



Università di Ain Shams
Facoltà di Lingue (Al-Asun)
Dipartimento di Italianistica

Tesi di Master

Analisi dei segnali discorsivi nel corpus di parlato spontaneo

(C-ORAL-ROM)

Studio pragmatico

Presentata da:

Doaa Mortada Behery ElSayed

Assistente presso il Dipartimento d'Italianistica

Relatrice

Prof.ssa Sawsan Aly Zein El Abedin

Ordinario di linguistica italiana presso il Dipartimento d'Italianistica

Correlatrice

Prof.ssa Nermin Abd Al Hamid

Professore associato di linguistica italiana presso il Dipartimento d'Italianistica

Il Cairo 2019

Ringraziamenti

Innanzitutto ringrazio Dio di avermi concesso pazienza per concludere questa tesi.

Vorrei porgere i miei ringraziamenti alla relatrice, Prof.ssa Sawsan Aly Zein Al Abedin, che è riuscita a darmi la carica per proseguire e portare a termine questa tesi che non avrebbe mai visto la luce senza la Sua disponibilità.

Un grazie particolare va alla Prof. ssa Nermin Abd Al Hamid che è stata sempre disponibile a dirimere i miei dubbi durante la stesura di questo lavoro.

Ringrazio di cuore il prof. Sherif Abo El Makarem per avermi fornito testi importanti per la realizzazione della tesi e per aver accettato di percorrere tutta quella distanza per partecipare alla discussione.

Ringraziamenti infiniti alla prof.ssa Maha Mohammed Abdel Aziz per aver accettato di partecipare alla discussione e per l'impegno svolto nella lettura della tesi.

Desidero esprimere la mia gratitudine al prof. Alessandro Panunzi per la sua professionale assistenza per risolvere i problemi tecnici riscontrati nell'uso del corpus e della Banca Dati IPIC. Ringrazio pure la prof.ssa Ilaria Fiorentini, la prof.ssa Elisabetta Jafrancesco, Antonella Luzziconti e Filippo Panzuto per mettere a mia disposizione testi importanti per la realizzazione della tesi.

Vorrei porgere i miei ringraziamenti infiniti alla mia famiglia, soprattutto ai miei genitori e al marito Mahmoud Mohammad che non mancano mai di sostenermi in tutto e per tutto, e che mi sono sempre vicini.

Uno speciale ringraziamento va a Neama Abdel Aty che mi ha incoraggiato in tutti i modi possibili fin dalle prime fasi di elaborazione della tesi.

Ringrazio di cuore Gehad Mohammed e Zeinab Anas che mi sono state sempre vicine e mi hanno sostenuto nei momenti più difficili.

Ringrazio sentitamente e cordialmente le mie amiche Maria Oliveri, Mai Morsi e Rahma Shaarawy per il loro sostegno e incoraggiamento.

Abstract

La tesi si divide in due capitoli oltre all'introduzione e alle conclusioni.

L'introduzione evidenzia l'interesse dei linguisti verso il fenomeno dei SD, il nucleo dell'idea e lo scopo della tesi. Di seguito, fornisce una breve presentazione del corpus utilizzato, tesa a chiarirne la composizione e la strategia di organizzare i materiali raccolti in modo tale da facilitare sia il reperimento dei dati sia la loro analisi. Poi, espone la teoria adottata (Teoria della Lingua in Atto) che serve a fornire un quadro di riferimento per l'analisi dei SD.

Il primo capitolo mette in considerazione alcuni tratti teorici, come: la distinzione tra connettivi semantici e quelli pragmatici; un accenno sulla prospettiva pragmatica, che ritengo necessaria per trattare una categoria così funzionale (i SD); alcuni problemi teorici concernenti le varietà di denominazione e la distinzione tra codifica concettuale e quella procedurale. Poi, mette in rassegna gli approcci dei linguisti allo studio del fenomeno dalle sue prime fasi, dando più spazio al modello di Bazzanella che ha avuto un ruolo pionieristico in questo campo. Di seguito, vengono elencate le caratteristiche dei SD.

Il secondo capitolo è dedicato all'analisi stessa dei SD; si accentuano le funzioni la cui analisi non può prescindere dalle particolarità di questa classe funzionale. In questo punto, partendo da tre macrofunzioni, si sistematizzano le principali funzioni svolte dai SD nel corpus, determinando le differenze funzionali, strettamente legate al contesto in cui appaiono e ad altri parametri e mostrando come un elemento discorsivo con una determinata funzione possa acquisirne un'altra se cambiano certe condizioni prosodiche e contestuali.

Nelle conclusioni abbiamo messo in rilievo i risultati raggiunti.

Parole chiavi:

Parlato spontaneo - polifunzionalità - funzioni pragmatiche - funzioni interazionali-
funzioni metatestuali - funzioni cognitive - contesto.

Abbreviazioni e segni

DM	Discourse marker
MP	Marcatore pragmatico
S1	Il segmento precedente
S2	Il segmento seguente
SD	Segnale discorsivo
TU	Unità intonativa
UI	Unità informativa
v.	Vedi

Unità informative¹

ALL (Allocutivo): Si utilizza per richiamare l'interlocutore per assicurare la sua attenzione all'enunciato.

APC, APT (Appendice): Integra testualmente l'unità di *Comment* (APC) o di *Topic* (APT) con informazioni accessorie.

COM (*Comment*): Esprime la forza illocutiva dell'enunciato, cioè rappresenta lo scopo dell'enunciato stesso.

CNT (Conativo): Incita l'interlocutore a prendere parte allo scambio.

EMP : Indica che l'unità informativa è interrotta (come nei casi di false partenze o di interruzioni da parte dell'interlocutore).

EXP (Espressivo): Stimola l'interlocutore a condividere un comune punto di vista sull'enunciato.

PHA (Fatico): Regola il canale comunicativo.

¹ Le unità informative sono sigle di tre lettere che si scrivono in apice della parola e si utilizzano, secondo la teoria della lingua in atto, per analizzare i corpora di parlato spontaneo.

INT (Introduttore locutivo): Introduce il discorso diretto riportato, le esemplificazioni, le liste, ecc.

INP (*Incipit*): Segnala la presa di turno da parte del parlante.

PAR (Inciso): Inserisce una valutazione di tipo metalinguistico sul testo dell'enunciato.

SCA : Indica un'unità intonativa che non ha valore informativo necessario.

TMT (*Timetaking*): Indica che il locutore ha difficoltà nel pianificare il suo messaggio o nel prendere il turno.

TOP (*Topic*): Specifica il campo di applicazione della forza illocutiva espressa dal *Comment*.

UNC (*Unclassified*): Indica che l'unità informativa non è classificabile.

Segni di trascrizione

// (doppia sbarra): indica la fine di un enunciato, cioè una pausa prosodica conclusiva.

/ (sbarra semplice): segnala la fine di unità informativa, cioè una pausa prosodica non conclusiva.

+ : segna sequenze interrotte involontariamente, in quanto il programma del parlante viene interrotto.

[/]: identifica una pausa prosodica non conclusiva dovuta ad un falso inizio.

[//]: indica una pausa prosodica non conclusiva dovuta ad un falso inizio con ripetizione parziale di materiale linguistico.

[///]: segna una pausa prosodica non conclusiva dovuta ad un falso inizio senza ripetizione parziale di materiale linguistico.

(cannelletto): indica una pausa chiaramente percepibile e misurabile durante il flusso del discorso, o comunque, un silenzio di una durata più di 250 ms (# = pausa breve e ## = pausa lunga).

. . . (i tre puntini): segnalano la soppressione di un passaggio.

[] (le parentesi quadre): racchiudono i commenti dell'analista.

< > (le parentesi aguzze): segnano l'inizio e la fine di un testo sovrapposto del parlante.

[<]: apparisce all'inizio del secondo turno, immediatamente prima del testo sovrapposto.

&: segna frammenti di discorso.

hhh : indica le risate.

xxx: segna una parola incomprensibile.

yyy: segnala una parola non trascritta per ragioni di *privacy* o decenza.

Indice

Introduzione.....	12
Corpus di riferimento (C-ORAL-ROM Italia).....	14
Teoria della Lingua in Atto.....	23
Primo capitolo: Aspetti teorici e prototipici dei SD.....	32
1. Distinzione tra connettivi semantici e pragmatici.....	32
2. Verso una prospettiva pragmatica.....	35
3. Problemi teorici dei SD.....	36
3.1. Varietà di denominazioni e delimitazioni di campo.....	37
3.1.1. Varietà di etichette in italiano.....	37
3.2. Distinzione tra codifica concettuale e procedurale.....	38
4. Modelli dei linguisti sul fenomeno.....	39
4.1. Approcci generali al fenomeno nella trattazione internazionale.....	39
4.2. Studio del fenomeno in contesto italiano.....	43
a. Monica Berretta.....	43
b. Klaus Lichem.....	44
c. Edith Mara.....	44
d. Patrizia Manili.....	45
e. Sorin Stati.....	45
f. Piera Molinelli.....	46
g. Elizaveta Khachaturyan.....	48
h. Angela Ferrari.....	49
i. Carla Bazzanella.....	49
5. Caratteristiche dei SD.....	53
5.1. Eterogeneità categoriale.....	53
5.2. Indipendenza dei SD (Caratteri formali).....	54
5.2.1. Indipendenza morfologica.....	55
5.2.2. Indipendenza sintattica e semantica.....	56
5.2.2.1. Esteriorità al contenuto proposizionale.....	57
5.2.2.1.1. Eliminabilità.....	58
5.2.2.1.2. Intersostituibilità.....	62
5.2.3. Indipendenza prosodica.....	62
5.2.3.1. Caratteristiche distribuzionali.....	64
5.2.3.1.1. Varietà di collocazione.....	64
5.2.3.1.2. Ripetizioni.....	68
5.2.3.1.2.1. Tic verbali.....	68
5.2.3.1.2.2. Cumuli.....	78
5.2.3.1.2.3. Catene.....	79

5.3. Carattere funzionale.....	84
5.3.1. Polifunzionalità.....	84
5.3.1.1. Polifunzionalità a livello paradigmatico.....	84
5.3.1.2. Polifunzionalità a livello sintagmatico.....	86
Secondo capitolo: Funzioni pragmatiche dei SD.....	90
1. Funzioni pragmatiche dei SD.....	90
1.1. Macrofunzione interazionale.....	91
1.1.1. Dalla parte del parlante in corso.....	92
1.1.1.1. Presa di turno.....	92
1.1.1.2. Tenuta del turno.....	99
1.1.1.3. Riempitivi.....	103
1.1.1.4. Richiesta di attenzione.....	109
1.1.1.5. Fatismi.....	117
a. Termini di parentela.....	119
b. Titoli.....	120
c. Vocativi e connotazioni.....	121
1.1.1.6. Controllo della ricezione.....	122
1.1.1.7. Richiesta di accordo e / o conferma.....	126
1.1.1.8. Cessione del turno.....	127
1.1.2. Dalla parte dell'interlocutore.....	128
1.1.2.1. Meccanismi di interruzione.....	129
1.1.2.2. <i>Back channels</i>	132
1.1.2.3. Ricezione e acquisizione di conoscenza.....	134
1.1.2.4. Accordo e / o conferma.....	136
1.1.2.5. Richiesta di spiegazione.....	138
1.2. Macrofunzione metatestuale.....	140
1.2.1. Demarcativi.....	141
1.2.1.1. Articolazione in parti.....	141
1.2.1.1.1. Iniziatori.....	141
1.2.1.1.2. Proseguitori.....	144
1.2.1.1.3. Finalizzatori.....	147
1.2.1.1.4. Digressione.....	149
1.2.2. Indicatori di riformulazione.....	151
1.2.2.1. Indicatori di parafrasi.....	152
1.2.2.2. Indicatori di correzione.....	155
1.2.2.3. Indicatori di esemplificazione.....	157
1.3. Macrofunzione cognitiva.....	158
1.3.1. Meccanismi di modulazione (Modulatori).....	159
1.3.1.1. Mitigatori.....	159
1.3.1.2. Rafforzatori.....	162

Conclusioni.....	165
Bibliografia.....	171

Premessa

La presente tesi si divide in introduzione, due capitoli e conclusioni. Nell'introduzione ho evidenziato l'interesse dei linguisti verso il fenomeno dei segnali discorsivi (d'ora in poi SD), il nucleo dell'idea e lo scopo della tesi. Di seguito, ho fornito una breve presentazione del corpus utilizzato, tesa a chiarirne la composizione e la strategia di organizzare i materiali raccolti in modo tale da facilitare sia il reperimento dei dati sia la loro analisi. Poi, ho esposto la teoria adottata (Teoria della Lingua in Atto) che serve da base a fornire un quadro di riferimento per l'analisi dei SD e a tentare di identificarli e determinare le loro funzioni.

Nel primo capitolo ho inizialmente preso in considerazione alcuni tratti teorici, come: la distinzione tra connettivi semantici e quelli pragmatici; un accenno sulla prospettiva pragmatica, che ritengo necessaria per trattare una categoria così funzionale (i SD); alcuni problemi teorici concernenti le varietà di denominazione e la distinzione tra codifica concettuale e quella procedurale. Poi, ho puntualizzato gli approcci dei linguisti allo studio del fenomeno dalle sue prime fasi, dando più spazio al modello di Bazzanella che ha avuto un ruolo pionieristico in questo campo e che ritrovo con stima alle radici del mio percorso di ricerca. Di seguito, vengono elencate le caratteristiche dei SD.

Il secondo capitolo è dedicato all'analisi stessa dei SD; ho accentuato le funzioni la cui analisi non può prescindere dalle particolarità di questa classe funzionale. In questo punto, partendo da tre macrofunzioni, ho sistematizzato le principali funzioni svolte dai SD nel corpus ed ho determinato le differenze funzionali, strettamente legate al contesto in cui appaiono e ad altri parametri, mostrando come un elemento discorsivo con una determinata funzione possa acquisirne un'altra se cambiano certe condizioni prosodiche e contestuali.

Nelle conclusioni ho messo in rilievo i risultati raggiunti.

Introduzione

Quando si parla, per trasmettere il contenuto al destinatario, si devono non solo fornire informazioni sufficienti, ma anche elaborare il discorso per aiutarlo a riconoscere le varie parti del discorso e le connessioni tra di esse. Ciò si verifica utilizzando quegli elementi linguistici di carattere colloquiale, quali i SD. Questi elementi non hanno una funzione grammaticale, ma costituiscono l'organizzazione del discorso e possono essere chiamati elementi della lingua parlata reale poiché, come ribadisce Lichem (1985), quando un autore inserisce tali elementi nei suoi dialoghi, "è evidente che cerca di essere naturale e di avvicinarsi alla realtà della lingua parlata" (p. 214).

Con questo lavoro mi propongo di affrontare queste espressioni linguistiche spesso trascurate dalla grammatica e dalla linguistica tradizionali, orientate per lo più all'analisi della lingua scritta. A partire dagli anni Ottanta, i nuovi orizzonti - che la linguistica apre con l'analisi del testo e del discorso, delle funzioni linguistiche e della comunicazione verbale - affermano che ci si discosta sempre di più dalla morfosintassi formale tradizionale e si moltiplicano, soprattutto in ambito anglofono, gli studi sulle manifestazioni linguistiche colte nella dimensione pragmatica ed interazionale. Questo tipo di ricerche ha il fine di analizzare non solo la competenza linguistica dei parlanti, ma anche la loro competenza comunicativa. In Italia, non si riscontra un interesse analogo nei confronti della lingua intesa come strumento d'interazione e per lungo tempo gli studi di linguistica non danno il giusto peso a questo tipo di fenomeni. Solo recentemente, l'analisi conversazionale focalizza l'attenzione sull'uso di queste espressioni (cfr. Businaro 2002: 471 e Molinelli 2014: 208).

I SD attirano l'attenzione degli studiosi dopo la metà del secolo scorso nel periodo della nascita della linguistica testuale e della pragmatica. Pur avendo attirato solo recentemente l'attenzione degli studiosi, questi elementi svolgono in realtà funzioni essenziali dal punto di vista discorsivo ed interazionale. Così, utilizzando certi SD, il parlante può manifestare l'intenzione di prendere, mantenere oppure cedere il turno, cosicché gli altri partecipanti sappiano quando è possibile intervenire nel rispetto delle regole conversazionali. Ad es., se pensiamo a: *sai, ecco, praticamente, cioè, beh, insomma, voglio dire, eh, esatto, ecc.*, troviamo che il parlante, che si serve di questi SD, ha uno scopo che va al di là della semplice partecipazione all'interazione e intende contribuire alla trasmissione dell'informazione, facilitando lo scambio interattivo.

L'idea di questa tesi nasce dall'osservazione che l'italiano parlato presenta tanti usi pragmatici di SD che sembrano distanti dalla semantica. La frequenza di questi usi è tale da rendere importanti questioni, quali le caratteristiche di questi elementi, le loro funzioni nonché la relazione tra loro e la semantica. Inoltre, mi sembra interessante indagare sull'identità di questi elementi che non hanno semanticamente un significato vero e proprio. Studiare questi aspetti contribuisce alla comprensione degli usi tipici dei SD, considerati come strumenti necessari nell'architettura della comunicazione, in quanto arricchiscono la lingua parlata, rendendola non solo più espressiva e spontanea, ma anche più personale.

Lo scopo del presente lavoro consiste nel proporre uno studio descrittivo approfondito sulle caratteristiche e sulle funzioni dei SD. Mi sono concentrata sulle funzioni che hanno un ruolo essenziale nella pianificazione del discorso. Per circoscrivere questo fenomeno che coinvolge la comunicazione interattiva, ho pensato di ricorrere all'uso dei corpora di parlato che si sta imponendo recentemente all'attenzione degli studiosi, in quanto forniscono un'opportunità eccellente per lo studio dell'interazione parlata. Il corpus italiano da cui ha preso avvio questo lavoro è C-ORAL-ROM Italia sul quale è basata la mia analisi dei dati.

La scelta del corpus deve essere effettuata in base ai fenomeni che vengono analizzati. Dunque, per l'analisi di fenomeni che appartengono alla lingua parlata (il caso dei SD), secondo Mara (1986), è indispensabile ricorrere a "testi orali autentici e non a materiale in cui buona parte dell'inventario linguistico messo in atto nella strategia interazionale è addirittura assente, perché spesso è solo nel parlato che sono tipicamente presenti le condizioni che ne rendono possibile l'uso" (p. 180). La scelta di corpora parlati piuttosto che scritti è dovuta anche al fatto che gli usi pragmatici ed interazionali sono più facilmente reperibili nell'italiano parlato, e quindi è garantita la frequenza dei SD in diversi contesti. La lingua dei romanzi, invece, è così, come sostiene Ferrini (1985: 44), più studiata e corretta dall'autore dove vengono a mancare fenomeni tipici del parlato, come false partenze, mutamenti di pianificazione sintattica, esitazioni, *impasse*, affermazioni e smentite anche in rapida successione, correzioni, ecc.

Il parlato spontaneo non è neanche come il dialogo teatrale che, anche se è un parlato *sui generis*, è un dialogo programmato dall'autore in cui non possono essere presenti tutti i fattori che caratterizzano il dialogo in situazione reale, come fenomeni di ridondanza, spreco, autocorrezione, interruzione, ecc. (cfr. Manili 1983: 180). Quindi, data la natura dei SD, il cui uso dipende dall'aspetto interattivo

dello scambio linguistico, in situazioni informali poco controllate, si rivela di vitale importanza avvalersi di un corpus di parlato spontaneo.

Una grande quantità di ricerche si svolge in base ai corpora di parlato. Secondo Bennet¹ (cit. in Al Makoshi 2014: 22-23), questo tipo di studi è classificato sotto ‘Corpus Linguistics’, cioè lo studio della lingua attraverso un metodo di analisi linguistica utilizzando una larga collezione di esempi naturali di discorsi scritti o parlati spontaneamente e raccolti elettronicamente. Questa collezione elettronica di testi è conosciuta come corpus o corpora (plurale). Come affermano Aijmer e Stenström (2005), queste ricerche si avvalgono del vantaggio che questi corpora, marcati prosodicamente, “rappresentano diversi tipi di testi della vita quotidiana e si usano per analizzare fenomeni tipici del parlato naturale come *discourse markers, hedge, tags, backchannels e ellipsis*” (p. 1743).

Lo sviluppo nell’uso dei corpora negli anni recenti è dovuto alla loro accessibilità dei dati a causa del diffondersi delle tecnologie multimediali e delle tecniche di annotazione più raffinate. Bazzanella e Bosco (2001: 70) sostengono che è necessaria un’annotazione accurata del contesto per codificare le informazioni multimodali² ed extralinguistiche, utili alla comprensione e all’identificazione del significato inteso dal parlante. Inoltre, per scegliere il corpus, mi sono basata sull’opinione di Bossaglia, Mello e Raso (2016) che affermano che - al contrario degli studi precedenti con prospettive lessicali - la prosodia è l’essenziale caratteristica formale impiegata nell’identificazione dei DM e delle loro funzioni: “this study would not be possible without perceptually segmented and text-to-speech aligned corpora, which allow the observation of how prosodic patterns convey specific functions” (p. 92).

Corpus di riferimento (C-ORAL-ROM Italia)

I dati che verranno esemplificati per la mia analisi dei SD sono ricavati dal corpus linguistico di italiano parlato, C-ORAL-ROM Italia, il sottocorpus italiano di C-ORAL-ROM (Integrated Reference Corpora for Spoken Romance Languages)¹. È un progetto di studio linguistico che raccoglie diversi campioni di produzioni di parlato spontaneo, registrati in situazioni naturali, delle quattro lingue romanze tradizionalmente considerate principali (italiano, francese, spagnolo e portoghese). È costituito da 36 ore di registrazione e trascrizione per ogni lingua (precisamente 36h 16’ 10”), dell’ampiezza di circa 308.815 occorrenze

¹ Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.

² Le informazioni multimodali sono per es. prosodia, sguardi e movimenti del viso e del corpo.

lessicali e 37.235 enunciati, 451 parlanti, per un totale di 204 sessioni per ciascuna lingua.

Questo corpus fornisce alla comunità linguistica una collezione multilingue di corpora comparabili di parlato spontaneo, che può essere utilizzata sia a fine di ricerca sia per l'insegnamento delle lingue. Il sito web ufficiale di questo corpus e della sua documentazione è <http://lablita.dit.unifi.it/coralrom>. Gli editori di questo lavoro sono Emanuela Cresti e Massimo Moneglia. C-ORAL-ROM è a disposizione della comunità scientifica a partire dalla fine del 2004 in due forme:

- In libreria attraverso un'edizione multimediale su un DVD, forma nella quale l'informazione testuale e il materiale sonoro sono sincronizzati (testo sincronizzato alla sorgente sonora) per permettere il pieno apprezzamento dell'informazione sia testuale sia acustica. Il DVD è accompagnato da un volume illustrativo.
- Attraverso licenze di utilizzo a fini di ricerca o di ricerca e sviluppo tecnologico, in un pacchetto di 9 DVD. In questa forma il corpus può essere copiato per usi interni, quali l'utilizzo simultaneo su più macchine nei laboratori di lingua o per scopi di sviluppo (cfr. Moneglia e Cresti 2005: 11).

Lo schema seguente è tratto dal corpus italiano e rivela il corpus design, che sottolinea la variazione dei contesti usati:

¹ C-ORAL-ROM è un progetto finanziato nel programma IST del V° Programma Quadro della UE, coordinato dal Laboratorio Linguistico del Dipartimento di italianistica dell'Università di Firenze (LABLITA). È concluso nel marzo 2004 e realizza un corpus di riferimento per il parlato romano. Il progetto comincia in gennaio 2001, è completata la realizzazione in tre anni e mezzo (cfr. <http://lablita.dit.unifi.it/italian/preprint.html>), cliccato il 21 / 01 / 2017.



Nelle tabelle seguenti proposte da Moneglia (2003: 3-5), si vedano le variazioni dei contesti da cui verranno tratti gli esempi:

Variazione di formalità	Variazione di canale	Variazione di ambito sociologico di uso	Variazione di struttura dell'evento comunicativo
Informale Formale	Parlato naturale Trasmesso (media / telefono)	Famiglia / Privato Pubblico	Monologo Dialogo Conversazione

Le tavole riguardanti il formale sono come segue:

Formale in contesti naturali	Legge	Dibattito politico	Religione	Scuola conferenze	Spiegazioni professionali	Affari
-------------------------------------	-------	--------------------	-----------	-------------------	---------------------------	--------