

Lessico dell'intervista, lessico degli intervistati : l'articolazione tra domande e risposte nell'analisi di Alceste

Carlo Tomasetto¹, Patrizia Selleri²

¹ Università di Bologna (Facoltà di Psicologia) – Piazza A. Moro, 90 –
47023 Cesena (FC) – Italia
tomasetto@psice.unibo.it

² Università di Bologna (Facoltà di Psicologia, Dipartimento di Scienze dell'Educazione) –
via Zamboni, 34 – 40126 Bologna – Italia

Abstract

Statistical analysis of responses to narrative or semi-structured open interviews challenge researchers either from a statistical or from a linguistic point of view. The dialogical aspects of interviews are often faced as threats to the analysis validity ; Instead of looking for strategies addressed to neutralize these threats, we suggest to emphasize the interactive dimension of the context in which these corpora are produced. In particular we propose some strategy to detect the effects of the interviewer-interviewed interaction by means of Alceste software.

Riassunto

L'analisi statistica di *corpora* testuali composti dalle risposte ad interviste aperte, narrative o semi-strutturate pone diversi problemi sia dal punto di vista statistico che linguistico. Anziché proporre strategie per neutralizzare le minacce alla validità dell'analisi derivate dalla natura dialogica del corpus, qui proponiamo di individuare strategie di analisi che mettano in evidenza il contesto relazionale in cui il testo è stato prodotto. In particolare proporremo alcune strategie per individuare gli effetti dell'interazione intervistatore-intervistato attraverso l'uso del software Alceste.

Parole-chiave : software Alceste, interviste semi-strutturate, analisi statistica di dati testuali, interazione, contesto discorsivo.

1. Introduzione : quando il corpus nasce da un'intervista

Le risposte a domande aperte, vuoi proposte come illustrazione di questionari strutturati, vuoi tratte da interviste o focus-group, costituiscono ormai un settore di elezione per le applicazioni della statistica testuale. Eppure, analizzare un corpus testuale ottenuto dalle risposte di molti intervistati pone una serie di problemi non superficiali (Della Ratta, 2000) ; Lebart e Salem (1988) ammettono ad esempio che questo tipo di testi può facilmente disorientare “*à la fois les statisticiens et les spécialistes des études textuelles*” (Salem, 1988 : 9) : i primi, per il carattere “*imprécis et multiforme*” del materiale a disposizione, i secondi soprattutto per l'artificialità del contesto e per la ridondanza (altrettanto artificiale) del vocabolario (ibid.), dal momento che tutti i soggetti hanno dovuto forzatamente confrontarsi con le stesse domande.

Ad incrementare il disorientamento interviene poi l'acquisizione, ormai condivisa sul piano metodologico, delle implicazioni relazionali dell'intervista stessa. L'intervistato non è un passivo portatore di informazioni e l'intervistatore non ne è il collettore neutrale (Gubrium e Holstein, 2002), quale che sia la strategia poi impiegata in fase di analisi dei dati : elaborare i testi raccolti nel corso di un'intervista senza prendere in considerazione il suo carattere di attività conversazionale situata (Atkinson e Coffrey, 2002) rischia di minare considerevol-

mente la validità delle conclusioni che se ne traggono, e questo ha giustificato in anni recenti critiche radicali all'utilizzo dell'intervista stessa come metodo di ricerca sociale. Senza arrivare a tali estremi, è stato comunque sottolineato che il materiale testuale proveniente dalle interviste libere va trattato con estrema cautela, e che uno dei problemi ai quali porre la maggiore attenzione è proprio quello del ruolo e dell'influenza esercitata dall'intervistatore nel corso dell'interazione (Hammersley, 2003).

Un corpus raccolto attraverso una o più interviste emerge, di fatto, nel corso di interazioni nelle quali entrambi i partecipanti negoziano la loro reciproca posizione (Py e Grossen, 1997), con gli effetti pragmatici e linguistici che tutto ciò comporta. Proprio perché un'analisi statistica si basa sul lessico delle risposte date dagli intervistati, dobbiamo tenere presente che anche l'intervistatore interviene nella gestione dell'interazione discorsiva: introduce i temi, chiede spiegazioni ed esempi, riformula concetti per rilanciare un discorso interrotto e così via. Soprattutto, l'intervistatore si serve di un lessico particolare nella formulazione stessa delle domande ed è con tali scelte lessicali che l'intervistato si confronta nel momento in cui costruisce il proprio contributo al mantenimento dell'interazione.

Per quanto Lebart e Salem riconoscano nell'"artificiosità del contesto" uno dei problemi sollevati dalle risposte a domande aperte, difficilmente ci si è posti il problema di far emergere da un'analisi statistica la componente dialogica di questi testi, mettendo in risalto i particolari meccanismi di interazione conversazionale che ne orientano la produzione. L'interesse non è naturalmente quello di utilizzare software e procedure di trattamento statistico per proporre un'analisi conversazionale, studiare gli scambi e le prese di turno, né studiare esaustivamente l'intervista in quanto pratica sociale. D'altro canto, non consideriamo soddisfacente nemmeno la neutralizzazione della dimensione relazionale, vista come minaccia alla validità dell'analisi statistica dei testi. Riteniamo invece che le tracce dello scambio in cui gli intervistati si sono impegnati siano un elemento con cui confrontarsi consapevolmente, sforzandosi (se possibile) di farne emergere le implicazioni per il lavoro interpretativo del ricercatore.

Rinunciando ad ogni pretesa di esaustività, affronteremo l'argomento limitandoci alle procedure e ai risultati offerti dal software ALCESTE (Reinert, 1986 e 1998), coscienti che la riflessione potrebbe essere estesa a procedure e software diversi.

L'analisi delle interviste con ALCESTE

Applicando una procedura di Classificazione Discendente Gerarchica, ALCESTE raggruppa gli enunciati di un corpus testuale in modo da creare classi il più possibile omogenee al loro interno (cioè composte da enunciati in cui si ripete sistematicamente la co-occorrenza delle stesse parole), ed il più possibile diverse l'una dall'altra. Procedendo nell'analisi, ALCESTE determina il vocabolario specifico di ciascuna classe, composto dalle parole significativamente più presenti negli enunciati della classe rispetto a tutto il resto del corpus. L'unità di analisi principale è quindi l'enunciato, al quale solo in un secondo tempo si abbina l'analisi delle singole parole. Per questo l'autore del software definisce il suo metodo più vicino alle procedure di analisi del contenuto che agli strumenti lessicometrici (Reinert, 1997). A differenza delle tradizionali analisi del contenuto, però, ALCESTE propone un'analisi della struttura interna di un corpus che non è determinata dalle griglie di post-codifica decise dal ricercatore, ma scaturisce dalla distribuzione delle parole scelte dai locutori per parlare di un certo argomento. Per riprendere la definizione di Reinert (1993), le classi omogenee di enunciati rappresentano la traccia osservabile di "*mondi lessicali*", ovvero di repertori lessicali ai quali i locutori attingono, più o meno consapevolmente, nel discutere di un certo argomento.

Da un punto di vista linguistico e statistico sono state però espresse perplessità sull'impiego di ALCESTE nell'analisi di testi compositi, nati dall'accorpamento delle risposte di soggetti

diversi (Della Ratta, 2000), non essendo a priori possibile sapere se locutori diversi possano considerarsi portatori degli stessi mondi lessicali.

L'opinione di Reinert è naturalmente diversa (Fenoglio e Reinert, 1998), sia dal punto di vista statistico, sia da quello linguistico. Le tracce dei mondi lessicali, infatti, non sono di per sé disponibili ed evidenti sulla superficie del testo, ma possono emergere soltanto in virtù della ripetizione sistematica delle stesse tracce lessicali in contesti di enunciazione diversi: tali regolarità si possono quindi cogliere solo su base statistica (Reinert, 2001), e diventano tanto più informative quanto più si riesce a farle emergere dalle produzioni linguistiche di soggetti diversi. Se poi si considera l'atto linguistico come il prodotto di regole socialmente costruite, anonime e condivise da una comunità di locutori (Aubert-Lotarski e Capdevielle-Mougnibas, 2002), ad essere al centro dell'attenzione non è più soltanto l'individuo che produce un corpus testuale, ma la sua inserzione sociale in una comunità linguistica e in determinati gruppi sociali. Non a caso, lo stesso Reinert definisce i mondi lessicali come luoghi che stanno a metà tra "*les représentations individuelles et les pre-construits sociaux*" (Reinert, 1993: 12). Interrogare più soggetti sugli stessi temi è dunque una strategia appropriata se si vuole far emergere la condivisione sociale dei mondi lessicali.

Tuttavia il principio secondo cui i mondi lessicali si impongono come fondamento topico dei discorsi, indipendentemente dalla volontà dei locutori, orientandone le scelte linguistiche verso alcune parole e non verso altre (Reinert, 2001), può essere condivisibile soltanto quando noi analizziamo discorsi spontanei (opere letterarie, produzioni scritte, discorsi pubblici...), o al massimo nel caso di interviste narrative nelle quali l'intervistatore assuma un ruolo volutamente periferico. Quando il corpus analizzato è invece composto dalle risposte a interviste semi-strutturate, nelle quali le sollecitazioni dell'intervistatore rivestono un peso innegabile, appare semplicistico attribuire soltanto agli intervistati la responsabilità delle scelte lessicali, senza prendere in considerazione il contesto conversazionale nel quale il corpus è stato raccolto.

Nella seconda parte di questo contributo presenteremo brevemente il contenuto delle classi emerse da un'analisi di interviste realizzate con insegnanti italiani, sforzandoci poi di trovare, tra gli indizi forniti dal programma, le tracce dell'interazione discorsiva tra gli insegnanti stessi e l'intervistatore.

2. Un esempio di ricerca : tracce di interazione nelle interviste ad un campione di insegnanti

Il corpus qui descritto è composto dalle risposte di 29 insegnanti italiani a interviste semi-strutturate di durata variabile tra i 45 e i 90 minuti. Le domande vertevano sulle riforme scolastiche promosse dal Ministero dell'Istruzione nel periodo 1997-2000¹, e lo scopo principale della ricerca era quello di identificare, nel discorso prodotto da un campione di insegnanti, l'articolazione tra le novità delle riforme scolastiche e il cambiamento nelle pratiche didattiche quotidiane.

La tabella 1 illustra brevemente l'andamento della Classificazione Discendente Gerarchica e il vocabolario specifico delle classi.

¹ I risultati dettagliati della ricerca sono in fase di pubblicazione sulla rivista *Psicologia dell'Educazione e della Formazione*.

Corpus intero							
χ^2	Classe 1 (261 u.c.e.)	χ^2	Classe 2 (440 u.c.e.)	χ^2	Classe 3 (101 u.c.e.)	χ^2	Classe 4 (115 u.c.e.)
70.35	Scuola	40.76	Bambini	50.64	Orari	65.83	Riunioni
20.43	Nuovo	17.83	Insegnante/insegnare	48.17	Tradizionale	53.65	Organizzazione
16.30	Modello	15.43	Cambiare	47.01	Lezioni	49.04	Organizzativo
12.64	Perplessi	14.81	Difficoltà	40.16	Minuti	24.28	Finanziari
12.47	Autonomia	14.42	Ragazzi	33.71	Quadrimestre	20.86	Elaborare
10.10	Privata	12.65	Genitori	24.41	Flessibilità_oraria	20.86	Svolgere
8.93	Azienda	12.26	Alunni	18.40	Recupero	20.86	Fondi
8.70	Cambiamenti	7.15	Responsabilità	17.53	Scelte	20.51	Progetti
8.58	Politica	7.09	Colleghi	16.78	Pomeriggi	9.40	Consigli_d'Istituto
6.56	Idea	6.55	Dialogo	16.78	Sperimentale	8.44	Problemi
6.40	Pubblica	5.45	Spiegazione	15.68	Decisioni	5.14	Risolvere
6.40	Contrari	4.43	Fatica	9.51	Bimestre	5.14	Costi
4.33	Tradizione	4.02	Ascolto	6.23	Interrogazioni	4.79	Iniziative
4.27	Favorevoli	3.59	Classe	4.90	Noi	4.09	Attività

Tabella 1. Vocabolario specifico delle classi. Il corpus conta 44.465 occorrenze, con 5.702 forme distinte; delle 1171 u.c.e. inizialmente estratte, 917 (pari al 78.31%) risultano stabilmente classificate.

L'analisi produce in primo luogo una completa scissione tra due piani discorsivi: quello legato alle riforme e al cambiamento scolastico (classi 1 e 2), e quello sulla gestione organizzativa della vita della scuola (classi 3 e 4). La prima macro-classe comprende a sua volta due nuclei lessicali ben distinti tra loro. La classe 1 riguarda le riforme propriamente dette, con enunciati che pongono in primo piano l'aspetto ideologico e politico del dibattito sulle riforme varate negli ultimi anni². Contrapposto al lessico sulle riforme, nella classe 2 troviamo l'ambito discorsivo centrato sul cambiamento nelle pratiche di insegnamento. Il contesto di riferimento non è più quello socio-politico, ma quello più strettamente interpersonale: le figure di riferimento fanno tutte parte del contesto interpersonale più immediato del docente (colleghi, dirigenti, genitori) ancorando il discorso al mondo delle interazioni faccia-a-faccia che caratterizzano la vita della singola scuola.

Il secondo dei due macro-contesti fa riferimento all'organizzazione della vita quotidiana della scuola. Troviamo in questo caso enunciati relativi alla scansione temporale delle attività di routine di ogni scuola (classe 3), oppure riferimenti a compiti gestionali e burocratici, tra i quali l'impiego delle esigue risorse finanziarie (classe 4). Tutto questo, però, senza espliciti riferimenti lessicali alle riforme che hanno modificato la gestione delle scuole, né alle ripercussioni didattiche delle decisioni organizzative.

Dal profilo delle classi al lessico dell'interazione

Dopo aver descritto brevemente il vocabolario specifico e l'articolazione tra le classi possiamo chiederci: quanto ha inciso l'interazione con l'intervistatore nel far emergere proprio queste classi, e nello stabilire proprio questi rapporti lessicali tra i diversi ambiti discorsivi?

Naturalmente le domande dell'intervistatore non vengono prese in considerazione nell'analisi ALCESTE. Nel caso di interviste fortemente strutturate, o di domande aperte inserite in un

² Tra le riforme prese in considerazione nel corso dell'intervista ricordiamo: l'autonomia scolastica (ampliamento delle competenze gestionali e didattiche delle singole scuole), il riordino dei cicli (accorpamento della scuola elementare e secondaria inferiore in un unico ciclo di base di 7 anni), il nuovo obbligo scolastico e formativo (innalzamento della frequenza scolastica a 15 anni e della formazione obbligatoria a 18 anni).

questionario standardizzato, si può operazionalizzare l'interazione tra risposte e domande trasformando queste ultime in variabili illustrative (ad esempio codificandole come domanda 1, domanda 2, ecc.) : in questo modo è relativamente semplice andare a verificare se determinate ricorrenze lessicali siano o meno associate al contenuto specifico di una o più domande. Tra l'altro, ALCESTE offre diverse possibilità di analisi supplementari (Analisi Fattoriale delle Corrispondenze, Classificazioni Discendenti Gerarchiche *tri-croisées*, ecc.) attraverso le quali le variabili-domande potrebbero essere sistematicamente poste in relazione con il profilo lessicale delle risposte ottenute.

Nel caso di interviste più aperte, narrative o semi-strutturate, questa strada non è più percorribile, perché l'intervistatore, oltre a introdurre nella conversazione una serie di temi predefiniti, interviene con sollecitazioni ("*mi può spiegare meglio cosa intende con questo...?*", "*può farmi un esempio?*", ecc.) o riformulazioni ("*quindi Lei intende dire che...?*", ecc.), e allo stesso modo l'intervistato può condurre la sua narrazione verso temi del tutto originali rispetto a quelli previsti nella griglia dell'intervista. In questo caso, gli interventi dell'intervistatore non possono essere categorizzati in variabili illustrative, data la loro imprevedibilità, né possono entrare a far parte del corpus testuale analizzato dal software, anche perché la loro ripetitività si rivelerebbe del tutto ridondante dal punto di vista lessicale.

Ciò non toglie che l'esigenza di ricondurre l'output statistico di ALCESTE ad una dimensione più ecologica e interattiva, che tenga conto dei processi di co-costruzione del testo nel corso dell'intervista, sia altrettanto (se non maggiormente) condivisibile a mano a mano che l'interazione discorsiva si fa più libera e imprevedibile. Quelle che proponiamo qui di seguito non sono che due tra le strade percorribili quando il corpus proviene da interviste poco direttive : si tratta in entrambi i casi di piste esploratorie, che non necessitano di ulteriori interventi o manipolazioni e passano semplicemente attraverso un'attenta revisione dei materiali già a disposizione del ricercatore interessato.

La prima strada percorribile è quella di un esame indiretto, a posteriori, del materiale testuale analizzato : una volta descritte le classi prodotte dalla Classificazione Discendente Gerarchica, possiamo riesaminarne i contenuti e l'articolazione interna facendo riferimento alla griglia utilizzata per condurre l'intervista. Non crediamo possa creare scandalo il ricorso a inferenze extra-testuali per guidare il lavoro interpretativo ; è del resto risaputo che il trattamento statistico dei testi implica molti successivi passaggi (a volte frutto di tentativi ed errori) prima di approdare alla preparazione del corpus, alla sua elaborazione numerica e all'interpretazione dei risultati ottenuti (Bolasco, 1999).

Rileggendo le classi alla luce della griglia utilizzata per guidare l'intervista sembra emergere chiaramente che tutte le richieste di pareri e commenti esplicitamente legati alle riforme e ai cambiamenti didattici hanno dato luogo a risposte poi confluite nelle classi 1 e 2, mentre le numerose richieste di esemplificazione legate ad aspetti logistici e organizzativi della vita scolastica hanno dato luogo ad enunciati isolati poi nelle classi 3 e 4. Prima di imputare la scissione delle classi alle scelte lessicali degli intervistati, dobbiamo chiederci in che modo tali scelte siano maturate ; in effetti eravamo noi a sollecitare esemplificazioni riferite alle routines organizzative, perché era nostra convinzione che l'impianto legislativo delle riforme avrebbe comportato ripercussioni concrete ed osservabili proprio a questo livello del funzionamento scolastico.

Il percorso di delle classi mette in luce il fatto che gli insegnanti, almeno a livello lessicale, non hanno costruito lo stesso legame tra i due piani concettuali. Di fronte alle nostre sollecitazioni, gli intervistati hanno enumerato incombenze e scadenze della gestione quotidiana della scuola, ma senza attivare legami lessicali con i temi della riforma e del cambiamento. Se lo avessero fatto avremmo potuto reperire enunciati in cui gli intervistati citavano sia parole relative alle riforme, sia parole relative alle routines gestionali. Semplificando, avremmo

potuto incontrare enunciati come “*con l'autonomia abbiamo modificato gli orari scolastici*”, oppure “*con l'autonomia non abbiamo modificato gli orari scolastici*” : in entrambi i casi ALCESTE avrebbe raggruppato gli enunciati in una classe comune, vista co-occorrenza tra i riferimenti lessicali all'una e all'altra sfera discorsiva. Nel corso delle interviste, però, ciò non è avvenuto : gli insegnanti hanno ampiamente parlato tanto delle riforme quanto delle routines organizzative, ma senza stabilire alcuna contiguità tra le une e le altre all'interno degli stessi enunciati.

Lo scambio verbale che qui riportiamo mette bene in luce l'andamento conversazionale che ha portato a segmentare il discorso sul cambiamento rispetto al discorso organizzativo :

D₁ : PARLIAMO DELL'AUTONOMIA...

R₁ : Lo dicevo anche prima, ha liberato delle energie nella scuola. ad esempio ne parlavo in commissione... abbiamo concretizzato, ha liberato un protagonismo da parte degli insegnanti...

D₂ : SI ASPETTA CHE POSSA ANCORA EVOLVERE NEI PROSSIMI ANNI?

R₂ : spero di sì, perché è una grande carta secondo me, il che vuol dire che la scuola, con il POF, propone certe cose alla società... e quindi è un peccato se vogliamo. Può legarsi bene ai bisogni della società. [...].

D₃ : RISPETTO ALL'ORGANIZZAZIONE DELLE LEZIONI O DEGLI ORARI AVETE MODIFICATO QUALCOSA CON L'AUTONOMIA?

R₃ : C'è qualche margine in cui si può decidere. Abbiamo fatto la flessibilità del sabato, per dire, che consiste nel chiudere la scuola alle dieci e mezza per il biennio. Adesso probabilmente ci presteremo per fare la settimana corta, senza il sabato (*matematica, superiori, oltre 50 anni, maschio, nessun incarico*).

Quadro 1. Esempio di scambio sulle conseguenze organizzative di una riforma.

Lo scambio appare in qualche misura forzato : di fronte alla suggestione sulla riforma dell'autonomia, l'insegnante esprime la sua opinione sulle opportunità di cambiamento, sulla relazione scuola-società e sulle aspettative future (D₁-R₁ e D₂-R₂) ; quando noi domandiamo se ci sono state ripercussioni organizzative e gestionali (D₃), l'insegnante risponde con eguale disponibilità, ma il discorso scivola verso un altro ambito di contenuto, che è quello dei tempi della scuola e delle relative negoziazioni (R₃) : in effetti, quella che per noi era una semplice richiesta di esemplificazione, per l'intervistato ha comportato una ridefinizione del tema della conversazione.

La distanza tra i due ambiti discorsivi si traduce, sul piano metrico e lessicale, in una netta separazione tra classi i cui enunciati non presentano riferimenti in comune. La genesi di questa separazione va riconosciuta nell'interazione tra una rappresentazione del fenomeno introdotta dall'intervistatore (le cui domande non sono mai neutre), una rappresentazione costruita dall'intervistato (che forse non aveva mai esplicitamente pensato ad un legame tra riforme e routines organizzative), e un contratto di comunicazione asimmetrico che lega i due protagonisti (l'intervistatore chiede, il soggetto risponde).

Lo stralcio riportato presuppone il ricorso a risorse diverse da quelle fornite dal software, in particolare una rilettura attenta delle trascrizioni originali delle interviste. La seconda strada che proponiamo di seguire prende invece in considerazione alcuni elementi forniti direttamente dall'output statistico di ALCESTE.

Abbiamo detto che il testo delle domande viene volutamente escluso dalle analisi, vista la sua ridondanza lessicale, e per ottenere tale esclusione è sufficiente trascrivere gli interventi dell'intervistatore in caratteri maiuscoli. Nel corso della suddivisione automatica degli enunciati può accadere che alcuni segmenti di domande vengano inclusi nelle unità di testo sottoposte ad analisi. In questo caso il testo delle domande, scritto in stampatello, non concorre all'analisi e alla formazione delle classi, ma viene ripreso in un secondo momento come

variabile supplementare : in particolare, il software calcola a posteriori la distribuzione delle parole scritte in stampatello e, nel caso tali parole compaiano più frequentemente negli enunciati di una classe rispetto al resto del corpus, provvede ad inserirle in coda al vocabolario specifico della classe. Nelle tabelle 2 e 3 sono elencate le parole in maiuscolo (cioè provenienti da spezzoni di domande) confluite rispettivamente nel vocabolario delle classi 1 e 2 ; nelle tabelle riportiamo anche l'eventuale presenza degli stessi vocaboli nel lessico delle risposte.

CLASSE 1			
Domande	Risposte	Domande	Risposte
ASPETTA	aspett+	PAROLA	---
AUTONOMIA	autonomi+	PROSSIMI	---
OBIETTIVI	obiettiv<	RIORDINO	---
PROPOSTO	propo<	SITUAZIONE	---
DALLA	---	STUDIATO	---
ENTRATA	---	VIGORE	---

Tabella 2. Parole utilizzate nelle domande e associate al vocabolario specifico delle classi 1 e 2 ; in corsivo le stesse parole reperite nel vocabolario specifico delle classi. Le parole provenienti dalle domande non vengono lemmatizzate da ALCESTE, dal momento che la loro attribuzione alle classi ha finalità puramente illustrative.

CLASSE 2			
Domande	Risposte	Domande	Risposte
ALUNNI	alunn+	RAGAZZI	ragazz<
BAMBINI	bambin<	RAGAZZINI	
CAMBIARE		RIFERIMENTO	riferimento
CAMBIATI	cambi+	BENE	---
CAMBIATO		DIRIGENTE	---
CAMBIEREBBE		FARE	---
CHIEDE	chied+	INCERTEZZA	---
CHIESTO		INTERNO	---
COLLEGA	collegh+	IPOTETICO	---
GENITORE	genitor<	LAVORARE	---
GENITORI		METODO	---
INSEGNAMENTO		PICCOLI	---
INSEGNANTE	insegn+	SCUOLA	---
INSEGNANTI	[professo<]	SCUOLE	---
INSEGNARE		SETTEENNIO	---

Tabella 3. Parole utilizzate nelle domande e associate al vocabolario specifico delle classi 1 e 2 ; in corsivo le stesse parole reperite nel vocabolario specifico delle classi.

Limitiamo l'esemplificazione alle prime due classi. È evidente che parte del lessico utilizzato dagli intervistati è stato introdotto dal ricercatore attraverso la formulazione delle domande, anche se la ripetizione di forme identiche si limita tutto sommato a pochi vocaboli.

Per quanto utile, l'esame dei frammenti di domande non fornisce indizi del tutto esaustivi, dal momento che l'inclusione di tali frammenti negli enunciati analizzati riveste carattere del tutto aleatorio³. Non solo, ma fermarsi al confronto del lessico delle domande con quello delle

³ ALCESTE offre la possibilità di introdurre le parole in maiuscolo come variabili supplementari nell'Analisi Fattoriale delle Corrispondenze, rendendo così visibile e quantificabile il legame tra lessico delle domande (o almeno di quella porzione di domande incidentalmente inclusa nell'analisi) e lessico del corpus analizzato. Si tratta di una possibilità di approfondimento potenzialmente molto promettente, che però al momento incontra almeno due limiti nella sua applicazione. In primo luogo, come detto, l'inclusione di spezzoni di domande nelle u.c.e. riveste carattere alquanto aleatorio, e non si presta quindi ad un'analisi sistematica e del tutto affidabile. In secondo luogo, la visualizzazione dei piani fattoriali attualmente permessa da software non appare sufficiente-

risposte può indurre a stabilire nessi causali non corretti, e comunque incompatibili con una visione dell'intervista come co-costruzione situata : così come non vogliamo ignorare il ruolo attivo dell'intervistatore, allo stesso modo non possiamo immaginarci un intervistato passivo e manipolato nelle sue risposte.

Se, anziché limitare il confronto alle singole parole, spostiamo il confronto al livello degli enunciati, che costituiscono a tutti gli effetti l'unità di analisi di ALCESTE, scopriamo che la stessa domanda dell'intervistatore può attivare scelte lessicali del tutto diverse :

U.C.E. 317 : SECONDO LEI, NEL PROSSIMO FUTURO, IL RUOLO DELL' INSEGNANTE È DESTINATO A CAMBIARE? sì, l' #insegnante deve diventare un tuttologo e deve diventare #sempre #di-più un #riferimento per la vita quotidiana dell' #alunno, visto che ormai tutti i #genitori lavorano. (*matematica, medie, 36-50 anni, femmina, collaboratrice del dirigente ; u.c.e. della CLASSE 2*).

U.C.E. 1018 : SECONDO LEI IL RUOLO DELL' INSEGNANTE È DESTINATO A CAMBIARE? #secondo-me bisogna #vedere #qual è il progetto della #scuola #che #viene fuori, perché se un tempo il ruolo #sociale dell'insegnante era un ruolo #comunque #riconosciuto... (*lettere, superiori, 36-50 anni, femmina, nessun incarico ; u.c.e. della CLASSE 1*).

Quadro 2. Esempi di scambio ; u.c.e. estratte a partire dalla stessa domanda.

Entrambe le risposte riprendono un vocabolo presente nella domanda (*insegnante*), ma lo sviluppo che ne segue va in direzioni del tutto diverse, attivando (nei termini di ALCESTE) due mondi lessicali distinti. Nel primo caso l'insieme dei riferimenti lessicali richiama il mondo delle relazioni interpersonali che ruotano intorno alla scuola e nelle quali è immerso l'"insegnante" evocato dall'intervistatore ("*genitori*", "*alunno*", ecc.) ; in effetti, la risposta confluisce negli enunciati che compongono la classe 2. Nel secondo caso la prospettiva evocata non è più quella delle relazioni immediate, ma è quella che potremmo definire sovra-ordinata e che dall'"insegnante" arriva a comprendere la "*scuola*" e la "*società*" nel suo complesso. La risposta, in questo caso, va naturalmente a confluire tra gli enunciati della classe 1.

3. Conclusioni

Utilizzando un software statistico (nel nostro caso ALCESTE) per analizzare testi per loro natura dialogici, quali sono quelli prodotti nel corso di una o più interviste, ci siamo chiesti quali cautele dovessero essere prese per rendere conto, almeno in parte, della dimensione ecologica del corpus raccolto, soprattutto per quel che concerne le sue condizioni di produzione. Per sua natura, l'intervista costituisce una pratica discorsiva che coinvolge almeno due soggetti, ciascuno dei quali è attivo costruttore di conoscenza nell'ambito di uno specifico contratto di comunicazione ; l'attività di co-costruzione nella quale i due attori sono impegnati si manifesta a livello pragmatico, contenutistico ed anche sul piano prettamente lessicale, incidendo sui risultati che l'elaborazione statistica del testo fornisce alla nostra interpretazione.

La prima strada che abbiamo proposto per un riesame critico dei risultati di ALCESTE prevede il ritorno al testo originario, confrontando criticamente la segmentazione degli enunciati ottenuta attraverso la Classificazione Discendente Gerarchica, con la scansione dei temi proposti dal ricercatore nel corso dell'intervista. Il secondo percorso di revisione prende invece in considerazione indizi presenti nell'output statistico fornito dal software, ed in particolare si concentra sulle eventuali corrispondenze tra lessico impiegato nella formulazione delle

mente chiara e richiede un cospicuo lavoro di ripulitura da parte dell'utente. L'AFC ci sembra comunque una delle strade più promettenti in vista di un più rigoroso esame su base statistica delle interazioni discorsive.

domande (inserito a posteriori nei risultati come elemento illustrativo supplementare), e lessico utilizzato dagli intervistati nelle loro risposte. Questa seconda opzione presenta elementi estremamente aleatori e deve essere senz'altro raffinata ; ciò non di meno entrambe le strade consentono di individuare alcune forzature nelle scelte lessicali degli intervistati.

Il legame tra domande e risposte non va visto nè in termini neutrali, nè in termini deterministici : non ci sembra corretto analizzare un corpus di interviste ignorando o sopprimendo *tout court* il ruolo dell'intervistatore, né immaginare che le risposte degli intervistati seguano pedissequamente le suggestioni delle domande. Assumere una prospettiva dialogica e interazionista permette non solo di studiarne l'organizzazione interna di un testo analizzato su base statistica, ma soprattutto di individuare i punti in cui questa organizzazione si discosta da quella immaginata a priori (e implicitamente veicolata) dal ricercatore. Un interessante sviluppo della ricerca in questa direzione potrebbe partire dal confronto sistematico (nella forma, nella struttura e nel contenuto) di *corpora* prodotti nel corso di interazioni dialogiche gestite in modi diversi, ad esempio comparando tra loro interviste condotte con un grado di strutturazione più o meno forte. A questo scopo, diversi approcci di indagine, a partire dai classici strumenti della lessicometria (orientati allo studio della ricchezza e della diversificazione del lessico prodotto in situazioni diverse), per arrivare agli sviluppi più recenti delle analisi discriminanti e delle funzioni di regressione, potrebbero fornire indizi importanti per un ulteriore approfondimento teorico e metodologico .

Naturalmente, quelle che abbiamo illustrato non sono che due tra le strade percorribili da chi voglia ancorare l'analisi statistica di un testo alle concrete condizioni nelle quali il testo stesso è stato prodotto. Abbiamo proposto, a titolo di esempio, due strategie che qualsiasi utente di ALCESTE può impiegare a partire da un riesame critico dell'output statistico e/o del corpus di partenza, ben sapendo che il problema affrontato meriterebbe un approfondimento ben maggiore. Un ulteriore sviluppo nelle tecniche multidimensionali per il trattamento di dati testuali potrà senz'altro offrire al ricercatore interessato strumenti di analisi più sistematici, rendendo più agevole e proficuo l'esame congiunto di un corpus testuale e dell'interazione sociale che ad esso ha dato origine.

Riferimenti bibliografici

- Atkinson P. e Coffrey A. (2002). Revisiting the relationship between participant observation and interviewing. In Gubrium J.F. e Holstein J.A. (Eds), *Handbook of interview research*. Thousand Oaks, Sage : 800-814.
- Aubert-Lotarski A. e Capdevielle-Mougnibas V. (2002). Dialogue méthodologique autour de l'utilisation du logiciel Alceste en sciences humaines et sociales : 'lisibilité' du corpus et interprétation des résultats. In *Actes des JADT 2002* : 45-56.
- Bolasco S. (1999). *Analisi multidimensionale dei dati*. Carocci.
- Della Ratta F. (2000). L'analisi testuale : uno strumento per la ricerca sociale. *Sociologia e ricerca sociale*, vol. (61) : 102-127.
- Fenoglio E. e Reinert M. (1998). Application d'*Alceste* à un corpus d'entretiens : de l'analyse de données à l'analyse de l'énonciation. In *Actes des JADT 1998* : 223-229.
- Gubrium J.F. e Holstein J.A. (2002). Introduction. In Gubrium J.F. e Holstein J.A. (Eds), *Handbook of interview research*. Thousand Oaks, Sage : 3-32.
- Hammersley M. (2003). Recent radical criticism of interview studies : any implication for the sociology of education? *British Journal of Sociology of Education*, vol. (24/1) : 119-126.
- Lebart L. e Salem A. (1988). *Analyse statistique des données textuelles : questions ouvertes et lexicométrie*. Dunod.
- Lebart L. e Salem A. (1994). *Statistique Textuelle*. Dunod.

- Py B. e Grossen M. (1997). Introduction Générale : Interactions, médiations et pratiques sociales. In Grossen M. e Py B. (sous la direction de), *Pratiques sociales et médiations symboliques*. Peter Lang.
- Reinert M. (1986). Un logiciel d'analyse lexicale. *Les Cahiers de l'Analyse des Données*, vol. (4) : 471-484.
- Reinert M. (1993). Les « mondes lexicaux » et leur « logique » à travers l'analyse statistique d'un corpus de récits de cauchemars. *Langage et Société*, vol. (66) : 5-37.
- Reinert M. (1997a). Les "mondes lexicaux" de six numéros de la revue « Le Surréalisme au service de la Révolution ». *Mélusine, L'Age d'Homme*, vol. (XVI) : 270-302.
- Reinert M. (1998). Mondes lexicaux et topoi dans l'approche Alceste. In Mellet S. e Vuillaume M. (sous la direction de), *Mots chiffrés et déchiffrés*. Honoré Champion.
- Reinert M. (2001). *Alceste*, une méthode statistique et sémiotique d'analyse de discours ; application aux « *Rêveries du promeneur solitaire* ». Papier présenté au *XIIème colloque de Psychiatrie de Marseille*, 17-19 Mai 2001.