'parlato trasmesso') e il parlato conversazionale spontaneo dall'altro. Il parlato non trasmesso comprende due categorie: conversazioni spontanee e interazioni transazionali. La scelta dei lessemi da considerare per l'analisi è stata fatta manualmente. I lessemi selezionati come regionalismi (più di 780 occorrenze) sono stati sempre considerati nella forma base. In alcuni casi le entrate lessicali sono costituite da unità polilessematiche (cosiddette 'polirematiche'), per es. *incidente della circolazione, perequazione finanziaria, dipartimento delle istituzioni, disegno animato, mandato di prestazione*, ecc. Si sono catalogati i prestiti dal dialetto e, distinguendoli dai prestiti non adattati, le inserzioni mistilingui del dialetto nell'italiano. Abbiamo considerato tali tutti i casi in cui gli elementi dialettali erano costituiti da espressioni o formule di più di una parola, integrate in una struttura sintattica frasale italiana. Non sono state, invece, computate le vere e proprie commutazioni di codice. Come detto abbiamo considerato nel computo delle occorrenze anche i dialettismi (D), intendendo per essi parole o brevi espressioni che mantengono la forma del prestito integrale (per es.: *sctria*,² *inisci*, ecc.). Sono stati computati come regionalismi (R) anche termini che non compaiono nelle liste 'classiche' riportate nella bibliografia sul tema, seguendo un criterio di percezione personale della scrivente e verificando, quando possibile, presenza e significato sui De Mauro (2005) e Zingarelli (2005). Pertanto la regionalità di alcuni o molti lessemi qui classificati come regionalismi può essere certamente discutibile (per es. nel caso di sintagmi come: *valli superiori, mandato di prestazione, canone d'acqua*, ecc). Sono stati considerati anche alcuni fatti morfosintattici, limitatamente però a usi idiomatici e a reggenze preposizionali. Abbiamo selezionato quindi reggenze verbali difforni da quelle dell'italiano standard come per es.: *pensare qualcosa e aver bisogno qualcosa* (con presumibile interferenza del dialetto), *mettere sotto discussione, chinarsi su + complemento astratto* (per es.: *il mondo politico si china su queste problematiche*);

² Rendiamo con il digramma *sc-* la realizzazione palatalizzata della sibilante sorda anche davanti a consonante.

o certi costrutti fissi come *sul venerdì, sul mezzogiorno, da parte a*, ecc. In qualche caso è interessata anche la morfologia nominale, per es. in diverse assegnazioni di genere come *la meteo, la giudice, la fine settimana*, ecc. Non sono stati computati, invece, termini dell'italiano regionale ticinese, in genere appoggiati al dialetto, che siano anche ampiamente diffusi nelle varietà regionali lombarde o siano rinvenibili in genere negli italiani regionali settentrionali, per es. verbi sintagmatici come *lavar giù* o reggenze come *capace a, obiettare a* (+ inf.), ecc. Dal punto di vista sociolinguistico, le peculiarità lessicali dell'IRT tratte dal corpus sono state classificate, ove possibile e pertinente, anche secondo le dimensioni di variazione (diacronia, diafasia, diastratia). Dai dati così trattati sono state ricavate liste di frequenza assoluta e liste d'uso, mediante il calcolo dell'indice di dispersione secondo la formula impiegata nel LIF e nel LIP.

3 Analisi dei dati

La tabella 1 mostra la presenza di regionalismi e dialettismi nel corpus.

REGIONALISMI				
	F. ASS TOT	% SUL N. PAROLE	FREQ/PAROLE	FREQ/TEMPO
R	514	0,24		
D/DM	273	0,13		
TOT	787	0,37	271	2:17

Tabella 1. Regionalismi nel corpus (corpus di riferimento: 213.007 parole, 29 ore 53 minuti).

La frequenza dei regionalismi e dialettismi nel parlato rappresenta lo 0,37% del numero delle parole, vale a dire un regionalismo ogni 271 parole, o, in riferimento al tempo, un regionalismo ogni 2 minuti e 17 secondi di produzione orale. La presenza nel discorso effettivo di termini regionalmente marcati sembra essere quindi esigua. Tuttavia è pur sempre vero che ogni poco più di due minuti un parlante ticinese rivela linguisticamente la propria origine. A questo proposito si deve tuttavia osservare che manca ogni metro di paragone, data l'assenza di ricerche con questa impostazione quantitativa per gli altri italiani regionali d'Italia. Non è quindi possibile dire se l'italiano del Ticino sia, dal punto di vista della frequenza lessicale, più o meno marcato nell'uso rispetto agli altri italiani regionali di Italia. Le 787 occorrenze rimandano a un totale di 229 types, da attribuire alla somma di regionalismi e dialettismi (vd. tabella 2).

	TOKENS	TYPES	TOKENS/TYPES
D	273	31	8,81
R	514	198	2,60
TOT	787	229	3,43 (globale media)

Tabella 2. Tokens/types (corpus di riferimento: 213.007 parole, 29 ore e 53 minuti)

Il rapporto tra *tokens* e *types* risulta per i regionalismi piuttosto basso, pari a 2,60, vale a dire che ogni parola diversa tende a comparire meno di tre volte nel corpus. Tale rapporto è invece molto più alto (8,81) per i dialettismi. Questo potrebbe voler indicare che i termini

dialettali presentano una maggiore concentrazione su un numero relativamente ristretto di termini, e la presenza per quasi due terzi di *hapax*. Nel nostro caso la differenza è fondamentalmente da ricondurre alla frequenza eccezionale del segnale discorsivo *bon* (ben 228 occorrenze). Nella tabella 3 che segue è espressa la suddivisione dei regionalismi e dialettismi nei tipi di situazioni comunicative considerate.³

787	TOT	22	421	31	2	92	27	3	77	51	61
TOT		VFC	VIC	VFT	VIT	RN	RFD	RIA	TFD	TIA	TN
266	D	2	234	11	0	0	2	1	2	13	1
7	DM	0	5	0	0	0	0	0	0	2	0
514	R	20	182	20	2	92	25	2	75	36	60

Tabella 3. Regionalismi e dialettismi e situazione comunicativa
(*corpus* di riferimento: 213.007 parole, 29 ore 53 minuti)

La maggior concentrazione di termini propri dell'IRT (R) si ha nella conversazione amichevole informale (VIC, 421 occorrenze), ma i termini sono distribuiti, se pur in misura minore, in tutti in tipi di situazione comunicativa. I regionalismi non appaiono quindi fatti linguistici limitati a situazioni informali o specifiche, ma compaiono trasversalmente in tutto l'IRT e lo caratterizzano come italiano regionale del Canton Ticino. I dialettismi (D e DM), invece, compaiono prevalentemente nelle situazioni caratterizzate da informalità, e sono poco presenti nel parlato radiofonico e quello televisivo. Unico dialettismo che compare, in pochi casi comunque, anche in situazioni più formali (2 occorrenze in RFT e 2 in VFC, 11 in VFT) di *bon*, usatissimo nel parlato ticinese.

4 Liste di frequenza

RG F. ASS.	RG USO	LEMMA	L.	F. ASS.	USO
1	1	<i>bon</i>	D	228	152,601
3	2	evidente [generale per 'chiaro']	R	14	10,740
5	3	Gran consiglio [Governo cantonale - potere legislativo]	R	12	4,702
2	4	controprogetto	R	20	6,067
10	5	dipartimento (ministero)	R	11	5,099
5	6	cantonale (predizia)	R	12	4,954
12	7	avere bisogno (quantezza)	R	8	3,837
14	8	meteo (la)	R	7	3,559
5	9	educazione (istruzione)	R	12	3,347
4	10	conferenza (rinnicome)	R	13	3,288
19	11	rispettivamente [sia ...sia, o ...o]	R	6	3,120

³ VFC: intervista, parlato formale, dibattito; VIC: intervista, parlato informale, conversazione amichevole; VFT: intervista, parlato formale, interazione transazionale; VIT: intervista, parlato informale, interazione transazionale; RN: trasmissione radiofonica, notiziari; RFD: trasmissione radiofonica, parlato formale, dibattito; RIA: trasmissione radiofonica, parlato informale, intrattenimento; TFD: dibattito televisivo, parlato formale, dibattito; TIA: trasmissione televisiva, parlato informale, intrattenimento; TN: trasmissione televisiva, notiziario.

19	11	rispettivamente [sia ...sia, o ...o]	R	6	3,120
14	12	Consiglio di stato [Governo cantonale - potere esecutivo]	R	7	3,040
12	13	pù di quel tanto [pù di tanto]	R	8	2,730
24	14	a dipendenza di [a dipendenza da]	R	5	2,519
5	15	Friburgo [Friburgo]	R	12	2,485
5	16	riservare	R	12	2,375
19	17	annunciarsi [presentarsi, iscriversi]	R	6	2,363
19	18	scuole maggiori [scuole post-elem. prima della riforma]	R	6	2,152
24	19	Consiglio nazionale [Camera Parlamento federale]	R	5	2,125
30	20	per rapporto a [in rapporto a]	R	4	2,090
11	21	espe (prova in classe)	R	9	1,998
30	22	mantella [cappotto]	R	4	1,893
19	23	cassa malati [mutua]	R	6	1,796
42	24	da ... via [da ... in poi]	R	3	1,647
42	25	germanico [tedesco]	R	3	1,539
42	26	incidente della circolazione [incidente stradale]	R	3	1,395
30	27	monitore [istruttore]	R	4	1,369
42	28	Natel [(telefono) cellulare]	R	3	1,181
42	29	nucleo [centro storico di un paese]	R	3	1,147
30	30	patrizio [membro corporazione locale]	R	4	1,118
24	31	urino [del Canton Uri]	R	5	1,117
42	32	ancora [già]	R	3	1,014
24	33	buco tedesco [tedesco di Germania]	R	5	0,991
24	34	da parte (a) [a lato, in fianco]	R	5	0,963
42	35	corso di ripetizione [richiamo serv. militare]	R	3	0,960
30	36	consigliera (la)	R	4	0,907
42	37	consigliere di stato	R	3	0,903
14	38	domicilio [residenza]	R	7	0,855
14	39	ciola [esprime meraviglia, assuefazione, sicurezza, disapprovazione]	D	7	0,792
42	40	emissione [trasmissione radiofonica]	R	3	0,788

Tabella 4. lista di frequenza assoluta e uso ordinata secondo l'uso, fino al rango 40.

Le glosse, in parentesi quadre, accanto ai lemmi segnalano il corrispondente termine usato in italiano d'Italia o forniscono indicazioni sul valore del termine in IRT.

Ad una prima scorsa alla lista generale di frequenza spicca appunto la posizione di *bon*, con ben 228 occorrenze. In generale va segnalata nel nostro *corpus* una certa dispersione dei regionalismi: ben 103 *types* su 198 compaiono con una sola occorrenza e 36 con due, cosicché solo 59 *types* presentano almeno tre occorrenze. La graduatoria per uso mostra relativamente poche differenze rispetto alla frequenza assoluta. Il primo posto in entrambe le liste rimane occupato da *bon* che, come abbiamo detto, è diffuso in quasi tutti i testi esaminati. I ranghi appena sotto il primo subiscono variazioni più o meno significative: per es. *riservare* "prenotare" scende dal rango 5 al 16 nella lista d'uso, *espe* "prova in classe" dal rango 11 al rango 21. All'inverso *evidente*, nel valore contestuale che ha in IRT, sale dal rango 3 al rango 2, *meteo* sale dal rango 14 al 8. I primi regionalismi che compaiono nella lista delle frequen-

ze assolute, oltre a *controprogetto* (al primo posto con 20 occorrenze, ma che scende al rango 4 nella lista d'uso), sono lessemi afferenti alla sfera semantica burocratico-amministrativa: *conferenza* 13 occorrenze, *cantonale (polizia)* ed *educazione* 12 occorrenze. Un fatto che può stupire è che molti dei ticinesismi più 'classici', ampiamente riportati dalla bibliografia, non compaiono affatto nel nostro *corpus*. Non troviamo infatti nella lista per esempio *azione*, *trattanda*, *istoriato*, *bucal lettere*, ecc. Su 621 entrate di un dizionario di riferimento allestito sommando le entrate lessicali presenti in Petralli 1990 e *Lo Svizzionario* 2004, solo 67 sono i regionalismi che sono stati rilevati nel mio *corpus*: circa il dieci per cento. Ciò significa anche che sul totale dei regionalismi (*types*) tratti dal mio *corpus*, 198, ben 131 non compaiono nelle liste 'classiche' della letteratura sul tema. Una ragione per le assenze, del resto inerente alle ricerche su *corpora* determinati, sta nella casualità almeno parziale in cui sempre incorre la raccolta di materiali spontanei. Il riferimento al giudizio soggettivo della scrivente nella selezione dei regionalismi dal *corpus*, se certamente rende per così dire variabile e non ben definito l'inventario delle peculiarità, non basta a dar conto di una differenza così netta. Sarà necessario trovarne le ragioni in altri fattori, forse anche nel fatto che i lavori contenenti ticinesismi si sono spesso basati in buona parte su fonti scritte.

5 Conclusioni

Dei pochi risultati dell'indagine che abbiamo potuto presentare qui, a conclusione notiamo anzitutto come un approccio di lessicografia computazionale rechi un contributo in parte nuovo all'esame di fatti per lo più considerati solo in un'ottica sociolinguistica qualitativa. In un certo senso la quantificazione statistica consente di 'misurare' la caratterizzazione regionale (o, nel caso del Canton Ticino, 'statale', cfr. Muljačić 1987) del lessico di una determinata varietà diatopica di lingua. Infatti, se a prima vista le cifre e le percentuali dei regionalismi appaiono certamente basse, dal punto di vista della presenza nel discorso, calcolata sia in termini di quantità di occorrenze per unità di tempo sia in rapporto al numero totale delle parole, un ticinese produce parlando mediamente un elemento localmente marcato ogni 2 minuti e 17 secondi e ogni 271 parole grafiche. Un numero che, pur in mancanza di riferimenti comparabili, si può ritenere non trascurabile. Tale valutazione è rafforzata da un'ulteriore considerazione che possiamo avanzare confrontando i nostri dati con quelli del LIP relativi ai forestierismi. Si tratta ovviamente di un paragone indiretto, che proponiamo in maniera meramente indicativa. Il LIP registra nel suo corpus una percentuale dello 0,30 di forestierismi, mentre i ticinesismi (compresi i dialettismi) incontrati nel nostro corpus sono lo 0,37%. Dunque compaiono nel parlato più ticinesismi in IRT che forestierismi in italiano d'Italia. Si può quindi concludere, provvisoriamente, che un parlato ticinese è lessicalmente riconoscibile come tale in misura non irrilevante, anche se le cifre assolute sono naturalmente basse. Questo fatto va visto anche in relazione con il basso indice *tokens / types*, cioè con l'esistenza di un nucleo di regionalismi relativamente frequenti e una notevole quantità di regionalismi poco frequenti, occasionalismi che però concorrono a caratterizzare l'IRT. C'è poi da tener in conto che spesso a contrassegnare l'italiano del Canton Ticino è una leggera devianza semantica, di selezione o contestuale (sia di situazione sia di 'collocazione' nel contesto linguistico) di molti lessemi nei confronti dell'italiano d'Italia. Si è anche visto che, a parte il caso

particolare di *bon*, i primi termini a comparire nella lista di frequenza dei regionalismi e dei dialettismi sono quelli più o meno strettamente legati alla sfera semantica burocratico – amministrativa. Questo sembrerebbe confermare l'affermazione di Lurati (1976:199-206) circa una limitata marcatezza regionale nell'IRT nell'uso comune, dove l'IRT sia scritto sia parlato è vicino alla varietà regionale lombarda. In realtà, a nostro avviso, l'IRT ha avuto e continua ad avere una sua peculiarità come varietà regionale dell'italiano anche nell'uso comune e non solo, come del resto è del tutto normale, nella denominazione autonoma di istituzioni, procedure, politica ecc. legate ad un sistema politico – amministrativo diverso da quello dell'Italia. Stupisce inoltre la ridotta coincidenza fra l'insieme dei particolarismi lessicali schedati da Petralli (1990) e Savoia-Vitale (2004) e quelli che sono emersi nell'analisi condotta nel presente lavoro sono solo un terzo dei regionalismi considerati come tali da chi scrive nello spoglio del corpus è presente nella lista di riferimento ricavata dalle due opere citate, e viceversa.

Bibliografia

A. Dizionari

- De Mauro, T. (2005), *Il dizionario della lingua italiana*, Torino, Paravia.
Zingarelli, 2005 [2004], *Vocabolario della lingua italiana*, Bologna, Zanichelli.

B. Altri testi

- Bianconi, S. (2001), *Lingue di frontiera. Una storia linguistica della Svizzera italiana dal Medioevo al 2000*, Bellinzona, Casagrande.
Bortolini, U., Tagliavini, C., Zampolli, A. (1972), *Lessico di frequenza della lingua italiana contemporanea*, Milano, Garzanti-IBM (LIF).
Cresti, E., Moneglia, M. (eds.) (2005), *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
De Mauro, T., Mancini, F., Vedovelli, M. e Voghera, M. (1993), *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri (LIP).
De Mauro, T., Chiari, I. (a cura di) (2005), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne.
Lurati, O. (1976), *Dialecto e italiano regionale nella Svizzera italiana*, Lugano, Banca Solari e Blum.
MacWhinney, B. (1997), *Il Progetto CHILDES – Strumenti per l'analisi del linguaggio*, Edizione italiana a cura di Elena Pizzuto e Umberta Bortolini, Pisa, Ed. del Cerro.
Petralli, A. (1990), *L'italiano in un cantone. Le parole dell'italiano regionale ticinese in prospettiva sociolinguistica*, Milano, Franco Angeli.
Savoia, S., Vitale, E. (2004), *Lo Svizzionario 2004. Splendori e miserie della lingua italiana*, Linguanostra, Lugano.
Taddei Gheiler, F. (2005), *La lingua degli anziani*, Locarno, Dadò.
Muljačić, Ž. (1987), "Statalismes" sémantiques dans la Romania, in AA.VV., *Aspects of Language. Studies in Honour of Mario Alinei. Vol. II. Theoretical and Applied Semantics*, Amsterdam, Rodopi, pp. 339-356.
Sabatini, F. (1997), 'Prove per l'italiano "trasmesso" (e auspici di un parlato serio semplice)', in AA.VV., *Gli italiani trasmessi. La radio*, Accademia della Crusca, Firenze, pp. 11-30.
Taddei Gheiler, F. (2004), 'L'italiano in Ticino. Dalla "Questione della lingua" alla linguistica ticinese: un secolo di dibattiti', in Moretti, B. (a cura di), *La terza lingua. Aspetti dell'italiano in Svizzera agli inizi del terzo millennio. Vol. I. Norma e varietà di lingua in Ticino*, Osservatorio linguistico della Svizzera italiana, Locarno, Dadò, pp. 17-182.