Obiettivi

Il nostro lavoro si propone di mostrare la rilevanza e la ricchezza euristica dell'interazione tra due discipline tradizionalmente distanti: la linguistica e la storia. In particolare intendiamo sottolineare la produttività di strumenti linguistici come la linguistica dei corpora e l'analisi della conversazione nell'analisi di corpora di narrazioni conversazionali, fonti su cui si basa la storia orale.

Metodo

Il corpus è costituito dall'insieme delle interviste raccolte nell'arco dell'ultimo decennio a uomini e donne che hanno partecipato in varie sedi universitarie italiane al Sessantotto. Si tratta di interviste orali non strutturate che non hanno alle spalle la riflessione di chi scrive, corregge, pensa e ritorna sul suo testo. Le interviste sono state tutte digitalizzate e sono 63, pari a 842.378 occorrenze e a 37.443 parole; a ciascuna intervista sono associate una serie di note etnografiche riferite alla persona intervistata: il genere sessuale, l'anno di nascita, il luogo di nascita, la sede universitaria nel 68, la facoltà d'appartenenza, il titolo di studio superiore, la professione del padre, la professione della madre, la condizione professionale o lavorativa al momento dell'intervista, le origini regionali. Il corpus è stato analizzato per mezzo di un software (TaLTaC2) di analisi automatica del testo. Tale analisi consente di dare delle rappresentazioni del fenomeno studiato su base quantitativa sia a livello di unità di testo (parole) sia a livello di unità di contesto (frammenti/documenti), quindi come linguaggio utilizzato e come contenuti trattati nel testo. Il programma è stato ideato e elaborato da un team internazionale coordinato da Sergio Bolasco. TaLTaC2 utilizza risorse sia di tipo statistico sia di tipo linguistico, altamente integrate fra loro, e consente un'indagine a due livelli, lessicale e testuale, ovvero l'analisi del testo (text analysis) e il recupero e l'estrazione d'informazione, secondo i principi del data mining e del text mining. Il primo dato prodotto dal software è il vocabolario complessivo delle interviste: un elenco di tutte le parole presenti nel corpus, ordinato in una tabella secondo l'ordine di frequenza, ovvero secondo il numero di volte in cui quelle parole ricorrono all'interno del corpus complessivo delle interviste. Il vocabolario caratteristico o vocabolario specifico è un sottoprodotto del vocabolario complessivo del corpus e del suo incrocio con I tratti etnografici, ovvero il vocabolario specifico dei gruppi in cui può essere suddiviso il corpus: è la lista delle parole specifiche di quel gruppo, quelle che lo caratterizzano rispetto agli altri. È basato su un algoritmo che ci dice se la presenza di una determinata parola in quella determinata partizione è sovradimensionata rispetto alle attese (ricorre con una frequenza più alta che altrove) senza che questo risultato sia dovuto al puro effetto del caso (sulla base di un valore che si chiama p-value, valore di probabilità: più basso è questo valore, più alta è la specificità della parola). Sono insomma le parole proprie di quel gruppo. Il vocabolario specifico è il dato primario su cui si è fondata la nostra analisi. Ogni volta sono state prese in considerazione le prime 500 parole di ciascuna lista delle specificità. Per ciascuna di quelle parole, attraverso lo strumento delle concordanze, è possibile risalire rapidamente al contesto testuale in cui è stata pronunciata e al nome della persona intervistata. È il momento in cui l'anonimato del vocabolario specifico di un gruppo si ritrae per restituire al soggetto la sua individualità e il suo racconto. Altri strumenti di analisi sono le co-occorrenze e il linguaggio peculiare. Le co-occorrenze sono le principali associazioni tra una coppia di termini che compaiono vicini nel testo entro un ridotto numero di parole: i due termini della coppia possono essere contigui o intercalati da altre parole. Le co-occorrenze ci dicono quali associazioni di parole sono più frequenti in un testo, o in un insieme di testi. Il linguaggio peculiare poggia sul valore di peculiarità che è dato, per ogni parola del corpus, dal confronto con un altro vocabolario sulla base dei rispettivi indici di frequenza. Nel caso di TaLTaC2 il vocabolario di confronto è costituito da dieci annate del quotidiano «la Repubblica» degli anni '90: un corpus quantitativamente rappresentativo di una forma di comunicazione (il testo giornalistico scritto) che più di altri si presta a rispecchiare con una certa ampiezza gli usi standard medio-alti dell'italiano. Se per collocazione sociale degli intervistati la lingua del corpus è sicuramente medio-alta, diverso è invece il registro comunicativo tra i due vocabolari: di fronte a questo limite sostanziale del confronto con il linguaggio peculiare, sono stati utilizzati i risultati delle peculiarità solo come valori indicativi. La trascrizione delle interviste, pur non avendo fatto ricorso alla trascrizione Jefferson, tipica dell'analisi della conversazione, si ispira agli stessi criteri filologici di quella. È stata adottata una trascrizione ortografica che permette la leggibilità del testo e sono stati trascritti i fenomeni paralinguistici delle pause, dei cambi di progetto come le false partenze o le interruzioni, e fenomeni caratteristici del parlato come le ripetizioni. Ciascuna trascrizione è integrata con delle glosse e dei commenti

che descrivono comportamenti non verbali come le risate e i pianti nel continuum che va dalla risatina al riso pieno, dal silenzio di commozione al pianto.

Oggetto dell'indagine

Nell'analisi dei trascritti, attraverso il vocabolario specifico per appartenenza di genere, è emersa una occorrenza maggiore di risate nelle interviste alla donne, tanto che nelle specificità femminile compaiono "ride" al secondo posto e "ridono" al ventiseiesimo. Diverse le ipotesi interpretative: un segno di leggero distanziamento da ciò che si dice, un modo di sottrarsi a toni autocelebrativi, oppure una implicita, non verbalizzata complicità di genere, tra l'intervistatrice e le intervistate? In ogni caso un tratto distintivo su cui intendiamo riflettere nel presente lavoro. Ci occuperemo pertanto del riso nelle diverse posizioni nei turni all'interno del dialogo prodotte dai partecipanti all'interazione, ovvero le intervistate, gli intervistati e l'intervistatrice. L'analisi della conversazione ha messo in evidenza il carattere estremamente pervasivo della risata nell'interazione umana e il suo ruolo multifunzionale a seconda delle posizione all'interno dei turni della sequenza interazionale. In base al carattere condiviso o meno del riso, la risata può essere interpretata come espressione di affiliazione, intimità, resistenza, distacco e rappresentazione delle identità. Sulla base di un attento esame che, partendo dalle glosse di commento, ritorna sul dato primario della registrazione dell'intervista, si cercherà di individuare il ruolo svolto dall'espressione della commozione nel corpus orale analizzato. Il nostro obiettivo è pertanto mostrare il contributo che l'analisi di un fenomeno così fine dell'interazione può fornire nell'individuazione delle identità e della soggettività che intervistate e intervistati costruiscono attraverso il loro racconto, in un procedimento per cui dal microcosmo dell'interazione si risale al macrocosmo della società in un determinato periodo storico così come è vissuta nella memoria del singolo soggetto.

Bibliografia

P. Glenn e E. Holt eds., *Studies on Laughter in Interaction*, London: Bloomsbury, 2013

A. Portelli, *Storie orali, Racconto, immaginazione, dialogo*, Roma: Donzelli, 2017

S. Bolasco, *L'analisi automatica dei testi. Fare ricerca con il text mining*, Carocci: Roma, 2013