



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
Filologia, Letteratura italiana e Linguistica
CICLO XXXIV

COORDINATRICE Prof.ssa Paola Manni

Corpus multimodale dell'italiano parlato: basi metodologiche per la creazione di un prototipo

Settore Scientifico Disciplinare L-LIN/01

Dottorando
Dott. Luca Lo Re

Tutor
Prof. Alessandro Panunzi

Coordinatrice
Prof.ssa Paola Manni

Anni 2018/2022

A Monica.

Incipe, parve puer, risu cognoscere matrem [...].

Virgilio, Bucoliche, Ec. IV-60.

Indice

| | |
|--|------------|
| INDICE | 4 |
| INTRODUZIONE | 6 |
| CAPITOLO 1 – IL PARLATO..... | 12 |
| 1. NOZIONE E CARATTERISTICHE DEL PARLATO..... | 12 |
| 1.1 <i>Cos'è il parlato</i> | <i>12</i> |
| 1.2 <i>Le caratteristiche testuali del parlato.....</i> | <i>16</i> |
| 2. IL CANALE ACUSTICO E LA PROSODIA | 18 |
| 3. L'ANALISI DEL PARLATO | 23 |
| 3.1 <i>La Teoria della Lingua in Atto</i> | <i>23</i> |
| 3.2 <i>L'analisi conversazionale.....</i> | <i>26</i> |
| 3.3 <i>Il modello autosegmentale-metrico.....</i> | <i>27</i> |
| 4. CONCLUSIONI..... | 28 |
| CAPITOLO 2 – LA GESTUALITÀ..... | 30 |
| 1. CENNI STORICI SUGLI STUDI SUL GESTO..... | 30 |
| 2. LA NASCITA DEGLI <i>GESTURE STUDIES</i> : I CONTRIBUTI DI ADAM KENDON E DAVID McNEILL..... | 39 |
| 2.1 <i>Il gesto come linguaggio non verbale</i> | <i>39</i> |
| 2.2 <i>Adam Kendon: dalle unità del gesto all'enunciato come azione visibile</i> | <i>41</i> |
| 2.3 <i>David McNeill: gesto, parlato e growth point.....</i> | <i>46</i> |
| 3. LE FUNZIONI DEL GESTO | 53 |
| 3.1 <i>Gesto, cervello e cognizione.....</i> | <i>54</i> |
| 3.2 <i>Il gesto tra funzioni comunicative e pragmatiche.....</i> | <i>59</i> |
| CAPITOLO 3 – LA LINGUA MULTIMODALE..... | 65 |
| 1. INTRODUZIONE..... | 65 |
| 2. LA NOZIONE DI MULTIMODALITÀ..... | 66 |
| 3. LINGUAGGIO E MULTIMODALITÀ | 71 |
| 4. I PROBLEMI SEMIOTICI DELLA MULTIMODALITÀ | 76 |
| 5. IL CONTRIBUTO DELLA LINGUISTICA COGNITIVA | 85 |
| 6. CONCLUSIONE | 94 |
| CAPITOLO 4 - L'ANALISI DEL SIGNIFICATO GESTUALE | 96 |
| 1. INTRODUZIONE..... | 96 |
| 2. L'ANALISI DEI GESTI COME ESPRESSIONE DELLA CONCETTUALIZZAZIONE | 104 |
| 3. L'ANALISI DA UNA PROSPETTIVA LINGUISTICA: IL VALORE DELLA FORMA | 108 |
| 3.1 <i>Per una grammatica dei gesti: l'analisi dei metodi di significazione.....</i> | <i>112</i> |
| 3.2 <i>Le funzioni linguistiche dei gesti</i> | <i>117</i> |
| 4. IL CORPO NELL'INTERAZIONE: L'APPROCCIO CONVERSAZIONALE | 119 |
| 5. IL CORPO COME SUBSTRATO: L'APPROCCIO KINESIOLOGICO | 125 |
| 6. CONCLUSIONE..... | 131 |
| CAPITOLO 5 – IL CORPUS | 134 |
| 1. INTRODUZIONE..... | 134 |
| 2. DAI CORPORA <i>MONOMODALI</i> AI CORPORA <i>MULTIMODALI</i> : CARATTERISTICHE, PROCEDURE E OBIETTIVI ... | 135 |
| 2.1 <i>La registrazione e la raccolta dei dati</i> | <i>137</i> |
| 2.2 <i>La trascrizione dei dati multimodali.....</i> | <i>142</i> |

| | | |
|--|--|------------|
| 2.3 | <i>L'annotazione dei dati multimodali</i> | 143 |
| 3. | IL CORPUS CORMIP | 146 |
| 3.1 | <i>La raccolta dei dati: metodi e strumenti</i> | 146 |
| 3.2 | <i>Trascrizione e annotazione dei dati</i> | 151 |
| 3.2.1 | La trascrizione e l'annotazione del parlato | 153 |
| 3.2.2 | Trascrizione e annotazione del gesto..... | 161 |
| 3.3 | <i>Dall'azione linguistica all'analisi multimodale</i> | 167 |
| CAPITOLO 6 – IL CASO DI STUDIO: L'AZIONE LINGUISTICA TRA GESTO E PAROLA | | 172 |
| 1. | INTRODUZIONE | 172 |
| 2. | GESTI E PRAGMATICA | 173 |
| 3. | I DATI | 177 |
| 4. | DISCUSSIONE DEI DATI | 186 |
| 5. | CONCLUSIONE..... | 196 |
| CONCLUSIONI | | 198 |
| BIBLIOGRAFIA..... | | 201 |

Introduzione

Il lavoro di ricerca presentato in questa tesi ha lo scopo di creare un corpus multimodale pilota dell'italiano parlato in contesto spontaneo. Questa ricerca nasce dall'esigenza, largamente condivisa, di approcciarsi alla lingua come sistema multimodale che costruisce ed esprime i significati attraverso diversi indici e canali. Negli ultimi anni si è assistito a un interesse sempre crescente verso gli studi sul gesto (*gesture studies*) attraverso l'uso di approcci e metodi diversi così che, ad oggi, manca uno standard condiviso rispetto alla raccolta e al trattamento dei dati. Inoltre, la complessità dell'oggetto di studio richiede un grande sforzo nella compilazione dei corpora portando, molto spesso, a una quantità di dati ridotta, a dati elicitati in laboratorio o a raccolte annotate per specifici obiettivi di ricerca.

L'obiettivo che ci siamo posti porta con sé una serie di questioni le cui radici affondano sia nella teoria linguistica, e nella nozione stessa di lingua, che su questioni pratiche strettamente legate al trattamento dei dati. La prima di queste questioni riguarda la nozione di lingua multimodale. Parlare di linguaggio come sistema multimodale implica tenere in considerazione diverse modalità espressive che vanno dal canale parlato ai gesti, dall'espressione facciale alla prossemica fino all'uso degli oggetti del contesto in cui avviene l'evento linguistico. Di conseguenza, la nozione di lingua multimodale porta a ripensare la definizione tradizionale di lingua. Infatti, l'idea di lingua che gli studi linguistici hanno tradizionalmente portato avanti esclude tutti gli aspetti e le caratteristiche che rendono la lingua sistema multimodale, ponendo così questioni che riguardano il metodo di analisi, di segmentazione del flusso del parlato e della gestualità, il riconoscimento del valore linguistico degli aspetti dell'attività del parlato considerate, fino a non molto tempo fa, para-verbali e non linguistiche (McNeil 1985; Pietrandrea 2012).

Il nostro lavoro di ricerca ha l'intento di proporre una possibile soluzione ai problemi che riguardano la compilazione dei corpora multimodali, con la consapevolezza che il nostro contributo non potrà di certo essere pienamente esaustivo ma potrà, a suo modo, contribuire al dibattito assumendo un punto di vista finora poco esplorato. In modo particolare, abbiamo cercato di basare questa ricerca su un metodo che definiamo pragmatico e percettivo, e che consiste nel concepire la lingua parlata come un fenomeno che si realizza attraverso l'uso di diversi canali che risultano segmentabili sulla base della percezione umana. Infatti, così come si può parlare di lingua che si esprime attraverso il canale fonico (*lingua parlata*) ed è percepita dall'interlocutore attraverso il proprio canale acustico (*lingua audita*), allo stesso modo possiamo parlare di lingua espressa attraverso la corporeità (*lingua gestita*) che è percepita dall'interlocutore attraverso il proprio canale visivo (*lingua vista*) (De Mauro 2000). Di conseguenza, abbiamo ritenuto indispensabile fare riferimento al quadro teorico della Teoria della Lingua in Atto (Cresti, 2000; Cresti e Moneglia, 2010; Moneglia & Raso, 2014) e agli studi sul gesto di Adam Kendon (2004). Come vedremo nel corso di questo lavoro, il nostro approccio permette di indagare la lingua nei principali aspetti multimodali attraverso l'analisi dell'interazione tra le diverse unità di riferimento - le unità del parlato e le unità gestuali - e restituisce un metodo di trascrizione utilizzabile da diversi approcci teorici.

La compilazione di un corpus richiede una riflessione sul metodo di raccolta dati, sulla strutturazione del data-set e sulla possibilità di utilizzo. Rispetto a ciò, il nostro lavoro cerca di garantire la spontaneità dei dati basando la raccolta su interazioni naturali registrate attraverso strumentazione poco invasiva. Inoltre, il data-set è stato strutturato in modo tale da poter garantire una diversità di generi interazionali e di varietà di lingua. In questo modo abbiamo costruito un corpus che raccoglie tre diversi tipi di scambi comunicativi - uno monologico, uno dialogico e uno

conversazionale (più di due persone) – e in due diverse città italiane, Firenze e Catania. Il campione non è stato individuato con l'intento di essere rappresentativo delle varietà di parlato italiano, perché avrebbe richiesto un dispendio di energie e di strumenti di gran lunga superiore alle disponibilità di un percorso dottorale. Per questa ragione abbiamo creato un corpus pilota con l'intento di sperimentare l'efficacia del metodo da noi proposto.

Il corpus è conservato nei server dell'Università di Firenze, presso il laboratorio di Informatica Umanistica del Dipartimento di Lettere e Filosofia ed è consultabile al seguente link: <https://drive.google.com/drive/folders/1xI8Ymni7DamXsv6xw1sF5WBkyziY4yQI?usp=sharing>

La tesi è divisa in due parti. La prima parte introduce i problemi teorici legati alla nozione di lingua e all'approccio multimodale. Il primo capitolo è dedicato alla definizione di lingua parlata con le caratteristiche testuali che ne determinano la complessità. In particolare, ci siamo soffermati sulla nozione di lingua parlata come canale di comunicazione acustico e sulla nozione di prosodia con i suoi elementi caratterizzanti e le sue funzioni. Il capitolo, infine, si chiude con una rassegna su tre principali approcci allo studio del parlato. In particolare, sono posti in rassegna l'approccio della Teoria della Lingua in Atto, l'Analisi Conversazionale e il modello Autosegmentale-Metrico.

Nel secondo capitolo vengono introdotti gli studi sulla gestualità. Partendo da un excursus storico dei principali studi sul gesto, viene poi introdotta la nascita dei moderni *gesture studies* grazie al contributo di Adam Kendon e David McNeill. In particolare, si pone l'attenzione al dibattito intorno alla verbalità del gesto, che vede protagonista McNeill, e alle unità gestuali individuate da Kendon. Il capitolo si chiude con un'analisi delle principali funzioni che sono state riconosciute al gesto.

Nello specifico, il ruolo che gioca la gestualità nella cognizione e come il cervello gestisce la gesticolazione, e le funzioni pragmatiche e comunicative del gesto con le caratteristiche contestuali che ne determinano l'uso e la funzionalità.

La prima parte si chiude con un capitolo dedicato alla lingua come sistema multimodale che si compone di due principali canali analizzati nei capitoli precedenti. Così il terzo capitolo introduce la nozione di multimodalità con le principali correnti e approcci che rendono il campo di studio un'area di interesse multidisciplinare. Il capitolo prosegue con un paragrafo dedicato alla visione del sistema lingua come sistema multimodale, focalizzandosi sulle evidenze che portano a questa definizione e sugli elementi che caratterizzano il linguaggio multimodale. Vengono indagate anche le principali nozioni di unità che fanno riferimento alla multimodalità come *utterance as visible action* (Kendon, 2004), il *Growth Point* (McNeill, 2005), il *composite utterance* (Enfield, 2009) e le prove che emergono dalla lingua dei segni e che corroborano la visione del linguaggio come sistema multimodale.

Ma se da un lato diversi studi sostengono che la lingua è intrinsecamente multimodale, dall'altro lato ciò implica una revisione della nozione stessa di lingua. Così, il quarto paragrafo è dedicato alle questioni semiotiche legate alla multimodalità linguistica e alle caratteristiche che definiscono la *linguisticità* di un sistema di comunicazione. Sulla base delle nozioni tradizionali di lingua proposte da Hockett (1960) e Martinet (1960), si introduce il dibattito intorno al riconoscimento delle lingue dei segni come sistemi linguistici (De Mauro, 1998; Pietrandrea, 1998; Pietrandrea, 2021). La relazione che intercorre tra la gestualità e i segni, che vede due posizioni contrapposte tra chi vi riconosce una continuità (Müller, 2018) e chi li vede come elementi radicalmente diversi (Goldin-Meadow-Brentari, 2017), rappresenta un'opportunità per estendere il dibattito sulla linguisticità delle lingue dei segni alla gestualità. Successivamente, il capitolo si

concentra sul contributo dato dalla linguistica cognitiva per una nozione di lingua estesa e rinnovata sulla base delle teorie embodied (Cienki, 2017). Il capitolo si chiude con una nostra riflessione sulla nozione di lingua come sistema multimodale e sul conseguente riconoscimento del valore linguistico dei gesti.

La seconda parte della tesi è dedicata alla costruzione del corpus. In particolare, il quarto capitolo è dedicato all'analisi del significato gestuale. Il primo paragrafo riporta la letteratura riguardante l'individuazione dei parametri e delle unità di analisi proposte da Kendon (1988) e McNeill (1992), e successivamente viene riportata una rassegna delle principali classificazioni gestuali definite su base funzionale o semiotica (Wundt, 1972; Efron, 1971; Ekman e Friesen, 1969). Il paragrafo successivo riporta l'approccio cognitivista all'analisi del gesto, che viene visto come espressione del processo di concettualizzazione (Cienki, 2013). Segue il paragrafo dedicato alla prospettiva linguistica e di come, attraverso un approccio *form-based* di Bressem (2013), sia possibile definire una grammatica dei gesti (Ladewing, 2012) e tracciare un quadro delle funzioni linguistiche espresse (Müller, 2013). Il quarto paragrafo riporta l'approccio conversazionalista al gesto, che indaga e valorizza le funzioni interazionali sia nella costruzione dei turni e negli scambi che al legame che esprimono con il contesto (Mondada, 2013-2011). Infine, il capitolo si chiude con l'approccio kinesiologico sviluppato da Boutet (2018) e che è incentrato sulla valorizzazione del corpo inteso come *substrato* del fenomeno gestuale che ne determina l'espressività per mezzo dei limiti fisici e delle potenzialità.

Il quinto capitolo, dopo una rassegna delle caratteristiche e delle procedure che la letteratura indica come necessarie per la costruzione di un corpus (Knight e Adolphs, 2006), illustra il metodo utilizzato per la compilazione del corpus CORMIP. Vengono illustrate le fasi di raccolta dei dati, la loro trascrizione e annotazione e infine, l'ultimo paragrafo, illustra

l'ipotesi di utilizzare l'annotazione dell'azione linguistica come mezzo per analizzare la multimodalità del linguaggio. Questa tesi si chiude con un capitolo dedicato a un *case study*. L'analisi ha due principali finalità: da un lato quello di illustrare un possibile uso del corpus CORMIP, e dall'altro verificare la nostra ipotesi di utilizzare l'azione linguistica come mezzo di analisi del sistema multimodale.

Il lavoro di ricerca che abbiamo svolto, e che riportiamo in queste pagine, ha cercato di proporre un metodo per la costruzione di un corpus multimodale dell'italiano parlato, inserendosi nella tradizione dei corpora di italiano parlato compilati dal gruppo di ricerca LABLITA. Abbiamo cercato di raggiungere questo obiettivo attraverso la costruzione di un corpus pilota, così da poter sperimentare il metodo proposto e porre le basi per ulteriori ricerche future.

Capitolo 1 – Il parlato

1. Nozione e caratteristiche del parlato

1.1 Cos'è il *parlato*

Pensare alla lingua e allo studio dei fenomeni linguistici vuol dire pensare e interessarsi a una quantità di fenomeni tendenzialmente infinita, ciascuno dei quali si caratterizza per tipo di lingua, canale e modalità di trasmissione, contesto d'uso e tante altre caratteristiche che, sinteticamente, definiamo sociali. Nella moltitudine dei fenomeni e delle loro caratterizzazioni è ormai assodata la consueta divisione dei fenomeni linguistici in due grandi tipi, la lingua scritta e la lingua parlata.

Il parlato è stato oggetto di studio da parte di linguisti di ogni epoca, con strumenti e metodi diversi: dapprima la sua osservazione si è basata sui testi scritti che riportavano il parlato (testi teatrali o novelle) e successivamente, quando lo sviluppo tecnologico ha reso disponibile l'opportunità di registrare l'audio, un approccio basato sullo studio di dati a tutti gli effetti riconducibili al parlato (Nencioni 1976). La lingua parlata caratterizza gran parte delle interazioni comunicative e, in relazione alla dicotomia scritto-parlato, è senza dubbio filogeneticamente nata prima¹. Inoltre, ogni essere umano impara prima a parlare e successivamente a scrivere e soprattutto non tutte le persone imparano a scrivere, mentre tutte imparano a parlare (Bazzanella 1994).

Il parlato si distingue dallo scritto sia in relazione alla realizzazione - e quindi codificazione e canale utilizzato - che in relazione ai processi psicocognitivi coinvolti nelle diverse attività *parlare-ascoltare* e *scrivere-leggere* (Sornicola 1985). In questo lavoro, e in questo capitolo nello specifico,

¹ Alcuni studiosi ipotizzano che il linguaggio debba la propria origine filogenetica al gesto (Corballis, 2020).

ci occuperemo quasi esclusivamente degli aspetti legati alla realizzazione più che agli aspetti cognitivi. In particolare, cercheremo di mettere in rassegna le prospettive che la linguistica teorica e la pragmatica hanno assunto sull'argomento, cercando di comprendere meglio il dibattito intorno alle unità di analisi, alle caratteristiche testuali e alle modellazioni della comunicazione parlata per comprendere l'agire dei parlanti nella costruzione dei significati.

La lingua parlata, dunque, è caratteristica precipua dell'essere umano così inevitabilmente è diventata oggetto di numerose riflessioni in ambiti di ricerca molto diversi (retorica, stilistica, semiologia, etnoantropologia, psicologia, psicolinguistica) arrivando a occupare un posto di rilievo negli studi linguistici a partire dagli anni Ottanta del Novecento (Voghera, 1992). I diversi ambiti di ricerca che nel corso degli anni si sono interessati alla lingua parlata, hanno basato le loro ricerche su nozioni e aspetti diversi di uno stesso fenomeno.

Se a un livello più superficiale, come abbiamo appena accennato, è possibile identificare la nozione di parlato sulla base della diversa configurazione della produzione rispetto allo scritto la questione si complica, invece, se si cerca di delimitare una nozione di parlato basata su tratti linguistici. La difficoltà non sta nell'individuare una serie di strutture linguistiche proprie della lingua parlata – che nel corso degli anni sono state pressoché identificate – ma si tratta piuttosto di comprendere se con il termine parlato intendiamo un insieme di parole e costrutti, una determinata pratica di interazione oppure un canale di trasmissione di segni verbali. Alla luce di ciò, Miriam Voghera (1992) identifica tre livelli di specificità della nozione di parlato:

- come modalità semiotica e quindi si parla di *oralità*;
- come modalità fisica di trasmissione e quindi si parla di *sistema fonico-uditiva*;

- come modalità d'uso di una lingua e quindi si parla di *varietà di lingua*.

Il termine *oralità* appartiene all'ambito degli studi storici e antropologici e indica il complesso di abitudini comunicative di una società o comunità che non conosce la tecnica della scrittura. La differenza delle culture esclusivamente orali porta, di conseguenza, a una serie di differenze anche a livello sociale e psicocognitivo in quanto il carattere effimero dell'*oralità* porta a una diversa organizzazione della vita sociale, della conservazione e trasmissione del sapere e dei modelli educativi (Ong, 2011). Così è possibile parlare di strategie semiotiche che caratterizzano l'*oralità* e sono la soggettività, la dialogicità, la stretta dipendenza con il contesto di enunciazione, le diverse modalità di significazione, un massiccio uso della ripetizione, un andamento tematico epicicloidale e la ridondanza sintattica.

Il *sistema fonico-uditivo* porta con sé una serie di caratteristiche. Innanzitutto, i rapporti tra gli elementi sono temporali e non spaziali, i suoni cioè si susseguono nel tempo avendo delle conseguenze sia nel sistema di ricezione che nella pianificazione da parte del parlante. Ciò implica la caratteristica della non ripetibilità e non permanenza, le produzioni foniche sono effimere e quindi non è possibile tornare sopra a un enunciato appena detto. Il *sistema fonico-uditivo* si caratterizza anche per la contemporaneità dell'emissione e della ricezione dei suoni che incide direttamente sulla velocità e sul tipo di pianificazione.

La concezione del parlato come modalità d'uso di una lingua porta ad approcciarsi all'oggetto dello studio come *varietà di lingua*, identificando la lingua parlata come altra rispetto alla lingua scritta. Gli studi sul parlato inteso come *varietà di lingua* hanno indagato e rintracciato gli elementi strutturali che distinguono il parlato dalle altre porzioni del sistema linguistico, fino a ipotizzare l'esistenza di una grammatica diversa rispetto a quella della lingua scritta (Berutto, 1985 – Sabatini, 1985).

Questi tre livelli di specificità della nozione di *parlato*, come abbiamo visto, possono essere visti e concepiti come categorie di riferimento per le diverse analisi linguistiche, per questo può risultare così difficile determinare un'unica nozione di *parlato*. Piuttosto, le proprietà di un testo parlato dipenderanno dal modo in cui queste caratteristiche entrano in relazione tra loro concependo così il *parlato* come un insieme di rapporti.

«L'espressione *lingua parlata* identifica un insieme di caratteristiche strutturali e funzionali che si manifestano primariamente, ma non in modo esclusivo, quando si usa la lingua attraverso il canale fonico-uditivo in condizioni naturali e spontanee. La lingua parlata è quindi costituita da un insieme di usi linguistici prodotti dalle specifiche condizioni enunciative del parlare, non (o solo parzialmente) osservabili in altre modalità di trasmissione» (Voghera, 2010)

In conclusione, possiamo parlare non tanto di un'univoca nozione di parlato ma di un'idea di parlato prototipico caratterizzato da elementi propri che lo differenziano dall'idea di scritto prototipico e che sintetizziamo nella tabella seguente (Bernardelli & Pellerey, 1999):

| PARLATO PROTOTIPICO | SCRITTO PROTOTIPICO |
|---|--|
| <ul style="list-style-type: none"> • Produzione ed elaborazione in tempo reale • Compresenza parlante e interlocutore • Dipendenza dal contesto di produzione e ricezione • Evanescenza dell'enunciato • Irreversibilità dell'enunciazione • Impossibilità dell'interlocutore di tornare indietro • Ordine e tempo di ricezione obbligati • Impossibilità di verifica | <ul style="list-style-type: none"> • Pianificazione del testo • Non compresenza scrittore e lettore • Autonomia dal contesto di produzione e ricezione • Permanenza del testo • Possibilità di correzione • Possibilità del lettore di muoversi avanti e indietro nel testo • Autonomia dei tempi di fruizione • Verifica delle affermazioni |

Nelle pagine di questo capitolo, approfondiremo le caratteristiche di cui qui abbiamo fatto cenno. Lo faremo per inquadrare meglio gli obiettivi e le

basi su cui si fonda questa ricerca, perché l'oggetto di studio di cui ci occupiamo è per sua natura complesso e intricato e riteniamo indispensabile riuscire a fare ordine sia dal punto di vista della teoria che dell'analisi che ne consegue.

1.2 Le caratteristiche testuali del parlato

L'interazione comunicativa può assumere forme diverse, nel paragrafo precedente abbiamo visto le differenze tra la forma di lingua parlata e quella di lingua scritta. Le caratteristiche prototipiche del parlato si rivelano in modo tangibile nella costruzione dell'interazione e su più livelli.

A livello fonico il parlante tende ad articolare in modo economico garantendo lo sforzo necessario alla comprensione di ciò che sta dicendo, così elimina ogni movimento articolatorio che non sia indispensabile (Brown, 1990). Questo porta alla produzione di segnali poco specificati al livello segmentale e di conseguenza ai fenomeni di fusione - derivanti dalla coarticolazione o dall'indebolimento di articolazioni consonantiche - ai fenomeni di elisione di foni o sillabe o ai mutamenti del timbro vocalico (Albano Leoni & Maturi 1992).

Inoltre, la pianificazione e la produzione in tempo reale, il coinvolgimento di un interlocutore portano all'emergere di quelle che vengono chiamate disfluenze e che riguardano l'organizzazione del discorso. La natura delle disfluenze è fortemente legata sia a motivi cognitivi che a motivi comunicativi assumendo così la natura di segnali piuttosto che di errori.

Tra i fenomeni di disfluenza possiamo distinguere le disfluenze di natura fonetica - che includono le interruzioni o le alterazioni della catena fonica lasciando inalterata la sequenza verbale - e le disfluenze di natura testuale - che invece alterano la sequenza verbale lasciando inalterata la catena

fonica. Alle prime appartengono le pause² piene realizzate con laringalizzazioni, nasalizzazioni, allungamenti di vocali o consonanti, vocalizzazioni varie (Pettorino & Giannini 2005); alle seconde invece le false partenze, le autocorrezioni, le autoripetizioni e normalmente richiedono un'operazione di ricostruzione testuale da parte dell'ascoltatore.

Dall'aspetto interattivo del parlato scaturiscono una serie di elementi che possiamo raggruppare sotto la nozione di dialogicità, cioè la forma più diffusa della lingua parlata e anche la prima con cui ogni essere umano entra in contatto. Uno di questi elementi è la turnazione, infatti la comunicazione parlata tra più persone, comunemente definita conversazione, si costruisce sull'alternanza tra parlanti e ogni presa di parola viene identificata come turno dialogico. Altri elementi tipici dello scambio dialogico sono le sovrapposizioni – che possono ricorrere come errori nel meccanismo della turnazione creando partenze simultanee a seguito di un turno concluso o una sovrapposizione in un turno in corso – e le interruzioni caratterizzate da un aspetto più competitivo e che consistono nel togliere la parola all'interlocutore. L'interattività dello scambio dialogico è caratterizzata anche dal fenomeno del backchannels, brevi enunciati realizzati da un interlocutore che sta assumendo il ruolo di ascoltatore durante il turno di parola dell'altro interlocutore. Questo fenomeno assume diverse funzioni: può esprimere l'interesse dell'ascoltatore, può essere un segnale di comprensione, può avere una funzione collaborativa nella costruzione del discorso e può anche essere di aiuto nella ricerca di una parola (Schergloff, 1982 – Ford & Thompson, 1996). Infine, la conversazione si caratterizza anche per la sua sintassi che può essere distribuita su più turni – quando il parlante completa la propria enunciazione in un turno successivo all'intromissione dell'interlocutore –, può essere co-costruita quando un parlante completa la frase dell'interlocutore (Voghera, 1992), oppure può essere caratterizzata dal

² Le pause possono assumere due distinte funzioni associate alla loro durata. Si distinguono pause brevi dette di 'articolazione' e pause lunghe di 'esitazione' (Goldman & Eisler 1968).

fenomeno delle espansioni – quando un parlante induce l'interlocutore a espandere la propria enunciazione chiedendo delle specificazioni.

Altra caratteristica del parlato è l'integrazione di elementi non orali alle espressioni verbali. Ciò avviene attraverso il richiamo agli elementi della situazione comunicativa e attraverso l'uso di altre modalità comunicative come la mimica, i gesti e i movimenti del corpo. La condivisione della situazione facilita il riferirsi al contesto a discapito dell'uso di parole per far riferimento agli elementi della situazione. Ciò avviene sia attraverso l'uso dei deittici verbali che attraverso l'integrazione di gesti di indicazione agli enunciati. L'uso dei riferimenti diretti agli elementi del contesto permette ai parlanti di accelerare e ottimizzare i tempi di produzione e ricezione delle informazioni garantendo la brevità. L'uso di elementi verbali in concomitanza con le espressioni facciali, i gesti e i movimenti del corpo permette al parlante di comunicare su più piani contemporaneamente. Come vedremo più approfonditamente nel corso di questa tesi, l'uso della gestualità non si limita a una mera funzione di accompagnamento della sequenza verbale ma gioca ruolo nella creazione dei significati diventando parte integrante dell'enunciazione (Magno Caldognetto, Ursini & Poggi 2004), ed è così che si parla di interazione multimodale (cfr. capitolo 3).

2. Il canale acustico e la prosodia

La comunicazione parlata si definisce principalmente per l'uso del canale acustico non soltanto come esclusivo mezzo di codificazione, ma anche come strumento di significazione. Trattare il parlato da un punto di vista acustico vuol dire approcciarsi alla sua fisicità e da questo punto di vista il parlato è suono, cioè un insieme di onde prodotte dal movimento ciclico di compressione e decompressione dell'aria i cui principali fenomeni sonori hanno rilevanza per l'analisi linguistica. I parlanti, infatti, trasmettono l'informazione sia attraverso l'uso del lessico e della sintassi, sia per mezzo di una serie di fattori soprasegmentali. L'accento, il ritmo, il tono e

l'intonazione forniscono così informazioni extralinguistiche (sesso, età, stati emotivi), informazioni paralinguistiche (registro, provenienza e fenomeni di pianificazione) e informazioni linguistiche (forza illocutiva, struttura sintattica e struttura informativa).

Assumendo questo punto di vista, si rende possibile una descrizione e un'analisi dei fenomeni sia dal lato percettivo che dal lato strumentale. L'analisi strumentale si basa sulla misurazione dei parametri acustici che sono la durata, l'intensità e l'altezza; elementi che interagiscono tra loro nella realizzazione dei processi prosodici. La durata misura la lunghezza temporale di un suono; l'intensità si deve all'ampiezza delle onde sonore emesse e da un punto di vista della produzione è data dall'energia articolatoria, dalla pressione dell'aria che arriva dai polmoni e dal grado di tensione delle corde vocali, mentre dal punto di vista percettivo si fa riferimento al volume, si misura in decibel e ha carattere relativo; la frequenza fondamentale si riferisce al numero di vibrazioni glottidali compiute in un secondo, si misura in Hertz e di conseguenza è presente solamente nei suoni sonori. A livello percettivo la frequenza fondamentale è responsabile della differenza di altezza dei suoni, così se la frequenza fondamentale è alta si percepiranno suoni più acuti, viceversa se la frequenza fondamentale è bassa si percepiranno suoni più gravi. Inoltre, la frequenza fondamentale è il correlato fisico responsabile delle variazioni intonative tipiche del parlato.

La relazione tra analisi strumentale e analisi percettiva non è perfettamente biunivoca; infatti, sono state individuate delle soglie minime al di sotto delle quali le unità non sono più percettivamente pertinenti. Per la durata la soglia minima è in media 30ms, per quanto riguarda l'intensità la soglia minima percepita dal sistema uditivo umano è invece di ± 1 dB al di sopra del quale progressivamente la capacità uditiva diminuisce. Per quanto riguarda la frequenza fondamentale è particolarmente difficile riuscire a individuare una soglia minima: nella percezione della frequenza interagiscono diversi fattori, tra cui il livello frequenziale entro il quale si

pone lo stimolo sonoro, la sua durata e la velocità di variazione di frequenza. Nel valutare l'interazione di tutti questi fattori, si è arrivati a calcolare che in una comunicazione reale il limite si colloca tra 1,5 e 2 ST ('t Hart, 1981).

L'intonazione è l'elemento prosodico che maggiormente caratterizza il parlato e può essere definito come l'andamento melodico dell'enunciato, un tratto comune a tutte le lingue. È un fenomeno che ha come dominio l'intero enunciato e svolge un ruolo importante nell'organizzazione sintattica, semantica e pragmatica del messaggio trasmesso dai parlanti. La funzione primaria dell'intonazione è quella di segmentare il flusso del parlato in gruppi di parole coesi e coerenti sul piano testuale. A ciascuno di questi segmenti viene dato il nome di unità intonativa o unità tonale ed è un costituente prosodico composto da un accento intonativo nucleare, da possibili altri accenti e limitato da una marca di confine. L'intonazione è il mezzo attraverso cui si esprime la forza illocutiva, la modalità di un enunciato e differenzia i diversi tipi di frasi tra assertivi e non assertivi. Svolge anche una funzione paralinguistica in quanto esprime tutti quei valori legati agli aspetti non linguistici come l'espressività e l'emotività. Dal punto di vista fonetico la melodia del parlato è data dalle variazioni dell'altezza tonale con cui vengono realizzati i suoni risultando così più acuti o più gravi. Il parametro acustico coinvolto è la frequenza fondamentale che dal punto di vista articolatorio è data dal numero di vibrazioni glottidali delle corde vocali in un determinato periodo di tempo.

All'interno del contorno intonativo si individuano alcuni fenomeni messi in atto dal parlante per esprimere le diverse funzioni comunicative. L'estensione tonale (o *pitch range*) è l'intervallo di frequenza tra il valore massimo e il valore minimo in Hertz. Tendenzialmente il *pitch range* risulta alto quando il parlante cerca di esprimere uno stato d'animo enfatico, sia per rabbia o per sorpresa.

Il fenomeno di *resetting* è la rimodulazione dell'andamento della frequenza fondamentale che, nella sua naturale tendenza alla declinazione,

subisce un innalzamento, riformulando l'andamento intonativo. Per valutare la presenza di un *resetting* bisogna tenere in considerazione il picco iniziale dell'unità prosodica, la fase discendente che si presenterà come un avvallamento che sarà seguito da un nuovo picco di frequenza all'inizio della successiva unità intonativa. A questo fenomeno è stato riconosciuto da alcuni una natura puramente fisiologica determinata dalla respirazione (Lieberman, 1967) altri studi, invece, vedono nell'origine del *resetting* l'organizzazione sintattica. Studi più recenti escludono, di contro, un nesso tra questo fenomeno e la struttura sintattica; piuttosto si è vista una stretta correlazione con l'espressione di valori pragmatici. Infatti, è stato notato che un significativo innalzamento della frequenza in concomitanza con un'estensione del *pitch range* corrisponde all'inizio di un nuovo argomento discorsivo.

La *prominenza*, che è definita tradizionalmente come picco di intensità nella realizzazione di parti della catena fonica, è un fenomeno che riguarda il grado di salienza percettiva che il parlante assegna ad alcune sillabe o a parte dell'enunciato ed è un fenomeno linguistico locale relativo in quanto può essere percepito solo se posto a confronto con altri elementi vicini non prominenti (Lieberman & Pierrehumbert, 1984). Da un punto di vista fonetico la *prominenza* è data dall'attivazione di diversi indici acustici come la frequenza fondamentale, la durata, l'intensità e l'escursione frequenziale. I diversi indici svolgono un ruolo diverso nella realizzazione della *prominenza* in relazione al sistema linguistico, in italiano la *prominenza* è dovuta principalmente all'incremento della durata.

Il parlante, dunque, assegna una determinata salienza prosodica a quelle parti dell'enunciato che intende porre in evidenza in quanto caratterizzate informativamente dal valore di novità. Così, sul piano fonetico, l'elemento dell'enunciato focalizzato risulterà temporalmente allungato e sarà associato a un'escursione melodica.

Come abbiamo già accennato, l'intonazione ha anche una funzione delimitativa che risulta nella divisione del parlato in unità tonali. Il margine

di ciascuna unità temporale emerge da un particolare fenomeno di demarcazione che viene chiamato confine intonativo. L'individuazione di un confine intonativo si avvale sia di criteri interni che di criteri esterni. Dei primi, che sono parametri fisici, fanno parte il cambio di altezza tonale rispetto a ciò che è contiguo, l'estensione frequenziale di questo movimento, la ricorrenza di *reset* e il cambio di velocità elocutiva. Dei secondi invece fanno parte tutti quei fenomeni relativi alla qualità della voce e al calo di intensità sonora. In un quadro così complesso, l'interazione tra parametri esterni e interni spesso non garantisce un'individuazione univoca e oggettiva dei confini intonativi. È stato visto che nella segmentazione prosodica, anche trascrittori esperti non riescono a raggiungere dei risultati sufficientemente congruenti (Sorianello, 2014).

La complessità dell'intonazione ha innescato un lungo dibattito sulla sua natura, portando gli studiosi a dividersi tra coloro che ne hanno riconosciuto una natura discreta e coloro che invece ne riconoscono un carattere continuo. I due schieramenti teorici che si sono scaturiti hanno creato modelli teorici opposti, da un lato c'è l'approccio per configurazioni - nato in seno scuola britannica e che vede l'intonazione come una sequenza di movimenti tonali globali - e dall'altro c'è l'approccio per livelli - di tradizione statunitense che identifica nell'intonazione una serie di livelli tonali distintivi. Nel quadro di questi studi spiccano, a partire dagli anni Settanta, le ricerche e il modello del gruppo di ricerca olandese dell'*Instituut voor Perceptie Onderzoek*. La teoria che hanno elaborato cerca di fare sintesi tra il metodo di analisi fonologico, di natura astratta, e il metodo di analisi fonetico, di natura concreta. Questa mediazione è stata resa possibile per via dell'adozione di un metodo percettivo, tenendo così in considerazione il punto di vista dell'ascoltatore e assumendo che il contorno intonativo sia costituito da unità discrete ('t Hart, Collier e Cohen, 1990). Questo metodo verrà discusso più approfonditamente nelle pagine successive, in quanto è stato adottato per l'annotazione del corpus oggetto di questa tesi.

3. L'analisi del parlato

Abbiamo visto fin qui la complessità del parlato derivante dalle sue caratteristiche fisiche e dal contesto di realizzazione che il più delle volte è interazionale. Così nel corso degli anni sono stati sviluppati diversi modelli teorici con l'obiettivo di riuscire a studiare e analizzare il parlato nella sua interezza e nella sua complessità dell'uso. Ciascuno di questi framework, come vedremo, si basa su aspetti diversi del parlato identificandone unità di riferimento diversi che vanno, come vedremo, dall'enunciato al turno dialogico. In questo paragrafo faremo una rassegna dei principali modelli teorici.

3.1 La Teoria della Lingua in Atto

Il modello della **Teoria della Lingua in Atto** (Cresti, 2000; Cresti & Moneglia, 2010; Moneglia & Raso, 2014) si sviluppa nel contesto del rinnovato interesse per lo studio della lingua parlata della seconda metà del Novecento, come estensione della Teoria degli Atti Linguistici di J. L. Austin (1962) e ha per oggetto il parlato spontaneo. Il modello della Teoria della Lingua in Atto pone a suo fondamento l'azione linguistica realizzata dal parlante e la prosodia, il mezzo formale attraverso cui il parlante produce l'Enunciato e l'interlocutore ne decodifica l'informazione.

L'unità di riferimento del parlato è identificata nell'Enunciato, unità puramente pragmatica e autonoma rispetto alla sintassi; è infatti il corrispettivo linguistico di un atto, autonomo e interpretabile isolatamente, al di là della sua forma e consistenza sintattica.

È individuato percettivamente e corrisponde a una sequenza acustica terminale demarcata da un break prosodico riconoscibile come conclusivo, risultando così prosodicamente identificabile nel continuum parlato. La percezione acustica permette quindi all'ascoltatore di riconoscere le unità

del flusso di parlato in base all'interazione e la co-variazione di più parametri: il primo è la frequenza fondamentale (f_0), a cui si aggiungono durata e intensità.

Le analisi empiriche condotte in seno al modello hanno permesso di individuare un alto grado di isomorfismo tra la struttura informativa dell'Enunciato e quella intonativa così da considerare indispensabile il dato acustico nell'analisi delle strutture del parlato.

L'informazione, quindi, risulta organizzata sulla base del principio illocutivo, secondo cui ogni Enunciato esprime un valore illocutivo che emerge nella struttura del linguaggio segmentando il flusso del parlato in sequenze autonome, e il principio della struttura informativa, per cui l'Enunciato consiste in una struttura di unità informative corrispondenti alle unità prosodiche.

Per la Teoria della Lingua in Atto è centrale l'osservazione della lingua parlata nella sua realizzazione spontanea in eventi comunicativi diversi fra loro, calati in ambienti e situazioni differenti; è infatti costitutivamente ancorata a osservazioni empiriche su corpora di parlato spontaneo (Cresti & Moneglia, 2005). La segmentazione e l'annotazione intonativa si basa sul riconoscimento dei break prosodici, che dividono il continuum sonoro in unità discrete in corrispondenza con le variazioni prosodiche percettivamente rilevanti del flusso orale; ogni parlante madrelingua li riconosce e percepisce e, anche in base a questi, interpreta il discorso. La segmentazione del parlato avviene dunque grazie al riconoscimento dei break terminali, percepiti come conclusivi e che indicano confini di Enunciato, e i break non terminali, percepiti invece come non conclusivi e che indicano i confini delle unità tonali in cui si articola l'Enunciato.

L'Enunciato può essere formato da una o più unità tonali. C'è l'unità Comment (COM) che è quella deputata al compimento dell'illocuzione sulla base della sua autonomia pragmatica ed è l'unica necessariamente presente all'interno di un enunciato; le altre unità, opzionali, si distribuiscono in relazione a questa, assumendo funzioni di complemento. Il Comment è

prosodicamente autonomo e, in quanto tale, sufficiente a costituire un Enunciato intonativamente concluso; viene realizzato attraverso un'unità tonale di tipo root ('t Hart et al., 1990), secondo diversi profili prosodici che ne determinano il valore illocutivo. La Teoria della Lingua in Atto ha sviluppato una classificazione dettagliata (Cresti, 2000; 2017; 2020), rielaborando le tassonomie di atti linguistici precedenti (vedi tra gli altri Austin, 1962; Searle, 1969) al fine di classificare i valori illocutivi espressi dalle singole unità di Comment. Tale classificazione distingue cinque classi illocutive generali, determinate dall'atteggiamento del parlante e dal suo rapporto con l'interlocutore.

Come accennato sopra, la segmentazione delle unità tonali garantisce unità formale all'Enunciato attraverso il riconoscimento di break prosodici terminali. L'Enunciato può essere costituito da un unico pattern tonale, che corrisponde alla singola unità di COM venendo a formare, in questo caso, un Enunciato semplice. Di contro, l'Enunciato complesso vede al suo interno più unità tonali separate e riconoscibili nel contorno melodico tramite i break prosodici non terminali. Dal punto di vista informativo, l'Enunciato complesso presenta, oltre al Comment, unità informative di supporto al compimento dell'illocuzione e allo scambio comunicativo. Queste altre unità di supporto sono catalogate all'interno del framework teorico sulla base della forma prosodica e della funzione.

Si identifica così un tagset informativo articolato in tre classi: le Unità Informative di tipo Testuale, che strutturano l'Enunciato e ne determinano la portata semantica (il COM è una di queste); le Unità Dialogiche, di ausilio al progredire del discorso, che non contribuiscono al contenuto semantico dell'Enunciato e si rivolgono direttamente all'interlocutore; le Unità senza valore informativo.

I Comment che compongono il pattern possiedono una propria caratterizzazione azionale e sono interpretabili pragmaticamente; ciononostante possono non corrispondere né a frasi né a sintagmi verbali. Di tali pattern è stata proposta una classificazione in modelli meta-illocutivi

che comprende, fra i più frequenti, le relazioni di Lista, Comparazione, Rinforzo, Richiamo funzionale, Relazione Necessaria e Alternativa.

3.2 L'analisi conversazionale

L'analisi conversazionale è un settore di ricerca che si è sviluppato negli anni Sessanta grazie agli studi e alle analisi di conversazioni fatte dai sociologi Sacks e Schegloff. Il loro approccio al parlato interazionale ha assunto un valore di novità perché ha reso la struttura della conversazione oggetto di studio e di interesse. L'unità di analisi di riferimento è il turno degli scambi parlati e le unità costitutive del turno (Turn Construction Unit, TCU, unità sintattiche individuate tramite strumenti intonazionali) (Sacks, Schegloff & Jefferson, 1974). In seno a questo approccio si è spesso sviluppato un ampio dibattito sulla definizione dei limiti dei turni e della loro struttura interna, sviluppando così i concetti di multi-unit turn (Sacks, 1992), completion/continuation (Du Bois et al., 1993) e increment (Schegloff, 1996). La formazione delle strutture linguistiche dipende dall'attenzione del parlante e della comprensione dell'interlocutore, il turno dunque risulterà costruito interattivamente dai partecipanti lungo tutta l'interazione (Goodwin, 1979; sul ruolo interattivo del destinatario cfr. anche Chafe, 1993).

Il metodo dell'analisi conversazionale si basa sullo studio di conversazioni reali, considerando ogni evento come unico nonostante sia strutturato con metodi che hanno carattere generale. Un principio fondante di questo approccio è il concetto di consequenzialità per cui dato un segmento di interazione si crea la condizione perché ve ne sia un altro in risposta, inoltre, le diverse unità si distribuiscono nella loro sequenza seguendo un preciso schema per il quale a ogni tipo di unità corrisponde una precisa posizione (Schegloff, 1972).

Nell'organizzazione della conversazione i conversationalisti riconoscono due livelli, uno che agisce su base locale nei diversi turni e uno

che considera la conversazione come un'unità globale che ha un inizio e una fine. Un'altra nozione importante che ha modellato il metodo dell'analisi conversazionale è quella di contesto concepito come limitato, definito da Schegloff (1992) contesto intra-interazionale cioè tutto ciò che precede o segue il fenomeno conversazionale analizzato. Di conseguenza, i dati sociali e situazionali non vengono presi in considerazione in quanto irrilevanti per la strutturazione della conversazione che è gestita a livello locale.

3.3 Il modello autosegmentale-metrico

Il modello Autosegmentale-Metrico mette in primo piano la prospettiva prosodica e ha come punto cardine la rappresentazione dei toni su un livello diverso e indipendente da quello segmentale (Ladd, 1996). Questo modello definisce delle regole di associazione tra il piano tonale e quello prosodico tramite il sistema di trascrizione ToBI (Tone and Break Indices), basato principalmente sul lavoro di Pierrehumbert (1980) e i successivi sviluppi (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988).

L'intonazione è vista come l'interazione tra tre diversi componenti fonologici:

- La grammatica dei toni di una lingua, che ha la funzione di generare le sequenze tonali;
- La rappresentazione metrica;
- Le regole che specificano le modalità di associazione tra la melodia e il testo.

La rappresentazione fonologica dell'intonazione ha come risultato la rappresentazione melodica, cioè la curva della frequenza fondamentale. Il contorno intonativo viene concepito come una sequenza lineare di accenti intonativi (*Pitch Accents*), che sono definiti come unità discrete che si allineano per mezzo di regole di associazione al livello segmentale, gli

accenti possono associarsi solamente alle sillabe metricamente prominenti. Questo modello, oltre ai *Pitch Accents*, riconosce dei toni di confine che vengono associati ai margini di un costituente intonativo. A livello prosodico vengono riconosciuti due costituenti gerarchicamente organizzati: l'*Intonation phrase* (Sintagma Intonativo), gerarchicamente superiore, può inglobare uno o più *Intermediate Phrase* (Sintagma Intermedio), che è un costituente più breve del Sintagma Intonativo e formato solitamente da uno o due costituenti sintattici (Beckmann & Pierrehumbert 1986).

4. Conclusioni

In questo breve capitolo iniziale abbiamo visto come sia difficile riuscire a definire in modo univoco cosa si intenda con il concetto di *lingua parlata*. Ci si può riferire al parlato come una specifica modalità semiotica adottata dagli uomini per la loro comunicazione, oppure ci si può riferire alla specificità fisica della lingua e dunque avvicinarsi allo studio del parlato per indagare le articolazioni fonetiche e i correlati acustici di trasmissione del messaggio. Infine, si può utilizzare la nozione di parlato per riferirsi a una varietà di lingua per differenza alla varietà scritta e così studiare quali processi e quali differenze determinano le due varietà da un punto di vista grammaticale e linguistico.

In questo lavoro facciamo riferimento al *parlato* in due diverse accezioni. La prima è di carattere generale e si riferisce all'insieme delle attività che i parlanti realizzano e che strutturano gli eventi comunicativi attraverso l'uso dei canali fonico-uditivo e gestuale-visivo. La seconda è di carattere specifico e fa riferimento al solo canale fonico-uditivo e alle sue caratteristiche testuali, semiotiche e funzionali a cui faremo riferimento anche con i termini oralità e/o verbalità. A questa sovrapposizione terminologica non siamo riusciti a trovare una soluzione adatta poiché, come vedremo di seguito (cfr. cap. 2-4), le questioni intorno alla

multimodalità della lingua che porta a estendere il concetto di lingua riconoscendo il gesto come elemento pienamente verbale e linguistico non ci appaiono al momento risolte.

Per meglio comprendere la portata delle questioni appena accennate, nel capitolo seguente indagheremo gli studi sul gesto cercando di comprendere quali posizioni e approcci hanno determinato il dibattito di oggi. Pur essendo una questione di natura generale e teorica, riteniamo che sia un passaggio fondamentale per gli obiettivi di questo lavoro. Infatti, la costruzione di un corpus multimodale che possa rispondere alle esigenze di ricerca - e che quindi possa essere uno strumento che rappresenti la lingua parlata caratterizzata nella sua complessità attraverso la definizione di unità di analisi chiare e coerenti - implica avere ben chiare le dinamiche del dibattito.

Capitolo 2 – La gestualità

1. Cenni storici sugli studi sul gesto

La gestualità è un elemento caratterizzante della comunicazione parlata e assume un ruolo determinante nell'interazione, di conseguenza l'interesse sul gesto è condiviso da diversi campi e discipline, ciascuno con propri interrogativi e propri metodi di indagine. Sono state studiate le caratteristiche fisiche e semantiche del gesto, le loro correlazioni con il parlato e la loro potenza comunicativa, sia per arricchirne l'espressività e per controllarla al meglio, sia per comprenderne gli usi in un determinato sistema culturale, con l'obiettivo a volte di controllarlo, che per analizzarne i correlati clinici e psichici.

Ma con l'uso della parola *gesto* non tutti intendono la stessa cosa, c'è infatti chi con gesto identifica esclusivamente l'uso delle mani e/o delle braccia e chi, invece, considera gesto qualsiasi movimento del corpo. Inoltre, il gesto è stato da alcuni considerato archetipo del linguaggio, e per questa ragione è stato visto come potente strumento comunicativo universalmente condiviso, mentre altri lo hanno identificato come azione primitiva che tradisce ignoranza. Di certo intorno al gesto non sono mancate stereotipizzazioni conseguenti a ideologie e convinzioni che nel corso degli anni hanno portato anche a stigmatizzazione e ghettizzazione. Ma ciò che da sempre ha contraddistinto l'approccio allo studio della gesticolazione è la considerazione che i gesti siano espressione del *linguaggio del corpo* o *linguaggio non verbale* (Ricci Bitti - Cortesi 1977).

In questo paragrafo cercheremo di fare un excursus storico dei diversi approcci che si sono susseguiti nel tempo nell'indagine sul gesto, con l'obiettivo di comprendere e situare meglio i lavori di Adam Kendon e David

McNeill che senza dubbio possiamo considerare pionieri e fondatori degli *gesture studies* moderni.

Già greci e romani si mostrarono interessati all'uso della gesticolazione. Gli obiettivi che hanno spinto gli antichi a studiare e a dibattere sull'uso del gesto erano ben diversi da quelli attuali e da quelli che nel corso del tempo si sono susseguiti. L'interesse degli antichi era rivolto all'uso tecnico dei gesti nell'oratoria. Aristotele vedeva il gesto come tecnica utilizzata dagli oratori pubblici per aggraziarsi la folla, la gestualità era considerata una tecnica che sminuiva l'arte oratoria che doveva basarsi esclusivamente sull'uso del ragionamento. Nella tradizione romana, invece, i gesti sono stati rivalutati come tecniche oratorie efficaci e degne di essere coltivate e curate nel loro uso. Già Cicerone ne fece accenno nelle sue opere sull'oratoria enfatizzando la necessità di regolarne l'uso attraverso modalità misurate e definite. Ma sarà Quintiliano con la sua opera *Institutio oratoria* a dedicare una trattazione sistematica sull'uso della gesticolazione da parte dell'oratore. Nella terza parte dell'undicesimo libro della sua opera sulla formazione dell'oratore, egli dedica la sua attenzione all'uso della gesticolazione. Così come Cicerone, Quintiliano divide l'*actio* (cioè l'azione oratoria) in due componenti: la voce e il gesto (Quintiliano). Nell'affrontare il discorso sul *gestus* l'oratore sottolinea il primato della voce rispetto al gesto; infatti, scrive che «[...] bisogna parlare prima della voce, alla quale si conforma anche il gesto» (ibidem). Quello che viene chiamato *gestus*, e che noi traduciamo come gesto, fa riferimento a tutti i movimenti del corpo (quindi mani, braccia, postura, espressione facciale, movimenti della testa e sguardo). È molto interessante come Quintiliano inizia la trattazione sulle mani, egli infatti attribuisce alle mani una grande forza comunicativa facendo un'associazione con quelle che noi oggi chiamiamo azioni linguistiche e ipotizzando una capacità comunicativa interlinguistica:

«Per quanto riguarda le mani, poi, senza le quali l'azione oratoria sarebbe mutila e debole, è quasi impossibile dire quanti movimenti possiedano, perché essi eguagliano quasi il numero stesso delle parole. Infatti, le altre parti del corpo aiutano chi parla, queste, starei per dire, parlano da sé. Non è forse vero che con esse chiediamo, promettiamo, chiamiamo, congediamo, minacciamo, supplichiamo, abominiamo esprimiamo timore, interroghiamo, neghiamo, esprimiamo gioia, tristezza, dubbio, ammissione, pentimento, misura, quantità, numero, tempo? Non sono forse le mani a incitare, a trattenere, ad approvare, a esprimere meraviglia, e vergogna? Non svolgono forse la funzione degli avverbi e dei pronomi nell'indicare luoghi e persone? – al punto che in una così grande diversità di lingue parlate da tutti i popoli e da tutte le nazioni, mi sembra che questo sia l'unico linguaggio comune a tutti gli uomini.» (Libro XI, III.85-87)

Per l'oratore sono da evitare quei gesti che descrivono gli oggetti attraverso un'azione descrittiva o una pantomima, perché i gesti dell'oratore devono essere a servizio del pensiero piuttosto che delle parole. Per Quintiliano è bene che l'oratore usi il gesto per trasmettere la forza di ciò che dice e per indicare ciò che pensa, e non per sostituire ciò che dice a parole. Così fa un elenco di gesti che egli raccomanda di usare. Si tratta di diciotto gesti specifici di cui le modalità d'uso e *loci* dell'orazione in cui devono essere impiegati evidenziando che il movimento delle mani deve iniziare e concludersi con il pensiero che viene espresso. Purtroppo, non abbiamo testimonianze che possano dirci come le indicazioni di Quintiliano siano state recepite dai suoi coevi. Sappiamo però che la sua opera fu ben accolta e valorizzata all'inizio del XVII secolo in cui iniziarono ad apparire diversi trattati sulla retorica in cui veniva dedicato ampio spazio all'uso del gesto.

L'attenzione nei confronti del gesto e del comportamento corporeo crebbe nel periodo medioevale in quanto le azioni gestuali avevano raggiunto un alto livello di formalizzazione tale da diventare rituali legali di grande rilievo. A tal proposito Schmitt (1990), nell'introdurre il suo

saggio sul gesto nel medioevo, racconta un curioso aneddoto³ tra Ugo Capeto e l'imperatore che testimonierebbe come nel periodo medioevale il gesto avesse assunto un valore simbolico-formale di grande importanza sia sociale che politico. Ma nel periodo medioevale l'importanza riconosciuta alla corporeità e ai gesti non era legata semplicemente alla sfera legale, ma permeava soprattutto la vita religiosa. Proprio in questo periodo si assiste alla scrittura di veri e propri trattati sulla gestione del corpo e delle mani durante la preghiera e altri riti religiosi, finanche la predicazione.

Ma è la riscoperta dei testi latini, e soprattutto i testi di Cicerone e Quintiliano, avvenuta nei circoli umanisti a riportare in auge la figura dell'oratore come uomo dedito alla filosofia e all'azione politica. In questo modo la retorica ritorna a essere vista come una importante pratica per la formazione degli intellettuali. Questa tendenza arriva fino al sedicesimo secolo, dove la predicazione assume un ruolo importante sia per i protestanti che per i cattolici al punto da essere acquisita in modo sistematico nelle università, specialmente quelle gestite dall'ordine dei Gesuiti.

Il contemporaneo sviluppo dell'ideale di gentiluomo e cortigiano ha fatto in modo che gli interessi intorno al comportamento corporeo e gestuale riguardassero anche altre sfere della vita sociale oltre a quella religiosa. Alcune opere di riferimento e di grande circolazione in quell'epoca - come *Il libro del cortegiano* di Castiglione e *Il Galateo* di Della Casa - davano indicazioni sui comportamenti da assumere per poter diventare dei cortigiani basandosi sull'idea che fosse la condotta a influenzare lo status di una persona piuttosto che la sua origine sociale. Inoltre, i nuovi e sempre maggiori contatti tra le popolazioni europee e gli abitanti nativi delle nuove terre scoperte hanno portato a una rinnovata consapevolezza sull'uso dei gesti come lingua universalmente condivisa. Tutti questi fattori hanno

³ Durante un colloquio con Ugo Capeto (in quel periodo duca), l'imperatore Ottone II fece in modo che il duca gli portasse la spada (posta volutamente su una sedia). Questa azione, svolta davanti agli occhi dei presenti, assumeva un valore simbolico facendo così diventare Ugo Capeto uomo dell'imperatore.

contribuito a sistematizzare l'interesse sulla gestualità che si è tradotto nelle prime opere dedicate esclusivamente al gesto. Così nel 1571 Arias Montanus ha condotto uno studio sui gesti all'interno del suo commentario biblico, mentre nel 1616 Giovanni Bonifacio pubblica il suo *L'Arte de' Cenni* in cui propone anche una ricognizione di tutti i segni che è possibile realizzare con il corpo. Poco più di un decennio dopo a Milano, nel 1627, Francesco Bartolomeo Ferrari pubblica il suo *De veterum acclamationibus et plausu libri septem*, un saggio sui gesti di acclamazione. Mentre in Inghilterra John Bulwer nel 1644 pubblica il suo *Chirologia or the Naturall Language of the Hand and Chironomia or the Art of Manual Rhetoricke* (pensati come due volumi ma successivamente pubblicati insieme).

L'Arte de' Cenni di Bonifacio è uno dei primi libri pubblicati dedicati interamente alla gestualità, e altro non è che il tentativo di descrivere tutti i segni realizzabili dal corpo ma anche tutti i significati che possono essere trasmessi attraverso l'uso dell'abbigliamento. L'autore sosteneva che i segni realizzati attraverso l'uso del corpo rivelassero in modo più veritiero i sentimenti e le intenzioni delle persone. Il libro è diviso in due parti: nella prima parte sono descritti tutti i gesti che possono essere realizzati attraverso le varie parti del corpo, mentre nella seconda parte vengono analizzati i gesti e i segni usati nelle diverse professioni. Inoltre, Bonifacio sostiene l'idea che i gesti potrebbero diventare un linguaggio universale che potrebbe sostituire la confusione generata dalle diverse lingue parlate.

Così, come accennato poco prima, anche il volume dell'inglese Bulwer rappresenta una pubblicazione fondamentale nella storia degli studi sul gesto. L'autore dichiara di essere stato ispirato da un passaggio dell'opera di Bacone⁴ in cui il filosofo nota che Aristotele, nonostante abbia discusso sull'apparenza fisica, non ha espresso nulla rispetto alla gestualità. Per Bulwer è importante riconoscere nell'uso artificiale del gesto (quello indicato da Quintiliano) il fondamento naturale. Se infatti il significato dei gesti è giustificato dalla loro origine naturale, il loro impiego deve essere regolato

⁴ Si tratta del trattato *Sull'avanzamento e sul progresso del sapere umano e divino*, 1605.

dall'appropriatezza, grazia e decoro (aggiunge alla sua opera un elenco di gesti naturali che però sono considerati volgari e inappropriati). Il "linguaggio delle mani" era visto dall'autore come naturale e dunque strumento utile per comprendere la diretta relazione tra forma e significato; per Bulwer il gesto è un sintomo dello stato d'animo.

L'idea che la gestualità fosse un linguaggio molto più vicino alla natura rispetto a quanto lo fosse la lingua parlata (e che perciò è una forma di comunicazione comune a tutti i bambini), si è diffusa ed è stata alimentata per tutto il diciottesimo secolo spinta dai filosofi illuministi che si distanziavano in modo netto dall'idea dell'origine divina della lingua. Secondo l'idea di alcuni filosofi di quel periodo, lo studio dei gesti dava la possibilità di indagare le modalità della transizione da un segno naturale a un segno istituzionalizzato e condiviso (dando origine a una lingua) all'interno di una forma di comunicazione simbolica. Cioè i gesti rappresentavano, a loro avviso, una forma di espressione non ancora distorta dalla convenzione della lingua e credevano che la forma dell'espressione della gestualità fosse condivisa da tutti i bambini rappresentando così un potenziale linguaggio universale. Giambattista Vico è stato il primo filosofo del diciottesimo secolo a portare avanti queste idee nella sua opera *Scienza Nuova* (1725). È interessante la concezione di Vico sulla natura del segno linguistico. Per Vico il segno linguistico deriva dal processo di estensione metaforica delle rappresentazioni di immagini create attraverso l'immaginazione, un processo principalmente di tipo visuale, così i gesti rappresentano la prima forma di espressione linguistica.

Un altro importante contributo è arrivato da Etienne Bonnot de Condillac che nel suo *Essay sur l'origine des connaissances humaines* (1746) sostiene l'ipotesi secondo cui la lingua ha iniziato a emergere come una forma di istituzionalizzazione dei pianti sollecitati dalle emozioni e combinati con le azioni corporee, che vengono eseguite nei momenti di necessità nel tentativo di soddisfarle. Questo passaggio sarebbe avvenuto in forma progressiva, attraverso un periodo di copresenza del linguaggio di azione

con il linguaggio parlato. In questo modo i gesti rappresentavano per il filosofo un oggetto di indagine di grande interesse che avrebbe potuto rappresentare la chiave di volta nello studio dell'origine del linguaggio, a causa della loro ipotetica vicinanza alla prima forma di espressione linguistica.

Anche Denis Diderot si è occupato della questione in una delle sue lettere dedicate alla conoscenza delle persone private di uno dei maggiori sensi. Nella sua *Lettre sur les sourds et muets* (1751). Diderot sostiene che la natura lineare del linguaggio parlato abbia imposto dei vincoli all'espressione e afferma che la struttura del pensiero riflette la struttura della lingua usata. Così, in questa sua lettera, sostiene che se si analizzasse l'espressione gestuale di un sordo si potrebbe scoprire l'ordine naturale o originale dell'espressione del pensiero. Poiché, secondo la sua visione, una persona sorda non ha pregiudizi rispetto al modo di esprimere i suoi pensieri. I gesti per Diderot rappresentano un mezzo più diretto e meno artificiale per l'espressione del pensiero.

La prospettiva illuminista che vede nei gesti la prima forma di linguaggio ha alimentato un interesse più descrittivo che durerà fino a tutto l'Ottocento. Proprio nel diciannovesimo secolo iniziano a diffondersi le prime indagini etnografiche e antropologiche che avevano lo scopo di descrivere le variazioni culturali tipiche di quei territori scoperti nei secoli precedenti. Così, l'interesse sul linguaggio è divenuto centrale sia per la necessità degli etnografi e degli antropologi di comunicare con sistemi culturali e linguistici poco conosciuti, che per l'interesse verso l'origine della facoltà simbolica.

La più importante opera etnografica dedicata interamente alla gestualità è *La mimica degli antichi investigata nel gestire napoletano* di Andrea de Jorio (1832). Il testo si presenta come una collezione di gesti napoletani con l'obiettivo di dimostrarne la diretta discendenza da quelli greci. Questo approccio si basa su un confronto con delle testimonianze presenti nelle pitture dei vasi greci e di altri reperti archeologici. L'opera di de Jorio si

caratterizza anche per l'approccio teorico utilizzato nell'analisi dei gesti. Infatti, egli sostiene che un gesto può avere più di un significato e che questa variabilità semantica è data dalle correlazioni di altri fattori come l'espressione facciale, i movimenti della testa e soprattutto il parlato. Risulta inoltre interessante la definizione di gesto data da de Jorio: un particolare tipo di azione caratterizzato da due inseparabili componenti, quella fisica e quella semantica.

Nello stesso secolo i lavori di Tylor e Mallery diedero un importante contributo al riconoscimento della natura linguistica dei gesti. I due studiosi, infatti, con le loro ricerche sulla lingua dei segni degli indiani d'America, evidenziarono le capacità comunicative e simboliche dei gesti. Tylor, nel suo *Researches into the Early History of Mankind* (1865) sostiene che lo studio della lingua dei gesti possa aiutare a comprendere la natura del linguaggio e della sua origine. Ma se Tylor non si espone a sostenere la teoria che i gesti siano alla base dello sviluppo del linguaggio, Mallery - nel suo *Sign Language Among North American Indians Compared with that Among Other Peoples and Deaf Mutes* (1881) - conclude che sia i gesti che la voce sono stati utilizzati sin dall'inizio dell'origine del linguaggio, ma i gesti hanno giocato molto probabilmente un ruolo più importante.

Alla fine del diciannovesimo secolo si è assistito a un declino dell'interesse sui gesti. Le cause sono da rintracciare da una parte al divieto dato dalle Società linguistiche di Parigi di trattare il tema dell'origine del linguaggio, a causa delle mancate certezze scientifiche, e dall'altro dall'invito - del Congresso internazionale degli educatori sordi avvenuto a Milano nel 1880 - a non comunicare con i segni poiché questi avrebbero ostacolato l'apprendimento della lingua parlata e il conseguente inserimento nella società. Ma nel Novecento l'interesse per la gesticolazione e il dibattito sulla sua natura ritorneranno ad alimentare ricerche e scambi scientifici, coinvolgendo sia la psicologia che la linguistica, che condurranno alla nascita di quello che noi oggi chiamiamo *gesture studies*. Di queste questioni tratteremo nel paragrafo successivo, dove affronteremo

sia il dibattito sulla natura verbale dei gesti, che vede protagonista David McNeill, che i primi studi di tipo linguistico portati avanti da Adam Kendon.

2. La nascita degli *gesture studies*: i contributi di Adam Kendon e David McNeill

2.1 Il gesto come linguaggio non verbale

Nel 1941, periodo in cui si è assistito al drastico calo di interesse nei confronti della gestualità, ha visto la luce l'importante lavoro di David Efron, dal titolo *Gesture, race and culture*. Lo studio antropologico nasce nel pieno periodo dalle idee razziste del nazifascismo e ha l'obiettivo di dimostrare come i fattori culturali siano del tutto disgiunti dai fattori biologici. Efron ha provato che la gestualità è determinata anche da fattori culturali che inevitabilmente si trasformano a seguito del contatto tra popolazioni diverse: lo studioso ha monitorato le abitudini gestuali di emigrati italiani ed ebrei dimostrando che tendono a cambiare, fino quasi a scomparire, man mano che i membri delle comunità si inseriscono nella nuova comunità di destinazione. Questo studio è uno degli ultimi a utilizzare il metodo etnoantropologico infatti, sempre agli inizi del Novecento, si assiste all'incremento delle indagini di tipo psicologico. Sia il comportamentismo che la psicoanalisi hanno enfatizzato l'interesse per quei comportamenti che andavano oltre il controllo conscio dell'individuo. Quest'attenzione segna una svolta anche nel metodo di indagine, se infatti gli antropologi e gli etnologi erano interessati allo studio del gesto nell'ambiente della sua realizzazione, gli psicologi hanno iniziato gli studi basandosi su esperimenti di tipo laboratoriale al fine di acquisire dati per poter così modellizzare i comportamenti umani attraverso lo studio di campioni di popolazione. Questo interesse vede i gesti, insieme ad altri elementi, come fattori costituenti la comunicazione non-verbale. Ciò fa in modo che i gesti delle mani non siano più oggetto di analisi specifica, ma diventino uno dei tanti elementi della comunicazione corporea. Gregory

Bateson, insieme ad altri colleghi⁵, fu il primo a indagare la comunicazione non-verbale sia in interazioni quotidiane che in seduta psicoterapica. I primi risultati hanno dimostrato che tante informazioni passavano attraverso l'uso di canali diversi rispetto al canale enunciativo. Una parte cospicua della comunicazione passava attraverso il tono della voce, le modalità di esitazione, lo stile dell'eloquio, la struttura intonativa, la qualità vocalica, la postura del corpo, i movimenti del corpo, lo sguardo e l'espressione facciale.

In quegli anni si cercherà di estendere anche all'interazione umana la teoria dell'informazione, questo farà in modo che si arrivi a una distinzione teorica fondamentale al punto di influenzare in modo sostanziale la divisione tra comunicazione verbale e non-verbale. Il comportamento umano venne concepito come codice all'interno di un sistema di trasmissione di informazione, portando a distinguere tra un codice analogico - di cui fanno parte l'espressione facciale e qualsiasi comportamento corporeo (quindi anche i gesti) e che stabiliscono e mantengono le relazioni interpersonali - e un codice digitale - di cui fa parte l'espressione verbale che trasmette l'informazione proposizionale (Watzlawick et al. 1971).

Nonostante in quegli anni l'interesse coinvolgeva lo studio della comunicazione non-verbale tralasciando di prestare particolare attenzione allo studio dei gesti, si assiste a qualche tentativo. Uno dei più importanti è quello fatto da Ekman e Friesen nel 1969 che hanno cercato di dare una schematicità agli studi sulla comunicazione non-verbale proponendo una classificazione in riferimento al grado di intenzione comunicativa e al tipo di significato veicolato e che è stata ispirata anche dagli studi di Efron, ponendo di conseguenza particolare attenzione sui gesti.

⁵ Gregory Bateson insieme a Margaret Mead ha pubblicato *Balinese Character* nel 1942, uno studio sul comportamento quotidiano della popolazione Balinese. Mentre nel 1956 pubblica, insieme a Don D. Jackson, Jay Haley and John H. Weakland un articolo dal titolo *Toward a theory of schizophrenia*, in cui venivano analizzati i comportamenti interazionali in famiglie con familiari schizofrenici.

Nonostante il Novecento sia il secolo che ha visto la nascita della linguistica moderna e nonostante nell'ambito della psicologia crescesse l'interesse per la comunicazione, in particolare per quella non-verbale, proprio in questo secolo si è registrato un forte disinteresse nei confronti dei gesti. Così per molto tempo la comunicazione è stata vista dagli studiosi come nettamente separata in verbale e non-verbale, come se i parlanti esercitassero due tipi di comunicazione in modo distinto e per scopi diversi (uno proposizionale e uno relazionale ed emotivo).

Solo nella seconda metà del ventesimo secolo, altri studiosi, come Argyle (1991) o De Mauro (2000), hanno iniziato a evidenziare i limiti derivanti da questa netta divisione tra verbale e non-verbale e la convinzione che il canale non-verbale fosse comune con le altre specie animali. A queste posizioni si aggiungono gli studi pionieristici di Virginia Volterra e Elizabeth Bates che hanno dimostrato la centralità dei primi gesti nello sviluppo linguistico dei bambini (Bates, Camaioni, Volterra 1975).

In questo clima iniziano a svilupparsi gli studi moderni sul gesto, dando inizio a quelli che oggi chiamiamo *gesture studies*. Sarà, infatti, David McNeill nel 1985 a dare inizio a un dibattito sulla natura verbale dei gesti di cui tratteremo ampiamente nei paragrafi successivi.

2.2 Adam Kendon: dalle unità del gesto all'enunciato come azione visibile

Adam Kendon può essere definito il pioniere degli studi sul gesto. È stato uno dei primi studiosi ad approcciarsi alla gestualità come elemento fondamentale dell'enunciazione, riuscendo a definirne unità e struttura. La portata del suo contributo è tale che a lui si deve gran parte del lessico proprio degli *gesture studies*.

Uno dei suoi primi contributi, pubblicato nel 1972, è uno studio sulla relazione tra il parlato e i movimenti del corpo (che egli definisce *gesticulation* facendo riferimento alle mani, alla testa e alle braccia). In questo suo studio analizza un minuto e mezzo di video e audio registrati all'interno

di un pub di Londra e che vede un uomo chiacchierare con altre persone presenti. L'intento dello studio di Kendon è quello di comprendere come siano organizzate le configurazioni del movimento e come si correlino con le strutture del parlato. Infatti, la sua analisi parte dalla trascrizione e segmentazione delle unità del parlato⁶ e delle correlazioni con i movimenti. Così Kendon, nonostante la ridotta quantità di dati analizzati, giunge a cinque conclusioni: 1) così come il flusso del parlato è divisibile gerarchicamente in unità, allo stesso modo può avvenire per i movimenti del corpo che a esso si associano; come se ogni unità del parlato avesse un proprio "equivalente"⁷ nei movimenti; 2) ogni unità di parlato si distingue anche per la struttura del movimento o per la parte del corpo che viene impiegata; 3) prima di ogni unità di parlato è stato notato un cambio di posizione di una o più parti del corpo, questo tipo di movimento è chiamato da Kendon *speech preparatory*; 4) il movimento *speech preparatory* inizia in anticipo rispetto all'unità di parlato in relazione alla sua grandezza, cioè più grande è l'unità e prima e più estesi sono i movimenti preparatori; 5) Kendon nota che i cambiamenti della struttura del movimento sono coordinati con i cambiamenti della struttura del suono e che nella corrispondenza movimenti e lessico, il movimento inizia prima ma finisce insieme all'elemento lessicale.

Con il suo contributo Kendon sembra definire un approccio prettamente linguistico allo studio del gesto, tralasciando interpretazioni di carattere psicologico o antropologico che avevano caratterizzato l'interesse sulla gesticolazione di quegli anni. L'approccio di Kendon alla materia lo ha condotto a definirne unità e a dare un'interpretazione della gestione del contenuto e dei modi di rappresentazione dell'idea espressa. Lo studioso

⁶ L'unità più piccola considerata da Kendon è di natura fonetica, è chiamata Phrase e si basa sulle definizioni di Tone Group di Boomer 1965, Halliday 1963 e Lieberman 1967. Le Phrase fonetiche vicine possono formare delle Phrase Cluster. L'unità superiore è la Locution che tende a essere la frase completa. Le unità Locution adiacenti che si distinguono dalle altre vicine per qualche elemento condiviso formano le Locution Groups. Queste ultime possono formare un raggruppamento ancora superiore che viene definito Locution Cluster (Kendon 1972).

⁷ Kendon stesso mette il termine *equivalente* tra virgolette.

nel sostenere che gesti e parlato appaiono insieme e quindi rappresentano un unico atto di enunciazione, pone il presupposto che i movimenti associati al parlato siano fenomenicamente distinti dagli altri esperimenti (Kendon 1972). Nell'analizzare la struttura del flusso della gesticolazione, Kendon nel 1980 ha utilizzato il concetto di Gestural Phrase⁸, delimitata da momenti di movimento e momenti di riposo, in cui è possibile riconoscere diverse unità gerarchicamente ordinate. All'interno dell'escursione, cioè dal momento in cui l'arto superiore inizia a muoversi fino a quando ritorna alla posizione di riposo, si possono rintracciare più di una Gesture Phrase che viene definita e distinta per ogni fase di movimento di escursione in cui l'arto (o una sua parte) raggiunge il picco di sforzo. Il momento di picco viene definito *stroke* e solitamente può essere preceduto dalla fase di *preparation*, cioè la fase in cui l'arto passa dalla posizione di riposo alla posizione in cui inizia la fase di picco (lo *stroke*), e può essere seguito dalla fase di ritorno alla posizione di riposto che viene chiamata *retraction* (Kendon 1980).

Nel lavoro del 2004, in cui Kendon raccoglie in modo sistematico i risultati dei lavori portati avanti negli anni, definisce in modo più chiaro la divisione in unità del flusso gestuale. In primis le unità da considerare per l'azione gestuale sono due: la *gesture unit* e la *gesture phrase*. Le *gesture phrases* sono unità di azioni visibili del corpo identificate da caratteristiche kinesiche che corrispondono a unità significative di azioni come la deissi, la rappresentazione, la pantomima o la realizzazione di un gesto convenzionalizzato. Se nel 1980 Kendon faceva riferimento alla *gestural phrase* per indentificare l'intera escursione del movimento degli arti, nel 2004 – così come appena accennato – egli parla di *gesture units* all'interno della quale è possibile distinguere le fasi del movimento di cui la più importante, in quanto portatrice di significato, è lo *stroke* che può essere anticipato o seguito dalle altre due *phases*. È proprio lo *stroke* a definire l'unità *gesture phrase*. Riassumendo, l'unità più grande è la *gesture units* che

⁸ *Gesticular Phrase* lo traduciamo con *Frase gestuale*.

è definita dall'intera escursione del movimento che può contenere una o più unità di *gesture phrase* che è definita dal cambio di direzione o dal tipo di gesto. Si può generalmente affermare che la *gesture phrase* è quella che riconosciamo come singolo gesto. Una *gesture phrase* a sua volta è composta dalle diverse *gesture phases* (*preparation, stroke e retraction*)⁹.

Finora studiosi ed esperti si erano approcciati alla gestualità come elemento ancillare della comunicazione, Kendon invece inaugura un metodo che si avvicina allo studio del gesto riconoscendone la natura linguistica e vedendo nella gesticolazione una componente dell'atto enunciativo. Questa sua intuizione viene già descritta nel 1972 e viene ancor meglio definita nel suo lavoro del 2004 in cui parla dell'enunciato¹⁰ come un oggetto composto da materiale gestuale e di parlato, elementi che possono dirsi componenti costitutivi dell'enunciato stesso. La componente parlata fa riferimento a tutte le attività vocali impiegate nella lingua (e quindi dal lessico agli elementi prosodici), mentre la componente gestuale a tutte le attività visibili del corpo che esprimono valore comunicativo.

Se nell'organizzazione della struttura gestuale non è di certo riconoscibile una diretta corrispondenza con la struttura lessicale e sintattica del parlato, è sicuramente osservabile una correlazione a livello del contenuto espresso dall'enunciato. Così lo studioso ci ricorda che gesticolazione e parlato lavorano insieme in una relazione intima e complessa di grande flessibilità per esprimere il contenuto dell'enunciato. Kendon identifica tre tipi di movimenti gestuali sulla base delle modalità usate per esprimere il contenuto: gesti che assumono la forma di oggetti, movimenti che consistono in azioni direttamente descritte dalla componente parlata e gesti che suggeriscono la disposizione spaziale o lo spostamento di un oggetto nello spazio (Kendon 1980). I modi di codifica

⁹ Kita aggiungerà un'altra *gesture phase* che chiama *hold*. Si approfondirà nei paragrafi successivi.

¹⁰ Adam Kendon definisce l'enunciato come «[...] the ensemble of actions, whether composed of speech alone, of visible action alone, or of a combination of the two, that counts for participants as a 'turn' or 'contribution' or 'move' within the occasion of interaction in which they are engaged.» (Kendon, 2004).

usati dal gesto cambiano: spesso i gesti trasmettono il significato in modo globale e sintetico, è il caso dei gesti spontanei (la gesticolazione), oppure possono utilizzare forme ben determinate a cui si associa un significato (gli emblemi) così come avviene per le parole. Ma tra le due modalità non è possibile fare una netta distinzione, ma è possibile osservare la trasformazione di un gesto spontaneo e globale in un gesto dalla forma stabile e codificata. Quando ciò avviene si assiste anche a un cambio del significato che solitamente diventa più astratto: non codifica più informazioni relative a una specifica azione da descrivere, o alla disposizione spaziale o a una caratteristica visibile di un oggetto. In gesto in questione, nello stabilizzare il rapporto tra forma e contenuto, diventa come una parola, come una vera e propria lessicalizzazione. Inoltre, Kendon ha notato che quando i gesti accompagnano il parlato molto spesso servono per rappresentare aspetti del contenuto attraverso immagini o pantomime; mentre quando sono usati in modo alternativo al parlato realizzano la stessa funzione delle parole o di una frase. In questo senso è possibile parlare di gesti che diventano parole (Kendon 1988; 2004).

Tenendo in considerazione elementi cinetici e semantici della componente gestuale, Kendon sviluppa il concetto di *gesture families*. Le famiglie gestuali sono raggruppamenti di espressioni gestuali che hanno in comune una o più caratteristiche cinetiche o formazionali. Cioè sono gesti raggruppabili o su base semantica o perché formati su una stessa struttura di movimento (Kendon 2004). La relazione tra le due componenti e le forme assunte dai diversi gesti dipende sia dagli obiettivi comunicativi del parlante che dalla situazione comunicativa. In riferimento all'interazione semantica tra gesto e parlato, Kendon classifica i gesti in tre macrocategorie sulla base della funzione che esprimono. Ci sono i gesti che esprimono una funzione *referenziale* e che possibile dividere in gesti *rappresentativi*, cioè quelli che rappresentano un aspetto del contenuto dell'enunciato, e gesti *deittici*¹¹ che indicano un oggetto a cui si fa riferimento nel discorso. Poi ci

¹¹ Generalmente i *pointing gestures*.

sono i gesti con valore *pragmatico*¹² e possono essere di tre tipi, i gesti con funzione *modale* che alterano ciò che viene detto esprimendo come l'enunciato deve essere interpretato; i gesti con funzione *performativa* che indicano il tipo di atto linguistico o di mossa interazionale realizzata e infine i gesti con funzione di *parsing*, quelli solitamente utilizzati per puntualizzare le parti del discorso. E infine ci sono i gesti *interattivi* o *interpersonali*, sono quei gesti che regolano il turno nelle conversazioni o per specificare a chi è indirizzato l'enunciato (Kendon 2004).

I lavori di Adam Kendon hanno contribuito alla nascita degli studi moderni sul gesto sotto tanti punti di vista: hanno innanzitutto dato un fondamento teorico svincolando la gestualità dall'ibrido mondo della paraverbalità e riconoscendone il ruolo nell'atto enunciativo e soprattutto hanno impostato un metodo di analisi basato sull'individuazione delle unità di riferimento del gesto. Ma come vedremo di seguito, gli *gesture studies* devono molto anche a un altro pioniere in questo campo di indagine, David McNeill.

2.3 David McNeill: gesto, parlato e *growth point*

Nel 1985 David McNeill pubblica un articolo dall'eloquente titolo *So You Think Gestures Are Nonverbal?* aprendo un dibattito sulla natura verbale del gesto. Come è emerso nei paragrafi precedenti, il gesto è sempre stato concepito dagli studiosi che se ne sono occupati - dall'antichità agli anni Ottanta del Novecento - come elemento para-verbale. Portando la disciplina linguistica a sviluppare con ritardo l'interesse nei confronti della gestualità e della multimodalità. Così il lavoro di McNeill, che si ispira alle indagini di Kendon, contribuisce al dibattito e butta le basi per le prime indagini linguistiche sul ruolo del gesto nella comunicazione umana.

L'obiettivo che McNeill dichiara di porsi, ad apertura del suo lavoro monografico del 1992, è quello di «embracing in one theoretical system two

¹² Sul valore pragmatico dei gesti si rimanda ai paragrafi successivi.

forms of expression, speech and action» (McNeill 1992) in modo da dimostrare come gesto e parlato siano collegati e in cosa differiscono. Per lo studioso significa concepire parlato e gesto come un unico sistema integrato, staccandosi nettamente dall'idea di "linguaggio del corpo". In modo particolare egli si è interessato all'indagine di quei gesti che esibiscono un'immagine. La scelta è motivata dal fatto che, secondo McNeill, le persone nel realizzare questa tipologia di gesti rendono visibili i loro pensieri interni e il modo in cui comprendono gli eventi della realtà (ibidem).

McNeill classifica i gesti in cinque tipi. I gesti iconici che si caratterizzano per la stretta relazione formale con il contenuto del parlato, in questa co-espressività si realizza anche la complementarità semantica tra le due parti. Parlato e gesto possono riferirsi allo stesso evento e possono parzialmente sovrapporsi, ma le immagini che rappresentano sono diverse. I gesti metaforici che sono come i gesti iconici ma il cui contenuto illustrato è astratto, questi gesti rappresentano un'immagine dell'invisibile. Un terzo tipo sono i beat (o batonici), chiamati in questo modo perché si muovono in accordo con il ritmo del parlato e sono distinti dagli altri gesti anche perché hanno due fasi del movimento: dentro/fuori, su/giù. Il valore semiotico di un gesto beat sta nel fatto che enfatizza la parola o la frase che accompagna, non per il valore semantico espresso, ma per il suo contenuto discorsivo-pragmatico. Un'altra tipologia individuata da McNeill sono i gesti coesivi, cioè quei movimenti che aiutano a tenere insieme tematicamente parti diverse del discorso che sono temporalmente separate. Mentre i gesti beat evidenziano la discontinuità, i gesti coesivi enfatizzano la continuità discorsiva e sono molto eclettici nella forma, ma esprimono il valore coesivo attraverso la ripetizione della forma, del movimento o del luogo di gesticolazione. Infine, ci sono i deittici che svolgono la funzione di indicare oggetti o eventi nel mondo concreto o uno spazio in cui non c'è nulla esprimendo così un riferimento astratto.

Tuttavia le tipologie dei movimenti che accompagnano il parlato si distinguono anche sulla base della relazione che intercorre con il parlato (Kendon 1988). McNeill, ispirandosi al contributo di Adam Kendon, ordina i movimenti con valore comunicativo su un continuum che chiama appunto Kendon's continuum:

*Gesticulation -> Speech-linked gestures -> Emblems -> Pantomimes -> Sign Language*¹³

Muovendosi da sinistra verso destra del continuum si nota il declino dell'obbligatorietà della presenza del parlato, crescono le proprietà caratteristiche del linguaggio e i gesti idiosincratici vengono sostituiti da gesti socialmente codificati. La *gesticolazione* (che McNeill chiama anche con il termine *gesture*) comprende i movimenti spontanei e idiosincratici delle mani che accompagnano il parlato e quasi mai occorrono in sua assenza. Sotto l'etichetta *Language like Gestures* sono inclusi tutti quei movimenti che nella forma sono simili alla gesticolazione ma che esprimono una funzione grammaticale ben precisa nell'enunciato sostituendo una parte di parlato (un esempio è: "i genitori stavano bene ma i bambini [gesto]¹⁴"; in questa frase il gesto completa il contenuto della frase occupando lo spazio di un aggettivo). Nelle *pantomime* le mani rappresentano un oggetto o un'azione senza che la presenza del parlato sia obbligatoria. Gli *emblemi* hanno uno standard rispetto alla forma e posseggono la proprietà di poter sostituire il parlato, un esempio è il gesto per dire "OK". Infine, la *lingua dei segni* (come l'ASL o la LIS) è un sistema pienamente linguistico con la proprietà della segmentazione, composizionalità, con lessico e sintassi propri, arbitrarietà, uno standard e una comunità di segnanti (McNeill 1992 - 2005).

¹³ Nella prima versione del continuum, in McNeill 1992, la categoria "Speech-linked gestures" viene denominata "Language-like Gestures". Inoltre, le pantomime precedevano gli emblemi specificando che «whiche one - emblem or pantomime - belongs more to the right probably is arbitrary».

¹⁴ L'esempio è tratto e tradotto da McNeill 1992.

Il *Kendon's continuum* può essere visto come un sistema complesso di differenti continua basati su diversi elementi analitici che dispongono in modo differenziato i tipi di movimenti. Il primo continuum, come abbiamo visto, prende in considerazione la relazione dei movimenti con il parlato. Il secondo continuum si basa sulla presenza o assenza delle caratteristiche semiotiche proprie di un sistema linguistico: la gesticolazione e la lingua dei segni mantengono la stessa posizione, ciò che avviene è un cambio tra pantomime ed emblemi come è possibile vedere nella tabella che segue.

Tabella 1 *Kendon's continuum*

| Gesticulation | Pantomime | Emblems | Sign Language |
|-----------------------------------|-----------------------------------|---|------------------------------------|
| Assenza di proprietà linguistiche | Assenza di proprietà linguistiche | Presenza di qualche proprietà linguistica | Presenza di proprietà linguistiche |

Dalla comparazione tra il primo e il secondo continuum si evince che la presenza del parlato è direttamente connessa con l'espressione delle proprietà linguistiche. Infatti, quando la modalità manuale non esprime il carattere linguistico, la presenza del parlato è obbligatoria in quanto portatrice di quei valori. Quando, viceversa, la modalità manuale possiede le caratteristiche linguistiche, il parlato è assente perché siamo in presenza di una lingua dei segni. Così si può affermare che gesto e parlato si combinano in un sistema in cui ciascuna modalità esprime le proprie funzioni supportandosi (McNeill 2005).

Il terzo continuum si basa sul grado di convenzionalità, cioè quanto la relazione tra la forma di un gesto e il suo significato siano socialmente costruiti o rappresentino uno standard collettivo. La tabella seguente riporta il terzo continuum, dove è possibile notare all'estremità di destra la gesticolazione che è costituita dall'assenza di convenzionalità in netto contrasto con i segni della lingua dei segni.

Tabella 2 Kendon's continuum e convenzionalità

| Gesticulation | Pantomime | Emblems | Sign Language |
|------------------------|------------------------|-----------------------------|-------------------------------|
| Non convenzionalizzato | Non convenzionalizzato | In parte convenzionalizzato | Pienamente convenzionalizzato |

Mentre il quarto continuum riguarda le differenze semiotiche, in questo modo è possibile apprezzare la ricchezza che emerge dalla combinazione di due sistemi semiotici diversi come la gesticolazione e il parlato.

Tabella 3 Kendon's continuum e differenze semiotiche

| Gesticulation | Pantomime | Emblems | Sign Language |
|---------------------|---------------------|------------------------|------------------------|
| Globale e sintetico | Globale e sintetico | Segmentato e sintetico | Segmentato e analitico |

Le argomentazioni che McNeill porta a sostegno della natura verbale del gesto sono cinque e si basano sulla dimostrazione che gesto e parlato formano un unico sistema. Quello che per Kendon sono due componenti dell'enunciato, per McNeill sono due aspetti di un singolo processo cognitivo (McNeill 1985).

La prima argomentazione a sostegno della stretta unione tra gesto e parlato è che i gesti si realizzano solamente durante il parlato. Fatta eccezione per emblemi e pantomime che possono comparire anche in assenza del parlato, le altre tipologie di gesti compare per il 90% delle volte durante l'enunciazione. La seconda argomentazione dice che gesti e parlato sono semanticamente e pragmaticamente co-espressivi. Cioè, i gesti che accompagnano gli enunciati esprimono lo stesso significato della parte parlata e realizzano la stessa funzione pragmatica. Presentano la stessa immagine del parlato come un oggetto, o uno spazio o altre realtà fisiche. Il terzo argomento è che gesti e parlato sono sincroni, cioè i segmenti linguistici e le unità gestuali si realizzano contemporaneamente. La quarta argomentazione riguarda lo sviluppo del linguaggio nei bambini, McNeill

infatti nota che gesti e parlato si sviluppano insieme nei bambini. Nello specifico i primi gesti nei bambini sono deittici concreti (indicano referenti concreti) e alcuni tipi di gesti iconici. Successivamente verranno sviluppati altri tipi di gesti iconici, batonici, metaforici e in fine i gesti deittici astratti. Si nota quindi che lo sviluppo dei gesti va di pari passo con lo sviluppo del linguaggio che va da costruzioni linguistiche di tipo referenziali e che arriva alla strutturazione di discorsi e alle costruzioni metaforiche e astratte, passando da elaborazioni descrittive. Infine, la quinta argomentazione riguarda le afasie: gesto e parlato scompaiono insieme nelle afasie. Le afasie che interessano l'area di Broca consistono tendenzialmente nella compromissione dell'abilità di combinazione di termini in unità grammaticali più grandi, ma rimane pressoché intatta la capacità d'uso dei termini di riferimento. Per quanto riguarda i gesti si assiste a una abbondanza di gesti iconici e a una maggiore discretezza, con la perdita dei gesti batonici e metaforici. Mentre, per quanto riguarda l'afasia di Wernicke si caratterizza per un parlato fluente a cui manca la coerenza semantica. Di conseguenza il gesto appare grande, mobile ma con un significato non interpretabile (McNeill 1992).

La necessità di dover portare argomentazioni a favore della tesi secondo cui gesto e parlato sono due aspetti di una stessa azione, nasce dalle differenze visibili che intercorrono tra gesto e parlato e che lo stesso McNeill mette in evidenza. Gesto e parlato, infatti, trasmettono il significato attraverso modalità diverse. Il parlato si caratterizza per la capacità di segmentare e linearizzare gli eventi e le idee, organizzando i diversi segmenti in stringhe di parola gerarchicamente strutturate. Mentre il gesto riesce a trasmettere diversi significati senza sottostare alla segmentazione o linearizzazione, in modo globale, sintetico e non gerarchico. Di conseguenza i gesti sono non combinatori, due gesti che vengono prodotti insieme non formano un gesto più grande e più complesso. Questa caratteristica è in netto contrasto con la struttura del parlato che per antonomasia è costituita da unità lessicali che si compongono fra loro per la

formazione di unità più grandi e più complesse come le frasi e i discorsi. Inoltre, se per il sistema linguistico del parlato esiste una forma standard, naturalmente diversa per ogni lingua, così non è per il gesto. I gesti di diversi parlanti possono veicolare lo stesso significato ma utilizzando forme molto diverse, fatta eccezione per gli emblemi. Infine, al gesto manca la proprietà della doppia articolazione che per McNeill è strettamente legata all'arbitrarietà. L'autore afferma che sia la forma cinetica che quella sonora sono livelli indipendenti dal linguaggio, ma la forma cinetica nel gesto è determinata dal significato da veicolare mentre la forma sonora si associa al significato in modo arbitrario, «if we explain the meaning of a gesture we explain the form» (McNeill 1992).

In sintesi, possiamo affermare che le argomentazioni e le tesi di McNeill sembrano illustrare una situazione che vede il parlato e il gesto come sistema unitario a livello cognitivo e psicologico, ma che a livello dell'espressione utilizzano modi e metodi diversi. Questo ha condotto lo stesso studioso alla concettualizzazione del *growth point*, per cui si intende un'unità psicologica (psychological predicates) o unità di "discorso interiore". È un'unità non scomponibile, sempre connessa al contesto dell'interazione, inclusi gli aspetti sociali, ed è il luogo in cui una nuova idea verbale nasce per poi essere esternalizzata come discorso seguendo le regole proprie della *langue*. Il *growth-point* è l'unità minima della dialettica tra il lato *imagery* della nostra cognizione e il la *lingua*. Rappresenta dunque una instabilità tra pensiero e attraverso la combinazione di elementi semioticamente opposti (l'immagine e la forma linguistica) attraverso cui viene garantita la co-espressività (McNeill 2005). Dunque, il *growth-point* può essere visto come l'unità cognitiva da cui scaturisce l'atto di enunciazione che si realizza attraverso due modalità distinte ma co-espressive.

David McNeill, insieme a Adam Kendon, possono essere definiti i fondatori dei moderni studi sul gesto. Le loro ricerche e i loro studi hanno contribuito a riconoscere il valore verbale e linguistico della gestualità

dando principio a ricerche sempre più approfondite sulla natura linguistica del gesto, sul suo ruolo all'interno dell'interazione comunicativa e le sue funzioni cognitive. Se da un lato Kendon sembra più incentrato sull'aspetto e sul ruolo comunicativo ed espressivo, McNeill dall'altro si interroga principalmente sull'aspetto psicologico e cognitivo. Certamente da questi studi, come dicevamo, sono scaturite molte altre ricerche di cui sinteticamente cercheremo di riportare le più salienti nei paragrafi successivi.

3. Le funzioni del gesto

Il dibattito intorno al ruolo del gesto nell'interazione si concentra principalmente sulla natura del legame con il parlato. La questione è se gesto e parlato fanno parte di uno stesso processo o, invece, il gesto è un sistema autonomo che coesiste con il parlato e la cui funzione è da collegare ai meccanismi cognitivi più che a quelli comunicativi. Pur avendo constatato le differenze strutturali, soprattutto a livello semiologico, tra gesto e parlato, riteniamo limitanti e superate le teorie che sostengono la loro netta separazione. Piuttosto, così come già affermato da Kendon (2000) e McNeill (1985), è necessario approfondire con ulteriori ricerche la questione e allargare di conseguenza il concetto di linguaggio (cfr. Capitolo 3).

In questo contesto è utile non dimenticare che le critiche mosse alla posizione assunta da McNeill nel 1985 con il suo articolo *So You Think Gestures Are Nonverbal?* e riprese nel suo volume *Hand and Mind* nel 1992, hanno sostenuto la netta separazione tra gesto e parlato basandosi sulla *Lexical Retrieval Hypothesis* (Feyereisen 1987), dimostrando che i gesti occorrono anche in assenza di parlato e basandosi sul non allineamento dei due elementi sia a livello temporale che semantico (Butterworth - Hadar 1989) e sulla difficoltà di interpretazione semantica univoca e condivisa di alcuni gesti sia in assenza del parlato corrispondente che in sua presenza (Feyereisen - Van de Wiele - Dubois 1988). La risposta di McNeill (1987),

alle critiche di Fayereisen e dei colleghi, si è limitata alla constatazione che le sue considerazioni riguardavano la gesticolazione mentre i critici avevano analizzato le pantomime e gli emblemi che non risultano sovrapponibili ai gesti.

Senza dubbio gli studi analizzati finora si basano su una visione della lingua ancora troppo statica, che non vede l'interazione linguistica come momento dinamico in cui il significato si costruisce sull'interazione tra elementi linguistici, culturali, sociali e cognitivi. Questa consapevolezza deve portare necessariamente a un approccio sempre più aperto e dinamico allo studio dell'interazione comunicativa. Per meglio comprendere la complessità della questione, in questo paragrafo faremo una sintesi dei maggiori studi che hanno indagato le funzioni del gesto sia da un punto di vista cognitivo che pragmatico.

3.1 Gesto, cervello e cognizione

Le ricerche intorno al ruolo svolto dalla gesticolazione nella cognizione umana passano, quasi in modo propedeutico, per lo studio della relazione tra i diversi tipi di gesti e le parti del cervello coinvolte nella realizzazione.

Nel 1992 David McNeil, nel suo volume *Hand and Mind*, riporta e approfondisce uno studio pubblicato due anni prima insieme ad altri colleghi (McNeill, Levy e Pedelty 1990) in cui presenta un'ipotesi su come il cervello controlli e metta in relazione gesto e linguaggio. Questo studio si interroga su tre aspetti: i gesti sono lateralizzati tutti allo stesso modo? Cosa avviene in presenza di afasia? Come si relazionano gesto e linguaggio nei pazienti split-brain?

Senza dubbio i gesti sono azioni manuali e in quanto tali possono apparire lateralizzati sulla base della mano dominante, ovvero dal fatto che una persona possa essere mancina o destrorsa. Ma i gesti partecipano anche al processo comunicativo e simbolico e quindi la lateralizzazione potrebbe

dipendere da fattori funzionali e dal ruolo svolto dal gesto all'interno del discorso.

Già Kimura (1976) aveva notato che i destrorsi tendenzialmente usano la mano destra per i gesti correlati al parlato. Mentre i mancini utilizzano entrambi le mani con una leggera preferenza per la sinistra. Da queste costatazioni Kimura ipotizza che il parlato e i gesti siano controllati da un comune sistema che è localizzato nell'emisfero sinistro.

McNeill, basandosi sull'analisi di pazienti afasici (sia fluenti che non fluenti) e di pazienti sottoposti a commissurotomia, evidenzia che l'emisfero non linguistico svolge un ruolo cruciale nell'integrazione dei processi verbali. Così la normale produzione del parlato coinvolge entrambi gli emisferi che forniscono contributi specifici. Lo studioso suggerisce che la convergenza di contributi provenienti dai due emisferi avviene nell'emisfero sinistro, mentre il destro svolge comunque un ruolo cruciale nel processo linguistico. Il risultato è quindi, per McNeill, un'unità costruita sia con aspetti "immagistici" - che fanno riferimenti all'aspetto gestuale - che con aspetti lineari-segmentali che fanno riferimento agli aspetti linguistici del parlato. L'origine di questa unità è identificata dallo studioso in quello che chiama *growth-point* (McNeill 1992).

Appaiono interessanti i risultati raggiunti da Hedda Lausberg e dal suo gruppo di ricerca nel grande progetto NEUROGES. Le loro ricerche si basano su due evidenze derivanti dagli studi sui pazienti con traumi al *corpus callosum*: emerge che la loro capacità nel pieno uso di entrambi le mani nella gesticolazione, inoltre, il monitoraggio di un gruppo di persone - di cui alcuni con una completa disconnessione tra emisferi, altri con una parziale disconnessione e altri soggetti sani - non ha evidenziato differenze per il numero di gesti realizzati con la mano destra e con la mano sinistra in un minuto. Così è possibile dedurre che i gesti spontanei correlati al parlato possano essere generati nell'emisfero destro indipendentemente dal fatto che l'emisfero sinistro gestisce il linguaggio. A partire da queste evidenze, Lausberg sostiene che la scelta della mano per differenti tipi di gesti riflette

la lateralizzazione di diverse funzioni cognitive ed emozionali (Lausberg H. – Sloetjes H. 2015).

Oltre alla posizione di McNeill sulla natura cognitiva del gesto, altri studiosi ne hanno avanzate altre. Quelle che hanno maggiormente influenzato il dibattito sono quella di Kita (2000) e quella delle studiose Hostetter e Alibali (2008). Secondo la teoria dell'*Information Packaging* di Kita, il parlante ha accesso a una gamma molto ampia di possibilità di organizzazione del pensiero. Questo sarebbe determinato dal fatto che il pensiero è costituito da due modalità (il pensiero analitico e quello spazio-motorio) che organizzano in modo diverso le informazioni. Cioè, il pensiero analitico – sottostante al parlato – organizza le informazioni attraverso degli schemi concettuali decontestualizzati e strutturati gerarchicamente; mentre il pensiero spazio-motorio organizza l'informazione attraverso schemi di azione e sulla base delle caratteristiche dell'ambiente. Il pensiero spazio-motorio organizza anche l'ambiente virtuale che è creato internamente come immagine, i gesti rappresentazionali sono azioni nell'ambiente virtuale e così la gesticolazione è organizzata dal pensiero spazio-motorio. Per Kita, quindi, il linguaggio scaturisce dalla collaborazione dei due pensieri che sono indipendenti nell'organizzazione delle informazioni, ma che collaborano quando il fine ultimo è parlare (Kita 2000; Kita e Özyürek 2003).

L'ipotesi di Hostetter e Alibali (2008), *Gestures as Simulated Action*, si incentra sul concetto di "immagine mentale". L'immagine mentale è una rappresentazione analogica di un oggetto percettivo o di un evento motorio può essere contrapposta alle rappresentazioni verbali, che sono descrizioni in linguaggio naturale. Le autrici distinguono tra immagini mentali visive, che rappresentano gli oggetti, e le immagini mentali motorie che invece rappresentano azioni.

Le immagini mentali dipendono dalle stesse relazioni che intercorrono tra processi percettivi e motori coinvolti nell'interazione con gli oggetti fisici. Le immagini mentali visive condividono i medesimi meccanismi

neurali delle percezioni visive, mentre le immagini mentali motorie condividono i meccanismi deputati al movimento poiché la creazione dell'immagine mentale si basa sul meccanismo di simulazione dei neuroni specchio (Rizzolati e Sinigaglia, 2019). Situati nella corteccia premotoria, i neuroni specchio si attivano sia quando viene prodotta un'azione che quando viene osservata. L'attivazione dei neuroni produce un movimento solamente quando l'attività passa dalla corteccia premotoria alla corteccia motoria. Così Hostetter e Alibali propongono un meccanismo simile con il loro quadro teorico. La simulazione fa scaturire l'attivazione degli stati d'azione nella parte premotoria, che ha il potenziale per diffondersi nelle aree motorie e realizzare un'azione palese. Un gesto nasce quando l'attivazione neurale premotoria riesce a diffondersi all'area motoria (Hostetter - Alibali 2008).

Alcuni studiosi hanno ipotizzato che la principale funzione dei gesti fosse il recupero lessicale. Questa teoria viene chiamata Lexical Retrieval Hypothesis e i ricercatori, alla luce degli studi portati avanti, sostengono che i parlanti riescono ad avere un maggior successo nel recupero lessicale quando gesticolano e questo porterebbe a una maggiore occorrenza di gesti referenziali durante le disfluenze (Butterworth and Hadar, 1989; Morrel-Samuels and Krauss, 1992; Rauscher et al., 1996; Krauss, 1998; Morsella and Krauss, 2005). Uno studio più recente ha invece dimostrato una maggiore correlazione tra il discorso fluente e i gesti, contrariamente a quanto affermato dall'ipotesi del recupero lessicale, e soprattutto ha evidenziato che molto spesso i gesti che occorrono nelle disfluenze sono trattenuti, rafforzando di conseguenza l'ipotesi che gesto e parlato fanno parte di un unico sistema (Graziano e Gullberg 2018).

Una posizione nettamente contraria è stata assunta dall'*image maintenance theory* (Wesp et al. 2001), secondo cui i gesti non sono impiegati direttamente nella ricerca lessicale dei parlanti, ma piuttosto aiutano l'attività della memoria attraverso il mantenimento del concetto non-lessicale. L'idea di base di questa teoria è che la rappresentazione

prelinguistica delle informazioni spaziali viene creata attraverso l'immaginazione spaziale e il mantenimento è facilitato e garantito dai gesti.

Infine, è interessante l'approccio *embedded/extended* alle funzioni cognitive dei gesti. Se l'approccio *embodied* (rappresentato dalla teoria GSA) ha un approccio maggiormente internalista alla cognizione, l'approccio *embedded/extended* vede l'attività cognitiva come una continua interazione tra cervello, corpo e mondo circostante. La visione del ruolo svolto dall'ambiente circostante nel processo di cognizione determina due visioni, una più radicale (quella *extended*) che vede il mondo circostante come parte della cognizione, e una visione meno radicale (quella *embedded*) che invece sostiene che il corpo e l'ambiente abbiano un inaspettato impatto nell'attività cognitiva ma non fanno parte del processo (Pouw et al. 2014). In questa visione si inserisce Clark (2013) il quale sostiene che l'atto fisico del gesticolare sia parte di una duplice azione in divenire, corporea e neurale, concependola come un processo esteso del pensiero. Cioè i gesti danno una forma fisica e stabile al processo cognitivo. Quindi i gesti non sono solamente un modo per esternalizzare il pensiero e il parlato ma permettono anche una temporanea stabilità cognitiva.

Alla luce degli studi che abbiamo appena posto in rassegna, risulta evidente l'importanza del ruolo svolto dai gesti nella cognizione dei parlanti. Come abbiamo visto, i gesti possono aiutare a trovare le parole, a organizzare il pensiero spazio-motorio o a rispecchiare ed esternare l'immagine mentale. Ma d'altro canto è necessario ricordare che i gesti svolgono anche un importante ruolo nell'interazione tra parlanti. Cercare di indagare e comprendere questo aspetto, ci porta a intraprendere un'analisi che deve necessariamente sdoppiarsi nel tentativo di capire sia la produzione dei gesti da parte di un parlante che la ricezione e interpretazione del destinatario. Su questi argomenti verranno approfonditi nella seconda parte della tesi, mentre nel paragrafo successivo vedremo alcune funzioni comunicative e pragmatiche dei gesti.

3.2 Il gesto tra funzioni comunicative e pragmatiche

I gesti non sono esclusivamente una rappresentazione fisica delle immagini mentali di un individuo, spesso subiscono variazioni nella forma e nelle funzioni che dipendono dal contesto in cui si svolge l'interazione (Campisi 2018). Alcuni studi, infatti, dimostrano che non sono solamente i parlanti a beneficiare della presenza dei gesti, ma anche i destinatari. Kelly e Barr (1999) hanno dimostrato che gli enunciati con una richiesta indiretta risultano più facilmente interpretabili se accompagnati da gesti rispetto a quando l'enunciato viene espresso esclusivamente attraverso il parlato. Cocks e colleghi hanno mostrato che questi benefici sono riscontrabili anche negli afasici. Sembrerebbe, infatti, che anche quando la comprensione delle parole è gravemente compromessa la presenza dei gesti riesca a facilitare gli afasici a capire il senso di un enunciato (Cocks et al. 2009).

La produzione dei gesti, in quanto atto comunicativo e interattivo, subisce modifiche in relazione al contesto in cui viene realizzata. Infatti, la gesticolazione sembra calibrarsi in relazione alla posizione e al numero dei parlanti. In uno studio di Özyürek (2002) si evince che i parlanti, attraverso l'uso dei gesti, cercano di creare uno spazio condiviso tra i partecipanti all'interazione e, nonostante il racconto verbalizzato rimanga pressoché uguale, modificano i gesti in relazione al numero dei presenti. Inoltre, uno studio di Bavelas e colleghi (2008) dimostra che anche la visibilità può interagire con la gesticolazione. Gli studiosi, indagando diverse situazioni che vanno dall'interazione faccia a faccia a quella telefonica, hanno osservato che nella condizione di visibilità la produzione dei gesti cresce sia nel numero di occorrenze che nelle caratteristiche dimensionali, diminuendo nella ridondanza in relazione alle informazioni espresse per mezzo del canale verbale.

L'analisi e lo studio dei gesti non può, dunque, prescindere dalla considerazione del contesto comunicativo in cui vengono realizzati. Se il contesto d'uso può senza dubbio rappresentare un ausilio per

l'interpretazione semantica di un gesto, bisogna di conseguenza cercare di capire come i gesti possono esprimere valori prettamente pragmatici e come si relazionano con le unità pragmatiche proprie del canale verbale.

In passato, il gesto veniva considerato una questione preminentemente di ambito pragmatico in quanto non veniva visto come elemento linguistico, questa visione tradizionale è nata dall'influenza della linguistica generativa (Cienki 2017). Come abbiamo già visto nei paragrafi precedenti, negli ultimi anni diversi studi hanno dimostrato che i gesti svolgono funzioni importanti nella comunicazione verbale e hanno sottolineato che i gesti giocano un ruolo cruciale sia nella parte cognitiva (McNeill 2005) che nella pragmatica del discorso (Kendon 2004).

Nell'analizzare alcuni gesti italiani, Kendon ha evidenziato alcune funzioni pragmatiche espresse da quattro forme gestuali dell'uso comune. La sua analisi ha portato a evidenziare che alcuni gesti marcano la forza illocutoria di un enunciato così da essere etichettati come *illocutionary marker gestures*, e gesti che hanno la funzione di indicare lo status dell'unità all'interno di un discorso così da parlare di *discourse unit marker gestures*. Lo studioso conclude affermando che:

«speakers may use gestures which can explicitly mark a given stretch of speech as being a particular type of speech act. Within a discourse, they can differentiate gesturally topic from comment, or indicate what units are 'focal' for their arguments»

definendo questi gesti pragmatici¹⁵ (Kendon 1995). Bressemer e Müller hanno individuato una lista di *gesti ricorrenti* in tedesco che esprimono funzioni pragmatiche (Bressemer e Müller 2014). Per *gesti ricorrenti* le autrici indentificano un tipo di co-speech gestures che mostra similarità sia da un punto di vista funzionale che strutturale con gli emblemi. A differenza di

¹⁵ Segnaliamo che Kendon – nel suo lavoro del 2004, *Gesture: visible action as utterance* – considera insoddisfacente il termine pragmatico. Inoltre, ciò che nel lavoro del 1995 è chiamato *illocutionary marker gesture* e *discourse structure marker gesture*, nel 2004 verrà definito rispettivamente definito come *performative gesture* e *parsing gesture*.

quest'ultimi però, i *recurrent gestures* non possono sostituire il parlato piuttosto collaborano con il canale verbale per la formazione del significato dell'enunciato multimodale. Vengono definiti *ricorrenti* poiché vengono usati abitualmente in diversi contesti e da diversi parlanti mantenendo le medesime caratteristiche semantiche e formazionali. Il metodo di analisi utilizzato dalle ricercatrici si basa sullo studio della forma assunta dai gesti per la ricostruzione sistematica del loro significato ed è stato applicato a un corpus costituito da 24 ore complessive di parlato da cui è stato estratto un repertorio di sedici tipi di *gesti ricorrenti*. Lo studio si è basato sulla classificazione searliana dell'illocuzione. Quello che è stato osservato è che tutti e sedici tipi di gesto esprimono un valore assertivo e che molto frequenti sono i *gesti ricorrenti* che assumo valore direttivo o espressivo, sono invece piuttosto rari i gesti che esprimono un atto commissivo. I *recurrent gestures* sembrano incorporare la forza illocutiva espressa o l'azione comunicativa che spesso rimane implicita a livello verbale, infatti, spesso marcano vari aspetti del discorso o della struttura del discorso. I tipi di *gesti ricorrenti* che riescono a incorporare il valore illocutivo sono i PDOH (Palm Down Open Hand) che vengono usati per marcare eventi, stati o per esprimere l'idea di incertezza e indeterminatezza, e poi ci sono i *shaking off* che vengono usati per marcare un oggetto o una situazione come potenzialmente pericolosa.

Enfield e colleghi (Enfield, Kita e De Ruiter 2007) - studiando la popolazione del Laos - hanno svolto uno studio sui gesti deittici, chiamati anche *pointing gestures*. Gli studiosi, approcciandosi all'analisi dei gesti concependoli come strutturalmente legati con il parlato formando un unico *composite utterance*, hanno distinto due tipi di *pointing gesture* sulla base della loro forma: *B-point* (grandi nella forma) e *S-point* (piccoli nella forma). Quello che hanno notato è l'esistenza di una distinzione della funzione espressa in relazione alla grandezza della loro forma assunta, infatti, il primo tipo di *pointing gesture* gioca un ruolo necessario all'interno dell'enunciato multimodale portando informazioni primarie, mentre il

parlato sembra svolgere solamente un ruolo di supporto al *B-point*. Invece, i gesti *S-point* mostrano una maggiore dipendenza nei confronti del parlato e sono di conseguenza più nascosti nella struttura informativa dell'enunciato; esprimono informazioni secondarie, di sfondo, che rispondono a una possibile ma incerta mancanza di conoscenze comuni tra gli interlocutori.

«While a B-Point is doing the primary work of the utterance, with speech playing a supporting role, an S-point adds a backgrounded modifier to an utterance in which speech is central» (Enfield, Kita e De Ruiter 2007).

Un *S-point gesture* rappresenta di fatto un'azione comunicativa a basso rischio, che potrebbe salvare il parlante da un costo sociale e interpersonale potenzialmente elevato (Enfield 2006). Queste analisi dimostrano come il gesto sia fortemente integrato con il parlato costituendo insieme un'unica azione comunicativa. Questa stretta collaborazione tra le due modalità è tale che la formazione dei gesti – nelle caratteristiche dimensionali e tipologiche – risulta influenzata dalle esigenze comunicative di tipo semantico, informativo e pragmatico.

Tutti questi lavori, nonostante i diversi approcci e punti di vista teorici adottati, possono senza dubbio contribuire a estendere l'idea di azione linguistica come un'azione multimodale. Perché i gesti svolgono un importante ruolo pragmatico (esprimendo diverse funzioni), è importante riconoscere che la parte gestuale non è una caratteristica di appendice dell'enunciato, ma i gesti realizzano – insieme al parlato – l'azione linguistica.

Infatti, come abbiamo visto, Kendon definisce il gesto come «a name for visible action when it is used as an utterance or as a part of utterance» (Kendon 2004) e vede l'enunciato come

«any unit of activity that is treated by those co-present as communicative 'move', 'turn', or contribution. Such units of activity may be constructed from

speech or from visible bodily action or from combinations of these two modalities» (Kendon 2004)

Bressemer e Müller hanno basato il loro studio sull'enunciato multimodale in senso kendoniano. Mentre Enfield parla di *composite utterance* definendolo «as a communicative move that incorporates multiple signs of multiple types» (Enfield 2009). L'enunciato composito ha un significato codificato – che consiste di valori lessicali e grammaticali (a esempio un segno linguistico convenzionale) – e un significato arricchito che può essere indessicale se spiega i riferimenti non trasparenti dell'enunciato – questo può essere realizzato sia esplicitamente (da un simbolo indessicale come "questo") che implicitamente (dalla copresenza nel tempo e nello spazio, per esempio un avviso per il divieto di fumare) – o implicazionale secondo il modello griceano – così il significato è raggiunto sia attraverso un sistema di codici che per mezzo di un'interpretazione basata su un terreno comune di conoscenze (Enfield 2009).

L'idea di un enunciato multimodale sembra essere un concetto teorico, basato su prove empiriche, ma difficile da trasformare in un'unità utile all'analisi linguistica. Non c'è una definizione basata su caratteristiche fisiche o sull'enunciato parlato. Se da un lato Kendon non ha definito l'enunciato multimodale in modo pratico, dall'altro Enfield ha fatto riferimento al *composite utterance* partendo dell'unità di base dell'interazione sociale, definendola *mossa* secondo la teoria di Goffman che dice:

«a move may be defined as a recognizable unit contribution of communicative behavior constituting a single, complete pushing forward of an interactional sequence by means of making in some relevant social action recognizable (e.g. requesting the salt, passing it, saying thanks) » (Enfield 2009)

Concepire l'azione linguistica multimodale fino a concepire l'enunciato come unità multimodale, porta con sé una serie di questioni relative all'idea di lingua tradizionalmente condivisa. Inoltre, quando si parla di

multimodalità bisogna tenere presente che si parla di un campo di studi multidisciplinare ampio che, tra punti di vista differenti e oggetti di studio diversi, contribuisce in modo sempre più proficuo allo studio della comunicazione umana. Tutte queste questioni, che riportiamo in chiusura di questo capitolo, saranno approfondite nel capitolo successivo dedicato alla multimodalità.

Capitolo 3 – La lingua multimodale

1. Introduzione

I parlanti utilizzano diversi canali e diversi metodi per creare e trasmettere i significati. Da un lato il parlato che per mezzo del lessico, della sintassi e della prosodia riesce a esprimere sfumature di significati e azioni linguistiche molto precise e dettagliate. Dall'altro lato la gestualità che può giocare un ruolo complementare al parlato o in totale autonomia. Nella sua complessità, anche il gesto esprime funzioni e significati diversi integrandosi pienamente nel sistema verbale, nonostante l'autonomia comunicativa che può assumere per mezzo della convenzionalizzazione. La gestualità non è un sistema comunicativo separato dal parlato che funge da ausilio, ma è un elemento intrinseco del sistema verbale che svolge diverse funzioni, da quelle cognitive a quelle pragmatiche. Pertanto, si può pensare alla lingua come sistema intrinsecamente multimodale dove ogni espediente semiotico interagisce con l'altro nella realizzazione dell'azione linguistica.

Il concetto di multimodalità è legato a un nuovo e ampio campo di studi sulla comunicazione. Come vedremo in questo capitolo, la comunicazione multimodale può essere definita partendo da diversi punti di vista: sia in relazione alla comunicazione verbale che alla comunicazione scritta. Ma il presente lavoro mira alla creazione di un corpus che raccolga il linguaggio parlato spontaneo quindi, in questo capitolo, cercheremo di definire dapprima il campo degli studi multimodali, per poi mostrare come è possibile approcciarsi al linguaggio come sistema multimodale. Affronteremo quindi la questione semiologica legata alla nozione dinamica di linguaggio per cercare alla fine di proporre una soluzione basata sulla

linguistica cognitiva e sulla pragmatica che rappresenterà il punto di partenza del nostro approccio.

2. La nozione di multimodalità

Il linguaggio è sempre stato visto, a partire dal prezioso contributo di Ferdinand De Saussure, come un sistema binario e coerente. Questa visione ha sostenuto l'idea che un'interazione tra parlanti sia semplicemente un fatto di codifica e decodifica di un messaggio (Aristotele; Locke). Ma, con gli studi di Grice (1967), Austin (1962) e Searl (1969), è stato possibile riconoscere che il linguaggio è un sistema più complesso rispetto a un semplice sistema di codifica. La complessità del sistema lingua si evince dal fatto che quando parliamo, creiamo i significati utilizzando diverse fonti come la voce, il contesto, il gesto, l'espressione facciale, le interazioni sociali e le caratteristiche culturali. Così, negli ultimi anni, abbiamo assistito a un netto aumento degli studi che vedono la lingua come un sistema multimodale, che si diramano in diversi interessi al punto da diventare un campo di studi multidisciplinare.

Il termine "multimodalità", coniato a metà degli anni '90, oggi è ampiamente usato da diversi studiosi che ad oggi non sono riusciti a costruire una definizione condivisa e unica dell'oggetto di studio della multimodalità. In modo generico è possibile affermare che l'oggetto di interesse è l'uso di diverse modalità utilizzate per la creazione del significato.

Il termine è stato utilizzato per la prima volta in due diversi studi da diversi accademici. Da Charles Goodwin in 1998 (Goodwin 2000) negli Stati Uniti, e da Gunther Kress e Theo van Leeuwen (2001) nel Regno Unito. Gli studiosi utilizzano il termine secondo tradizioni di studi diversi, da un lato Goodwin si inserisce all'interno degli studi di etnometodologia e della Conversation Analysis, dall'altro Kress e van Leeuwen si inseriscono nella tradizione degli studi sociosemiotici.

Dunque, se fino alla fine degli anni Novanta del Novecento l'interesse per i diversi mezzi messi in opera per la produzione di significati è rimasto diviso in diversi settori e discipline (separando di fatto lo studio dello scritto dal parlato, dell'immagine dallo scritto, del gesto alla voce), con l'inizio delle indagini multimodali si è cerca di comprendere come i diversi mezzi vengono utilizzati insieme dai parlanti per creare il significato. Approcciarsi alla Multimodalità come campo di studi vuol dire «recognition of the need for studying how different kinds of meaning making are combined into and *integrated, multimodal whole* that scholars attempted to highlight when they started using the term 'multimodality'» (Jewitt et al. 2016). L'obiettivo è quello di sviluppare teorie e metodi per indagare il modo in cui usiamo gesti, parole, scritte, immagini e altri mezzi insieme per realizzare oggetti o eventi dotati di un significato complesso.

In questo panorama lo sviluppo di nuove tecnologie gioca un duplice ruolo, da un lato ha favorito nuovi metodi e spazi per la creazione del significato – l'uso del computer e dei social media – e dall'altro ha permesso approcci sempre più efficaci per l'analisi degli eventi di comunicazione (la creazione dei corpora e l'uso di videocamere e tecnologie di *motion capture*).

Jewitt e colleghi, nel loro lavoro del 2016, hanno formulato le tre premesse chiave per l'approccio alla multimodalità:

- il significato si crea con diverse risorse semiotiche, ciascuna delle quali offre potenzialità e limiti distinti;
- la creazione di significati implica la produzione di insiemi multimodali;
- se si vuole studiare il significato, è necessario prestare attenzione a tutte le risorse semiotiche che vengono utilizzate per creare un insieme completo.

Queste premesse rappresentano la base metodologica largamente condivisa dai diversi approcci allo studio della multimodalità, ma al

momento rappresenta solamente una piccola nozione teorica condivisa che deve fare i conti con una larga schiera di approcci diversi. Per riuscire a comprendere a pieno le sfaccettature dei diversi approcci allo studio della multimodalità, è necessario fare alcune precisazioni. Fino a questo momento abbiamo utilizzato la locuzione *meaning making* (creazione di significato), ma non tutti coloro i quali si approcciano agli studi multimodali usano questa locuzione, dipende dal quadro teorico di riferimento, e non tutti usano il termine *mode*, qualcuno preferisce parlare di *resource*. Inoltre, l'interesse per la relazione tra i diversi mezzi e/o modi usati dai parlanti, per la creazione del significato, è antecedente rispetto alla formulazione della nozione di multimodalità. Gli studi di Barthes (1964) sulla relazione tra immagine e scrittura e di Kendon (2004) sulla relazione tra gesto e parlato.

Un'altra questione riguarda i diversi approcci teorici con cui i diversi studiosi conducono le proprie indagini sulla multimodalità. La mancanza di una univoca definizione di multimodalità e di un framework teorico associato è legata alle diverse necessità epistemologiche in riferimento alle diverse modalità studiate. Così è possibile affermare che la nozione di multimodalità si sta sviluppando in questi ultimi anni, ma ciò che rimane a fondamento di tutti i diversi approcci sono le tre premesse chiave riportate sopra.

Come abbiamo già affermato, nel campo della multimodalità ci sono diversi approcci con diversi metodi e diverse teorie del significato. Le tre principali discipline che si sono avvicinate alla multimodalità sono: la Linguistica Funzionale Sistemica, la Semiotica Sociale e l'Analisi Conversazionale. La Linguistica Funzionale Sistemica, sviluppata da Michael Halliday (1994), mira a comprendere come le diverse funzioni sociali dei parlanti influenzino il sistema lingua nella sua espressione e vede il linguaggio come una risorsa semiotica a natura sociale. Le risorse semiotiche sociali, centrali per gli studi multidimodali nella Linguistica Funzionale Sistemica, rappresentano quattro funzioni sociali: costruire la

nostra esperienza del mondo, fare collegamenti logici nel mondo esperito, realizzare legami sociali e organizzare il messaggio. L'oggetto di studi della LFS è la relazione tra testo e immagine nei libri di testo scolastici.

L'approccio della Semiotica Sociale, iniziato da Gunther Kress e Bob Hodge (1988) sulla base della linguistica critica, della semiotica e della teoria sociale, mira a evidenziare l'organizzazione degli attori sociali e la relazione tra le relazioni sociali. La nozione di segno motivato, così come è suggerito da Saussure, è centrale in questo approccio. Il 'modo' è definito «a set of socially and culturally shaped of resource for making meaning that has distinct 'affordances'» (Kress 2014). L'interesse dell'approccio della Semiotica Sociale è rendere conto dell'interazione sociale registrata in video attraverso le indagini sul campo.

L'approccio della Conversation Analysis, sviluppato negli Stati Uniti da Schegloff, Sacks e Jefferson (1974), negli ultimi anni ha rivolto il suo interesse all'interazione multimodale. La nozione di significato si basa sulla sequenzialità, infatti l'interazione è vista come un'azione che si svolge nel tempo.

Il 'significato' è quindi un'azione, e può essere compreso in relazione all'azione che l'ha preceduta e seguita. I principali oggetti di indagine di questo approccio riguardano l'uso del corpo, degli oggetti e dell'ambiente costruito come mezzo per realizzare l'interazione.

Alla luce dei principali approcci allo studio della multimodalità appena descritti, crediamo sia necessario specificare quale sia l'aspetto della multimodalità a cui è interessato il presente lavoro. Il nostro interesse si rivolge all'italiano parlato spontaneo, concependo la lingua come un'azione che si realizza attraverso diversi mezzi: prosodia, parola, gesto, espressioni facciali e postura. E ci basiamo sui risultati e i riferimenti teorici degli studi sul gesto di Kendon, McNeill, Müller ed Enfield.

Questi lavori possono contribuire a estendere l'idea di azione linguistica come un'azione multimodale. Infatti, Kendon definisce il gesto come «a name for visible action when it is used as an utterance or as a part of an

utterance» (Kendon 2004). Inoltre, Kendon vede l'enunciato come «any unit of activity that is treated by those co-present as a communicative 'move', 'turn' or contribution. Such units of activity may be constructed from speech or from visible bodily action or from combinations of these two modalities» (*ibidem*). Invece Enfield parla di *composite utterance* definendolo «as a communicative move that incorporates multiple signs of multiple types» (Enfield 2013). Il *composite utterance* ha un significato codificato - che consiste in valori lessicali e grammaticali (es. segno linguistico convenzionale) - e un significato arricchito che può essere indessicale se spiega i riferimenti dell'enunciato non chiari - che può essere realizzato sia esplicitamente (attraverso un simbolo indessicale come "questo") che implicitamente (dalla copresenza nel tempo e nello spazio, come un cartello per il divieto di fumo) - o implicazionale secondo il modello griceano - così il significato è raggiunto sia attraverso un sistema di codici che da un'interpretazione basata su conoscenze comuni e contesti condivisi (Enfield 2013).

Così l'idea di un enunciato multimodale sembra essere un concetto teorico, basato su prove empiriche, ma che non può essere considerato come unità di riferimento per l'analisi linguistica. Infatti, non esiste alcuna definizione basata su caratteristiche pratiche oltre che sull'enunciato parlato. Se da un lato Kendon non ha definito quali elementi determinano e delineano l'enunciato multimodale, Enfield si riferisce al *composite utterance* come all'unità di base dell'interazione sociale, che ha chiamato mossa secondo la teoria di Groffman:

«a move may be defined as a recognizable unit contribution of communicative behaviour constituting a single, complete pushing forward of an interactional sequence by means of making in some relevant social action recognizable (e.g., request the salt, passing it, saying thanks)» (Enfield 2013)

Considerando questo quadro teorico, con questo lavoro intendiamo disegnare un approccio pragmatico al fine di costruire un corpus

multimodale che possa garantire ricerche sull'interazione parlata nei suoi aspetti multimodali. Ma concepire il sistema linguistico come sistema multimodale implica affrontare l'importante questione riguardante la nozione di lingua stessa, e precisamente rispetto alla sua natura semiologica e al ruolo svolto dalla gesticolazione.

3. Linguaggio e multimodalità

Abbiamo visto come il carattere multimodale sia riscontrabile in diversi aspetti della comunicazione umana che si esplica nella capacità di creare significati attraverso l'uso di diversi canali semiotici, come immagine-scrittura, immagine-suono e suono-gesto. L'interesse del nostro lavoro, come già dichiarato in precedenza, è proprio la capacità umana di sincronizzare gesto e parlato negli eventi comunicativi. Ciò vuol dire che in queste pagine la lingua viene vista come sistema multimodale e di conseguenza il gesto, inteso in questo caso come qualsiasi movimento espressivo del corpo umano, assume un valore linguistico e verbale. Questa nostra assunzione deriva dalle evidenze testimoniate dagli studiosi e dalle studiose degli *gesture studies* (di cui abbiamo in parte trattato nel capitolo 2 e di cui approfondiremo nel capitolo 4). Per diversi anni, diverse discipline si sono approcciate al gesto come elemento non verbale, legando il valore di linguisticità alla verbalità come se la caratteristica dell'oralità fosse elemento costitutivo di un sistema linguistico. In questo paragrafo cercheremo di introdurre il dibattito sul valore linguistico e verbale del gesto, portando ulteriori prove alla nozione di lingua come sistema multimodale.

Kendon è stato il pioniere della ricerca sul gesto come elemento comunicativo integrato al sistema lingua, e nei primi anni Settanta ha individuato i modelli di comportamento corporeo interattivo; i suoi studi hanno sostenuto che le azioni corporee sono altamente strutturate, legate al parlato e significative. Ma negli anni Settanta l'influenza della grammatica

generativa ha portato la ricerca linguistica a perdere l'interesse per il gesto come elemento linguistico. Così è stata la psicologia ad alimentare l'interesse per la gesticolazione, concependola come elemento della comunicazione non verbale. Sarà negli anni Ottanta che David McNeill aprirà il dibattito intorno alla natura del gesto sfidando la convinzione della psicolinguistica secondo la quale i gesti farebbero parte della dimensione non verbale della comunicazione. In particolare, McNeill ha portato cinque prove a difesa della sua posizione: «a) gestures occur only during speech; b) they have semantics and pragmatics functions that parallel those of speech; c) they are synchronized with linguistic units in speech; d) they dissolve together with speech in aphasia; e) they develop together with speech in children» (McNeill 1985). Secondo McNeill il gesto e il discorso sono due facce di una stessa medaglia e nascono da un'unica azione cognitiva che chiama *Growth Point*.

Il *Growth Point* «is a package that has both linguistic categorial and imagistic components, and it has these components irreducibly» (McNeill 2005). Questa unità complessa rappresenta un modello per spiegare ciò che McNeill chiama dialettica immaginario-linguistica. Infatti, da una parte c'è il linguaggio visto come dimensione statica secondo il quadro teorico saussuriano (una dimensione segnica), dall'altra c'è l'attività del linguaggio che è la dimensione dinamica, «on this dimension, utterances come and go, emerge and disperse in real time, and this stands in contrast to the abstraction from time of linguistic objects in *langue*» (McNeill, 2013). In questo modo, il *Growth Point* può spiegare la sincronia gesto-parlato, la loro espressività e il fatto che parlato e gesto incarnino la stessa idea ma in due modalità semiotiche diverse.

La posizione di McNeill si affianca agli studi di Adam Kendon (anche se gli obiettivi e le premesse sono diversi). Infatti, Kendon assume un approccio più descrittivo nutrendo interessi di tipo linguistico e comunicativo. Kendon afferma che il gesto «is the name for visible action when it is used as an utterance or as a part of an utterance» (Kendon 2004).

La nozione di enunciato a cui fa riferimento Kendon si basa sull'idea di Goffman sull'interazione sociale tra persone che stanno nello stesso luogo (Goffman 1963). Così, l'enunciato può essere definito come qualsiasi attività che viene interpretata dai compresenti come un contributo comunicativo. Molte di queste unità enunciative possono essere formate da unità di parola, da azioni corporee visibili o dalle combinazioni di queste due diverse modalità, quindi «'gesture' is the visible bodily action that as a role in such units of actions» (*ibidem*). Le due modalità lavorano insieme per raggiungere l'intenzione comunicativa di un oratore, «in creating an utterance that uses both modes of expression, the speaker creates an *ensemble* in which gesture and speech are employed together as *partners* in a single rhetorical enterprise» (*ibidem*). La relazione tra le componenti gestuali e parlate non è dunque casuale, ma i parlanti possono controllarle e organizzarle secondo l'occasione e l'intenzione comunicativa. Quindi è possibile affermare che anche Kendon vede il linguaggio come un sistema multimodale, lo concepisce come un'azione composta da elementi che trasmettono con diverse modalità di significazione.

Tra coloro che si sono interessati alla questione del linguaggio multimodale c'è Enfield che, come abbiamo già visto, ha sostenuto una definizione di *composite utterance* disegnando l'anatomia del significato. Egli considera il significato come una nozione composita, infatti, il significato dell'enunciato è «unified product of multiple sources of information» (Enfield 2009). Il *composite utterance* può essere formato da tre tipi di segni: segni convenzionali, segni non convenzionali e segni simbolici indessicali. Un segno convenzionale è caratterizzato da una relazione arbitraria e stretta tra il significante e il significato che è riconosciuto da tutti i membri di una comunità. La categoria dei segni convenzionali include parole, modi di dire e gesti emblematici. Un segno non convenzionale è caratterizzato invece da una relazione tra significante e significato che non è determinata di un evento interpretativo, per esempio i gesti rappresentativi delle mani. Il segno simbolico indessicale è un ibrido dei due tipi appena descritti. «These

include anything comes under the rubric of deixis, that is, form-meaning mappings whose proper interpretation depends partly on convention and partly on context» (*ibidem*), per esempio l'uso di un pronome "lui" in un contesto in cui il referente deve essere indicato per chiarire il referente del pronome. L'interpretazione del significato di un *composite utterance* dipende da un contesto e da un'altra differenza: «we must carefully to distinguish between token meaning (enriched, context-situated), type meaning (raw, context-independent, pre-packaged), and sheer form (no necessary meaning at all outside of a particular context in which it is taken to have meaning). These distinctions may apply to signs in any modality» (*ibidem*). Così, Enfield definisce il *composite utterance* come una mossa comunicativa che include più segni di diverso tipo e che deve essere interpretato attraverso il riconoscimento dell'insieme di questi diversi segni come unità pragmatica.

La capacità di comunicare in modo multimodale è una specificità che caratterizza la comunicazione umana in tutte le sue espressioni, anche nella lingua dei segni. Secondo un'interessante ricerca – condotta da Fontana (2009) – è possibile riconoscere nelle lingue dei segni aspetti appartenenti all'oralità che svolgono funzioni analoghe a quelle svolte dal gesto nella lingua parlata. La ricercatrice, nello specifico, ha studiato le funzioni dei tratti orali nella lingua dei segni, in particolare le labializzazioni¹⁶ e i gesti labiali¹⁷ nella LIS (Lingua Italiana dei Segni). Questi due fenomeni sono considerati da Fontana come una forma gestuale all'interno del sistema della lingua dei segni. Il gesto nella lingua dei segni è legato sia alla lingua dei segni stessa che alla lingua parlata, in quanto la comunità dei segnanti è perennemente esposta al contatto con la comunità dei parlanti. In questo modo l'oralità gioca un ruolo importante nella significazione della lingua dei segni – come il gesto nel parlato – e di conseguenza la struttura del gesto

¹⁶ «[...] le *labializzazioni* (mouthings) consistono nella riproduzioni di elementi fonetici rilevanti di una data parola della lingua orale, con o senza emissione di voce» (Fontana, 2009: 209).

¹⁷ «[...] i *gesti labiali* (mouth gestures) sono componenti orali che non hanno alcun rapporto con il parlato» (*ibidem*).

orale¹⁸ nella lingua dei segni risulta definita da due sistemi diversi: il sistema della lingua orale e il sistema concettuale basato sull'esperienza e sulle categorizzazioni prelinguistiche. Ma non è possibile affermare lo stesso per il gesto nella lingua parlata che non è influenzato dalla lingua dei segni. Entrambe le caratteristiche orali – la *labializzazione* e i gesti labiali – sono caratterizzate dalla co-espressività, ma differiscono per le loro funzioni. Infatti, i *gesti labiali* esprimono la dimensione fonosillabica legata alla concettualizzazione prelinguistica che è la stessa per le persone sorde e udenti. Mentre le *labializzazioni* svolgono funzioni a livello sintagmatico – attraverso il mantenimento della coerenza semantica dell'enunciato – e a livello paradigmatico – attraverso la distinzione di due omonimi, supportando i neologismi. Così, Fontana ha dimostrato che i gesti manuali co-verbali e i gesti orali co-segnici sono due forme diverse assunte dal gesto nella comunicazione umana, sia essa orale o segnica.

La sua ipotesi si basa sull'idea che il gesto è principalmente un'azione del pensiero che supporta la costruzione dell'espressione linguistica sia nel parlante che nel segnante (Fontana 2009):

«[...] i gesti co-verbali sembrano veicolare prevalentemente aspetti relativi all'*imagery* come rivela più volte McNeill, i gesti co-segnici sembrano privilegiare da una parte, quella *imagery* fonosimbolica del tutto assente nella materia gestuale delle lingue dei segni, dall'altra sfruttano elementi della lingua orale che interfacciano il livello segnico con obiettivi di ridondanza, coesione e stabilizzazione del messaggio» (*ibidem*)

Dunque, possiamo affermare che la comunicazione umana è multimodale nella sua natura e nel suo realizzarsi. Così diventa necessario – oltre che evidente – l'esigenza di considerare il gesto come elemento linguistico oltre che verbale. Infatti, come abbiamo visto finora, il gesto e il parlato sono due aspetti fortemente connessi e che insieme costruiscono

¹⁸ Con l'espressione *gesto orale* si fa riferimento agli elementi di oralità co-segnica presenti nella lingua dei segni (labializzazioni e gesti labiali) che svolgono funzioni analoghe a quelle svolte dai gesti manuali nella lingua parlata.

l'intero sistema del linguaggio. Considerare il linguaggio come multimodale implica dunque un ripensamento della nozione di lingua. Come descritto da Kendon, McNeill; Enfield e Fontana il linguaggio è un sistema dinamico che si sviluppa attraverso diverse modalità ciò si scontra con la tradizionale nozione di lingua.

4. I problemi semiotici della multimodalità

La nozione di linguaggio come sistema multimodale ci porta a riconoscere che la natura del gesto è verbale (inteso come elemento della comunicazione orale) e che è un canale che esprime diversi valori linguistici. Prima che l'interesse di ricerca di molti studiosi si indirizzasse al gesto come elemento verbale e linguistico, la nozione di linguaggio si basava esclusivamente su una visione statica degli eventi linguistici. L'interesse degli studiosi di linguistica, negli anni passati, si è prevalentemente rivolto allo studio della lingua nelle sue manifestazioni scritte e/o orali escludendo tutti gli eventi legati all'uso della lingua che sono stati in parte inglobati dalla pragmatica.

Nello specifico possiamo notare come Hockett (1960) individua tredici parametri per classificare i sistemi di comunicazione umana e animale:

1. il canale fono-articolatorio;
2. trasmissione e ricezione direzionale;
3. evanescenza del segnale;
4. la possibilità di riprodurre più volte un messaggio all'interno dello stesso sistema;
5. il totale controllo del messaggio prodotto da parte di chi lo realizza;
6. la specializzazione del segnale;
7. il valore semantico del segnale;
8. l'arbitrarietà;
9. la discretezza;

10. la possibilità di parlare di referenti lontani nel tempo e nello spazio;
11. la produttività;
12. la trasmissione del codice da parlante a parlante;
13. la doppia articolazione.

Nell'applicare questi parametri nella classificazione di alcuni sistemi di comunicazione umana e non-umana, lo studioso nota come tutte queste proprietà sono rintracciabili solamente nella lingua verbale mentre i linguaggi animali e dei gesti ne presentano solamente alcuni, aggiungendo che le lingue dei segni sono simili alle lingue verbali eccetto l'uso del canale fonico e la presenza dell'arbitrarietà. Mentre Martinet (1960) ha incentrato la sua definizione di lingua sulla nozione di doppia articolazione. Infatti, per lui, questa caratteristica è fondamentale perché sostiene l'economia linguistica: i parlanti usano un piccolo numero di unità per codificare infiniti significati riuscendo a creare un vero e proprio sistema.

La questione intorno alla definizione di lingua ha avuto grandi ripercussioni per le lingue dei segni che per troppo tempo non sono state considerate nella loro dignità di sistemi linguistici autonomi, portando anche a discriminazioni e ghettizzazione delle comunità dei sordi. Da un punto di vista meramente teorico, riteniamo che le questioni intorno alla linguisticità dei gesti e della lingua dei segni siano di fatto affini. Infatti, molti studi si sono occupati anche della contiguità tra segni e gesti.

Intorno alla relazione tra gesto e segno linguistico si è sviluppato negli anni un ricco dibattito che vede il contributo di due visioni opposte: chi vede il gesto e il segno legati e chi vede gesto e segno nettamente divisi. Il primo si basa su considerazioni storiche, funzionali e materiali e definisce dinamicamente questa relazione. Kendon, che è il principale sostenitore di questo punto di vista, usa il termine gesto come copertura per ogni forma kinesiologica di espressione, comprese le forme cinesiche convenzionali e non convenzionali. Egli considera la connessione storica tra gesto e segno come una lessicalizzazione dei gesti. Questo sviluppo storico coinvolge la

pantomima o i movimenti descrittivi, essi cambiano in una forma arbitraria riducendo la loro iconicità (Kendon 2004). Dal punto di vista funzionale Kendon sostiene che il gesto può essere usato come le parole anche in un enunciato vocale, questa interazione tra gesto e discorso può essere di due tipi: semantica o pragmatica. Mentre dal punto di vista materiale, emerge che sia il segno che il gesto utilizzano lo stesso mezzo cinesico. Inoltre, parlante e segnante usano lo spazio in modo simile e i classificatori della lingua dei segni hanno molte caratteristiche in comune con la tecnica di rappresentazione del significato del gesto (Kendon 2004).

L'altra posizione opposta all'interno del dibattito si basa sull'idea mcneilliana secondo cui il segno e il gesto sono nettamente divisi, in quanto considera impossibile qualsiasi processo di convenzionalizzazione. McNeill ha concentrato la sua ricerca sui gesti singolari – tralasciando i gesti emblematici – perché li concepisce come creazioni spontanee dei parlanti che riescono a mostrare il funzionamento della mente.

«Whith these kinds of gestures, people unwittingly display their inner thoughts and ways of understanding events in the world. These gestures are the person's memories and though rendered visible» (McNeill 1992)

McNeill, come abbiamo già detto, vede il pensiero del parlante come una medaglia composta da un lato immaginativo¹⁹, che nasce dai gesti, e il lato linguistico, che è la parte lineare-segmentata che emerge e si rende visibile nel parlato. Quindi, i gesti convenzionali rappresentano una categoria di elementi comunicativi più vicini al parlato, e dunque convenzionalizzato e controllato, perdendo di fatto la "spontaneità espressiva" e la capacità di aprire una finestra sulla mente del parlante. Per questo motivo, McNeill segna una netta differenza tra segno e gesto, mettendo in relazione questo contrasto con la capacità di rivelare la cognizione del parlante. La stessa posizione è stata espressa recentemente da Goldin-Meadow e Brentari

¹⁹ Il termine usato da McNeill è *imagistic* che abbiamo tradotto con immaginativo.

(2017), che sostengono la loro opinione attraverso un esperimento in cui hanno sottoposto i parlanti a due condizioni diverse: raccontare un evento prima con il parlato e poi senza. Gli autori hanno notato che nella condizione di soppressione del parlato i gesti erano più elaborati e più discreti, diventando più simili alla lingua e più segmentati:

«The gestures without speech immediately took on sign-like properties – they were discrete form, with gestures forming segmented word-like units that were concatenated into strings characterized by consistent (non-English) order» (Goldin-Meadow & Brentari 2017)

L'argomento principale portato, da Goldin-Meadow e Brentari, a sostegno della netta divisione tra gesto e segno linguistico, è la constatazione del fatto che quando un parlante si trova nella condizione di dover comunicare senza l'uso del parlato cambia la tipologia dei gesti usati, i quali sembrano diventare segni di una lingua al fine di soddisfare la necessità comunicativa.

La critica maggiore a questa posizione arriva da Cornelia Müller (2018), che ribatte sostenendo che per poter parlare di lingua che è emerse è necessario che ci sia «the understanding, the reflexivity and intersubjectivity of meaning shared within a moment of discourse or across a community of speakers/signers» (*ibidem*). Inoltre, Müller mostra che la loro nozione ristretta di gesto (che include solo il gesto spontaneo) ha un'importante implicazione per il diverso concetto di linguaggio e quindi anche per le opinioni sulla relazione tra gesto e segno.

Müller sostiene l'idea che ci sia una relazione dinamica tra gesti e segni. Vede il gesto come movimento significativo del corpo il cui significato è fondato sull'esperienza *embodied* che è dinamica e intersoggettiva. Inoltre, i gesti sono azioni deliberate e hanno diversi gradi di convenzionalizzazione. Con questa nozione di gesticolazione riesce a inglobare, come Kendon, tutti i tipi di gesto anche le forme convenzionalmente ibride, chiamate *recurrent gestures*, che si collocano tra i gesti singolari e quelli emblematici. In questo

modo è chiaro che sia il gesto convenzionale che quello non convenzionale possono essere impiegati dai parlanti allo stesso modo, sulla base di esigenze comunicative inerenti all'uso, e dunque tutte queste diverse forme di convenzionalizzazione dei gesti non possono essere descritte come nettamente separate. Infatti, la Müller ha sostenuto che da un punto di vista storico si possa parlare di trasformazione gestuale che lei definisce come una «historical dynamic of gesture and sign»:

«Gestural forms (through repeated usages) and in some case, undergo processes of lexicalization and grammaticalization and transform into signs within a signed language» (Müller, 2018)

In fine, confrontando la lingua dei segni e la gestualità nella lingua parlata sottolinea due sfaccettature che rivelano dei punti in comune: il mezzo di espressione condiviso e il contatto linguistico. Questi argomenti portano a sostenere l'idea che tra gesto e segno esiste una relazione dinamica.

Il dibattito intorno alla natura linguistica delle lingue dei segni ha portato gli studiosi a riflettere sulla nozione di lingua, infatti, ci sono diverse differenze tra la lingua parlata e la lingua dei segni: una riguarda la dimensione sintattica – la lingua parlata è lineare mentre la lingua dei segni usa quattro dimensioni (le tre dimensioni spaziali e la dimensione temporale) – e l'iconicità. Inoltre, anche la doppia articolazione rappresenta un problema rispetto al gesto, a causa della natura globale e non discreta (Kendon 2014).

Il punto di vista multimodale sul linguaggio implica che il sistema linguistico sia dinamico e semiologicamente eterogeneo. Infatti, la significatività di un atto linguistico multimodale nasce da una continua dialettica con il contesto in cui si realizza e dalla concomitanza di diverse modalità di significazione che non hanno lo stesso grado di convenzionalizzazione. Quindi il sistema lingua non può essere definito su parametri come la doppia articolazione e l'arbitrarietà, poiché si

escluderebbero i gesti da questo sistema riducendoli a elementi accessori del linguaggio. La lingua appare come un sistema composto da diversi tipi di modalità semiologiche che interagiscono tra loro per realizzare un'unità globale. Come abbiamo visto sopra, questa unità globale può essere chiamata *visible bodily action*, *composite utterance* o *growth point*. Vediamo dunque come alcuni studiosi di gesti vedono il sistema lingua.

McNeill (2005) delinea una nozione elaborata di linguaggio basata su un sistema dinamico e dialettico e che concilia la visione statica della lingua con una visione dinamica che guarda al linguaggio come a un processo e non come a un oggetto. Quindi abbiamo da un lato la dimensione dinamica che possiamo definire come il processo cognitivo che processa l'emergere dell'*imagery*, e dall'altro lato la dimensione statica che è costituita dalla forma linguistica che emerge come sistema di regole per la grammaticalità e di buona-formazione delle forme. Così, nella realizzazione linguistica si innesca una dialettica tra le due dimensioni in cui l'*intuizione* gioca un ruolo fondamentale nel bloccare il processo dialettico. McNeill definisce l'*intuizione* come esperienza individuale del sistema linguistico, la percezione diretta della forma linguistica. Così, le intuizioni giocano un ruolo importante nel parlare «*by stopping the dialectic processes: intuitions of well-formedness provide stopping points. There is a growing sense of completeness that functions as a kind stop order, as a dialectic moves forward to its state of repose*» (McNeill 2005).

In questa complessa visione del linguaggio, di natura eminentemente cognitiva, che ci restituisce McNeill sembrerebbe scorgersi la possibilità di superare la nozione tradizionale di linguaggio attraverso l'assunzione di un punto di vista dinamico che interseca il significato globale del gesto con il significato sintetico del linguaggio. Ma la nozione di *Growth Point*, punto di origine della dialettica, è un'ipotesi cognitiva che non trova riscontro empirico per una definizione di una unità di riferimento per gli eventi linguistici. Il concetto mcneilliano di linguaggio non ci permette di descrivere e definire il linguaggio come un sistema ben strutturato, pur

proponendo un'interessante ipotesi su come possa funzionare il processo cognitivo che genera la sincronizzazione delle diverse modalità usate dai parlanti.

Da questo punto di vista è interessante il contributo dato da Tullio De Mauro e dal gruppo di ricerca dell'Istituto di Psicologia del CNR di Roma guidato da Virginia Volterra (Pietrandrea 2012). La nozione di linguaggio elaborata da De Mauro ha rappresentato un importante punto di partenza per il dibattito. Egli ha sostenuto che non è il carattere dell'audioralità a definire le lingue umane; piuttosto esse si distinguono dagli altri codici comunicativi per la capacità di produrre nuovi significati e nuove parole. Questa possibilità è data dalla capacità, del sistema lingua, di riarticolare i propri segni sia a livello semantico che a livello del significante. L'indeterminatezza del segno permette la riformulazione della relazione tra significante e significato per mezzo del contesto d'uso; infatti, una delle caratteristiche della lingua elencate da De Mauro (2000) è il carattere locale del funzionamento dei segni e la necessità della realizzazione di intese tra i parlanti («la natura temporalmente e antropicamente circoscritta della funzionalità di ciascuna lingua, utilizzata in un tempo dato da singoli gruppi o sotto-gruppi umani» De Mauro, 2000). Questa flessibilità, quindi, si realizza nell'uso del linguaggio e rende possibile la rideterminatezza dei segni linguistici così da estendere il loro significato fino al raggiungimento dell'autoreferenzialità; ed è attraverso l'uso metalinguistico che i parlanti controllano e gestiscono la duttilità del segno nella negoziazione sociale. Così lo studioso afferma:

«Come ogni altra forma di comunicazione tra i viventi, non solo tra gli umani, una lingua è un codice semiologico, un insieme di strumenti che permette di regolare la interazione tra viventi mediante segni. Essa prevede, come altri codici, che i suoi segni siano articolati in parti le quali sono dotate di un loro significato. [...] Non tutti i codici prevedono segni scomponibili in "morfi" [...]. In particolare, ai nostri fini, è importante ricordare che non prevedono ciò anche alcuni linguaggi gesticolati e segnati, come il linguaggio cosiddetto

napoletano dei segni. Ricco, efficace, ma con segni non decomponibili in sottosegni *significativi*» (De Mauro, 2000)

Così, il linguaggio si caratterizza per la sua indeterminatezza determinata dall'arbitrarietà e dall'uso sociale. Ma come può l'arbitrarietà determinare l'indeterminatezza delle lingue dei segni che sono fortemente legate all'iconicità? Nonostante ciò, è possibile affermare che l'arbitrarietà è essenziale anche per le lingue dei segni. Cioè, l'iconicità dipende dall'arbitrarietà: da un lato un segno è legato alla realtà extralinguistica per mezzo dell'iconicità, dall'altro si forma sulla base dell'arbitrarietà che determina la relazione tra significato e significante (le cui unità articolatorie da utilizzare sono definite dal paradigma del sistema lingua). Inoltre, il segnante usa l'arbitrarietà per scegliere quale aspetto del referente deve essere usato per formare un nuovo segno; infatti, è impossibile prevedere la relazione tra un referente e il suo specifico segno (Pietrandrea 2000).

Così, De Mauro ha sostenuto che il linguaggio è uno strumento sociale che riafferma continuamente la relazione tra significante e significato nell'uso attraverso l'arbitrarietà e la negoziazione sociale. Ogni sistema linguistico utilizza la sua modalità, la lingua dei segni con il segno e la lingua orale con le parole, ma, come abbiamo visto sopra, ogni sistema linguistico è multimodale.

«Come nelle lingue segnate la digitazione alfabetica di taluni nomi e parole integra elementi di derivazione orale nella gestualità, questa accompagna e intride, ben più ampiamente, le realizzazioni orali. È questa una prova sincronica, attuale e permanente dell'unità profonda della capacità del linguaggio» (De Mauro 2000)

De Mauro offre una nozione dinamica di linguaggio in cui il gesto può vedere riconosciuto il suo ruolo e le sue funzioni. In particolare, lo studioso ne individua quattro: 1) appoggio extrafunzionale (inteso come ruolo di scansione e sottolineatura, «un ruolo di appoggio alla scansione sintattica e

alla determinazione del senso complessivo di un enunciato» (*ibidem*)); 2) integrazione alla semantica di singoli lessemi o gruppi sintagmatici di lessemi; 3) sostituzione semioticamente equivalente; 4) sostituzione semioticamente equivalente (linguaggi speciali degli operatori di borsa, di pescatori, degli addetti aeroportuali ecc.).

Come abbiamo già visto anche Enfield cerca di superare la staticità della nozione tradizionale di lingua attraverso un approccio socio-interazionale e con la concettualizzazione del *composite utterance*. L'enunciato è visto come una unità del comportamento sociale che ha una ben definita relazione causa-effetto e ogni mossa comunicativa scaturisce da determinati dalle presupposizioni e dai *commitments* che vengono riportati nello scambio dei turni dialogici e di cui l'interlocutore è tenuto a rispondere. È proprio il riconoscimento dell'anatomia casuale/condizionale e normativa delle sequenze di interazione, in cui ogni mossa porta a un nuovo obiettivo con conseguenze per i parlanti coinvolte, che permette a Enfield il superamento della staticità del segno linguistico sassuriano. La dinamicità semantica del *composite utterance* è determinata anche dalla diversa natura dei segni che lo compongono; segni convenzionali, segni non convenzionali e segni simbolici indessicali (Enfield 2009).

La questione della multimodalità del linguaggio ha portato a rivedere la definizione di lingua da un punto di vista semiotico-linguistico, ma ci porta anche a estendere la questione anche all'aspetto cognitivo del linguaggio. Lo stesso De Mauro, a nostro avviso, aveva accennato a questo bisogno affermando:

«[...] il linguaggio è fatto di modalità dell'organizzazione delle nostre esperienze nei significati e nei segni, di strutture profonde, e non della materialità cui ricorriamo per rendere percepibili i significanti dei segni. Uno stesso istinto linguistico guida la fonazione e i gesti, non c'è un dio maggiore e uno minore, ma, se c'è e per quel che c'è, è un unico dio» (De Mauro, 1998)

Così nel paragrafo successivo vedremo qual è il contributo portato dalla linguistica cognitiva, in particolare nel suo approccio embodied, e vedremo come una visione cognitiva possa aiutare a comprendere meglio l'interazione tra i diversi codici semiotici che, nonostante la loro diversa natura, riescono a interagire e integrarsi in un unico sistema lingua.

5. Il contributo della linguistica cognitiva

Il contributo degli studi cognitivisti alla linguistica ha portato a una visione diversa del linguaggio e dell'interazione tra parlanti al punto da segnare un vero e proprio cambio di rotta.

Come abbiamo visto nel paragrafo precedente, l'approccio tradizionale basa l'idea di lingua sul concetto di arbitrarietà e vede la lingua come sistema monolitico strutturato sul rapporto vicendevole di vari elementi. Abbiamo anche visto come, davanti alla multimodalità del linguaggio e agli studi sulla lingua dei segni, alcuni studiosi abbiano esteso la nozione di lingua riconoscendole la natura poliesemiotica e socialmente dinamica. Con la nascita della linguistica cognitiva, all'interno della tendenza degli studi funzionalisti e pragmatici della seconda metà del XX secolo, si approfondiscono sempre di più gli studi sulla relazione tra lingua e contesto concentrandosi soprattutto sul processo della comprensione del linguaggio visto come processo dinamico. La linguistica cognitiva si focalizza sui meccanismi cognitivi che stanno alla base di una lingua e del comportamento dei parlanti concentrandosi appunto sulla cognizione²⁰ che si sviluppa attraverso l'esperienza, i sensi, le emozioni, il pensiero e la lingua.

L'iniziatore della linguistica cognitiva Langacker sostiene che esiste una stretta relazione tra il linguaggio e tutti gli altri aspetti della cognizione umana, assunto che si estende a tutti i livelli di analisi linguistica. Questo

²⁰ Il termine cognizione racchiude in sé tutti i processi di acquisizione, elaborazione, produzione e comprensione.

approccio è strettamente influenzato dal concetto di *embodiment*²¹ secondo cui la mente è condizionata sia dalla fisicità del corpo e del cervello che dall'esperienza soggettiva e culturale dei parlanti, mettendo in evidenza il valore dell'iconicità nel sistema lingua.

L'approccio delineato dalla linguistica cognitiva favorisce così il delinearsi di una nuova idea di comprensione e interazione tra parlanti che non si basa più sul modello di codifica e decodifica ma che affonda le proprie radici sulla simulazione mentale²². Ma nello specifico, qual è il ruolo giocato dalla simulazione nel processo di comprensione del linguaggio? E soprattutto che relazione si instaura tra simulazione, comprensione e multimodalità?

Alan Cienki ha cercato di dare una risposta a queste domande e nel farlo ha messo in rassegna quattro diversi approcci alla comprensione del linguaggio: il modello di codifica-decodifica, il modello dell'interazione dialogica, l'approccio pragmatico e infine quello multimodale.

Il modello di codifica-decodifica vede la produzione linguistica come codificazione di un determinato significato attraverso un determinato segnale che il ricevente deve decodificare per comprenderlo. Da ciò, dice Cienki (2017), si potrebbe dedurre che nella comunicazione interpersonale l'oggetto della comprensione non è la frase in sé stessa, ma il significato così come è inteso dal locutore. Questa affermazione, oltre ad aprire molte questioni relative alla semantica, solleva una domanda: quanto la simulazione mentale è coinvolta nel processo di produzione linguistica? Lo sviluppo di un pensiero è un tipo di simulazione? Senza dubbio l'idea di *growth-point* sviluppata dai lavori di McNeill (1992) – che approfondiremo nei capitoli successivi – su come gesto e parlato si sviluppino insieme nel processo di produzione del linguaggio, ci porta a pensare a questo tipo di

²¹ «[...] la forma e le capacità motorie ascrivibili a un corpo sono fattori determinanti rispetto alla genesi e allo sviluppo di un sistema cognitivo» Zipoli Caiani, 2016.

²² Secondo Bazzanella 2014 la *simulazione* è la rappresentazione concettuale dell'atto che si potrebbe eventualmente compiere. Cienki 2017: «The basic idea is that people who are comprehending language to some degree simulate the perceptual, the motoric, and the emotional qualities of the described scenes».

simulazione come un modo dinamico in cui un'idea si evolve come se stessimo producendo sia il parlato e che il gesto nello stesso tempo (Cienki 2017).

Per il modello dell'interazione dialogica la comprensione del linguaggio attraverso la simulazione è costruita in modo incrementale durante lo scambio dialogico tra i parlanti, così le simulazioni di uno dei parlanti sono modulate anche sulla base dei feedback ricevuti dall'interlocutore. In questo modo la comprensione, oltre a essere in parte un'attività di decodifica dei segnali comunicativi dell'interlocutore, è anche un'attività condivisa tra i parlanti che condividono conoscenze comuni, il medesimo contesto comunicativo e soprattutto la possibilità di assestare l'informazione attraverso feedback e correzioni.

Il terzo approccio considerato da Cienki (2017) è quello pragmatico che vede la comprensione come attività basata sulle inferenze pragmatiche che un ascoltatore deriva dalle informazioni ricevute da un parlante (Grice 1967; Levinson 1983; Searle 1975). Da ciò lo studioso si chiede: fino a che punto la simulazione del contenuto di un messaggio agisce nel riconoscimento dell'intenzione di un parlante? Come si arriva da questa idea di simulazione a quelle di inferenza pragmatica? La simulazione ha un ruolo funzionale nella comprensione dell'inferenza?

Infine, dal punto di vista della multimodalità (che, come abbiamo visto, ingloba gli aspetti gestuali tra gli elementi della comunicazione verbale) rimane ancora poco chiaro come gesto e parlato si combinino nell'atto della comprensione. Ma Cienki, citando uno studio di Gallese (2007), sostiene che da un punto di vista multimodale si possa parlare di simulazione embodied, cioè il significato comunicativo dei gesti si fonde con l'articolazione dei suoni impiegati per la creazione del significato stesso. L'impiego della gesticolazione può portare diverse tipologie di significato che vanno dalla traiettoria di una determinata azione o dalla dimensione di un determinato gesto, fino al punto di vista del parlante in relazione a un evento raccontato (cfr. cap. 4).

Alla luce di questa breve rassegna su come la lingua viene compresa, come se ci fossero diversi livelli che agiscono nello stesso tempo, è lecito domandarsi come agisce la simulazione. Sono diverse simulazioni che interagiscono o è una simulazione integrata? (Cienki 2017).

Cienki sostiene che la simulazione sia guidata da un lato dalla lingua, e quindi basata sull'enunciato che viene espresso, e dall'altro lato vi è invece una simulazione legata all'aspetto sociale del contesto d'uso della lingua. In questo modo, da un punto di vista multimodale è possibile pensare che i gesti riescano a veicolare sia informazioni di tipo esclusivamente linguistico che anche sociale e contestuale. A questo si aggiunge che i parlanti riescono a raccogliere informazioni espresse esclusivamente dai gesti, senza che queste vengano espresse anche dal parlato.

Cienki ci ricorda che Hostetter e Alibali (2008) hanno proposto un modello sulla produzione di parlato e gesto partendo dal presupposto che quando si esegue un qualsiasi tipo di gesto, non scaturisce da un pensiero totalmente spontaneo, ma emerge da un processo che fa uso di mezzi comuni di interazione fisica con il mondo. Il modello da loro proposto riguarda l'elaborazione del linguaggio di parlanti che interagiscono con immagini motorie e visive e fa riferimento a un loop azione-percezione che può dare origine a un'attivazione motoria al fine di produrre un gesto. Ma non necessariamente per ogni enunciato vengono prodotti dei gesti. Anzi, Hostetter e Alibali dicono che c'è una soglia che deve essere superata affinché un gesto sia effettivamente prodotto. Il livello della soglia potrebbe dipendere dalla forza neurale della connessione tra le aree premotorie e motorie del cervello. L'area premotoria è una specie di area di pianificazione che porta alle aree motorie, che ci fa effettivamente produrre movimenti. In alcuni casi, la soglia appare essere più alta, e si dovrebbe avere una maggiore attivazione motoria per produrre effettivamente un gesto. Questo dimostra che i gesti dei parlanti possono giocare un ruolo cruciale nell'influenzare la simulazione mentale degli uditori.

Così, conclude Cienki, la simulazione può essere vista non esclusivamente come un processo mentale inconscio, ma anche come un processo che coinvolge diversi tipi di immaginazione e che possono essere usati in modo consapevole. Quindi è possibile pensare alla simulazione sia a livello inconscio che a livello conscio, variando in base alle diverse situazioni comunicative.

In alcuni casi abbiamo un'immaginazione più ricca su ciò che sta accadendo, e in altri casi, non ci stiamo concentrando tanto su quell'immaginazione. Ma ancora, se vediamo qualcuno gesticolare potrebbe, a livello cosciente, iniziare ad attivare una parte di quel tipo di immaginario motorio (Cienki 2017).

L'approccio cognitivo permette una comprensione e un'analisi degli eventi linguistici in modo globale e quanto più vicino possibile all'uso. In questo modo è possibile condurre sotto la stessa analisi le diverse modalità su come le informazioni e i significati vengano elaborati e interpretati dai parlanti. Senza dubbio le questioni fin qui illustrate rappresentano solamente un indirizzo degli studi e non esauriscono molte questioni che rimangono aperte. Ma anzi, come abbiamo visto, cercare di comprendere come i parlanti riescano, attraverso la simulazione, a comprendere i diversi livelli informativi nelle diverse modalità, porta a ulteriori domande aprendo un vasto campo di interesse da indagare.

Il principale interesse di questo nostro lavoro è quello di costruire un corpus multimodale dell'italiano parlato in contesto spontaneo e dunque una delle questioni di maggiore interesse è comprendere come segmentare il flusso dell'azione linguistica multimodale. E in particolare in questo capitolo, come superare la definizione di lingua tradizionale che non riesce a inglobare gli aspetti legati alla gestualità. Proprio su quest'ultima questione la Grammatica Cognitiva riesce a offrire modelli e metodi che possono risultare efficaci.

Per la linguistica cognitiva la grammatica di una lingua fornisce un repertorio di strutture simboliche, utilizzate dai parlanti sulla base delle

loro capacità di concettualizzazione e categorizzazione, che vengono viste come coppie formate da una struttura formale (che è costituita dalle associazioni suono-significato) e una struttura semantica (che è costituita dagli schemi di costruzione). Queste unità simboliche non si costituiscono in modo arbitrario, ma sono motivate da schemi esperienziali che poi vengono estesi dal concreto all'astratto, spesso attraverso l'uso della metafora o altri mezzi utili alla concettualizzazione.

Langacker (2012) sintetizza l'idea di unità simbolica con l'immagine riportata nella figura di seguito, dove è possibile vedere che sotto la struttura concettuale non ci sono esclusivamente i referenti concettuali ma anche altri elementi come la struttura informativa e la gestione del discorso. Mentre il canale espressivo è composto dal contenuto segmentabile (che potremmo ricondurre all'espressione verbale) e da aspetti co-verbali come la prosodia e i gesti.

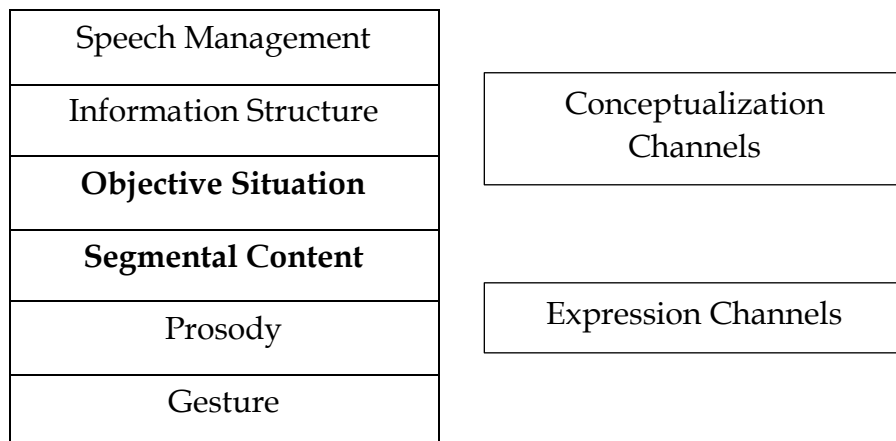


Figura 1 Unità simbolica di Langacker (2012)

Da ciò ne consegue che la linguisticità di un elemento dell'espressione non dipende dalla sua modalità (parlato o gestuale). Ma il potenziale grammaticale degli elementi co-verbali è relativo al loro grado di radicamento come strutture simboliche nella mente di un parlante e al grado di convenzionalità di queste strutture simboliche all'interno di comunità (Kok - Cienki 2016).

In un interessante articolo Kok e Cienki fanno un'analisi approfondita del rapporto tra la Grammatica Cognitiva e i gesti. I due studiosi, partendo dalla dimostrazione che i diversi tipi di gesti (emblemi, referenziali, ricorrenti e creativi²³) hanno un certo grado di schematicità formale-significato, analizzano come i gesti si relazionano con le diverse dimensioni del *construal* (specificità, prospettiva, focus, autonomia-dipendenza e le categorie grammaticali di base). Nonostante la Grammatica Cognitiva offra un approccio teorico e strumenti che ci restituiscono una nozione di lingua che ingloba la multimodalità, rimangono delle difficoltà legate allo studio sui gesti. La prima difficoltà è legata alla rappresentazione e annotazione dei gesti a causa della mancata corrispondenza biunivoca tra forma e funzione del gesto a cui si aggiunge la complessità della struttura morfosintattica derivante dalla modellazione della lingua come attività

²³ Quelli che Kendon definisce *gesticulation*.

dinamica e multimodale. Infatti, la struttura simbolica espressa da gesto e parlato solitamente appare come non organizzata né sequenzialmente né come modello. Infine, un'ultima questione è quella che Kok e Cienki chiamano *speaker-hearer asymmetry*. È stato visto come alcuni gesti abbiano una funzione cognitiva self-oriented, aiutano i parlanti a strutturare il processo cognitivo in atto attraverso l'esternalizzazione dell'immaginazione. Di conseguenza questi gesti sembrano sfuggire all'analisi linguistica in quanto non intenzionalmente comunicativi, pur essendo utilizzati dai parlanti e potenzialmente interpretabili dagli uditori (Kok & Cienki 2016).

La prospettiva cognitiva si è rivelata utile e fruttuosa per gli studi sulla lingua dei segni. Abbiamo visto precedentemente come in Italia, gli studi di De Mauro e del gruppo LACAM del CNR di Roma, abbiano dato un forte contributo al riconoscimento della linguisticità della LIS, grazie al superamento di una visione statica e stretta della nozione di lingua. Gli studi successivi hanno portato a un approccio cognitivo e sociosemiotico per la descrizione della LIS (Volterra et al. 2019) con il superamento del limite dato dall'uso di strutture e categorie proprie delle lingue vocali. In questo modo è stato possibile inglobare nell'analisi tutte le caratteristiche peculiari delle lingue dei segni.

Il punto di svolta è dato dal riconoscimento della continuità tra segno, gesto e parola che si fonda sul riconoscimento dell'origine comune rintracciabile nell'azione. È stato notato come i gesti fondamentali nelle prime fasi dell'acquisizione del linguaggio (sia per le lingue segnate che parlate) siano di fatto azioni. Da queste azioni si sviluppa successivamente la comunicazione intenzionale e simbolica fino all'acquisizione dei segni e delle parole; ed è stato inoltre dimostrato che i significati dei primi gesti sono gli stessi delle prime parole facendo emergere che gesti e parole condividono lo stesso spazio concettuale. E in un confronto tra i gesti utilizzati dai bambini udenti e quelli utilizzati dai bambini apprendenti la lingua dei segni americana sono state riconosciute lo stesso repertorio di

forme, questo ha confermato l'ipotesi dell'esistenza di un vincolo motorio nella realizzazione di gesti rappresentativi. Le somiglianze tra gesti e segni non si limitano solamente ai parametri di esecuzione ma coinvolgono anche le strategie rappresentative che sono fondamentalmente quattro: mimo (il corpo rappresenta l'azione o il personaggio), manipolazione (le mani assumono un tipo di afferramento che richiama la forma dell'oggetto), mano come oggetto (le mani assumono la forma dell'oggetto), forma e misura (le mani tracciano la forma o indicano le dimensioni) (Volterra & aliis 2019). Alla luce di questa continuità è utile evidenziare che la lingua dei segni utilizza come propri componenti il corpo, le mani e l'espressione che quotidianamente vengono impiegati per la realizzazione di diversi compiti come indicare, mostrare, dare, enumerare, manipolare e rappresentare. Questi componenti utilizzano le stesse strategie rappresentative appena elencate, non soltanto per la creazione dei segni ma anche per l'organizzazione stessa della lingua. L'approccio cognitivista adottato in questa nuova descrizione della LIS, che si basa appunto sulla continuità azione-gesto-segno-parola, ha portato a illustrare come i parametri formazionali interagiscono per la creazione dell'*unità di senso* che a loro volta possono essere distinte in Unità Deittiche, Unità Lessicali e Unità di Trasferimento²⁴ (vedi figura 2). Il rapporto tra forma e significato è sempre frutto di un compromesso tra le restrizioni fisiche imposte dal corpo, il tipo di significato che deve essere trasmesso e la strategia adottata per farlo (*ibidem*).

²⁴ Le Unità Deittiche sono i segni di indicazioni che in molte lingue dei segni sono utilizzate per riferirsi a persone, animali, oggetti o eventi; le Unità Lessicali sono segni cristallizzati che vanno a comporre il vocabolario di una lingua dei segni; le Unità di Trasferimento sono segni che rappresentano un oggetto o un evento sulla base di caratteristiche che il segnante o vuole mettere in risalto.

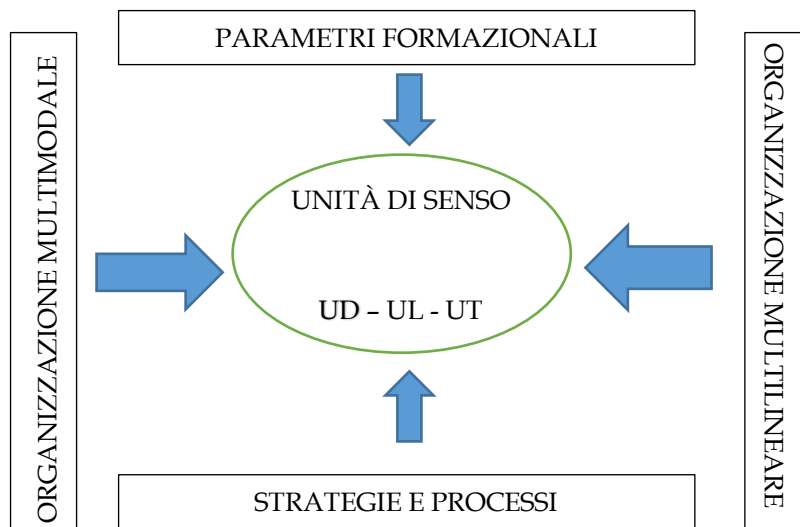


Figura 2 Meccanismi di significazione nella LIS (adattata da Volterra et al. 2019)

Questi studi ci confermano che la lingua, oltre a essere un sistema con proprie forme e paradigmi, è anche azione che i parlanti realizzano in una continua negoziazione sociale, grazie alla quale le strategie di significazione, basate sulla concettualizzazione e vincolate dalla corporeità e dal contesto d'uso, vengono rideterminate garantendo ai parlanti di comunicare.

6. Conclusione

L'essere umano è riuscito a sviluppare nel corso della sua evoluzione diversi canali di comunicazione che vanno dall'uso del canale fono-uditivo, all'uso dei gesti, dei segni grafici, delle immagini e dell'uso combinato di alcuni di questi canali. Senza dubbio il canale fono-uditivo, insieme ai gesti, è quello più usato dagli individui con l'uso concomitante dei gesti e di altri espedienti comunicativi che ne affinano le intenzioni comunicative e i significati trasmessi, rendendo la lingua un sistema multimodale complesso. La complessità del sistema lingua si mostra in modo evidente nello studio di fatti linguistici in cui appare in modo chiaro come la comprensione e la trasmissione dei significati non è un semplice fatto di

codifica e decodifica. Piuttosto emerge, in modo chiaro, che comunicare, per chi parla, vuol dire rendere pubbliche le proprie intenzioni comunicative, e per chi ascolta vuol dire riconoscere le comunicazioni dell'altro. Infatti, non possiamo comprendere il significato delle interazioni esclusivamente attraverso l'analisi dei significati delle parole o delle regole sintattiche. È necessario indagare in profondità come parole, intonazione, gesti ed espressioni facciali si coordinino per far in modo che i parlanti possano esprimere significati e intenzioni comunicative in modo efficace e definito.

La complessità del sistema lingua, data dalla sua natura multimodale, rappresenta un ostacolo per la compilazione di corpora che abbiano l'obiettivo di rappresentare la lingua in tutti i suoi canali espressivi e di significazione. La questione principale è data dal determinare quali unità di analisi costituiscono i diversi canali, come poterli mettere in relazione (per indagarne il funzionamento) e come poter definire un'unità unica che inglobi gesto e parlato. Il problema principale è, dunque, dato dalla segmentazione del flusso gestuale e del flusso parlato (cfr. cap. 1) e dalla definizione di una possibile unità multimodale.

Abbiamo visto in questo capitolo come da un punto di vista teorico siano state introdotte diverse proposte sulla definizione dell'enunciato multimodale. Ma ciò che rimane ancora da definire e indagare è la definizione e il riconoscimento di parametri e caratteristiche che permettano di rendere la nozione di unità multimodale una vera e propria un'unità di analisi. Nel costruire la nostra proposta, riteniamo indispensabile indagare i diversi approcci all'analisi del significato del gesto. Infatti, riteniamo che un metodo di trascrizione del gesto e del parlato non possa prescindere dal comprendere in che modo sia possibile approcciarsi al significato e alle funzioni del gesto.

Capitolo 4 - L'analisi del significato gestuale

1. Introduzione

Il secondo capitolo, dedicato all'introduzione agli studi sul gesto, ha messo in evidenza alcune funzioni espresse dalla gesticolazione, alcune categorizzazioni e l'individuazione della unità in cui è segmentabile il flusso gestuale. In questo capitolo, riprendendo le unità gestuali di Kendon, cercheremo dapprima di introdurre i parametri per l'analisi del gesto e successivamente faremo una rassegna delle principali categorizzazioni proposte.

La definizione di gesto non può prescindere dalla sua categorizzazione. Da sempre gli studi hanno operato una distinzione tra gesti che si accompagnano al parlato e gesti che non si accompagnano al parlato, ma la quantità di categorie gestuali potenzialmente distinguibili nell'uso dei parlanti ha portato David McNeill (1992), a seguito di una sua interpretazione di uno studio di Adam Kendon (1988), a distribuire i fenomeni gestuali lungo un continuum²⁵:

gesticulation – gesture language-linked – pantomime – emblems – sign language

All'estremo sinistro del continuum troviamo i movimenti delle mani e delle braccia totalmente indipendenti dal parlato mentre all'estremo destro ritroviamo le lingue dei segni.

Oggetto di questa tesi sono tutti quei movimenti spontanei che intenzionalmente esprimono un significato; quindi, la presente ricerca si

²⁵ Si rimanda al Capitolo 2 di questa tesi per un approfondimento sul *Kendon's continuum*.

focalizza su tutti quei fenomeni che si collocano sotto le categorie di *gesticulationi* e *gesture language-linked*.

Davide McNeill (1992) definisce i gesti²⁶ come movimenti spontanei e idiosincratici delle mani e delle braccia che accompagnano il parlato, mentre i *gestures language-linked* si distinguono dalla gesticolazione per la non obbligatorietà della presenza del parlato. Lo stretto legame tra gesto e parlato si ritrova anche in Kita *et alii* (1998) in cui la gesticolazione è definita come movimento del corpo prodotto spontaneamente in un contesto di parlato che ha un legame significativo con il parlato co-occorrente sia a livello semantico che pragmatico.

Adam Kendon, invece, non lega la definizione al parlato così racchiude sotto l'etichetta 'gesto' tutte le azioni che hanno elementi di deliberata espressività:

«They are those actions or those aspects of another's actions that, having these features, tend to be directly perceived as being under the guidance of the observed person's voluntary control and being done for purpose of expression rather than in the service of some practical aim. Participants in interaction readily recognize such actions and they tend to be accorded the status of actions for which the participants are held responsible» (Kendon, 2004)

Così, per Kendon non può essere data una definizione che possa delimitare in modo oggettivo gli elementi che distinguono un gesto da un'azione qualsiasi. Ciò che rende riconoscibile un movimento comunicativamente intenzionale è strettamente dipendente dal contesto interazionale e dai partecipanti all'interazione (*ibidem*).

La questione della codifica e della categorizzazione dei gesti, oltre a essere strettamente legate alla definizione stessa di gesto, implicano anche un approfondimento sulla segmentazione del flusso gestuale. Come

²⁶ Con il termine gesto faremo riferimento alle due categorie nell'estremità di sinistra del *Kenondon's continuum*, McNeill lo usa in riferimento alla sola gesticolazione.

abbiamo visto precedentemente (cfr. Capitolo 2), Kendon con i suoi studi ha proposto la segmentazione del flusso gestuale in unità gerarchicamente organizzate (Kendon, 1980):

- *Gesture unit*: è l'unità gerarchicamente maggiore e ingloba l'intera escursione del movimento, da quando le mani iniziano a muoversi fino a quando ritornano in posizione di riposo;
- *Gesture phrase*: è l'unità che suddivide la *gesture unit*. Può essere definita come l'unità a cui corrisponde un significato comunicativo e si identifica per un particolare movimento nello spazio o per una particolare configurazione delle mani;
- *Gesture phases*: sono le unità che compongono una *gesture phrase* e sono di tre tipi, la *preparation* che è la fase di preparazione del movimento, l'unità di *stroke* che è la fase culminante del gesto ed è l'unità necessaria e sufficiente e la *retraction*, che è la fase in cui le mani o le braccia tornano in una posizione di riposo.

La segmentazione in fasi presentata da Kendon è stata successivamente rivista da McNeill (1992) e Kita *et alii* (1998), i quali hanno aggiunto la fase di *hold*, che è la cessazione temporanea del movimento trattenendo la mano o il braccio nella forma assunta. Una fase di *hold* lo troviamo prima o dopo lo *stroke*, così è possibile distinguerlo in due tipi *pre* o *post stroke*.

Data la natura fisica e cinetica del movimento diventa necessario comprendere entro quale spazio le diverse unità si muovono e agiscono per la costruzione del significato e l'interazione con il parlato. McNeill identifica lo spazio gestuale definendolo come una semisfera davanti al parlante delimitata dall'estensione massima delle braccia in avanti, di lato e in alto²⁷. Questo spazio può essere diviso in settori che si sviluppano in un sistema di quadrati concentrici: il settore davanti al petto del parlante è il *centro-centro*, il primo quadrato concentrico circostante è il *center* che a sua volta è

²⁷ Nei bambini lo spazio gestuale solitamente è più esteso (McNeill, 1992).

inquadrato all'interno di un settore più grande che è la *periferia* (a sua volta è divisa in *upper, lower, right, left*), e infine l'*estrema periferia* (fig. 3) (*ibidem*).

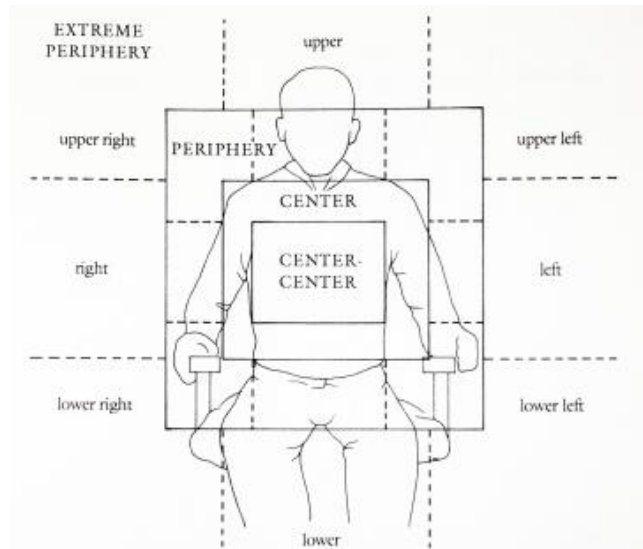


Figura 3 Lo spazio gestuale secondo McNeill (1992)

L'approccio allo studio dei gesti a partire da Kendon e McNeill ha senza dubbio cercato di indagare e comprendere i modi in cui la gesticolazione, in tutte le sue forme, riesca a esprimere significati e svolgere determinate funzioni nell'interazione tra parlanti. Nell'approntare i loro studi e le loro ricerche hanno cercato di individuare delle unità di analisi che permettessero di mettere le basi un approccio comune ed empirico allo studio sui gesti. Ma d'altro canto, come si evince in tutti gli approcci e le analisi che si sono sviluppate nel tempo da Quintiliano ai nostri giorni, emerge una comune esigenza di classificazione e categorizzazione del fenomeno gestuale. In piena coerenza con quanto emerge dagli studi di Kendon e McNeill, e cioè l'esigenza di un'analisi scientificamente solida che si basi su unità di analisi condivise, il ventesimo secolo si distingue dai secoli precedenti - che avevano un approccio al gesto finalizzato all'uso strumentale in contesti artistici o retorici - per aver sviluppato categorizzazioni e classificazioni gestuali con scopi scientifici (Kendon, 2004).

Wilhelm Wundt (1973), all'inizio del ventesimo secolo, propone una classificazione su base semiotica dividendo i gesti sulla base di come la forma dell'azione gestuale si correla al significato del gesto. Parte quindi da una macro-divisione tra gesti *dimostrativi* - che servono a rappresentare l'attenzione verso un oggetto presente, per indicare relazioni spaziali, per riferirsi a parti del discorso e per indicare parti del discorso - e gesti *descrittivi* che sono divisi in gesti *mimici*, *connotativi* e *simbolici*. I gesti *mimici* si caratterizzano perché imitano direttamente il loro referente, sia esso un oggetto o un'azione. I gesti *connotativi* sono quei gesti in cui è riconoscibile un richiamo a una parte di un oggetto come rappresentante per il suo tutto, i gesti che invece mostrano una relazione più complessa con il proprio referente sono quelli *simbolici*. La classificazione di Wundt si concentra esclusivamente sul gesto e sulla sua relazione con i referenti, cercando di dare un ordine a quello che lui stesso definisce linguaggio dei gesti, tralasciando di conseguenza la relazione tra gesto e parlato.

David Efron, nel suo volume *Gesture, Race and Culture* (1972), adotta uno schema di analisi che non si basa sulla categorizzazione delle diverse tipologie di gesti, ma sui diversi modi in cui possono essere osservati i gesti e i diversi modi in cui possono essere usati. Il principale oggetto di studio di Efron sono i movimenti delle mani e delle braccia e porta avanti la sua analisi su tre diverse prospettive:

- La prospettiva *spazio-temporale*: i gesti vengono visti come movimenti a prescindere dall'aspetto interattivo o referenziali e sono analizzati tenendo in considerazione il *raggio del movimento*, la *forma del movimento*, il *piano del movimento*, le *parti del corpo coinvolte* e il *tempo*;
- La prospettiva *interlocutoria*: viene analizzato l'aspetto interattivo del gesto, tralasciando l'aspetto referenziale, facendo riferimento alla *familiarità* tra gli interlocutori, ai *gesti simultanei*, ai

raggruppamenti conversazionali (quindi l'uso dello spazio tra parlanti) e i gesti realizzati *usando un oggetto*;

- La prospettiva *linguistica* in cui viene considerato il significato referenziale della gesticolazione e di conseguenza vengono distinti due grandi classe sulla base che abbiano un significato indipendente dalla parola o in unione con essa:
 - *Gesti logico-discorsivi*, di cui fanno parte tutti quei gesti che non si riferiscono a un oggetto, ma allo svolgersi del processo ideativo ponendo l'accento sul contenuto della modalità riferendosi al *come* e al *che cosa* dell'idea che interpretano. Di questa classe fanno parte i gesti *batonici* e gli *ideografici* (gesti che tracciano nell'aria il percorso del pensiero).
 - *Gesti oggettivi*, di cui fanno parte i gesti che hanno un significato indipendente da quello espresso dal canale verbale con cui possono co-occorrere o meno. In questa classe fanno parte i *deittici*, i *fisiografici* (che mostrano visivamente ciò che significano e che si dividono in *iconografici* - che suggeriscono la forma di un oggetto visibile - e *cinetografici* che suggeriscono un'azione) e gli *emblemi*.

Il metodo di analisi utilizzato da Efron riconosce in modo chiaro il modo in cui il gesto esprime le sue funzioni su livelli e con metodi diversi, sulla base delle esigenze comunicative del parlante. Senza dubbio l'approccio di Efron è uno dei primi che analizza il gesto in relazione al parlato, restituendo la multiformità del fenomeno gestuale e la conseguente necessità di adottare un approccio ampio e diversificato per riuscire a cogliere la sua complessità.

L'analisi di Efron è stata utilizzata anche Ekman e Friesen nel 1969, in un articolo in cui è stato proposto un loro sistema di categorizzazione del

“comportamento non verbale”. Gli autori definiscono “nonverbal behavior” qualsiasi movimento o posizione del viso e/o del corpo, questa definizione è direttamente connessa con il campo di interesse e di studio dei due ricercatori. Infatti, i due autori sono interessati a tutti quei movimenti della faccia e del corpo che giocano un ruolo nell’interazione in presenza. La categorizzazione proposta nell’articolo del 1969 parte dal bisogno di tenere in considerazione tre aspetti fondamentali di ogni fenomeno indagato, cioè è necessario comprenderne *l’origine* (che fa riferimento alla possibilità che un atto possa essere stato appreso o derivi da qualche forma azionale innata), *l’uso* (che fa riferimento ai condizionamenti esterni, alla relazione tra l’azione gestuale e il comportamento verbale che co-occorre, se un’azione è fatta in modo cosciente, il tipo di feedback che un’azione scatena nell’interlocutore e il tipo di informazione trasmessa dell’azione) e la *codificazione* (che riguarda il principio che governa la corrispondenza tra atto e il suo significato e che può essere arbitrario, iconico o intrinseco quando non c’è separazione tra l’atto e il suo significato). Così i due studiosi suggeriscono che «nonverbal behavior is not a single unified phenomenon with one type of usage, one origin and one of coding» (Ekman - Friesen, 1969), ci sono piuttosto diversi tipi di azioni non-verbali, loro indicano cinque categorie:

1. Gli *emblemi*, che includono i comportamenti non verbali che hanno una diretta traduzione nella modalità verbale;
2. Gli *illustratori* sono movimenti direttamente legati al parlato e che servono per illustrare quello che viene detto verbalmente;
3. Gli *affect displays* che sono principalmente le espressioni facciali che esprimono le emozioni;
4. I *regolatori*, quei gesti utilizzati nella gestione dei turni dell’interazione;
5. Gli *adattatori*, che sono movimenti che cercano di soddisfare i bisogni personali corporei, o che cercano di gestire le emozioni, o per

sviluppare e mantenere i contatti interpersonali; vengono distinti in *autoadattatori*, *alter-adattatori* e *oggetti-adattatori*.

Mentre David McNeill, nel suo lavoro del 1992, illustra una propria categorizzazione dei gesti che si basa sui vari modi in cui i gesti possono essere interpretati come atti simbolicamente espressivi in relazione al contesto creato dal contenuto del parlato che co-occorre. La classificazione parte da una prima distinzione tra gesti *imagistic* e *non-imagistic*, i primi sono quei gesti in cui è possibile riconoscere la forma di un oggetto, la rappresentazione di un'azione o di qualche schema di movimento mentre i gesti *non-imagistic* includono i deittici e i gesti che sembrano essere semplicemente ritmici. I gesti *imagistic* si dividono, inoltre, in *iconici* e *metaforici*.

Nel complesso panorama degli *gesture studies*, in cui negli anni si è giunti al riconoscimento di unità di analisi condivise (le unità gerarchiche individuate da Kendon), continua a mancare un approccio allo studio gestuale che si basi su metodi e punti di vista ampiamente condivisi dalla comunità scientifica. Si assiste, infatti, alla nascita di diversi approcci che, da un lato, hanno contribuito all'accrescimento delle conoscenze e all'apertura di nuove questioni intorno all'uso e alla funzione del gesto, e dell'altro lato hanno moltiplicato categorizzazioni, codificazioni, annotazioni e metodi rendendo di conseguenza complesso il dialogo tra i diversi punti di vista.

In questo paragrafo introduttivo abbiamo visto le principali classificazioni dei gesti del Novecento, nei paragrafi successivi faremo una rassegna dei principali approcci degli ultimi anni in modo da osservare come i diversi campi di studio si avvicinano al fenomeno gestuale e allo studio della sua significazione e delle sue funzioni. Questa ci permetterà di problematizzare la principale questione affrontata nella costruzione del corpus, la ricerca di un'unità di riferimento multimodale che possa rendere possibile la creazione di un corpus ampio e utilizzabile dai diversi campi di

studio. In questa rassegna vedremo un approccio definito linguistico e che si basa sulla decodifica del gesto per poterne comprendere significato e funzioni fino a riuscirne a definire una grammatica, l'approccio conversazionale che studia il ruolo della corporeità all'interno dell'interazione, l'approccio cognitivo che invece studia i metodi di significazione gestuale per comprendere il sistema di concettualizzazione e infine l'approccio Kinesiologico che studia la lingua come sistema multimodale integrando il ruolo del corpo con tutti i suoi vincoli e le sue potenzialità fisiche.

2. L'analisi dei gesti come espressione della concettualizzazione

La linguistica cognitiva, in tutte le sue linee di ricerca, riconosce nella semantica il punto di partenza per indagare e spiegare le strutture linguistiche approcciandosi al significato come forma di concettualizzazione (Cienki, 2013). La centralità del significato è strettamente legata all'analisi della lingua nel suo contesto d'uso, alla luce del fatto che per la linguistica cognitiva ogni significato può essere espresso attraverso l'uso di varie forme del comportamento (Langacker, 1988). Nel contesto degli *gesture studies*, l'assunzione che l'espressione del significato possa avvenire per mezzo di diverse forme di comportamento assume un'importanza rilevante perché di fatto ammette l'esistenza di forme di espressione oltre le parole stesse, che ricorrono negli eventi d'uso della lingua e che possono assumere lo status di unità simboliche (cfr Capitolo 3) (Cienki, 2013). Il termine concettualizzazione viene utilizzato in riferimento alla visione dinamica della natura del significato nella linguistica cognitiva (Langacker, 2008), nonostante questa visione sia largamente condivisa all'interno dell'approccio cognitivista rimane comunque oggetto di dibattito l'individuazione di ciò che dovrebbe essere incluso nella

concettualizzazione del linguaggio. In ogni caso, per la linguistica cognitiva la lingua non si riferisce direttamente al mondo, ma alla concettualizzazione che si basa sull'interpretazione del parlante di una data situazione (Cienki, 2013). Le principali nozioni teoriche del quadro teorico cognitivista poggiano sull'uso esclusivamente verbale della lingua ma, come vedremo, alcune ricerche hanno evidenziato l'importanza e l'intrinseco legame tra la linguistica cognitiva e gli *gestures studies*. L'approccio allo studio della concettualizzazione ha evidenziato come metodi e strumenti diversi influenzino la costruzione del significato negli eventi linguistici, nello specifico parliamo della metafora, della metonimia, della schematizzazione, della prospettiva della costruzione, degli spazi mentali e della simulazione mentale. Nel 2013, Cienki ha fatto una rassegna di come, da un punto di vista cognitivo, gesti e parlato esprimono la concettualizzazione attraverso l'uso degli espedienti appena elencati.

Gli studi sulla metafora concettuale rappresentano l'area di ricerca iniziale della linguistica cognitiva e vedono le metafore non come uno strumento linguistico ma come mezzo che i parlanti usano per la concettualizzazione; così le metafore concettuali possono essere espresse sia in parole scritte o parlate che per mezzo di altre forme del comportamento umano. Ci sono diversi modi in cui le metafore nella lingua parlata e nel gesto possono correlarsi; infatti, è possibile rintracciare casi in cui lo stesso immaginario appare sia nel gesto che nel parlato. Per esempio, quando qualcuno parla rispetto a prendere in considerazione alcuni aspetti allo stesso modo attraverso la metafora del bilanciamento e simultaneamente produce un gesto a due mani piatte con il palmo all'insù, e un movimento che vede le mani alzarsi e abbassarsi in modo simultaneo. Questo gesto è stato interpretato (Cienki, 1998) come il riflesso di una mappatura metaforica dal dominio di origine, la pesatura di oggetti, comparativamente al dominio di destinazione, considerare la loro importanza relativa. Questa corrispondenza non necessariamente deve avvenire in modo simultaneo, può avvenire sia consequenzialmente che parzialmente sovrapposte nel

tempo. Ci sono co-produzione di gesti e parole che esprimono diverse costruzioni metaforiche di uno stesso dominio target. In Cienki (2008) viene riportato l'esempio di uno studente che, parlando della distinzione morale tra giusto e sbagliato, utilizza la metafora concettuale di *bianco o nero* rendendola diversamente da un punto di vista gestuale. Infatti, il gesto che viene prodotto mettendo il bordo della sua mano destra, tesa e piatta nel piano verticale, contro il palmo della sua mano sinistra, piatto nel piano orizzontale e con il palmo rivolto verso l'alto. Il gesto di fatto crea una netta divisione spaziale in due riflettendo in modo iconico il concetto di *bianco e nero*. Inoltre, le metafore possono apparire o solo nel parlato o solo nei gesti: un esempio è riscontrabile nel gesto che scorre da sinistra verso destra mentre un parlante sta descrivendo una qualsiasi procedura.

La metonimia è un processo cognitivo in cui un'entità concettuale viene utilizzata come mezzo per accedere a un'altra entità concettuale all'interno di uno stesso modello cognitivo (Langacker, 1993). Gli studi portati avanti in questo campo di ricerca evidenziano una certa differenza tra la metonimia espressa verbalmente e quella prodotta dai gesti. Nell'espressione verbale il veicolo metonimico è un segno o una combinazione di segni che appartiene a un sistema convenzionalizzato di accoppiamenti forma-significato e che riduce di conseguenza la natura iconica delle espressioni. Al contrario, l'iconicità gioca un ruolo importante nei gesti referenziali anche se molto spesso appare in forme molto schematizzate. La costruzione metonimica nei gesti si realizza sia per mezzo dei gesti deittici che per quelli rappresentazionali. Nel primo caso avremo un gesto deittico che non indica direttamente il referente ma un indice riferibile al referente, mentre nei casi dei gesti rappresentazionali avremo dei gesti che rappresenteranno solamente una part o parti di qualche azione o entità (Cienki, 2013).

All'espedito cognitivo della schematicità è stato riconosciuto un ruolo importante nella modalità linguistica di rappresentazione del significato.

Gli *schemas*²⁸ sono strumenti che ci permettono di concettualizzare, memorizzare e riconoscere aspetti fisici del mondo che percepiamo permettendoci di utilizzare queste informazioni per i concetti astratti. Gli studi di linguistica cognitiva in relazione ai gesti si sono occupati di diversi argomenti, senza dubbio uno dei più interessanti è la relazione tra i gesti e gli *image schemas*: alcune ricerche si sono concentrate sulla ricezione, altre ricerche sulla produzione, come per esempio l'uso dell'immagine CYCLE intorno a cui i parlanti tedeschi hanno creato una *gesture families* (Ladewig, 2011; Kendon, 2004). Oltre agli *image schemas* le ricerche hanno individuato gruppi di gesti che schematizzano altri tipi di immagini come le strutture geometriche, in modo particolare se si tratta di rappresentazioni iconiche di oggetti fisici. Ci sono, inoltre, gesti in cui gli studiosi hanno riconosciuto la schematizzazione di azioni che può fornire una struttura applicabile a diversi contesti, in relazione a significati di diverso grado di astrazione e/o con diverse funzioni.

L'equiparazione presente nei dettami della linguistica cognitiva tra significato e concettualizzazione implica che essa venga realizzata sempre attraverso una determinata prospettiva. In relazione ai gesti, questa prospettiva può cambiare se colui che gesticola assume il punto di vista del personaggio o dell'osservatore (McNeil, 1992). Ma non sempre il punto di vista è direttamente legato alla prospettiva fisica che si assume nel descrivere una scena, infatti, un altro fenomeno di concettualizzazione è la descrizione di scene statiche in termini di processi dinamici e definito *fictive motion*. Questo dimostra che i gesti, nell'espressione del significato, vengono usati dai parlanti anche per segnalare il loro punto di vista o la prospettiva sull'argomento discusso in quel momento (Cienki, 2013).

La teoria dello spazio mentale prevede che alcune forme linguistiche (sia lessicali che grammaticali) rivelino l'uso che un parlante fa dei diversi domini di strutture referenziali all'interno del suo discorso. In questo modo,

²⁸ Utilizziamo il termine inglese per far riferimento agli studi sugli *schemas* al fine di evitare possibili confusioni concettuali.

le forme linguistiche, hanno la funzione di dare all'ascoltare delle indicazioni sulla costruzione dello spazio mentale in modo che possa essere vicino o coincidente con quello del parlante. Oltre che nelle forme linguistiche verbali, è possibile riscontrare la presenza di questi fenomeni anche in altre forme espressive come il gesto, lo sguardo e il movimento della testa (Cienki, 2013). Mentre, per la simulazione mentale, l'elaborazione concettuale utilizza ricostruzioni di stati senso-motori per rappresentare le categorie, così queste simulazioni sembrano essere delle vere e proprie ricostruzioni parziali (Barsalou, 2003). Alcuni recenti sviluppi, in quest'area di ricerca, sul parlato e sul gesto hanno dimostrato che anche i gesti possono assumere un importante ruolo nella simulazione mentale. Emblematico è l'esperimento di Cook e Tanenhaus (2008) che ha coinvolto alcuni partecipanti che dovevano risolvere la Torre di Hanoi, alcuni attraverso l'uso di un gioco fisico altri attraverso un computer, successivamente entrambi i gruppi dovevano descrivere il gioco ai nuovi partecipanti. Sono state riscontrate evidenti differenze sia nei gesti eseguiti dai due gruppi durante la spiegazione del gioco, ma soprattutto sui comportamenti assunti dai nuovi partecipanti che sono stati evidentemente influenzati dalle informazioni ricevute dai gesti del primo gruppo facendo quindi emergere la simulazione mentale suscitata dai gesti che avevano visto.

3. L'analisi da una prospettiva linguistica: il valore della forma

Jana Bressem (2013) ha assunto una prospettiva linguistica alla base della sua proposta di un sistema di trascrizione basato sulle caratteristiche formali del gesto. Secondo la prospettiva semiotico-linguistica, gesto e parlato sono elementi connessi in modo inseparabile, di conseguenza, il linguaggio è intrinsecamente multimodale. L'analisi linguistica del gesto

mira a indagare e scoprire le caratteristiche comuni al parlato e le specifiche differenze tra le due modalità, infatti, nonostante questo approccio definisca il linguaggio multimodale, considera parlato e gesto come due distinti sistemi che condividono proprietà comuni (*ibidem*).

L'analisi della forma del gesto si basa su quattro parametri sulla base dei criteri descrittivi delle lingue dei segni (Klima e Bellugi, 1979):

- Forma della mano;
- Orientamento;
- Movimento;
- Posizione.

Un approccio semiotico-linguistico al gesto assume che ciascuno di questi parametri possa avere un valore rilevante nella creazione del significato gestuale, di conseguenza, l'esclusione di uno dei parametri dalla descrizione di un evento può restituire una realizzazione semantica deficitaria di elementi e aspetti essenziali.

L'annotazione delle forme gestuali proposta da Bressemer (2013) assume una prospettiva fonetica, cioè mira a una rappresentazione articolatoria della forma partendo da una descrizione che sia moderatamente granulare. Questo tipo di annotazione è del tutto indipendente da una descrizione anatomica delle braccia o di altre parti del corpo, piuttosto, l'attenzione è rivolta alla descrizione anatomica delle mani. Secondo questa prospettiva, infatti, solamente le mani rappresentano l'elemento necessario per una descrizione delle forme gestuali, in quanto assumono un importante ruolo nella creazione del significato e, inoltre, alcune caratteristiche delle braccia sono deducibili dall'annotazione della forma delle mani. L'annotazione e la descrizione delle forme del gesto sono indipendenti dall'enunciato verbale così, come vedremo, le etichette dei quattro parametri notazionali sono state pensate per una trascrizione senza suono. Infine, la trascrizione secondo Bressemer deve evitare di parafrasare il significato e la funzione del gesto per garantire una descrizione delle forme gestuali che sia la più oggettiva possibile (Bressemer, 2013).

Il primo parametro dell'annotazione è la *forma della mano* in quanto caratteristica più prominente della forma gestuale. Infatti, la percezione e l'identificazione della forma della mano risulterebbe più facile e più stabile rispetto a quella del movimento. Secondo il sistema proposto da Bressemer, alla forma della mano si possono assegnare quattro categorie base: 1) Fist;

2) Flat hand; 3) Single fingers; 4) Combination of fingers. Queste categorie mostrano le quattro aree prominenti che determinano la forma assunta dalla mano. Inoltre, è necessario segnalare se la forma gestuale coinvolge entrambi le mani: se sono entrambe separate allora la descrizione si baserà sulle quattro categorie, sul numero delle dita coinvolte e sulla forma delle dita oppure indicate individualmente con l'etichetta 'hands interlocked'.

Per le etichette 'single fingers', 'combination of fingers' e per una forma gestuale che coinvolge entrambi le mani è necessario specificare il numero e la forma delle dita impiegate. Per identificare le dite delle mani, sono state numerate da 1 (pollice) a 5 (mignolo), mentre la forma delle dita può essere specificata attraverso l'uso di sei etichette che descrivono i diversi tipi di flessione che le dita possono assumere: 1) stretched; 2) bent; 3) crooked; 4) flapped down; 5) connected; 6) touching.

In breve, è possibile dire che l'annotazione del parametro 'forma della mano' passa per tre step: assegnare la forma della mano sulla base delle quattro categorie o classificarla come una forma che coinvolge entrambe le mani, numerare ciascun dito e infine specificare la forma delle dita (*ibidem*).

La codifica del parametro *orientamento* si basa sulla distinzione fatta da McNeill (1992) secondo cui l'orientamento della mano dipende dall'orientamento del palmo e dallo spazio gestuale. Per la definizione dell'orientamento del palmo della mano vengono utilizzate quattro etichette: 1) palm up; 2) palm down; 3) palm lateral; 4) palm vertical. A queste quattro categorie la Bressemer ha aggiunto il marcatore 'diagonal' che è utilizzato per differenziare le quattro angolazioni di base e per marcare un orientamento intermedio tra loro. Per gli orientamenti 'palm lateral', 'palm vertical' e per ogni orientamento marcato come laterale si aggiunge un'ulteriore differenziazione che riguarda lo spazio gestuale. Così si distinguono quattro tipi: 1) towards center; 2) away center; towards body; 4) away body.

Il terzo parametro nell'annotazione è il *movimento*, il più complesso. Si costituisce di tre aspetti che bisogna considerare in modo separato: il tipo di

movimento, la direzione del movimento e la qualità. Per *tipo di movimento* si intende la forma disegnata dell'andamento del movimento e se ne riconoscono sei: 1) straight; 2) arced movement; 3) circle; 4) zig-zag; 5) s-line; 6) spiral. Mentre per il movimento eseguito dal pugno si distinguono tre possibili tipologie: bending, raising e rotation. Un altro aspetto indispensabile per descrivere il movimento è la *direzione*. Sia che per il movimento delle braccia e delle spalle che per le singole dita è possibile distinguere tre principali direzioni: 1) movimenti lungo l'asse orizzontale (da destra a sinistra guardando dalla prospettiva di chi realizza il gesto); 2) movimenti lungo l'asse verticale; 3) movimenti lungo l'asse sagittale (dal corpo e verso il corpo). Per i movimenti circolari e spirali la direzione si distingue in 'senso orario' o 'senso antiorario'. Infine, c'è l'aspetto delle *qualità del movimento* che ingloba ulteriori caratteristiche che sono: 1) size (ridotto o ampio); 2) speed (decelerato o accelerato); 3) flusso del movimento (accentuato). L'aspetto qualitativo del movimento dà conto della sua marcatezza che diventa rilevante solamente in relazione ad altri movimenti. Infine, il quarto parametro è la *posizione*. Il sistema di annotazione di Bressemer si basa sul concetto di spazio gestuale di McNeill (1992) utilizzando le categorie usate dall'autore attraverso il metodo dei quadrati concentrici (Bressemer, 2013).

Questo metodo incentrato sull'analisi della forma del gesto nasce all'interno di un gruppo di ricerca guidato da Cornelia Müller. Come vedremo di seguito, la studiosa ha tracciato i lineamenti di una grammatica dei gesti, estendendo il senso di grammatica alla sistematicità delle forme e delle strutture dei gesti senza implicare una vera e propria grammaticalità gestuale.

3.1 Per una grammatica dei gesti: l'analisi dei metodi di significazione

L'approccio allo studio della gestualità da parte di Cornelia Müller pone al centro la forma del gesto: l'articolazione della forma, il movimento, la posizione e l'orientamento delle mani, delle dita e delle braccia sono dotate di significato. Ciò vuol dire che quando si osserva uno sforzo articolatorio delle mani, questo può essere percepito come uno sforzo comunicativo.

Il punto di partenza di questo approccio è da rintracciare nel concetto espresso da Adam Kendon di *deliberate of expressiveness*. Kendon (2004) ha dimostrato, attraverso un esperimento, che un gruppo di persone che osserva dei parlanti senza sentire ciò che si dicono, riesce a distinguere i movimenti comunicativi da quelli non comunicativi. Questi movimenti vengono percepiti come *intenzionali, consapevoli, e governati da un'intenzione comunicativa (ibidem)*. Gli osservatori hanno percepito questi gesti come parte di ciò che il parlante provava a dire, così per Kendon: «l'espressività intenzionale del movimento è risultato essere un movimento che ha un netto confine di inizio e di conclusione, piuttosto che risultare da qualsiasi cambio sostanziale di posizione». Inoltre, «se i movimenti sono realizzati in modo tale che abbiano una certa caratteristica dinamica, saranno percepiti come figure piuttosto che come parte di un altro movimento, e come movimento sarà visto come pienamente intenzionale e intenzionalmente comunicativo» (Kendon, 2004).

L'analisi gestuale della Müller, come già detto, si rifà alla supremazia della forma kendoniana ma ne differisce riguardo alla considerazione del significato che, a differenza di Kendon, non viene visto come una *visual action* ma come *form of embodied conceptualization*. Seguendo così il principio di Langacker che vuole che il significato risieda nella concettualizzazione (Müller et al).

Le ricerche di Ladewig (2012) hanno dimostrato che i gesti co-verbali possono sostituire unità del parlato e diventare parte della struttura sintattica di un enunciato. Queste evidenze sostengono la teoria di Kendon secondo cui i gesti e il parlato sono due lati di un unico processo di enunciazione (Kendon 1980). Müller e i suoi, nel completare questa teoria

con una prospettiva cognitivista (*embodied conceptualization*) sulla costruzione del significato dei gesti, si dimostrano debitori della teoria dei gesti di McNeill che vede la gestualità come la forma visibile del pensiero figurativo²⁹.

Come abbiamo visto precedentemente, i lavori di Kendon hanno dato un forte contributo anche allo studio delle relazioni sistematiche tra le diverse forme gestuali dotate di significato, facendo emergere delle comunanze e teorizzando l'esistenza delle *gesture families*. Questa ipotesi si basa sull'assunto che le forme dei gesti sono dotate di significato e che una *gesture family* si sviluppi intorno a un nucleo semantico o a un tema semantico incorporati in un particolare aspetto della forma. Così Müller insieme al suo gruppo di ricerca, ha indagato le modalità e le caratteristiche delle forme che stanno alla base del significato gestuale. Da queste premesse hanno portato avanti due diversi progetti: il "The Berlin Gesture project" e il progetto interdisciplinare "Towards a Grammar of Gesture: Evolution, Brain and Linguistic Structure" (ToGoG). Il termine "grammar of gesture" si riferisce alle proprietà delle forme dei gesti e della loro struttura e non implica che i gesti co-verbali abbiano qualcosa di grammaticalmente strutturato.

Le ricerche si sono tradotte in indagini sui principi di creazione del significato riscontrabili nei gesti e si basa su tre fondamentali presupposti:

- Il significato dei gesti è motivato;
- Le loro forme incarnano il significato in modo dinamico e appropriato;
- Le azioni manuali sono il nucleo base della creazione del significato dei gesti.

Le mani in movimento sono oggetti naturali e il loro habitat è lo spazio, e i movimenti e le azioni caratterizzano il loro agire e il loro essere nel

²⁹ In inglese *imagistic thought*.

mondo. Quando vengono impiegate nella gesticolazione, nella quasi totalità delle volte rappresentano o mimano un'azione o un oggetto tangibile e visibile nello spazio. Nel processo di formazione dei segni gestuali le mani vengono sottoposte a un processo di trasformazione che si basa sull'uso di un numero definito di tecniche basilari. Le modalità di rappresentazione basilari riconosciute dalla Müller sono quattro: *the hand acts; the hand molds; the draws; the hand represents*.

Nella modalità *acting* viene utilizzata la strategia del mimo. Alcuni esempi possono essere: afferrare qualcosa, la mano nell'atto di dare o ricevere qualcosa, aprire una finestra, accendere o spegnere un radiatore. Nella modalità *molding* le mani sembrano modellare una scultura transitoria, come ad esempio una ciotola. Nella modalità *drawing* le mani tracciano il contorno di un oggetto o la traiettoria di un movimento nello spazio. In questi casi i significati dei gesti sono motivati attraverso la ricreazione di un'azione, ma ciò non accade per la modalità *representing*. In questo caso le mani incarnano un oggetto come un tutto e diventano loro stesse un tipo di scultura manuale. Ad esempio, l'indice di una mano utilizzato per rappresentare la penna o la mano aperta che diventa un pezzo di carta su cui è scritto qualcosa. Mentre nelle prime tre modalità di rappresentazione la trasmissione del significato è affidata all'esperienza tattile e motoria dell'azione delle mani, nella quarta modalità probabilmente si tratta solamente di percezione visiva. Di conseguenza, Müller e i suoi, affermano che la derivazione da una particolare azione o una specifica "manipolazione" di oggetti non è data a priori ma, al contrario, i parlanti spesso usano diverse tecniche per creare gesti che si riferiscono allo stesso oggetto. La scelta della modalità da utilizzare nel creare un gesto è data dal parlante e dai suoi obiettivi comunicativi. Queste tecniche di rappresentazione implicano uno specifico orientamento dei parlanti verso specifiche sfaccettature di cose azioni percepite e concepite nel mondo. Da questo punto di vista, quindi, i gesti sono illusioni naturali e artefatti della realtà, creati dai parlanti nel flusso dell'interazione e sono

probabilmente i primi artifici mimetici che appaiono nel percorso dell'evoluzione umana (Müller 2013).

Nel riconsiderare queste quattro modalità di rappresentazione da un punto di vista prassiologico (facendo riferimento all'approccio di Streeck 2009), si può provare che, ad ogni modo, ci sono solamente due tecniche fondamentali: *acting* e *representing*. Nella prima modalità le mani si muovono come se realizzassero un'azione attuale, nella seconda modalità invece le mani si muovono come se loro stesse fossero gli oggetti che rappresentano. In particolare, entrambe le modalità implicano due processi cognitivi-semiotici: l'iconicità e l'indessicalità. Nell'*acting mode* l'indessicalità è il processo cognitivo-semanticale primario che media tra la forma gestuale e gli oggetti percepiti nel mondo, mentre nel *representig mode* è l'iconicità. Ciò farebbe riferimento all'indessicalità come processo cognitivo antecedente all'iconicità nella storia dell'evoluzione.

Müller e i suoi sostengono che non esistono relazioni automatiche o naturali tra gli oggetti e gli eventi che i parlanti percepiscono nel modo e la loro rappresentazione gestuale. I parlanti scelgono le modalità di rappresentazione o il mezzo mimetico e di conseguenza strutturano le concettualizzazioni astratte incarnate dalle mani che gesticolano.

Gli accademici suggeriscono che quando i parlanti impiegano maggiori dettagli nell'uso della forma delle mani, della postura e del viso è come se ricadessero in esperienze pienamente corporee avvenute in ambientazioni ricche. Mentre quando i gesti sono semanticamente ridotti e nessun'altra parte del corpo è coinvolta, i gesti rappresentano situazioni prototipiche e la loro esperienza corporea associata. Per esempio, alcuni parlanti rappresenteranno un pezzo di gomma da masticare appiccicoso esattamente sul punto dove è rimasto incollato, mentre altri eseguiranno semplicemente un movimento dell'indice e del pollice che illustra solo la natura e la qualità della viscosità. Così la Müller afferma che l'analisi da lei condotta degli elementi di formazione dei gesti, dei tipi e del numero degli articolatori coinvolti nella gesticolazione rivela se i soggetti

concettualizzano un evento specifico come percepito individualmente (gesto relativo all'oggetto) o come uno che è prototipicamente associato al significato di una particolare parola (gesti correlati all'interpretazione). Questa variazione può essere spiegata dal differire, nei diversi parlanti, del focus su cui concentrano la loro attenzione.

L'indagine sui gesti incentrata sulla forma, oltre a studiare le modalità adottate per la creazione del significato si è spinta a ricercare anche le funzioni linguistiche espresse e che riportiamo di seguito.

3.2 Le funzioni linguistiche dei gesti

Cornelia Müller (2013b) parte dal modello di Bühler per individuare e spiegare quali funzioni linguistiche vengono espresse dalla gestualità e come questo avviene.

Karl Bühler, psicologo tedesco del filone Gestalt, diede un importante contributo agli studi sullo sviluppo del linguaggio formulando l'ipotesi del *modello strumentale*, utilizzato da Müller.

Bühler per la sua teoria fa riferimento all'idea platonica del linguaggio, visto come uno strumento utilizzato da un parlante per comunicare qualcosa circa una cosa a un ascoltatore. La lingua, quindi, funziona da strumento per la creazione di una relazione tra tre fondamentali elementi degli eventi comunicativi: il parlante, il destinatario e l'oggetto. Collocando la lingua nella struttura base di ogni evento comunicativo stabilisce tre principali funzioni dell'enunciazione linguistica: la rappresentazione, l'espressione, e l'appello. Queste tre funzioni vengono concepite come fondamentali per la lingua e sono co-presenti in ogni evento comunicativo. I gesti realizzano pienamente queste tre funzioni e possono anche compierli simultaneamente; ciò è quello che rende la gestualità adatto alla lingua.

I gesti possono svolgere una funzione rappresentazionale illustrando azioni concrete o concetti metaforici. Nel primo basta pensare a qualcuno che sta raccontando di aver disfatto un bagaglio e nel parlarne gesticola

quasi come se stesse ripetendo l'azione, ma se un gesto simile viene realizzato per far riferimento a un segreto svelato ecco che la rappresentazione diviene metaforica. I gesti vengono utilizzati anche per rappresentare un percorso assumendo sia un valore concreto, in riferimento a un tragitto percorso, che metaforico in riferimento all'andamento di una relazione. La rappresentazione gestuale viene utilizzata dai parlanti anche per illustrare l'idea del tempo come spazio.

La funzione dell'espressione è rintracciabile nel sollevamento delle mani e delle braccia per esprimere gioia e vittoria, nell'azione di coprirsi il volto per esprimere tristezza o malessere o nel movimento del pugno per esprimere rabbia. Tutti questi gesti vengono messi in atto anche attraverso delle particolari espressioni facciali e possono arrivare a coinvolgere l'intero corpo (basti pensare all'esultanza dei calciatori dopo un goal).

Mentre i gesti che esprimono principalmente la funzione di appello sono utilizzati per regolare il comportamento degli altri in una conversazione. Per esempio, per riportare l'ordine quando si parla con più di un interlocutore.

Ma il modello di Bühler non riesce a inglobare tutti i tipi di gesti, in quanto modello principalmente semiotico. Diventa così un modello applicabile esclusivamente a uno scambio comunicativo ideale, non considerando la lingua come azione e perdendo, di conseguenza, l'aspetto pragmatico. Ci sono gesti la cui funzione primaria è quella di eseguire atti linguistici; sono gesti assimilabili ai verbi azionali per esempio il verbo *giurare*; infatti, quando qualcuno *giura* è solito enfatizzare e realizzare attraverso il gesto della mano alzata a palmo aperto. I gesti performativi sono molto comuni e possono anche convenzionalizzarsi come emblemi.

4. Il corpo nell'interazione: l'approccio conversazionale

Sin dagli anni Settanta l'analisi conversazionale iniziò ad approcciarsi alla lingua come fenomeno multimodale. Nel 1975, infatti, Emanuel Schegloff presentò uno studio all'American Anthropological Association sull'organizzazione strutturale di una azione corporea. Questo studio, frutto di un lavoro condiviso con Harvey Sacks, era nato in seno a una serie di conversazioni e di confronti avvenuti tra Charles e Marjorie Goodwins sull'analisi di conversazioni video registrate (Schegloff, 2002). Nata negli anni Sessanta, la Conversation Analysis, basa i propri studi sull'assunto che le società umane hanno come risorsa centrale l'organizzazione di interazioni formalizzata dall'uso del linguaggio. Questa organizzazione comprende diverse sotto-organizzazioni che riguardano le contingenze con cui i partecipanti a una interazione devono fare i conti, e sono: a) il problema della presa del turno (cioè la necessità di prendere o dare il turno uno alla volta in successione, un bisogno che è intrinsecamente legato alla linearità del discorso e alla sistematica inabilità degli umani di produrre e comprendere il discorso degli altri contemporaneamente), b) il problema della formazione delle azioni (cioè come le risorse del linguaggio sono modellate in conformazioni progettate per realizzare determinate azioni), c) il problema dell'organizzazione della sequenza (che è la necessità di disegnare i turni dell'interazione così che siano riconoscibilmente reattivi ad azioni precedenti e/o mostrino quali azioni di risposta cercano di suscitare) (Schegloff, 2007). In breve, la conversation analysis si avvicina all'interazione come azione sociale contribuendo allo studio dello scambio comunicativo attraverso l'indagine sulle modalità organizzative delle diverse risorse linguistiche per la realizzazione dell'azione sociale e per la costruzione dell'ordine interazionale, realizzando in questo modo l'intellegibilità pubblica dell'azione.

All'interno del quadro teorico dell'analisi conversazionale sono di fondamentale importanza la nozione di risorse e quella di consequenzialità. Con il termine risorsa si fa riferimento a tutti gli strumenti che un parlante può utilizzare per interagire, e quindi il linguaggio, lo sguardo, i gesti, la postura, i movimenti della testa, l'espressione facciale, la manipolazione degli oggetti, la tecnologia e la prossemica. La nozione di risorsa comprende una vasta gamma di forme che vanno da quelle più convenzionali, come la grammatica, a quelle meno standardizzate che sono usate dai partecipanti a una interazione per costruire una responsabilità intersoggettiva delle loro azioni. L'uso di una risorsa, da un punto di vista etnometodologico, è definito riflessivo; si riferisce al fatto che l'azione riflette le circostanze della sua organizzazione e realizzazione, sia adattandosi che trasformandole. Da ciò deriva che il significato e il valore di un'azione dipende dal contesto ma al tempo stesso è la risorsa che può costruire una precisa interpretazione di un determinato contesto.

La necessità dell'approccio conversazionalista di allargare l'analisi del linguaggio alla multimodalità non può fare a meno di tenere in considerazione la concettualizzazione di temporalità e di sequenzialità propri dell'analisi conversazionale e della linguistica interazionale. Questi due approcci hanno contribuito a rivedere il concetto di tempo in relazione al linguaggio andando ad analizzare la lingua nella sua realizzazione e nel contesto d'uso in tempo reale. Di conseguenza, un punto di vista di questo tipo che enfatizza l'analisi linguistica dei fenomeni nel loro realizzarsi, conduce a vedere la struttura linguistica in modo incrementale ed emergente. In questo modo, la costruzione dei turni (che rappresentano le unità di riferimento della Conversational Analysis, cfr. Capitolo 1) è frutto dell'iterazione tra i partecipanti, momento dopo momento, lungo il dispiegarsi dell'interazione. La formazione e l'organizzazione dei turni avviene attraverso una serie ordinata di posizioni sequenziali che vanno dal turno pre-inizio o iniziale, fino alle posizioni di pre-completamento, completamento e post-completamento (Schegloff, 1996). L'organizzazione

sequenziale è dunque frutto dell'interazione tra parlanti, che costruiscono così unità di discorso che sono dinamiche, negoziate e collettivamente formate (Mondada, 2013). Ciascuna delle risorse utilizzate costituisce l'organizzazione sequenziale embodied dei turni, inglobando sia gesti ma anche lo sguardo del co-partecipante verso il parlante. Alcune ricerche hanno dato un importante contributo su questo fronte, cercando di indagare il ruolo delle diverse risorse nella costruzione dell'interazione. Schegloff (1984) ha osservato come la prefigurazione, da parte dei gesti, del significato espresso dal parlato permette al destinatario di costruire la comprensione del turno step by step, venendo a creare uno spazio di proiezione che gli permette di rispondere per tempo e di agire in modo pertinente man mano che il turno si realizza. Goodwin (1981), ha dimostrato che la progressività dell'interazione può essere rallentata, il parlante può rallentare lo svolgersi del turno - attraverso la strategia del re-start - quando si accorge che il destinatario non gli sta rivolgendo la propria attenzione attraverso lo sguardo. Lo strumento del re-start, all'interno del turno, agisce creando un rallentamento - attraverso la sospensione della progressione - creando di conseguenza un richiamo di attenzione nei confronti del co-partecipante. Streek (1993) individua uno schema gesto-sguardo che consiste nel fatto che il parlante osserva il proprio gesto prima della parola che porta l'informazione chiave, e restituisce lo sguardo all'ascoltatore una volta che la parola chiave è stata pronunciata.

Di fronte all'esigenza, derivante dai primi studi dell'interazione, di superare lo studio delle relazioni tra le singole modalità e il discorso coordinato, molti studiosi si sono dedicati a indagare l'interazione in tutte le sue modalità e interezza. Di conseguenza si sono delineati campi di approfondimenti sulle risorse multimodali all'interno dell'interazione sociale in termini di materialità, spazialità e mobilità.

Approcciarsi allo studio delle interazioni in contesto spontaneo, così come fa l'analisi conversazionale, tenendo in considerazione tutte le risorse utilizzate dai partecipanti per interagire tra loro, richiede una forte

attenzione agli aspetti che caratterizzano l'ambiente in cui avviene l'interazione. In questo senso, lo studio di una interazione in termini di materialità pone la propria attenzione agli oggetti; in questo modo vengono indagati non soltanto le azioni delle mani sugli oggetti ma anche il modo in cui queste azioni si legano all'ambiente e il focus attenzionale creato tra i partecipanti dagli oggetti manipolati (Mondada, 2013). Goodwin (2003) evidenzia la necessità di tenere in considerazione l'uso integrato del linguaggio, del corpo e degli oggetti per riuscire ad analizzare un turno. L'attenzione ai corpi e agli oggetti comporta, di conseguenza, un interesse all'ambiente e alle modalità messe in atto dai comportamenti corporei che adattano, impiegano e trasformano le caratteristiche materiali che circondano l'azione. Da questo punto di vista, lo spazio non è visto come un mero ambiente preesistente ma diventa una configurazione creata dall'orientamento dei partecipanti e dalla loro disposizione nell'ambiente dell'interazione (Mondada, 2013). Nei primi studi sulla spazialità nell'interazione Goffman (1963) ha mostrato come i corpi si dispongono in modo tale da creare dei territori temporanei, che sono riconosciuti sia dai partecipanti all'interazione che dagli astanti, che materializzano il sistema fisico entro cui si realizza l'attività e che sono definiti dall'orientamento dei corpi (che viene definito mutuale) e dall'attenzione condivisa espressa dalle posizioni dei corpi, dagli sguardi, dalle posture e dai gesti. Più recentemente, Goodwin e Mondada hanno indagato la relazione tra le caratteristiche dell'ambiente fisico e l'uso dei corpi. Goodwin parla di "configurazione contestuale" partendo dall'assunto che lo studio dei gesti o delle posture non può essere sviluppato in isolamento ma deve descrivere il modo in cui le strutture e le caratteristiche dell'ambiente contribuiscono all'organizzazione dell'interazione (Goodwin, 2007). Della stessa opinione è la Mondada che parla di "spazio interazionale", come di uno spazio costruito attraverso le disposizioni dei corpi situate, reciprocamente regolate dei partecipanti nello spazio. Questo fa in modo che si crei una configurazione dello spazio che sia pertinente all'attività in cui i

partecipanti sono coinvolti, la loro mutua attenzione, il loro comune focus attenzionale e la manipolazione degli oggetti presenti nello spazio rispettivamente all'azione svolta (Mondada, 2011). Di conseguenza, dall'interesse sul ruolo della spazialità nell'interazione, è scaturito l'interesse per la mobilità dei corpi. Dall'osservazione di interazioni statiche si è passati allo studio delle interazioni in movimento, contesti come una passeggiata o interazioni durante viaggi in macchina o in aereo (Mondada, 2013).

Difronte alle questioni e agli interessi sviluppati all'interno dell'approccio dell'Analisi Conversazionale, Mondada (2016) ha proposto una convenzione per la trascrizione multimodale che permetta di annotare tutte le possibili azioni embodied rilevanti che possono occorrere all'interno dell'interazione sia in concomitanza del parlato che durante una pausa (come i gesti, gli sguardi, la postura del corpo, i movimenti). La convenzione mira all'economicità - cercando di evitare annotazioni ridondanti - e alla coerenza, riuscendo a restituire un sistema di trascrizione univocamente interpretabile. I principi su cui si basa la convenzione sono due: a) la descrizione della traiettoria temporale: ciascuna azione embodied è temporalmente localizzata all'interno del corso dell'attività multimodale ed è delimitata da due simboli che segnano l'inizio a sinistra e la sua conclusione sulla destra; b) descrizione dell'azione embodied: ogni azione è brevemente descritta.

Sulla base delle indicazioni della proposta di trascrizione, i partecipanti vengono identificati sia da uno pseudonimo che da un simbolo utilizzato per lo stesso partecipante in tutta la trascrizione. Nel caso in cui un partecipante dovesse realizzare diverse azioni embodied, verranno utilizzati simboli diversi per ciascuna risorsa usata.

Come accennato in precedenza, la traiettoria temporale delle azioni è delimitata da due simboli identici uno indicante l'inizio e l'altro indicante la fine. Questi simboli sono inseriti nella linea del parlato, così da permettere

1 BRII la maison qui se retrouve ici: A
Figura 4 Esempio riportato da Mondada 2016

una sincronizzazione tra il segmento parlato e quello azionale. La descrizione dell'azione è inserita tra questi due simboli. Se l'azione continua anche nella linea successiva o qualche linea dopo, la descrizione è seguita da una freccia che indica la direzione verso cui trovare il successivo simbolo che ne delimita la conclusione.

Rispetto al dispiegamento temporale dell'azione, vengono descritte tre fasi: a) fase preparatorio (per annotare questa fase vengono usati i puntini

```

1 LAU      +°ah oui°
  viv      +points---->
2          (0.5)
3 VIV      mais alors+ i- il y va pour les dé- pour les sortir
           -->+

```

Figura 5 Esempio riportato da Mondada 2016

“...”), b) la forma riconoscibile dell'azione (questa fase è annotata attraverso la descrizione dell'azione) e c) il ritiro dell'azione (l'ultima fase è annotata con una serie di virgole “,,,”).

```

1 VIV      sur des terres assez *euh ass- a: *- assez *bonnes*
  lau      *.....*points---* ,,,,,,*

```

Figura 6 Esempio riportato da Mondada 2016

Infine, rispetto alla descrizione dell'azione la Mondada (2016) suggerisce di evitare descrizioni fisiche e descrizioni della presunta intenzione dell'azione. È preferibile una descrizione breve che corrisponda con lo spazio della trascrizione, che risulti dunque leggibili e comprensibile. La scelta di quanto dettagliata debba essere una descrizione è lasciata al livello granularità analitica che si vuole adottare.

5. Il corpo come substrato: l'approccio kinesiologico

L'approccio kinesiologico (dal greco *kinēsis*, movimento e *logos*, discorso) è un quadro teorico e un metodo di analisi del gesto ideato da Dominique Boutet e che al momento rappresenta un unicum in tutto il panorama degli *gesture studies*. L'originalità dell'approccio di Boutet scaturisce dal vedere la gestualità non come un'appendice del parlato, di conseguenza, riconosce un ruolo imprescindibile alla materialità dei gesti, che è il corpo. Infatti, in questo quadro teorico la gestualità viene analizzata come elemento comunicativo plasmato dalla fisiologia corporea.

«Le sens de nos productions gestuelles est le fruit de notre corps, tel qu'il peut être naturellement articulé, empreint qu'il est de nos expériences passées, de sa mise en mouvement par rapport à l'espace dans lequel il se situe et de nos relations avec les autres. Il y a dans nos gestes une hybridation entre ce qui est universel, générique à travers la mise en mouvement du corps humain tel qu'il est articulé, et ce qui appartient à chaque individu, son développement, et son histoire» (Boutet – Morgenstern 2020)

Lo studio della gestualità si basa sui movimenti del corpo umano analizzati da un punto di vista biomeccanico e articolatorio, in questo modo i significati espressi dalla gestualità sono frutto del corpo. Così, se l'approccio semiotico non valorizza il corpo vedendolo esclusivamente come una modalità di espressione, l'approccio kinesiologico (come in una prospettiva embodied) si basa sul significante, sul corpo, sui suoi segmenti e su tutti i vincoli che creano il significato e ne rivelano la strutturazione. Questo approccio cerca di analizzare tutti gli elementi fisici del corpo che rendono possibile il movimento e che, al tempo stesso, rappresentano anche un limite cioè tutti quei limiti e quelle potenzialità che influenzano inevitabilmente la nostra espressività.

L'approccio kinesiologico mira, dunque, ad analizzare la lingua nel suo ambiente al fine di analizzare i suoi piani azionali e le sue funzioni simboliche all'interno di uno stesso movimento, studiando i corpi sia come substrato che come supporto del languaging, cioè l'uso multisemiotico del linguaggio. Il concetto di languaging, per il quale questo approccio accoglie la definizione di Linell (2009) («linguistics actions and activities in actual communication and thinking» al fine di realizzare l'atto di enunciazione multimodale del linguaggio nell'interazione) che permette di inscrivere l'approccio kinesiologico all'interno di un quadro teorico che vede il linguaggio come un'azione embodied piuttosto che considerarlo come un codice o un sistema simbolico.

Solitamente, l'analisi dei gesti co-verbali - che vengono definiti espressivi e interazionali - rischia di essere influenzata dal parlato che co-occorre. In molte categorizzazioni dei gesti è riscontrabile sia un'estensione delle categorie del parlato ai gesti che l'influenza dei valori semantici e delle funzioni espresse dal parlato nell'analisi dei gesti co-occorrenti e alla luce di ciò, l'approccio di Boutet enfatizza l'importanza dell'indipendenza dell'analisi gestuale dal flusso verbale. La gestualità, secondo il quadro teorico in analisi, è compositiva ma non in riferimento alla sua co-articolazione con il parlato e nemmeno al possibile ruolo svolto dall'espressione facciale e dallo sguardo. Il gesto è compositiva in quanto prodotto dagli arti superiori e dunque potenzialmente composto da movimenti delle spalle, delle braccia, degli avambracci e delle dita (Boutet - Morgenstern 2020). Si mostra quindi in modo chiaro la necessità di indagare l'organizzazione complessa del gesto che opera su un canale comunicativo multilineare del significato (Boutet, 2010), e al fine di comprendere ciò è necessario partire dall'assunto che il corpo non porta solamente la gestualità ma lo *informa*. Così il corpo non è un supporto della gestualità ma un substrato (Boutet & Cruxac, 2008 - Boutet, 2010).

Sulla base dell'indipendenza della gestualità dal parlato e dalla visione del corpo come substrato, Boutet invita a rivedere il Kendon's continuum

di McNeill (1992) e a invertilo di conseguenza. Infatti, alla luce dell'approccio kinesiologico, i gesti iconici e i batonici (considerati poco convenzionali e che appaiono visivamente informali o idiosincratichi) vengono considerati come schizzi di emblemi la cui produzione e gestione, scandita del flusso discorsivo, limita la loro piena realizzazione (Boutet 2010). Ci sarebbe, dunque, un vero e proprio rapporto formale tra i gesti maggiormente convenzionalizzati e quelli meno convenzionalizzati sulla base della loro struttura fisiologica, nonostante le differenze significative a livello visivo dovute alle parti del corpo coinvolte nel movimento. «Nos filtres culturels et notre appréhension restreinte à une sensorialité visuelle ont donc limité non analyses de la gestualité» (Boutet – Morgenstern, 2020).

Per l'approccio Kinesiologico, la descrizione e l'analisi della componente gestuale delle costruzioni multimodale si basano sulla forma dei gesti in modo da poter cogliere le differenze formali da cui scaturiscono le differenze di significato. In questo modo, si può parlare di un vero e proprio approccio formale che parte dall'idea che «les modalités de production semblent avoir été quelque peu oubliées dans l'analyse des phénomènes de la gestualité symbolique. Il semblerait que le corps soit considéré comme un lieu où les choses arrivent, où les mouvements apparaissent sans qu'aucune matérialité ne lui soit prêtée» (Boutet, 2008). La materialità del gesto riesce ad assumere la forma del nostro ambiente, dei nostri strumenti, degli oggetti e degli spazi che abitiamo e – secondo Boutet (2008) – la strutturazione di prodotti gestuali è strettamente legata alla gestualità prassica in piena continuità con i gesti simbolici. Infatti, referenti non sono direttamente associati ai gesti attraverso la capacità umana di realizzare analogie, ma il loro significato è modellato dai gesti prodotti e dalle loro dinamiche per questa ragione le unità gestuali vengono descritte, da questo approccio, sulla base delle loro caratteristiche formali e dei vincoli fisiologici piuttosto che sulla base della loro iconicità (Boutet – Morgersten 2020). Di conseguenza, Boutet propone un'analisi che si basi sulla modalità propriocettiva, piuttosto che su quella visuale, al fine di indagare la

gestualità come una fonologia articolare realizzata e costituita dalla fisiologia corporea.

La fonologia corporea si basa sull'analisi dei movimenti delle parti del corpo e sui vincoli che ne limitano e ne costituiscono la stessa mobilità. I concetti e gli elementi di base che diventano anche elementi di analisi sono i segmenti, i gradi di libertà e il flusso. Con segmento si definisce un'unità anatomica non deformabile situata tra due articolazioni, e si distinguono le spalle, le braccia, gli avambracci, le mani e le falangi. Le articolazioni che delimitano e definiscono i segmenti sono anche le parti fisiologiche che determinano il grado di libertà.

Un grado di libertà è definito come «on définit un degré de liberté comme le mouvement relative indépendant d'un segment par rapport à son homologue adjacent et plus proximal. Concrètement un degré de liberté répond à la rotation d'un segment autour d'un axe par rapport à un autre» (Boutet, 2018). Il grado di libertà dipende quindi dalla rotazione del segmento intorno all'asse in relazione anche ad altri assi, di conseguenza ogni grado di libertà è definito dalle possibili rotazioni di un asse. Ciò vuol dire che il movimento della mano è analizzata in rapporto alla posizione dell'avambraccio, in questo modo è possibile misurare il complesso dei movimenti di tutti i segmenti da un punto di vista dell'articolazione superando, in questo modo, la definizione di spazio gestuale di McNeil (1992). La definizione mcneilliana di spazio gestuale risulta, dal punto di vista kinesiologico, un riferimento tridimensionale egocentrico in cui i gesti vengono analizzati in modo olistico impedendo di fatto un'analisi dettagliata dei movimenti dei singoli segmenti coinvolti limitandone di conseguenza l'indagine dei vincoli che determinano le variazioni (Boutet – Morgersten, 2020). Si tratta dunque di un concetto di spazio gestuale – quello di McNeill – che rimane relativo e statico e che costituisce una base impressionista (Boutet, 2008).

In tutto ci sono ventotto gradi di libertà, ciascuno dei quali è unico, che determinano nel loro insieme tutti i movimenti manuali. Per esempio,

l'articolazione della spalla include tre gradi di libertà: abduzione/adduzione, flessione/estensione, rotazione interiore/esteriore. L'articolazione del gomito ha due gradi di libertà flessione/estensione e pronazione/supinazione.

Il gesto viene dunque analizzato come movimento che si propaga da un segmento all'altro e che viene descritto in riferimento al tronco distinguendo così parti del corpo prossimali (vicini al tronco) e parti del corpo distali (lontani dal tronco). La natura dei movimenti dei segmenti, distali o prossimali, permea i gesti di significato radicato nell'esperienza (Boutet - Morgersten, 2020). Il movimento può essere costituito da un movimento proprio del segmento o dallo spostamento di un segmento ma per un movimento non proprio. Nell'ottica della mobilità delle articolazioni e dei vincoli tra segmenti, quando si genera un movimento inevitabilmente si propaga ai segmenti limitrofi. Ma il significato è da ricercare nel segmento che realizza il movimento proprio (Boutet, 2018). In questo modo si distinguono due tipi di flusso del movimento: se si propaga dalla spalla alla fine delle dita, il flusso è chiamato prossimale-distale, viceversa è chiamato distale-proximale. Ma non tutti i movimenti del corpo implicano la presenza di un flusso; infatti, è possibile che si realizzi un gesto localizzato in una parte del corpo.

«La différence entre certains gestes relève uniquement du flux. Ils constituent ce qui semble le plus proche d'une paire minimale puisqu'en-deçà d'une différence d'orientation ou de configuration, la seule différence réside dans l'ordre de déroulement des mêmes traits. Le flux montre l'importance du substrat qui déroule le mouvement» (Boutet, 2018)

La centralità e l'importanza riconosciuta al movimento, all'interno dell'approccio kinesiologico, dimostra come all'interno di questo quadro teorico le forme non vengono considerate esclusivamente nella loro somiglianza con il referente ma soprattutto su come realizzano e costruiscono fisiologicamente il significato.

I movimenti si realizzano in relazione alle possibilità fisiologiche degli arti superiori, così la traiettoria, la forma e il modo in cui un gesto può essere prodotto è inevitabilmente influenzato dalle relazioni geometriche tra i diversi segmenti. Questi vincoli, che possiamo definire come controllo motorio, seguono quattro principi: 1) la velocità in relazione alla forma, 2) la legge dell'opposizione delle fasi, 3) il principio di isocronia e 4) il paradosso di Codman.

Il primo principio dice che la velocità di un movimento dipende dalla forma tracciata; più la forma è curva più aumenta la velocità. Il secondo principio mira all'economicità nella gestione dei possibili gradi di libertà coinvolti nella realizzazione di un movimento: nella realizzazione di un gesto, quando l'angolo che si crea tra il braccio e il corpo aumenta, l'angolo tra braccio e avambraccio varia in egual misura. Questo controllo della doppia articolazione riduce al minimo i gradi di libertà, in questo modo il nostro cervello controllerà un solo parametro in un possibile movimento complesso. Il terzo principio stabilisce che la velocità media di un gesto tra due punti cresce secondo la distanza tra quei due punti. Due punti distanti saranno collegati con un maggiore velocità rispetto a due punti vicini. Infine, il quarto principio - il paradosso di Codman - rivela che quando vengono eseguiti dei movimenti si producono delle rotazioni di articolazioni in modo involontario. Cioè, nel realizzare un movimento volontario in un ciclo - che in questo caso sono la flessione e l'adduzione del braccio - si verrà a produrre un movimento involontario che in questo caso è la rotazione dell'avambraccio tale per cui alla fine del ciclo avremo il pollice rivolto all'indietro nonostante all'inizio fosse rivolto in avanti (Boutet - Morgerstern - Cienki, 2018).

Come abbiamo visto, l'approccio di Boutet, basato sulla strutturazione fisiologica del gesto, vede il significato come fatto formale che si realizza attraverso un'articolazione di movimenti. Attraverso la ripetizione si giunge alla stabilizzazione di alcune forme. Ma nel riconoscere il carattere materiale del gesto, Boutet suggerisce un'analisi che veda la semioticità

gestuale emergere sia dalla persona che realizza il gesto ma anche da chi la percepisce. Grazie a un approccio che combine analisi del corpo e giudizio percettivo, si è stato possibile analizzare e isolare i gesti con forme simili e indagare una rete di relazioni basate sulla fisiologia corporea spiegandone i loro legami semantici (Boutet – Morgerstern, 2020). Ne è un esempio l'analisi dell'espressione dell'aspetto grammaticale in francese fatta da Boutet, Morgerstern e Cienki (2016) che ha dimostrato come la perfezionalità è semanticamente correlata alla propagazione del flusso del movimento.

6. Conclusione

Nella complessità del fenomeno gestuale è possibile, come abbiamo visto, rintracciare diverse modalità di significazione del gesto e diverse funzioni nell'interazione. Gli studiosi, che nel tempo hanno indagato la natura e la funzione del gesto, hanno proposto categorizzazioni dei fenomeni gestuali differenti. La diversità di queste classificazioni, derivante dai diversi approcci e interessi scientifici da cui scaturiscono, sembra generalmente essere indirizzata al raggiungimento di due obiettivi: da un lato la categorizzazione delle funzioni – come fa Kendon (2004) – e dall'altro la classificazione tipologica su base semiotica (Wundt 1973).

Il pionieristico studio di David Efron (1971) e la prospettiva di analisi adottata fa emergere in modo chiaro come l'azione comunicativa del gesto si realizzi su diversi livelli che racchiudono l'aspetto fisico, formale, temporale, interattivo e linguistico-funzionale. Pur nella sua limitatezza, dovuta alla natura pionieristica del lavoro, il saggio di Efron sembra anticipare tutti quegli aspetti che sono stati approfonditi dai quadri teorici successivi.

La fisicità del gesto, caratteristica costitutiva del fenomeno, è stata utilizzata sia dagli studi cognitivisti (Cienki, 2013) per analizzare la concettualizzazione del significato che dall'approccio semiotico-linguistico tracciato da Bressemer (2013) e Müller (2013). Questi due approcci, che

possiamo senza dubbio definire contigui, danno centralità alla forma del gesto analizzandola negli aspetti formazionali e semiotici, riconoscendone, di conseguenza, centralità nella funzione comunicativa. Mentre l'approccio kinesiological di Boutet si appropria alla fisicità del gesto come substrato, analizzandone gli aspetti articolatori nei loro aspetti vincolanti e qualitativi; così le caratteristiche fisiche del corpo trovano centralità nell'analisi funzionale e grammaticale dei gesti. La centralità e l'importanza riconosciuta al movimento, all'interno dell'approccio kinesiological, dimostra come all'interno di questo quadro teorico le forme non vengono considerate esclusivamente nella loro somiglianza con il referente ma nelle modalità di realizzazione e costruzione fisiologica del significato. L'approccio conversazionale, invece, riconosce centralità all'aspetto interazionale dei gesti indagando le modalità con cui i parlanti utilizzano braccia, sguardi, testa e postura al fine di creare uno spazio interazionale condiviso e costituire uno scambio ancorato ai significati verbale e al contesto enunciativo.

In sintesi, possiamo affermare che come la lingua, nella sua modalità verbale o scritta, articola e costruisce le informazioni e i contenuti semantici attraverso i diversi livelli di analisi (segmentali e soprasegmentali), allo stesso modo avviene per il gesto. Ma se il sistema lingua, da un punto di vista segmentale, organizza significati e funzioni attraverso la doppia articolazione e la linearità sintagmatica, il gesto utilizza stratagemmi ed espedienti intrinsecamente funzionali alla sua fisicità. In questo modo, la codifica del significato risulta dipendente sia dalla forma - che costituisce un livello di analisi assimilabile a quello lessicale - che alla qualità del movimento, che appare più vicino al livello pragmatico-illocutivo. Tutto ciò fa emergere la necessità di concepire un metodo di trascrizione e di analisi che cerchi da un lato di svincolarsi da qualsiasi classificazione tipologica, funzionale o semiotica e dall'altro di valorizzare il gesto nei suoi aspetti fisici, riconducibili alla forma e al movimento. Ciò vuol dire che è necessario un metodo di trascrizione e di analisi che possa individuare parametri

oggettivi formali cercando di limitare, per quanto possibile, l'interpretazione dei dati nella fase di trascrizione.

Questa nostra prospettiva non ci porta a vedere e a concepire la gestualità come un sistema di lingua parallelo alla modalità verbale. Piuttosto, la lingua in questa prospettiva è e rimane un unico sistema che esprime le azioni linguistiche costruendole sia su più livelli di lingua che attraverso l'uso di modalità diverse. Così, dunque, il corpo diventa strumento di espressione linguistica che usa tutte le sue potenzialità articolatorie che vanno da quelle più usate, studiate e articolate della verbalità che quelle meno studiate e meno articolate che riguardano l'espressività corporea.

Questo lavoro cerca di rispondere alle necessità appena descritte attraverso la proposta di un metodo di trascrizione che permetta di analizzare gesto e parlato per mezzo di unità di analisi oggettivamente riconoscibili. A ciò si aggiunge l'ulteriore esigenza di individuare un'unità di livello superiore che possa aiutarci a comprendere come i parlanti agiscono in modo multimodale nella costruzione di significati e azioni su più livelli linguistici espressi sia per mezzo della modalità parlata che gestuale.

Nel capitolo successivo descriveremo nel dettaglio la costruzione del corpus pilota in tutti i suoi aspetti, dalla raccolta dati al metodo di trascrizione e annotazione. In modo particolare, cercheremo di rispondere alle questioni emerse in queste pagine attraverso un approccio pragmatico sulla base degli studi sul parlato spontaneo condotti in seno al gruppo di ricerca LABLITA di Firenze.

Capitolo 5 – Il corpus

1. Introduzione

La lingua parlata è intrinsecamente multimodale. Questa constatazione, che si evince dagli studi posti in rassegna nelle pagine precedenti, porta alla necessità di grandi corpora che – sull’orma di quanto accaduto con i corpora di parlato e di scritto – consentano di indagare la lingua come sistema multimodale con dati ecologici e quantitativamente consistenti.

Di fronte a questa necessità, il nostro lavoro cerca di contribuire al dibattito, intorno alle questioni legate alla costruzione dei grandi corpora, attraverso la realizzazione di un corpus pilota del parlato italiano spontaneo. Le principali questioni da affrontare sono principalmente due: da un lato la questione di metodo, che riguarda la modalità della raccolta dei dati, la loro trascrizione e il metodo di annotazione, e dall’altro il riconoscimento e la segmentazione di una macro-unità di analisi multimodale. Le questioni e le problematiche da affrontare sono molteplici e cercano di rispondere alle esigenze di uno strumento di ricerca che possa essere qualitativamente alto e largamente condivisibile dai diversi approcci e indirizzi di studio. Come sarà possibile vedere di seguito, la nostra proposta cerca di rispondere in parte a queste questioni, concentrandosi maggiormente sul metodo di costruzione e cercando di proporre un approccio nuovo al problema al fine di alimentare un dibattito che possa rivelarsi presto risolutivo.

2. Dai corpora *monomodali* ai corpora *multimodali*: caratteristiche, procedure e obiettivi

L'approccio allo studio della lingua attraverso i corpora ha garantito il raggiungimento di enormi risultati, nei diversi campi della disciplina linguistica, sulla base di analisi empiriche. Il metodo tipico di questo approccio, che si basa sulla collezione di dati organizzati, affonda le sue radici in una lunga tradizione che va dagli studi lessicografici e grammaticali del linguaggio, fino agli studi biblici passando per le indagini sull'apprendimento del linguaggio (Knight, 2011). I primi studi su corpora si basavano su raccolte costruite e organizzate manualmente e dunque caratterizzate da dimensioni ristrette e un'alta possibilità di errore. L'avvento dell'elettronica ha rappresentato una vera e propria rivoluzione con la continua crescita del numero dei corpora e delle loro dimensioni. Così, la seconda generazione di corpora ha riscritto il metodo di indagine linguistica su corpus dando ai linguisti la possibilità di creare nuovi corpora in modo sistematico e consentendone l'interrogazione digitale, a differenza della faticosa operazione di sfoglio manuale dei primi corpora. Il continuo sviluppo delle tecnologie digitali ha scaturito il conseguente sviluppo delle capacità dei ricercatori di raccogliere e gestire, in modo più facile e veloce, una sempre crescente quantità di dati (Kennedy, 1998) così che è di fatto impossibile riuscire a fare una recensione dei corpora esistenti. Ma lo sviluppo di tecnologie sempre più avanzate e accessibili ha facilitato lo sviluppo di nuovi campi di studio in relazioni alla possibilità di acquisire e organizzare nuovi dati, come le immagini e i video dando così vita agli studi sulla multimodalità.

La continua crescita di tipologie di corpora è legata, oltre che allo sviluppo di nuove tecnologie, anche ai diversi interessi di ricerca o alle diverse correnti di studio del linguaggio. Questo fa in modo che l'organizzazione dei dati e la raccolta siano influenzate dagli interessi di

studio dei ricercatori che compilano i corpora e dei loro quadri teorici di riferimento; di conseguenza diventa difficile poter stabilire degli standard da adottare per la creazione di un corpus. Ma questa differenziazione nei metodi di raccolta e organizzazione dei dati è in ogni caso vincolata alla selezione di un campione autentico di linguaggio, pertanto, la progettazione di un corpus è strettamente vincolata a pratiche non ambigue, rigorose e documentate (Wynne, 2005). Così, nonostante la diversità di metodo, Sinclair (2005) offre dei suggerimenti per la costruzione dei corpora secondo precisi parametri di natura generale. Sulla base dei consigli di Sinclair, e anche sulla scorta dei contributi di Biber (1993), i dati raccolti in un corpus devono fornire campioni che siano *rappresentativi*, *bilanciati* e *omogenei* di eventi linguistici documentati nella loro complessità e corredati da ulteriori informazioni che completino le informazioni dei dati, i cosiddetti metadati. Inoltre, Sinclair sottolinea l'importanza e la necessità di documentare i dettagli della progettazione e della composizione del corpus descrivendo le ragioni e le giustificazioni delle decisioni prese nelle diverse fasi di costruzione. Questa chiarezza inerente al metodo garantisce un utilizzo appropriato da parte di altri ricercatori e una migliore interpretazione dei dati.

Queste indicazioni fanno riferimento a concetti teoricamente idealistici dato che gli eventi linguistici costituiscono una popolazione quantitativamente infinita che deve essere rappresentata all'interno di un corpus che è finito. Ciò, di fatto, non garantisce che un corpus possa raccogliere tutti i modelli di lingua in modo proporzionale (Sinclair, 2008). Essere consapevoli dell'inevitabile incompletezza dei corpora e della difficoltà di soddisfare pienamente i requisiti di rappresentatività, omogeneità e bilanciamento, deve condurre a riconoscere che la chiarezza del metodo adottato nelle fasi di costruzione del corpus e il raggiungimento della coerenza interna dell'organizzazione dei dati garantiscono il valore dei corpora come fondamentali strumenti della ricerca scientifica.

Nonostante la varietà di approcci e la difficoltà del raggiungimento di un campione rappresentativo, la raccolta di dati di tipo ecologico per la costruzione di un corpus si realizza attraverso quattro fondamentali fasi:

1. la registrazione dei dati;
2. la trascrizione;
3. la codifica o annotazione;
4. l'applicazione e presentazione dei dati.

Queste fasi, che caratterizzano la procedura di compilazione dei corpora, rappresentano un sistema di ricerca completo piuttosto che una procedura temporalmente ordinata e distinta in processi. Ciascuna di queste fasi sono dunque intrinsecamente legate tra loro; tuttavia, il modo in cui ciascuna di esse interagisce con l'altra è strettamente dipendente dagli approcci e dai metodi adottati (Knight, 2011).

Nei paragrafi successivi analizzeremo le differenti fasi del processo di costruzione di un corpus cercando di comprendere meglio quali siano le necessità e le questioni legate alla natura multimodale del nostro corpus.

2.1 La registrazione e la raccolta dei dati

La raccolta dei dati è il primo passo per la compilazione di un corpus e in quanto tale deve necessariamente essere ben pianificata. Da questo primo processo, infatti, dipenderà la qualità e la quantità dei dati che verranno successivamente analizzati. La registrazione dei dati può avvenire in laboratorio, attraverso l'uso di un metodo di elicitazione dei dati (come il map-task), o in contesto naturale. Questa scelta è naturalmente dipendente dagli obiettivi e dagli interessi di ricerca che il corpus intende soddisfare, e in ciascun contesto di raccolta dei dati è fondamentale porre attenzione ad accorgimenti che potrebbero compromettere la qualità e l'attendibilità dei dati raccolti.

In queste pagine porremo la nostra attenzione alle esigenze legate al contesto di raccolta di dati ecologici. In primis è fondamentale che tutte le registrazioni siano adatte e sufficientemente ricche di informazioni necessarie per un'indagine linguistica approfondita e che abbiano una qualità abbastanza alta. È importante riuscire a raccogliere una quantità di dati che possa essere quanto più possibile ricca di informazione in relazioni ai contenuti e al contesto interazionale, questo perché la mancanza o l'omissione di dati in un contesto naturale non può essere facilmente corretta o recuperata.

La progettazione del processo di raccolta deve, quindi, aiutare a determinare i tipi di soggetti che devono essere coinvolti (numero di partecipanti, sesso, provenienza geografica ecc.). Deve anche aiutare a definire l'intero processo di raccolta, determinando i tipi di dati che è necessario raccogliere, la quantità e gli argomenti. Inoltre, è importante definire anche le condizioni fisiche in cui la registrazione deve avere luogo, definendo a priori il setting, gli strumenti e le loro impostazioni (Thompson, 2005).

Per quanto riguarda la compilazione di corpora multimodali, come abbiamo già accennato, non esistono degli standard condivisi, così i corpora multimodali differiscono l'uno dall'altro anche nel metodo di costruzione (Knight et al., 2009). Tuttavia, il dibattito in corso sta cercando di definire standard che possano garantire un corpus-design che garantisca risultati affidabili e generalizzabili. Così, nella vasta varietà di interesse di ricerca della multimodalità e dei diversi strumenti messi in campo è possibile definire una lista di parametri entro cui si muovono quasi tutti i corpora multimodali (Knight, 2011):

- design e infrastruttura del corpus: il tipo di registrazioni, la metodologia utilizzata per raccogliere i dati, per la loro annotazione e la loro rappresentazione;

- dimensioni e ambito: la quantità di dati (misurata o con il numero di parole o con le ore registrate) e la variabilità della tipologia di dati inclusi sia in termini di contesti che di parlanti;
- spontaneità: in quali termini il dato viene concepito come naturale o autentico, se è pienamente naturale o parzialmente spontaneo;
- disponibilità e riusabilità: la regolamentazione per i diritti di accesso ai dati.

Nel corso degli anni, l'uso di audioregistrazioni per la compilazione dei corpora si è di fatto stabilizzata, mentre l'uso di videoregistrazioni risulta ancora in parte innovativo. I corpora multimodali hanno un portato di novità anche per quanto riguarda la qualità del dato, infatti, il set-up delle camere e le impostazioni video sono diventati parte integrante nello sviluppo del data-set; per cui la strumentazione diventa fondamentale nella progettazione della raccolta dati.

«The development of a multimodal corpus cannot proceed on the basis of the same assumptions that have prevailed in exclusively language-based corpora, which pick out a lemma in relation to a restricted environment in which the tokens occur [...] as we are not dealing with language alone but with the various multimodal semiotic resources». (Baldry & Thibault, 2001)

Le condizioni e le procedure utilizzate per la registrazione dei dati si basano pressoché tutte su un simile modello che prevede l'uso di una gamma di attrezzature altamente specializzate in un ambiente di registrazione standardizzato che risulta, di conseguenza, replicabile (Knight, 2011). È buona norma registrare l'audio separatamente con un microfono e un registratore di buona qualità e successivamente sincronizzare la traccia audio con quella video.

L'uso dei corpora di parlato nella ricerca linguistica era tradizionalmente legato alla nozione che l'analisi della lingua è lo studio dell'uso nel contesto spontaneo, di conseguenza i dati elicitati o 'artificiali' hanno avuto un ruolo

di controllo (McEnery & Wilson, 1996). La nozione di spontaneità o naturalezza dei dati rappresenta un obiettivo importante per i compilatori di corpora di parlato. Questo stesso scopo, pur essendo ugualmente importante per i corpora multimodali, risulta però maggiormente faticoso da raggiungere rispetto ai corpora monomodali. La nozione di spontaneità dei dati è per sua definizione astratta e soggetta a interpretazioni si potrebbe essere rappresentata dalla lingua della vita quotidiana che è libera e mutevole da un contesto all'altro. Nella costruzione di un corpus con dati spontanei il parametro della spontaneità è intrinsecamente legato alla metodologia adottata per la registrazione. Infatti, al fine di garantire dati naturali, senza alcuna influenza, è necessario che i parlanti siano il meno possibile condizionati sia dall'ambiente - con la possibilità di muoversi in un ambiente naturale - che dalla strumentazione attraverso l'uso di videocamere e audioregistratori poco invasivi. Ma una assoluta naturalità dei dati è pressoché impossibile da ottenere poiché, a causa del "paradosso dell'osservatore" (Labov, 1972), i partecipanti possono condizionare il loro comportamento in quanto consapevoli di essere registrati. Questa questione coinvolge anche problemi di natura etica, infatti, nel rispetto della privacy di ogni singola persona non è eticamente - oltre che legalmente - ammissibile registrare persone senza che ne siano a conoscenza.

Come già accennato, i metadati svolgono un'importante funzione nella costruzione e utilizzo del corpus. Data la natura selettiva dei corpora, i metadati archiviano tutte le informazioni inerenti questa selettività in cui è possibile dettagliare le tecniche di registrazione utilizzate e le informazioni sui parlanti coinvolti. In questo modo si facilita la fruizione del corpus permettendo ai ricercatori di contestualizzare al meglio i dati così da interpretarli in modo più consapevole. Le informazioni contenute nei metadati vengono integrati all'interno dei corpora durante il processo di annotazione e, così come accade per la codifica, ci sono vari modi in cui ciò avviene. Sebbene, anche per questo aspetto, non ci siano standard condivisi Burnard (2005) suggerisce di inserire nei metadati dettagli editoriali (chi lo

ha compilato, dove e quando), analitici (come i dati sono stati codificati e analizzati), descrittivi (come i dati sono stati taggati e classificati) e amministrativi relativamente al processo di composizione del corpus.

2.2 La trascrizione dei dati multimodali

La trascrizione dei dati è vista come un processo integrato nell'analisi dei dati linguistici. Senza dubbio la scelta di un determinato metodo di trascrizione e delle caratteristiche linguistiche da trascrivere rappresenta un indice degli obiettivi di ricerca e del quadro teorico a cui si fa riferimento. È proprio durante il processo della trascrizione che le parole pronunciate diventano dati linguistici riportati in un documento (Ochs, 1979), in questa la trascrizione è spesso vista come un'azione sia interpretativa che costruttiva. Come per le altre fasi di costruzione di un corpus, anche per la trascrizione non c'è nessun accordo tra i ricercatori, sebbene nel corso degli anni si sia registrato qualche sforzo in questa direzione³⁰. La necessità di uno standard comune è sempre più urgente, poiché consentirebbe la trasferibilità e il riutilizzo dei dati e la conseguente la possibilità di accedere a quantità di dati sempre più grandi, senza che ciò implichi un alto prezzo in termini di risorse economiche e di lavoro.

La trascrizione rappresenta una delle principali questioni legate alla costruzione dei corpora multimodali. Nello specifico, il ricercatore che decide di trascrivere dati multimodali deve decidere quali aspetti trascrivere (parlato, gesti, sguardo, movimento della testa ecc.), come trascriverli e soprattutto come riportare la relazione tra le diverse modalità. A ciò si aggiunge che nella fase di trascrizione dei dati, e quindi dal passaggio da uno stimolo audio-visivo a uno grafico-simbolico, alcune informazioni connesse alla natura del dato vengono perse. Questo aspetto è particolarmente enfatizzato per la trascrizione gestuale relativamente alla complessità del movimento (Kipp et al., 2007). La difficoltà della trascrizione gestuale è amplificata dal fatto che, nella loro natura motoria, a differenza della parola pronunciata non possono essere facilmente tradotti

³⁰ Un esempio è dato dal Transcription System di Gail Jefferson (Jefferson, 2004) anche se fortemente influenzato dall'approccio conversazionalista. Questo aspetto lo rende poco utilizzabile per i ricercatori con altri approcci teorici.

in unità testuali (Gu, 2006); di conseguenza manca un metodo testuale per la loro rappresentazione in forma di trascrizione. A queste questioni si aggiunge il conseguente problema della ricerca di unità gestuali attraverso i software per l'interrogazione che utilizzano il lessico come acceso alle ricerche dei dati.

Senza dubbio, qualsiasi tentativo di riportare determinate informazioni (legati agli aspetti sonori o visivi dell'azione linguistica) attraverso unità testuali è limitato. L'unica possibilità per superare questo problema è il modo in cui i dati vengono presentati e resi accessibili nei corpora. Rispetto ai corpora di parlato è importante l'esperienza di DB-IPIC (Panunzi - Gregori 2012) che accompagna le singole unità con le tracce audio corrispondenti. Mentre, per i corpora multimodali la Knight (2011) scrive:

«[...] one method of solving the challenge of transcribing the multimodal may perhaps be to restrict, as a 'reference point', the exploration of gestures to the visual medium rather than attempting to include references of these phenomena within the textual transcriptions»

Nonostante la studiosa individui un problema relativo alla quantificazione delle unità, che si rende possibile attraverso la trascrizione testuale, richiedendo al ricercatore lo sforzo di una consultazione manuale dei video, riteniamo che l'ausilio dei dati visivi e sonori sia di fondamentale importanza per i corpora multimodali. Il tutto deve necessariamente essere accompagnato da un metodo di annotazione che, come la stessa Knight suggerisce, riporti le informazioni del gesto o integrandole nell'annotazione del parlato o separatamente ma comunque allineata nel tempo.

2.3 L'annotazione dei dati multimodali

Al processo di trascrizione dei dati segue, o si associa, la fase di annotazione dell'informazione linguistica che consiste nell'associare a eventi o fenomeni un'etichetta sulla base di categorie stabilite, ed è

attraverso questo processo che gli aspetti qualitativi dei dati diventano analizzabili a livello quantitativo attraverso l'esplorazione dei corpora per mezzo di specifici software. Nell'annotazione abbiamo diversi livelli che vanno da quello basato sulla parola fino a un livello del discorso attraverso le annotazioni semantiche e pragmatiche.

Così come per le fasi precedenti, anche l'annotazione è fortemente influenzata dallo scopo del corpus e dal quadro teorico di riferimento e di conseguenza non esiste uno standard o un modo meccanico per scegliere quali etichette applicare a un determinato fenomeno (Leech, 1997). Ma, nonostante ciò, è riscontrabile una tendenza che accomuna il modo in cui integrare queste informazioni all'interno dei corpora. Codifiche e tag-set specifici vengono incorporati nel corpus, nella struttura sottostante, separando le informazioni testuali da quelle extra-testuali.

L'annotazione dei corpora multimodali, come abbiamo già accennato in altre pagine di questo lavoro, richiede un lungo lavoro di analisi per le individuazioni della unità e delle etichette da utilizzare. Questa fase di analisi vuol dire demarcare dove si realizzano determinati fenomeni gestuali e assegnare delle etichette sulla base di caratteristiche di tipo formale e, successivamente a questo procedimento di pre-analisi, inizia un'indagine di tipo ipotetico al fine di rispondere a domande specifiche sull'uso e la natura dei comportamenti gestuali e la loro interazione con il parlato (Knight, 2011). Questo ha rappresentato in parte il nostro obiettivo ma, a differenza di quanto affermato dalla Knight, senza mirare alla categorizzazione dei fenomeni gestuali; piuttosto cerchiamo di creare uno strumento che possa mettere in relazioni unità minime di analisi di parlato e gesto, percettivamente riconoscibili, per lo studio delle interazioni comunicative multimodali. Infatti, l'affidabilità e l'accuratezza delle classificazioni dei gesti dipendono dalla capacità dell'analista (Knight, 2011) e sono fortemente dipendenti anche dalla sua soggettività; in alcuni casi poi è possibile dire, di determinati eventi, 'ciò che non è' invece di 'ciò che è'.

Il problema della codifica gestuale è ben riassunto da Baldry e Thibault (2006):

«In spite of the important advances made in the past 30 or so years in the development of linguistic corpora and related techniques of analysis, a central and unexamined theoretical problem remains, namely that the methods adapted for collecting and coding texts isolate the linguistic semiotic from the other semiotic modalities with which language interacts [...] linguistics corpora as so far conceived remains intra-semiotic in orientation. [...] In contrast multimodal corpora are, by definition, inter-semiotic in their analytical procedures and theoretical orientations» (Baldry & Thibault, 2006)

Come abbiamo già visto, esistono diverse classificazioni e codifiche dei gesti in relazioni ad aspetti e funzioni specifiche. Ciò che continua a mancare è uno schema in grado di supportare in pieno la codificazione del gesto nel parlato o integrare le informazioni che arrivano dalle diverse modalità. Ci sono schemi che si concentrano maggiormente sulla codifica del movimento (McNeill, 1992; Boutet 2018) e altri approcci che mirano più alla classificazione delle funzioni o del rapporto semiotico o pragmatico tra il parlato e il gesto (Efron, 1941; Ekman & Friesen, 1968; Kendon, 2004; Müller, 2013). Da questo punto di vista, il contributo che arriva dalle ricerche relative alla trascrizione e annotazione delle lingue dei segni potrebbero dare una spinta significativa.

Negli ultimi anni il progetto ISLE 'International Standards for Language Engineering' sta gettando le basi per uno schema standard per affrontare la codifica cross-level e cross-modality dei dati linguistici naturali (Dybkjær & Ole Bernsen, 2004) e sono conosciuti come 'Natural Interaction and Multimodal Annotation Schemes' (NIMMs). Questi standard sono progettati per annotare il parlato, le espressioni facciali, lo sguardo, i gesti, la postura e l'uso di oggetti, ma al momento non sono stati formalmente presentati alla comunità scientifica.

Ad ogni modo, in attesa che la comunità scientifica raggiunga uno standard, è bene che, nella specificità degli approcci, gli schemi possano essere sistematicamente discriminabili, esaustivi e sistematicamente contrastivi (Edwards, 1993) così da garantire una coerenza del flusso dei dati e la loro processabilità.

3. Il corpus CORMIP

3.1 La raccolta dei dati: metodi e strumenti

Nel processo di compilazione del corpus, la dichiarazione dello scopo per cui viene creato è utile sia perché struttura e influisce il metodo adottato nella costruzione e sia perché facilita la consultazione dei dati. L'obiettivo che ci siamo posti è quello di cercare di rispondere alla necessità di ampliare il repertorio dei corpora di italiano con un corpus di italiano che raccogliesse dati spontanei e multimodali.

Il raggiungimento di questo obiettivo ci ha posto subito davanti ad alcune questioni che abbiamo cercato di affrontare, riuscendo però a trovare una proposta piuttosto che una soluzione. Nello specifico, ci siamo trovati in primo luogo davanti al problema di definizione di dati spontanei dell'italiano. Infatti, la rappresentatività di un campione è strettamente legata alla definizione di caratteristiche che specificano quel campione. Quindi voler rappresentare una realtà di per sé complessa, come gli eventi linguistici spontanei dell'italiano, implica una definizione di alcune caratteristiche minime che devono essere contenute all'interno del campione e soprattutto richiede una quantità di tempo e di risorse che mal si conciliano con un percorso dottorale. In secondo luogo, la seconda questione da dirimere è la definizione di dati multimodali. Abbiamo avuto già modo di approfondire in altre pagine di questo lavoro, come la prospettiva multimodale sui dati linguistici includa svariate caratteristiche

ed elementi linguistici. Un parlante che realizza un evento linguistico, infatti, trasmette le informazioni e i significati attraverso l'uso di diversi canali come quello sonoro (che si realizza attraverso l'enunciazione con tutte le sue caratteristiche fonetiche, fonologiche e prosodiche), quello gestuale (l'uso dei movimenti delle braccia e delle mani), l'espressione facciale e la postura. E ciascuna di queste modalità può essere indagata e approfondita su livelli di specificità diversi sulla base dell'interesse e della competenza in quel settore. Inoltre, l'approccio multimodale mira soprattutto a indagare l'interazione tra i diversi canali comunicativi rendendo ancora più complesso il raggiungimento dell'obiettivo della costruzione di un corpus che sia in pienezza multimodale. Di fronte alla complessità delle questioni da affrontare, si è ritenuto opportuno iniziare a elaborare una proposta metodologica attraverso l'uso di un corpus pilota che includa dati sulla gestualità, oltre che sul parlato inteso come canale, lasciando così cadere, in parte, gli obiettivi di rappresentatività del corpus.

Abbiamo progettato la raccolta dati cercando di garantire la diversificazione delle tipologie interazionali, una minima e indispensabile variabilità diatopica e la spontaneità degli eventi linguistici.

Per la tipologia interazionale abbiamo seguito il parametro del numero dei locutori «che determina l'ossatura dell'evento comunicativo e permette di distinguere i testi orali in monologhi, dialoghi e conversazioni» (Cresti, 2000). Abbiamo basato questa scelta su alcune evidenze che arrivano dagli studi sull'interazione tra la gesticolazione e il numero dei partecipanti all'interazione. Özyürek (2002) dimostra che i gesti realizzati dai parlanti sono influenzati dalla posizione degli interlocutori. A seguito di alcuni esperimenti, la ricercatrice ha dimostrato che i parlanti cambiano i gesti direzionali, che accompagnano preposizioni con contenuti spaziali di direzione, sulla base della variazione nella posizione dei destinatari. La studiosa ha dimostrato che ciò era dovuto al fatto che i cambiamenti della posizione dei destinatari cambiavano lo spazio condiviso, cioè

l'intersezione degli spazi gestuali dei parlanti e dei destinatari. Così, abbiamo ritenuto opportuno includere nei dati tre tipologie interazionali:

- Monologico, in cui ritroviamo solamente un parlante in un contesto interattivo dove l'interlocutore non può togliere il turno se non con esplicita richiesta (es. una lezione universitaria);
- Dialogico, in cui ritroviamo due parlanti che possono interagire fra loro prendendo il turno liberamente;
- Conversazionale, dove troviamo più di tre parlanti che interagiscono potendo prendere il turno liberamente.

Nella raccolta dati abbiamo però dovuto tralasciare il parametro della proiezione sociale del contesto comunicativo a causa della ridotta ampiezza del corpus. Ma la quasi totalità delle registrazioni è ascrivibile a un contesto sociale privato tra persone con un grado di conoscenza reciproca alto, fatta eccezione di un solo brano che riporta una lezione universitaria e dunque è possibile parlare di contesto comunicativo pubblico. Bisogna inoltre specificare che le interazioni sono, appunto, spontanee e quindi raccolte in un contesto naturale dove i parlanti hanno affrontato argomenti e questioni senza alcun tipo di stimolo.

Rispetto alla variabilità diatopica, il nostro interesse si rivolge alle varietà della lingua italiana che si diversificano in relazione all'origine e alla distribuzione geografica dei parlanti. Così, dunque, trattandosi di un corpus pilota, abbiamo incluso solamente due punti di raccolta dati riferibili a due città, Firenze e Catania. Per ciascuna delle due città abbiamo raccolto un brano per ogni tipologia interazionale. Questa nostra scelta non mira a rappresentare la variabilità diatopica dell'italiano, ma intende includere in minima questo livello di variazione linguistica, soprattutto per poter indagare (e nel caso testimoniare) una possibile variazione diatopica anche a livello gestuale. Kita e Özyürek (2003), hanno dimostrato che le forme gestuali dipendono anche dall'organizzazione dell'informazione dei diversi sistemi linguistici. Nel nostro caso, non si tratta di un confronto tra sistemi di lingua diversi ma di variazione regionale di una stessa lingua che, anche alla luce dello studio di Kita e Özyürek (2003), riteniamo sia utile

testimoniare e rendere indagabile una ipotetica variabilità diatopica dell'italiano anche a livello gestuale.

Come abbiamo già detto in precedenza, raccogliere dati spontanei vuol dire interfacciarsi con quello che Labov (1972) definì paradosso dell'osservatore. Per motivi etici e legali, la registrazione non può essere realizzata senza l'accordo dei partecipanti e, inoltre, la possibilità di nascondere i registratori rischia di compromettere la qualità dei dati. Così, il paradosso dell'osservatore non può essere del tutto superato ma è possibile attenuarne gli affetti. Nel nostro caso, abbiamo informato i parlanti (attraverso l'informativa sulla privacy) sul fatto che le registrazioni venivano realizzate per fini di ricerca, senza però specificare il campo di studi e gli elementi di interesse, e sulla conservazione e gestione dei dati e soprattutto sul loro uso. Abbiamo inoltre cercato di attenuare il paradosso dell'osservatore attraverso l'uso di strumentazione non invasiva e la costruzione di ambienti di registrazioni che possano essere familiari ai parlanti e che vedano i registratori posizionati in modo adeguato. Nello specifico abbiamo utilizzato una telecamera GoPro Hero 6 e un audio registratore Zoom H6 con un microfono panoramico (120°) posti in un ambiente conosciuto dai parlanti e in contesti comunicativi spontanei. I parlanti erano liberi di muoversi e di interagire con l'ambiente circostante, non sono state date loro regole o accorgimenti da seguire e gli argomenti trattati sono totalmente spontanei. Le registrazioni sono avvenute in eventi comunicativi reali come la pianificazione di un viaggio, una lezione universitaria, un incontro tra amici e una riunione organizzativa tra capi-scout.

In totale sono state registrate sei diverse situazioni comunicative, tre generi conversazionali per due diverse città. I partecipanti hanno un range di età che va dai vent'anni ai sessanta e il livello di istruzione più basso è il diploma di scuola media secondaria. Nella tabella di seguito viene illustrato il *dataset* del corpus.

Tabella 4 Data-set del corpus CORMIP

| Genere conversazionale | Evento comunicativo | Città | Durata |
|---------------------------|------------------------|---------|---------|
| Monologo | Lezione letteratura | Firenze | 5'.20'' |
| Dialogo | Dialogo Scout | Firenze | 5'.46'' |
| Conversazione | Arbitri pallamano | Firenze | 5'.42'' |
| Monologo | Racconto di vita | Catania | 5'.27'' |
| Dialogo | Conversazione studenti | Catania | 5'.52'' |
| Conversazione | Conversazione viaggio | Catania | 7'.12'' |

3.2 Trascrizione e annotazione dei dati

Diversi studi hanno disegnato sistemi di annotazione del gesto. Ognuno è caratterizzato dal suo metodo, dallo scopo della ricerca e dalla definizione dei tag. Alcuni esempi possono essere rappresentati da NEUROGES, CoGesT, e LASG. NEUROGES è un sistema di codifica basato sul presupposto che i gesti sono strettamente legati ai processi cognitivi, emotivi e interattivi. Questo sistema è ben organizzato e diviso in tre moduli (cinesica, relazione tra le mani e cognizione/emozione) e diversi passi. Questo sistema di codifica è a grana fine e presenta quindi decine di etichette (Lausberg 2013) (Lausberg e Sloetjes 2015).

CoGesT (Conversational Gesture Transcription), è stato creato per fornire un sistema di trascrizione per l'analisi linguistica e l'elaborazione automatica dei gesti. Questo sistema distingue i gesti in gesti Simplex e gesti Compound. Nel primo ci sono due tipi, luogo statico - un gesto che tiene una specifica configurazione della mano - e luogo dinamico dove i gesti sono caratterizzati da Sorgente, Traiettorie e Bersaglio; questi attributi sono rappresentati come un vettore (Gibbon et al. 2003) (Trippel et al. 2004).

LASG (Linguistic Annotation System for Gesture), offre un'annotazione dei gesti basata su un approccio linguistico cognitivo e si riferisce a un

approccio basato sulla forma per l'analisi dei gesti. Fornisce diversi livelli di annotazione: annotazione per il gesto che include sotto-livelli come unità determinanti, annotazione delle forme, motivazione della forma; annotazione del discorso che include come sotto-livello il turno e l'unità di intonazione; e annotazione del gesto sul discorso con altri sotto-livelli come prosodia, sintassi, semantica, e pragmatica (Bressem, Ladewig, e Müller 2013b). Tutti questi esempi sono ottimi sistemi di annotazione, ma presentano problemi per la creazione di un corpus di parlato spontaneo e per un sistema di annotazione che possa essere utilizzabile con dati spontanei e di grandi dimensioni. Il nostro obiettivo è quello di offrire un sistema di annotazione semplificato ed efficiente che possa evidenziare come il gesto e il parlato creino un enunciato multimodale. Pensiamo che sia necessario segmentare il flusso gestuale e vocale su unità di base che siano percettivamente rilevabili: le unità di intonazione per il discorso, e le unità di movimento per i gesti. Inoltre, consideriamo l'intonazione l'elemento cruciale dell'enunciato che è percettivamente ben definito e linguisticamente significativo. Loehr ha mostrato che il gesto e la prosodia sono strettamente connessi, entrambi i canali - gesto e discorso - lavorano insieme per costruire il discorso e regolare l'interazione. Questa relazione è stata trovata sia nella produzione che nella percezione, in tutte le età e in decine di lingue (Loehr 2007; 2014). Vogliamo unificare l'annotazione L-Act - che sottolinea il ruolo dell'intonazione nel discorso - con l'architettura del gesto progettata da Kendon che offre un'importante struttura del gesto composta dalla singola unità e fase del movimento gestuale (Kendon 1972).

L'idea è quella di creare un sistema di trascrizione e annotazione che possa identificare le unità strutturali su base percettiva. Questo obiettivo scaturisce dall'esigenza di dare centralità alla diversità delle modalità che andremo ad analizzare. Infatti, i diversi metodi di significazioni di ciascuna modalità (gesto e parlato) dipendono fortemente dalla loro fisicità, pur scaturendo entrambe da un medesimo processo cognitivo. Come ci ha ricordato Boutet (2018), la fisicità del gesto rappresenta un substrato che ne

determina le potenzialità e i limiti, allo stesso modo è possibile affermare per il canale orale del parlato. In questo modo, segmenteremo il flusso del parlato sulla base della percezione uditiva e il flusso gestuale sulla base della percezione visiva del movimento.

Nel concreto faremo riferimento al modello della Teoria della Lingua in Atto (Cresti, 2000) per il parlato, perché questo quadro teorico ci dà i mezzi e le prove per identificare gli enunciati e le unità di intonazione della modalità parlata e quindi sulla percezione uditiva. Per la trascrizione e l'annotazione gestuale, faremo riferimento agli studi di Kendon e McNeill che forniscono un'architettura del gesto che riesce a identificare le unità che compongono il movimento gestuale, senza correlare valori linguistici alle diverse unità. Ma appare chiaro, ormai, che esista un'unità necessaria e sufficiente rappresentata dalla fase dello *stroke*.

Abbiamo, così, ritenuto necessario tenere separate l'annotazione del parlato da quella dei gesti in modo da lasciare le informazioni linguistiche dei due canali indipendenti l'uno dall'altro, per poter così ridurre le possibili influenze (che possono in caso scaturire dal concepire il gesto come dipendente dal parlato) e per poter indagare in modo dettagliato come le due modalità si correlano. La multimodalità dell'azione linguistica emerge dall'annotazione dell'illocuzione, che rappresenta l'elemento linguistico che caratterizza a nostro avviso l'uso di elementi semantici, intonativi e gestuali.

3.2.1 La trascrizione e l'annotazione del parlato

Come abbiamo già accennato, per la trascrizione e l'annotazione del parlato faremo riferimento al modello della Teoria della Lingua in Atto (Cresti, 2000). Questo modello teorico si basa sulla teoria degli atti linguistici di Austin (1962) ed è stato elaborato su osservazioni empiriche dei corpora di parlato spontaneo. La teoria di Cresti si avvicina alle interazioni comunicative come eventi che scaturiscono dall'affetto del parlante verso il destinatario e che si realizzano in un atto linguistico con

valore pragmatico. In questo modello, la funzione pragmatica è considerata la funzione principale del parlato che gestisce e influenza gli elementi linguistici e la struttura sintattica. La proposta si basa su due tipi di unità di riferimento individuate prosodicamente: l'*utterance* e la *stanza*. L'*utterance* è l'unità linguistica minima e principale caratterizzata da un confine prosodico terminato e compie un unico atto linguistico; la *stanza*, invece, è formata da una sequenza di *comment* deboli che non corrispondono a una sequenza di enunciati. La *stanza* non è strettamente regolata da principi pragmatici ma segue piuttosto strategie di costruzione testuale (Moneglia & Raso, 2014) (Panunzi & Scarano 2009). Quindi, le unità di riferimento del discorso sono entità linguistiche basate su caratteristiche semantiche, pragmatiche e prosodiche e la loro identificazione avviene prosodicamente attraverso il riconoscimento percettivo dei confini terminati o non terminati da parte dell'annotatore. La prosodia gioca un ruolo importante all'interno della relazione tra atto illocutivo e locutivo, infatti, esprime la funzione pragmatica dell'atto linguistico rendendolo una vera e propria entità udibile. La struttura informativa del parlato è costruita intorno all'unità necessaria e sufficiente chiamata *Comment* e che potrebbe essere accompagnata da altre unità opzionali con le quali forma lo schema informativo. Le unità aggiuntive assumono diverse funzioni: *Topic*, *Parenthesis*, *Appendix*, *Locutive Introducer*, e *Discourse Markers*. La classificazione illocutiva L-AcT si basa sull'attivazione affettiva del parlante verso il destinatario e sull'analisi del corpus che porta i ricercatori a identificare cinque classi illocutive principali: rifiuto, affermazione, direzione, espressione e rito.

La centralità della prosodia all'interno della Teoria della Lingua in Atto si appoggia al modello prosodico elaborato dai lavori di IPO ('t Hart, Collier e Cohen 2006) dimostrando che tra la struttura informativa e la struttura prosodica esiste una corrispondenza (Moneglia e Raso 2014).

Il framework considera il parlato un'attività pragmatica svolta dal parlante:

« it [L-AcT (Language into Act Theory)] stressed that prosody plays a mandatory role in the performance of the utterance and its linguistic identification. Moreover, L-AcT foresees that the internal information organization of the utterance is governed by pragmatic principles and is crucially mediated by prosody» (Moneglia e Raso 2014)

Il quadro teorico della Teoria della Lingua in Atto, oltre a restituirci un metodo di analisi che non può prescindere dalle caratteristiche fisiche del parlato, ci fornisce gli strumenti utili per poter segmentare il flusso del parlato in unità prosodiche su base percettiva.

Abbiamo basato, dunque, la nostra trascrizione e la nostra annotazione del parlato sul modello appena illustrato. Ma se per la trascrizione abbiamo usato il metodo della Teoria della Lingua in Atto, per l'annotazione abbiamo invece adattato in parte il modello. Come abbiamo avuto modo di vedere nel primo capitolo di questa tesi, la lingua parlata è caratterizzata da diversi fenomeni specifici, alcuni dei quali sono legati all'interazione – ad esempio la sovrapposizione, la vocalizzazione e il retracting – altri fenomeni sono legati a caratteristiche linguistiche come l'intonazione. La trascrizione della lingua parlata non può prescindere da queste caratteristiche specifiche che permettono di rendere interpretabile un testo parlato.

I corpora compilati dal gruppo di ricerca LABLITA, dell'Università di Firenze, offrono un buon metodo di trascrizione basato sul modello CHAT adattato al quadro teorico della Teoria della Lingua in Atto (Cresti, 2000; Cresti & Moneglia, 2005). Il formato che ne deriva è denominato CHAT-LABLITA ed è stato creato in conformità con l'approccio teorico che implementa il formato CHAT, creato nell'ambito del progetto CHILDES, includendo l'intonazione e la sua funzione di demarcazione delle unità di enunciazione e di informazione. Il flusso del parlato è dunque segmentato percettivamente in unità tonali segnate da pause prosodiche che possono essere terminate o non terminate. L'unità prosodica terminata determina i confini dell'enunciato ed è rappresentata con due barre //; mentre l'unità

prosodicamente non terminata identifica le altre unità prosodiche all'interno dell'enunciato ed è rappresentata con una sola barra /. Per la trascrizione di altri fenomeni – come il suono non linguistico, i frammenti, le parole interrotte, ritratte e sovrapposte – il formato fornisce un repertorio completo come è illustrato nella tabella seguente.

Tabella 5 Simboli del sistema di trascrizione CHAT-LABLITA

| Simbolo | Valore |
|---------|--|
| // | Break prosodico terminale |
| ? | Break prosodico terminale con intonazione interrogativa |
| ... | Break prosodico terminale con intonazione sospensiva |
| + | Break prosodico terminale per sequenza interrotta |
| / | Break prosodico non terminale |
| [/] | Falsa partenza con ripetizione |
| & | Vocalizzazione o frammento di parola |
| hhh | Fenomeno paralinguistico o non linguistico come tosse o risata |
| xxx | Parola non comprensibile |

Nell'applicare il metodo di trascrizioni del modello CHAT-LABLITA, abbiamo scelto di non segnare le sovrapposizioni in quanto il software ELAN, che è stato utilizzato per la trascrizione annotazione e allineamento, restituisce l'informazione in automatico grazie all'interfaccia grafica.

La trascrizione del parlato viene realizzata ortograficamente e il flusso del parlato è segmentato su base prosodica attraverso l'applicazione dei segni indicati dal modello CHAT-LABLITA, questo permette di riportare in modo efficace tutti i fenomeni tipici del parlato. La natura dialogica dell'evento è stata riportata nella costruzione del *template* di annotazione, di cui discuteremo di seguito.

Il principio della percezione prosodica sta alla base anche dell'annotazione dei dati. Per il modello di Cresti la struttura informativa è pienamente corrispondente alla struttura prosodica, quindi le unità prosodiche, delimitate dai break prosodici, esprimono un valore informativo. Il modello prosodico a cui si fa riferimento è quello presentato in 't Hart et al. (1990), sviluppato dai lavori della scuola olandese di IPO, e che è stato successivamente sviluppato dal gruppo di ricerca LABLITA (Firenzuoli, 2003). Il punto di partenza del modello prosodico di IPO consiste nel riconoscere l'esistenza di variazioni di F0, nel flusso del parlato, a cui un ascoltatore risulta percettivamente sensibile. Queste variazioni vengono prodotte intenzionalmente dai parlanti e vengono chiamate *pitch movements* ('t Hart et. Al. 1990) e sono le unità su cui si basa la classificazione dei movimenti propsta da 't Hart:

«[...] the F0 changes that do contribute essentially to the perception of the speech melody are hust those changes that are programmed and voluntary executed by the speaker. Omission of any of these leads to a substantial change or distortion of the indentity of the speech melody» ('t Hart, 1990)

La teoria della scuola olandese riconosce l'esistenza di *configurazioni* costituite da movimenti che ricorrono sempre in particolari ambienti melodici e ne vengono identificate tre tipologia (*ibidem*):

- ROOT: sono configurazioni obbligatorie e non ricorsive, un pitch countour può contenere solamente una sola configurazione di questo tipo;
- PREFIX: sono configurazioni opzionali e ricorsive, precedono sempre un'altra configurazione PREFIX o una ROOT;
- SUFFIX: sono configurazioni opzionali e non ricorsive e seguono sempre una configurazione di tipo ROOT.

Queste configurazioni, come abbiamo visto, si distribuiscono sulla base di regole precise di concatenazione creando un contorno con una regola di composizione che può essere rappresentata come di seguito:

CONTOUR = (Prefix) ROOT (Suffix)

All'interno della teoria dell'articolazione informativa della Teoria della Lingua in Atto viene riconosciuta la possibilità delle unità tonali di tipo suffix di occorrere anche dopo unità di tipo prefix (Firenzuoli, 2003).

Nell'ottica della Teoria della Lingua in Atto è possibile rintracciare una corrispondenza tra le unità prosodiche di tipo root, che sono necessarie e sufficienti per la realizzazione di un pattern tonale, alle unità informative di tipo comment, anch'esse necessarie e sufficienti a costituire un enunciato. Mentre alle unità prosodiche di tipo prefix corrispondono le unità informative di topic e a quelle prosodiche di tipo suffix le unità informative di appendice (Firenzuoli, 2003; Moneglia & Raso, 2014). La corrispondenza tra *pattern tonale* e *pattern informativo* risulta rappresentate in modo efficace da una tabella presente in Moneglia e Raso 2014 e che riportiamo di seguito:

Tabella 6 Corrispondenza struttura prosodica e struttura intonativa (adattata da Moneglia-Raso 2014)

| Prosodic pattern | | Information pattern | |
|------------------|----------|-----------------------|------------|
| Root | | Comment | |
| (Prefix) | (Suffix) | (Topic) | (Appendix) |
| (Introducer) | | (Locutive Introducer) | |
| (Parenthetical) | | | |
| (Incipit) | (Phatic) | (Incipit) | (Phatic) |

Nel nostro tag-set del parlato possiamo dunque distinguere: l'etichetta ROOT per le unità prosodiche che realizzano la prominente prosodica e che

sono sufficienti a costituire un pattern; l'etichetta di PREFIX per le unità prosodiche con carattere opzionale, melodicamente subordinate e che occupano una posizione di precedenza temporale e che anticipano la realizzazione dell'unità root; l'etichetta di SUFFIX per le unità prosodiche che risultano intonativamente subordinate e temporalmente successive alle unità di Root o di Prefix; l'etichetta INCIPIT per le unità prosodiche che risultano intonativamente subordinate e che occorrono a inizio di turno o di enunciato e sono lessicalmente caratterizzate (per lo più avverbi; congiunzioni, pronomi personali o interiezioni); l'etichetta PHATIC per le unità prosodiche intonativamente subordinate, occorrono in qualsiasi posizione all'interno dell'enunciato e svolgono la funzione comunicativa di mantenimento dell'apertura del canale; l'etichetta PARENTHETICAL per le unità intonative con profilo prosodico basso e una velocità di realizzazione maggiore rispetto al resto dell'enunciato di cui rappresenta un'inserzione contenutistica.

Abbiamo ritenuto necessario aggiungere degli ulteriori tag per il discorso riportato, contrassegnando le diverse etichette con “_r”. E infine, abbiamo dedicato un'etichetta per le unità interrotte e una per le unità prosodiche non decifrabili. Nella tabella di seguito riportiamo il tag-set del nostro corpus per il parlato.

Tabella 7 Tag-set della struttura intonativa di CORMIP

| TAG | DEFINIZIONE |
|---------------|--|
| ROOT | unità prosodia prominente, necessaria e sufficienti per la realizzazione dell'enunciato |
| PREFIX | unità prosodica opzionale e subordinata, che occupa una posizione temporalmente antecedente a una Root, una Suffix o un'altra Prefix |

| | |
|------------------------|--|
| SUFFIX | unità prosodica opzionale e subordinata, temporalmente segue le unità di Root o di Prefix |
| INCIPIIT | unità prosodica intonativamente opzionale e subordinata, occorre a inizio turno o enunciato ed è lessicalmente caratterizzate |
| PHATIC | unità prosodiche subordinata e opzionale, occorre in qualsiasi posizione all'interno dell'enunciato e svolge la funzione comunicativa per il mantenimento dell'apertura del canale |
| PARENTHETICAL | unità intonative con profilo prosodico basso e realizzata con velocità maggiore rispetto al resto dell'enunciato di cui rappresenta un'inserzione contenutistica |
| ROOT_r | unità ROOT di parlato riportato |
| PREFIX_r | unità PREFIX di parlato riportato |
| SUFFIX_r | unità SUFFIX di parlato riportato |
| INCIPIIT_r | unità INCIPIIT di parlato riportato |
| PHATIC_r | unità PHATIC di parlato riportato |
| PARENTHETICAL_r | unità PARENTHETICAL di parlato riportato |
| INTERRUPTED | unità prosodica interrotta |
| EMPTY | unità prosodica non interpretabile |

Il tag-set appena riportato ci ha permesso di annotare il flusso del parlato, segmentato percettivamente, restituendo un elenco di unità di base la cui concatenazione, insieme alla sincronizzazione con gli altri canali di espressione linguistica, permette la codifica di valori informativi, illocutivi e semantici. Di seguito vedremo su quali basi abbiamo costruito la

trascrizione e l'annotazione del gesto, cercando di mantenere fede all'approccio percettivo e dando dunque risalto al movimento e al canale visivo. L'annotazione sul software ELAN apparirà come nell'immagine seguente.

| | | | |
|-------------------------------|---------------------------------|------|--------|
| LUI-UTTERANCE [79] | va be' / novantotto / arrivi // | | |
| | PREFIX | ROOT | SUFFIX |
| LUI-Intonation Units [119] | | | |

Figura 7 Esempio di annotazione del parlato di CORMIP su ELAN

3.2.2 Trascrizione e annotazione del gesto

Abbiamo visto in precedenza, come l'analisi del gesto possa essere condotta da diversi punti di vista che scaturiscono da altrettanti diversi approcci metodologici e analitici. E abbiamo anche mostrato come i diversi interessi di ricerche e i diversi riferimenti teorici influenzino i metodi di trascrizione e di annotazione, al punto che ancora oggi manca uno standard comune. In questo paragrafo illustreremo il metodo disegnato e adottato da noi per la costruzione del nostro corpus e che si basa sull'idea di segmentare e identificare le unità, che compongono gli eventi linguistici, su base percettiva.

Molto spesso le annotazioni dei gesti si basano sulle categorizzazioni delle funzioni svolte, del grado di convenzionalità o del rapporto con il parlato (cfr. capitolo 4). Con questo lavoro non intendiamo proporre nessun tipo di classificazione tipologica o categorizzazione funzionale per due ragioni: in primis perché riteniamo che qualsiasi tipo di proposta rappresenti una classificazione dei dati a priori e dunque elude l'obiettivo di un corpus che è quello di presentare dei dati che siano piuttosto organizzati; e in secundis perché crediamo che nella complessità degli eventi gestuali manchi la possibilità di distinguere in modo chiaro una tipologia gestuale da un'altra.

In un quadro linguisticamente complesso, come quello descritto nelle pagine di questo lavoro, riteniamo sia necessario affermare che il carattere multimodale della lingua – e dunque la sua natura polisemiotica – debba necessariamente indurci a indagare come i diversi sistemi semiotici riescano a formare un sistema unico e coerente. Già nel secondo e nel quarto capitolo abbiamo visto come all'interno dei gesti si possano distinguere diversi gradi di convenzionalità e iconicità che possono gradualmente giungere alla lessicalizzazione. Dal nostro punto di vista, e alla luce di quanto affermato da De Mauro (2000) (cfr. Capitolo 3), sosteniamo che i gesti esprimono diverse funzioni in relazione al loro grado di convenzionalizzazione a cui si associano un crescente grado di iconicità e articolazione. Così, al Kendon's continuum si potrebbe associare un continuum di funzioni che vede la corrispondenza tra un gesto iconico e una funzione lessicale e a gesti meno iconici, in cui è difficile distinguere elementi di articolazione, funzioni soprasegmentali. In questo modo, prima di addentrarci in categorizzazioni e classificazioni dei gesti, riteniamo possa essere più utile e significativo per la ricerca indirizzare la propria analisi a ricercare quali parametri e quali valori determinino la differenziazione di tipi e funzioni dei gesti. In questo lavoro ci limitiamo a tracciare un'ipotesi teorica da verificare attraverso future ricerche.

Quindi nell'attuare il principio di trascrizione e annotazione su base percettiva abbiamo scelto di utilizzare le unità di analisi di Kendon (1972, 1976, 2004). Se nel parlato il flusso fonico è percepito uditivamente, e quindi trascritto e annotato sulla base di ciò che l'annotatore percepisce³¹, per il gesto entrano in campo il movimento, l'iconicità e di conseguenza il canale visivo.

La segmentazione è avvenuta su diversi livelli, secondo l'architettura del gesto illustrata da Kendon (2004). Nel livello più alto abbiamo segmentato le *Gesture unit* (G-UNIT) che è l'unità gerarchicamente maggiore e ingloba l'intera escursione del movimento, ed è visivamente riconoscibile perché

³¹ Nei passaggi più complessi abbiamo risolto i dubbi attraverso l'uso di Praat.

definita dall'inizio del movimento delle mani fino al loro ritorno in posizione di riposo. Il livello di segmentazione più basso è rappresentato dalle *Gesture phrase* (G-PHR) ed è l'unità che suddivide la *gesture unit*. Può essere definita come l'unità a cui corrisponde un significato comunicativo e si identifica per un particolare movimento nello spazio o per una particolare configurazione delle mani. Il livello più basso è rappresentato dalle *Gesture phases*, cioè le unità che compongono una *gesture phrase* e si distinguono: la *preparation* che è la fase di preparazione del movimento, l'unità di *stroke* che è la fase culminante del gesto ed è l'unità necessaria e sufficiente e la *retraction*, che è la fase in cui le mani o le braccia tornano in una posizione di riposo. Inoltre, riconoscendo l'importanza assunta dall'espressione facciale nell'esprimere significati o nel modificarli attraverso l'espressione di un'attitudine o di un'azione linguistica, abbiamo ritenuto necessario aggiungere al primo livello di segmentazione un'etichetta generica che segni la presenza di un'espressione facciale.

La fase di segmentazione e trascrizione del gesto, che è stata fatta separatamente per la mano destra e per la mano sinistra, corrisponde a un'unica fase che si declina su tre livelli.

Nella tabella di seguito riportiamo in sintesi i livelli di trascrizione e i simboli utilizzati.

Tabella 8 Tag-set per la trascrizione gestuale di CORMIP

| LIVELLO | SIMBOLO | DEFINIZIONE |
|---------|----------------------|---|
| I | G | unità gerarchicamente maggiore e ingloba l'intera escursione del movimento |
| I | FACIAL EXPRESSION | espressione facciale |
| II | PHR | l'unità a cui corrisponde un significato comunicativo e si identifica per un particolare movimento nello spazio o per una particolare configurazione delle mani |

| | | |
|-----|-------------|--|
| III | STROKE | fase culminante del gesto ed è l'unità necessaria e sufficiente |
| III | PREPARATION | fase di preparazione del movimento |
| III | RETRACTION | la fase in cui le mani o le braccia tornano in una posizione di riposo |

Mentre di seguito mostriamo un estratto di trascrizione del gesto dall'interfaccia del software ELAN, in cui è possibile notare l'organizzazione dei tiers che riporta la natura gerarchica dell'architettura del gesto.

| | | | | |
|-------------------------|--|---------------|--------|------------|
| LUI-G-UNITS [16] | | G | | |
| LUI-G-Phrases [29] | | PHR | | |
| LUI-G-Phases-RG [67] | | STROKE | STROKE | RETRACTION |
| LUI-G-Phases-LF [64] | | | | |
| LUI-G-Type [29] | | Non-Pictorial | | |

Figura 8 Esempio di annotazione del gesto in CORMIP su ELAN

Come abbiamo detto in apertura di questo paragrafo, abbiamo scelto di non integrare nel nostro corpus un'annotazione gestuale che si basi su categorizzazioni o classificazioni di funzioni, di relazioni con il parlato e/o di gradi convenzionalità del gesto. Perciò abbiamo scelto di utilizzare etichette di natura generale che possano restituire minime informazioni di base utili ai ricercatori. Queste categorie si basano sull'importanza dell'iconicità per il gesto che si esprime in modo graduale anche in relazione alle funzioni espresse dal gesto. Tra gli altri (Efron 1972; Ekman e Friesen 1969) anche Kendon (2004) ha stilato le diverse funzioni esprimibili dai gesti distinguendo le funzioni pragmatiche (quelle di parsing, quelle modali che esprimono la modalità di interpretazione di un enunciato e quella performative che indicano il tipo di atto linguistico), funzioni interattive o interpersonali relative alle funzioni dialogiche interazionali e le funzioni referenziali.

Così abbiamo scelto di annotare i gesti dividendoli in *Pictorial*, *Non-Pictorial* e *Conventional*. Nello specifico, con l'etichetta *Pictorial* abbiamo cercato di raggruppare tutti i gesti che assumono visivamente una forma riconducibile a un'immagine, a un contorno di un oggetto o a un'azione riferibili a oggetti e cose del mondo reale. Sotto questa etichetta rientrano tutti i gesti che possono essere classificabili come iconici, pittografici, ideografici o rappresentativi e che quindi, sulla base delle strategie di rappresentazione individuate da Müller (2013), riportano l'idea di un'immagine. Mentre con l'etichetta *Non-Pictorial* raggruppiamo tutti quei gesti che assumono movimenti ritmici (come i batonici) o forme non assimilabili a oggetti nel mondo (come forme geometriche). In questa categoria abbiamo inserito anche i gesti deittici perché il criterio utilizzato è quello dell'iconicità, infatti, nonostante i gesti deittici facciano riferimento a elementi del mondo reale la modalità utilizzata non è iconica. Mentre con l'etichetta *Conventional* raggruppiamo tutti quei gesti che hanno raggiunto un grado di convenzionalità tale che, in un determinato sistema linguistico, a quel gesto è possibile associare un valore semantico; per esempio, il segno dell'OK con il pollice all'insù o il movimento della mano per il saluto. I gesti riferibili all'etichetta *Conventional* costituiscono un repertorio ben definito. Per quanto riguarda il nostro lavoro di annotazione dei gesti convenzionali abbiamo fatto fede alle competenze linguistiche dell'annotatore, alla luce del fatto che la quantità dei dati annotati non sia molto ampia da contenere un numero significativo di gesti convenzionali. Sarebbe senza dubbio auspicabile in un futuro ampliare il presente corpus con ulteriori dati spontanei e diatopicamente distribuiti, così da redigere un esaustivo repertorio di gesti convenzionali dell'italiano.

La natura complessa e dinamica del fenomeno gestuale non permette un'applicazione rigida della categorizzazione che abbiamo proposto. Infatti, è possibile ritrovarsi davanti a un gesto deittico in cui la mano riproduce la forma di un oggetto (per esempio un bicchiere). In questo caso l'annotatore ha cercato di sciogliere il dubbio attraverso l'interpretazione

dell'interazione della funzione principale espressa dal parlante. Se il parlante usa un gesto ibrido di questo tipo focalizzando maggiormente la deissi, allora si associa il tag NON-PICTORIAL, se invece il gesto viene utilizzato per focalizzare l'iconicità è associato il tag PICTORIAL.

Questo conferma quanto sostenuto nelle pagine precedenti, e cioè la necessità di trovare un metodo di analisi dei gesti che si basi su parametri oggettivi ed empirici. La categorizzazione su base iconica o funzionale non può garantire un risultato scevro da interpretazioni. Allo stesso tempo, una etichettatura di base rimane al momento necessaria per poter ampliare e indirizzare la ricerca verso la definizione di parametri oggettivi.

Nella tabella di seguito riportiamo il tagset del gesto.

Tabella 9 Tag-set per l'annotazione gestuale in CORMIP

| TAG | DEFINIZIONE |
|----------------------|--|
| PICTORIAL | gesti che riproducono una forma riconducibile a un'immagine o a un'azione |
| NON-PICTORIAL | gesti che riproducono movimenti ritmici o forme non assimilabili a oggetti nel mondo |
| CONVENTIONAL | gesti convenzionali a cui è possibile associare un valore semantico riconosciuto dalla comunità dei parlanti |

A nostro giudizio, il metodo di trascrizione e le etichette di annotazione gestuale utilizzate riescono a restituire un panorama sufficientemente completo del canale comunicativo gestuale. Come già accennato, l'intento del corpus non è quello di decifrare o classificare i gesti ma, piuttosto, è quello di creare uno strumento che permetta di indagare come la sincronizzazione dei diversi canali semiotici permettano ai parlanti di

comunicare. E crediamo, in fine, che indagare la lingua da un punto di vista multimodale debba tradursi nello studio del comportamento di unità di analisi dei diversi canali cercando di evitare categorizzazioni che rischiano di ridurre il gesto a mero ausilio del canale orale.

3.3 Dall'azione linguistica all'analisi multimodale

La Teoria della Lingua in Atto (Cresti, 2000), che rappresenta il quadro teorico entro cui si situa questo lavoro, si posiziona all'interno della tradizione di studi linguistici nati dal riconoscimento di Austin (1962) del fatto che parlare non è solamente dire ma anche fare. I discorsi che ogni parlante realizza quotidianamente hanno più o meno degli obiettivi che possono esprimere intenzioni o stati mentali e che possono produrre conseguenze di diverso tipo. L'analisi dei filosofi del linguaggio britannici – Austin e Searle nello specifico – ha preso in considerazione la complessità dell'azione del parlare distinguendo tre differenti atti: quello locutorio, quello illocutorio e quello perlocutorio (Sbisà, 2013).

Nel suo volume *How to do things with words*, in cui Austin illustra la sua teoria, lascia oscuri parecchi aspetti della nozione di atto linguistico. Primo fra tutti la definizione di atto illocutorio che definisce come forza dell'atto linguistico. Non chiarisce nemmeno la relazione tra le forze illocutorie e i verbi ed enunciati performativi; infine non chiarisce il problema relativo alla classificazione degli atti illocutori, in relazione soprattutto alla relazione tra le differenti classi e alla loro scalarità³² (Sbisà, 1994).

Senza addentrarci nel dibattito intorno alla nozione di atto linguistico e alla nozione di illocuzione, possiamo brevemente riassumere che per Austin (*ibidem*) l'illocuzione è quello che noi facciamo quando parliamo e che l'atto linguistico risulta dalla concomitanza di tre diversi atti che sono, appunto quello locutivo, quello illocutivo e quello perlocutivo. Per il

³² Per approfondimenti ulteriori rimandiamo a Sbisà 1972; 1983; 1989.

filosofo non sono tipi di uso della lingua che si escludono mutualmente, ma sono categorie che si applicano all'analisi di qualsiasi atto linguistico. Prendendo spunto dalla simultaneità degli atti, la Teoria della Lingua in Atto ha derivato la possibilità di porre in relazione l'atto linguistico con l'enunciato (Cresti, 2005). In questa relazione la prosodia è considerata l'interfaccia tra l'atto illocutivo e quello locutivo e rappresenta il mezzo necessario per trasdurre la concezione pragmatica in entità concreta e udibile, che è l'*utterance* (Cresti, 2020). Nella struttura dell'enunciato, la forza illocutiva è espressa dall'unità informativa di *Comment* e che prosodicamente corrisponde all'unità di *Root*.

Nel corso degli studi su corpora portati avanti dal gruppo di ricerca LABLITA, è stato costituito un repertorio di tipi illocutivi distribuiti su cinque classi generali identificate in:

- **Rifiuto:** un atteggiamento di libertà e separazione dal parlante dall'interlocutore, che permette uno scontro con quest'ultimo una richiesta di sua trasformazione;
- **Asserzione:** un atteggiamento di certezza del parlante nei confronti dell'interlocutore, sicurezza che consente di proporre giudizi, scoperte, valutazioni, rappresentazioni, come oggetti nuovi al mondo;
- **Espressione:** un atteggiamento di manifestazione "estetica" di stati d'animo, emozioni e credenze;
- **Rito:** un atteggiamento esterno di assolvimento di compiti linguistici che hanno effetti legali e sociali e che possono essere compiuti con la minima partecipazione affettiva.

Di seguito riportiamo la tabella con l'intero repertorio di tipi illocutivi identificati in seno alla Teoria della Lingua in Atto attraverso una ricerca basata su corpora:

Tabella 10 Repertorio dei valori illocutivi (Cresti 2020)

| ASSERTION | DIRECTION | EXPRESSION | RITE | REFUSAL |
|-------------------------------|--------------------------------------|---------------------------|------------------------|---------|
| Self-conclusion | Distal recall (non-visible addresse) | Contrast | Thanks | |
| On-going comment | Distal recall (vivable addresse) | Softening | Greetings | |
| Confirmation | Proximal recall | Expression of obviousness | Welcome | |
| Neutral assertion/explanation | Functional recall | Irony | Excuses | |
| Assertion taken for granted | Distal deixis (moving object) | Disbelief/doubt | Wishes | |
| Literal citation | Distal deixis (still object) | Admission | Congratulations | |
| Answer | Proximal deixis | Waiver/renouncement | Condolences | |
| Ascertainment | Prompt | Rhetorical question | Compliments | |
| Assertion of evidence | Event presentation | Protest | Legal declarations | |
| Hypothesis | Menatl deixis | Complain | Convictions | |
| | Instruction | Grumbling | Judgments | |
| | Introduction of person | Imprecation | Penalties | |
| | Request of agreement | Surprise/wonder | Results of examination | |
| | Self-correction | Wish/desire | Medical diagnoses | |
| | Reported speech | Easing | Dedications | |
| | Notification/warning | Approval/disapproval | Religious rite | |
| | Partial question | Derision | | |
| | Polar question | Provocation/challenge | | |
| | Alternative question | Reproach | | |

| | | | | |
|--|-----------------------------------|-----------------------------------|--|--|
| | Focalized question | Allusion/hint/negative suggestion | | |
| | Request of confirmation/agreement | Allowance/concession | | |
| | Order | Encouragement/support | | |
| | Interdiction | | | |
| | Prohibition | | | |
| | Invite | | | |
| | Offer | | | |
| | Agreement | | | |
| | Commitment | | | |
| | Proposal | | | |
| | Authorization | | | |

Abbiamo scelto di adottare l'annotazione dell'illocuzione anche per il nostro corpus ma operando, anche in questo caso, un adattamento. Abbiamo scelto di utilizzare le cinque classi generali senza entrare nel dettaglio della classificazione dei diversi tipi illocutivi, questa scelta è dettata dalla natura sperimentale dell'operazione. Infatti, l'annotazione del valore illocutivo non è stata realizzata per le sole unità di parlato ma abbiamo scelto di annotare il valore illocutivo che emerge dall'azione linguistica - il cui nucleo espressivo di riferimento rimane l'unità tonale di tipo Root - in quanto azione multimodale. Di seguito riportiamo il tag-set per l'annotazione delle classi illocutive.

Tabella 11 Tag-set dei valori illocutivi in CORMIP

| TAG | DEFINIZIONE |
|-------------------|-----------------------------|
| ASSERTION | Atto linguistico assertivo |
| DIRECTION | Atto linguistico direttivo |
| EXPRESSION | Atto linguistico espressivo |
| RITE | Atto linguistico rituale |

| | |
|----------------|-----------------------------|
| REFUSAL | Atto linguistico di rifiuto |
|----------------|-----------------------------|

Nello specifico la nostra annotazione dell'ilocuzione ha seguito tre principi:

- Se il parlante esprime l'azione linguistica attraverso l'uso esclusivo del canale verbale, viene annotato il valore illocutivo espresso dall'unità prosodica di tipo Root;
- Se il parlante esprime l'azione linguistica attraverso l'uso esclusivo del canale gestuale (compreso l'espressione faccia, o la co-occorrenza tra gesto manuale ed espressione facciale), viene annotato il valore illocutivo espresso dall'unità gestuale corrispondente;
- Se il parlante esprime l'azione linguistica attraverso l'uso concomitante del canale verbale e gestuale compreso l'espressione faccia, o la co-occorrenza tra gesto manuale ed espressione facciale, viene annotato il valore illocutivo espresso dalla co-occorrenza delle relative unità gestuale e verbali.

I principi appena elencati si basano su due presupposti. Il primo è che il valore illocutivo – che esprime l'intenzione comunicativa del parlante di natura affettiva – si trasforma fisicamente in azione linguistica attraverso il potenziale uso delle diverse modalità:

«Within L-Act³³ perspective, the ideation found in the speech origin is shaped by the speaker's affect toward the addressee and is physically transformed into a speech act containing a conventional pragmatic value via prosodic devices» (Cresti, 2020)

³³ L-Act (Language into Act Theory) è l'acronimo inglese utilizzato per riferirsi alla Teoria della Lingua in Atto.

E, aggiungiamo noi, anche attraverso l'uso dei gesti e dell'espressione facciale. Su questo argomento torneremo nel capitolo seguente, dove presenteremo un caso studio incentrato su questa tematica.

Il secondo presupposto è il giudizio dell'annotatore. Infatti, se da un lato la teoria della lingua in atto ha sistematizzato secondo quali parametri pragmatici, semiologici e cognitivi si definiscono i diversi valori illocutivi (Cresti, 2005; 2020) dall'altro mancano dei parametri corrispettivi per giudicare il valore illocutivi espresso dalla gestualità o dalla co-occorrenza di gesto e parlato. Pertanto, in questi casi, abbiamo fatto fede al giudizio analitico dell'annotatore in quanto, come già detto, ci avviciniamo all'annotazione dell'illocuzione nel nostro corpus in quanto sperimentale.

La natura sperimentale dell'operazione mira a utilizzare l'annotazione dell'azione linguistica come mezzo per indagare l'esistenza e le caratteristiche dell'unità multimodale. L'atto linguistico non rappresenta un'unità di analisi, ma un approccio metodologico di tipo pragmatico all'intenzione comunicativa espressa dai parlanti attraverso l'uso di diversi canali e/o della loro coordinazione. Al momento non siamo in grado di poter esprimere definizioni inerenti alle nozioni di atto linguistico multimodale, ma senza dubbio il nostro corpus rappresenta uno strumento utile a indagarlo.

Capitolo 6 – Il caso di studio: l'azione linguistica tra gesto e parola

1. Introduzione

In questo capitolo riportiamo uno studio che abbiamo realizzato analizzando i dati del nostro corpus. L'intento è duplice: da un lato mostrare come l'organizzazione del nostro corpus possa essere utile nell'analisi della lingua come sistema multimodale, e dall'altro vogliamo

portare una dimostrazione a sostegno della nostra proposta di utilizzare l'azione linguistica come approccio per indagare e definire l'unità di analisi multimodale. Abbiamo quindi scelto di indagare le correlazioni tra gli atti linguistici e i gesti nel nostro corpus. In particolare, vedremo come, da un punto di vista quantitativo, i gesti co-occorrono con i diversi valori illocutivi rintracciati nel corpus e soprattutto vedremo la distribuzione dei tipi gestuali in relazione alle cinque classi illocutive. Prima di analizzare i dati, ci soffermeremo sui principali contributi degli *gesture studies* sulle funzioni pragmatiche svolte dai gesti e sui loro valori illocutivi. Nello specifico vedremo i contributi di Kendon (1995; 2004), Bressemer e Müller (2014) e Enfield et al. (2007).

2. Gesti e pragmatica

I parlanti utilizzano diversi modi per realizzare le azioni linguistiche, in relazione al contesto e all'intenzione comunicativa. Come abbiamo visto nei capitoli precedenti, tra questi diversi modi rientrano anche i gesti sia nella loro espressione manuale che nell'espressività facciale. Si è soliti associare ai gesti una funzione sostantivale (Kendon, 1995) sulla base del ruolo giocato nella costruzione dei diversi aspetti del significato di un enunciato. Il contributo al significato enunciativo può assumere diverse forme e diverse metodologie semiotiche, possiamo avere sia un contributo più illustrativo e letterale che un contributo di tipo metaforico. La realizzazione dei gesti e la loro relazione con l'enunciato verbale non sono esclusivamente dipendenti dalla creazione del significato ma, come abbiamo visto nel Capitolo 2, diversi studi hanno dimostrato anche una forte dipendenza dei gesti con il loro contesto di realizzazione. Inoltre, è stato anche dimostrato che la presenza o l'assenza nell'interazione dei gesti influenza la ricezione e l'interpretazione degli enunciati da parte degli interlocutori. Possiamo dunque affermare che è possibile riconoscere ai gesti una funzione comunicativa di tipo pragmatica oltre che lessicale.

Già Kendon (1995) aveva individuato due differenti funzioni pragmatiche nei gesti convenzionalmente usati dai parlanti italiani, nello specifico campani³⁴. Le quattro principali strutture gestuali riconosciute dallo studioso sono le *mani a borsa*, le *mani giunte*, le *dita a grappolo*³⁵ e la *mano ad anello*, queste forme gestuali esprimono funzioni di marcatori illocutivi e di marcatori delle unità discorsive. I tipi di gesti seguono una precisa distribuzione in relazione alle funzioni: le *mani a borsa* e le *mani giunte* svolgono le funzioni di marcatori illocutivi, mentre le *dita a grappolo* e la *mano ad anello* svolgono funzioni di marcatori delle unità del discorso.

Alla prima struttura analizzata da Kendon (*ibidem*), la *mano a borsa*, è largamente riconosciuto la funzione di marcare una domanda come se servisse a rappresentare visivamente delle caratteristiche grammaticali o intonative del discorso. Ma il gesto può anche essere utilizzato in associazioni a enunciati che grammaticalmente non sono delle domande ma che tuttavia nel contesto conversazionale svolgono la funzione di domanda implicita. Inoltre, questa struttura gestuale può essere utilizzata per indicare una domanda in un discorso riportato. Queste funzioni vengono espresse sia dalla forma assunta dalla mano che dalla posa caratterizzante che consiste nella mano flessa sul polso in modo da avere il palmo rivolto verso il parlante. Infine, la *mano a borsa* può essere usata anche in assenza del parlato costituendo un enunciato, in questo caso Kendon riporta che viene utilizzato nel caso in cui un partecipante all'interazione si ritira da uno scambio di turni a causa di disapprovazione oppure per far volgere la discussione su un altro argomento. Le *mani giunte*, invece, accompagnano gli enunciati del parlato e rendono visibili alcune implicazioni che non possono essere esplicitate verbalmente. Questo gesto sembra esprimere principalmente la funzione di appello al destinatario al fine di accettare le condizioni implicite in ciò che viene detto, sollevando il parlante da certe responsabilità. In alcuni casi, il gesto serve come appello all'ascoltatore ad

³⁴ Il corpus di dati di Kendon è stato raccolto nel 1991 a Salerno in contesto di interazioni spontanee.

³⁵ Kendon usa il termine *finger bunch*.

accettare le conseguenze logiche di ciò che il parlante sta dicendo. Infine, le *mani giunte* sono usate per marcare l'implicazione del discorso di qualcun altro, che viene riportato.

Con l'etichetta *mano a grappolo*, Kendon fa riferimento a un gesto in cui la mano è tenuta con le dita unite, come per il gesto della *mano a borsa*, ma in questo caso si muove in avanti o verso il basso e lontano dal parlante, piuttosto che verso l'alto e verso il parlante. Molto spesso questo gesto co-occorre con la parte del discorso in cui il parlante sta specificando un *topic*, in questo caso la mano è spostata verso il basso e si apre in modo da estendere e allargare le dita. L'apertura della mano sembra avvenire in associazione con la parte del discorso che fornisce il *comment* del *topic* specificato. Nel gesto dell'anello la punta dell'indice è in contatto con quella del pollice, in modo che ci sia uno spazio circolare di forma di ovale tra di loro e le altre dita della mano sono, di solito, estese e divaricate. Il gesto dell'anello molto spesso occorre in associazione con un segmento di discorso che fornisce informazioni precise, o fa un riferimento specifico a qualcosa, crea un contrasto tra qualcosa di specifico o con qualcosa di più generale, o dà un esempio specifico di qualcosa (Kendon, 1995). Nei quattro gesti analizzati dallo studioso si possono facilmente rintracciare le due funzioni pragmatiche, quella di marcatori illocutivi (*mano a borsa* e *mani giunte*) e la funzione di marcatori di unità del discorso (*mano a grappolo* e il gesto dell'*anello*). Questi dati confermano quanto sostenuto da Kendon (1985, 1994, 1995) - e da noi stessi in questo lavoro - la necessità di considerare il gesto come una risorsa espressiva che può essere utilizzato in diversi modi rispetto agli obiettivi comunicativi del parlante.

Bressemer e Müller (2014) hanno individuato un repertorio di gesti ricorrenti nel tedesco che esprimono funzioni pragmatiche. I gesti presenti nel repertorio presentano diversi gradi di convenzionalizzazione e sono divisi in tre gruppi: a) i gesti che sono principalmente usati in co-occorrenza con il parlato e che possono assumere diverse varianti formali; b) i gesti che sono principalmente usati insieme al parlato ma che possono anche

sostituire il canale verbale e presentano anche un'alta variabilità formale; c) i gesti che mostrano un certo grado di emblematizzazione e sono usati sia in unione con il parlato che senza. I gesti ricorrenti del tedesco, inclusi in questo repertorio, svolgono anche la funzione di veicolare i valori illocutivi in relazione al loro contesto d'uso. Sulla base della classificazione degli atti illocutivi di Searle (1979), tutti questi gesti esprimono un valore di tipo assertivo, molto di frequente possono esprimere valori illocutivi di tipo direttivo e espressivo, mentre piuttosto di rado sono portatori di valori illocutivi di tipo commissivo. Oltre a esprimere i diversi valori illocutivi, i gesti ricorrenti evidenziano l'impatto e le conseguenze degli atti illocutori espressi (*ibidem*). Così, come Kendon (1995), Bressemer e Müller (2014) hanno individuato la possibilità di riconoscere alcuni gesti che incarnano la forza illocutiva dell'enunciato, esprimendo anche altre funzioni pragmatiche come marcatori di unità informative. Le studiose, quindi, affermano che la maggior parte dei gesti esprime più di una funzione pragmatica anche contemporaneamente.

Inoltre, ci sono gesti ricorrenti che incarnano la forza illocutiva dell'enunciato di chi parla e gesti che forniscono istruzioni per l'ascoltatore in relazione al suo agire, cioè su come tenere conto delle illocuzioni espresse agendo di fatto sul comportamento dell'interlocutore. A volte, i gesti possono esprimere anche il valore perlocutorio soprattutto negli usi in cui sostituiscono il parlato. In questo caso sono gesti che agiscono sul comportamento dell'altro esprimendo istruzioni su come agire. Così, dunque, un gesto può invitare l'interlocutore a non considerare ciò che è stato espresso oppure può esprimere la non conoscenza di determinate informazioni e chiedere dunque di concludere le domande (*ibidem*).

Enfield e colleghi (Enfield, Kita e De Ruiter 2007) – come abbiamo già accennato nel Capitolo 2 hanno studiato l'uso dei gesti deittici della popolazione del Laos. I ricercatori hanno distinto due tipi di gesti deittici sulla base della loro forma dividendoli in *B-point* (grandi nella forma) e *S-point* (piccoli nella forma). Questi gesti esprimono diverse funzioni in

relazione alla grandezza della loro forma. I gesti *B-point* trasmettono informazioni primarie rispetto al parlato, che invece sembra svolgere un ruolo di supporto al gesto. Mentre i gesti *S-point* mostrano una maggiore dipendenza nei confronti del parlato giocando un ruolo secondario nella struttura informativa dell'enunciato, Così dunque, esprimono informazioni secondarie, che rispondono a una possibile mancanza di conoscenze comuni tra gli interlocutori. Un *S-point gesture* rappresenta quindi un'azione comunicativa a basso rischio, che potrebbe salvare il parlante da un costo sociale e interpersonale potenzialmente elevato (Enfield 2006).

Questi lavori, pur nelle loro diversità metodologiche, dimostrano in modo chiaro le diverse funzioni pragmatiche espresse dal gesto. Da un lato Kendon (1995, 2004) e Bressemer e Müller (2014) individuano questa capacità espressiva nei gesti convenzionali, il cui grado di lessicalizzazione è alto. Possiamo quindi affermare che per questi studiosi la forza illocutiva espressa dal gesto emerge da valori semantici codificati in forme determinate. Mentre per Enfield e colleghi la capacità pragmatica, legata alla funzione di marcatore informativo del gesto, è espressa da una caratteristica formale del gesto che, in questo caso, è la grandezza interpretando il gesto come un elemento soprasegmentale.

3. I dati

Come abbiamo dichiarato nell'introduzione, l'obiettivo di questo studio è quello di analizzare la correlazione tra gesto e valore illocutivo all'interno dell'azione linguistica. Pertanto, abbiamo impostato il nostro studio partendo dalla distribuzione quantitativa delle illocuzioni all'interno del nostro corpus e del numero dei gesti. Il numero dei gesti è derivato dalla conta delle occorrenze delle *gesture phrases*, l'unità che individua il gesto nella sua completa esecuzione inglobando tutte le fasi che lo compongono.

Come è possibile notare nella tabella 12, la distribuzione quantitativa dei valori illocutivi segue i generi interazionali a causa della natura stessa dei tre diversi contesti comunicativi; infatti, il numero degli atti illocutivi aumenta con l'aumentare del numero dei parlanti presenti. Se da un lato la durata dello scambio comunicativo è mediamente uguale per tutte le registrazioni (cfr. capitolo 5) dall'altro le sovrapposizioni e il numero di scambi tra parlanti alimenta la crescita del numero di azioni compiute in ciascuna interazione. Così troviamo il 51% degli atti illocutivi nel genere conversazionale, il 30% nello scambio dialogico e il 19% nella comunicazione monologica.

Tabella 12 Numero di illocuzioni nel corpus

| ILLOCUZIONI NEL CORPUS | | |
|------------------------|-----|------|
| CONVERSAZIONALE | 464 | 51% |
| DIALOGICO | 272 | 30% |
| MONOLOGICO | 172 | 19% |
| Totale | 908 | 100% |

In questo studio terremo in considerazione tre delle cinque classi illocutive³⁶ poiché in CORMIP non sono stati riscontrati valori illocutivi riferibili alle classi RITE e REFUSAL. Come è possibile vedere nel grafico di seguito, abbiamo una notevole prevalenza di valori illocutivi di tipo assertivo che costituiscono il 61% dei valori presenti nel corpus e poi gli atti espressivi con il 23% e i direttivi con il 16%.

³⁶ Come riportato al capitolo 5, nel corpus sono stati annotati i valori illocutivi sulla base delle cinque classi generali individuate da Cresti (2000, 2005, 2020).

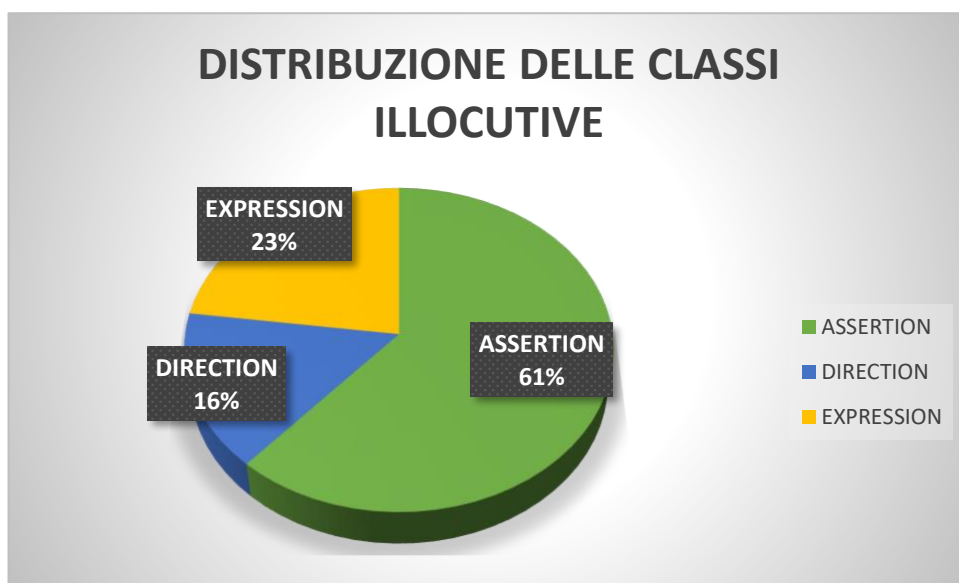


Figura 9 Distribuzione delle classi illocutive in CORMIP

Nella tabella 13 riportiamo la distribuzione delle classi azionali nelle diverse tipologie interazionali.

Tabella 13 Distribuzione delle classi illocutive in CORMIP

| | ASSERTION | DIRECTION | EXPRESSION |
|-----------------|-----------|-----------|------------|
| CONVERSAZIONALE | 26,9% | 10,7% | 13,5% |
| DIALOGICO | 17,6% | 4,3% | 8% |
| MONOLOGICO | 17,0% | 0,7% | 1% |

Per quanto riguarda i gesti la distribuzione non segue lo stesso andamento delle classi azionali. Come è riportato nella tabella 14, nella sezione conversazionale troviamo il 45% dei gesti del corpus, nello scambio dialogico il 23% e nella comunicazione monologica il 32%. Se la distribuzione delle forze illocutive vede decrescere il numero di occorrenze insieme al numero dei parlanti presenti nell'interazione, i gesti, invece, dimostrano una distribuzione diversa che concentra il maggior numero di occorrenze nello scambio conversazionale e una presenza maggiore nel genere monologico rispetto al dialogo. Questa differenza distribuzionale

tiene sicuramente conto dei contesti comunicazionali, riteniamo plausibile l'ipotesi che il contesto narrativo del brano MONO_Racconto-CT e istruttivo del brano MONO_Lezione-FI possano condizionare la produzione gestuale per fini illustrativi. Ma data la non rappresentatività del nostro campione crediamo che questa ipotesi meriti ulteriori approfondimenti.

Tabella 14 Distribuzione dei gesti nel corpus

| GESTI NEL CORPUS | | |
|------------------|-----|------|
| CONVERSAZIONALE | 272 | 45% |
| DIALOGICO | 143 | 23% |
| MONOLOGICO | 195 | 32% |
| Totale | 610 | 100% |

Nel grafico riportato in figura 10 è riportata la distribuzione dei tipi di gesti³⁷ all'interno del nostro corpus. Possiamo riscontrare che la quasi totalità dei gesti realizzati dai parlanti è di tipo non-pictorial e che i gesti con un'alta iconicità sono una esigua presenza insieme ai gesti convenzionali. Giudichiamo che questo dato sia indicativo, dimostrando che l'analisi e lo studio dei gesti necessiti di strumenti adatti all'analisi del movimento.

³⁷ Per i tipi di gesti rimandiamo al capitolo 5.

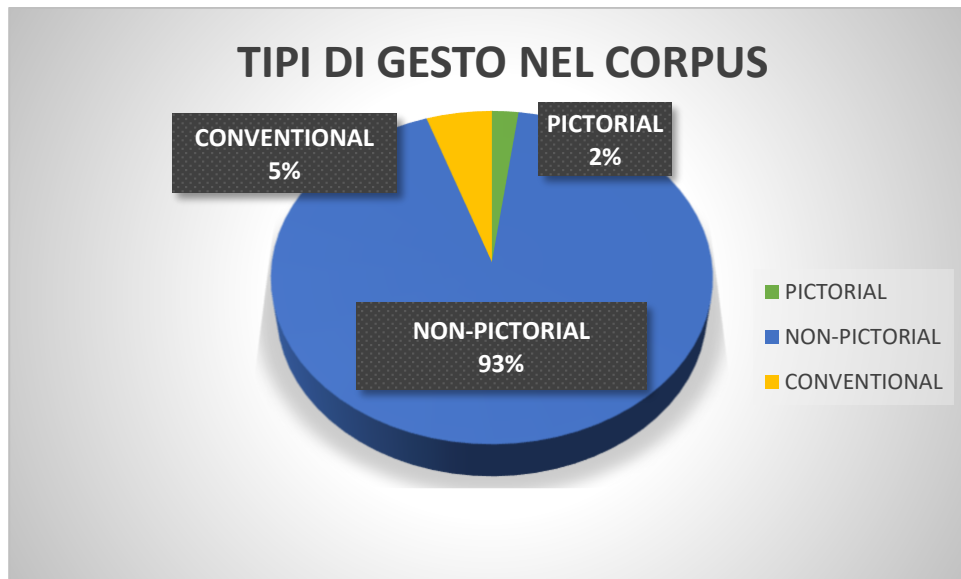


Figura 10 Distribuzione dei tipi gestuali in CORMIP

Nella tabella 15 riportiamo la distribuzione dei diversi tipi di gesti nei generi interazionali.

Tabella 15 Distribuzione dei tipi di gesto nei diversi tipi di interazioni in CORMIP

| | PICTORIAL | NON-PICTORIAL | CONVENTIONAL |
|-----------------|-----------|---------------|--------------|
| CONVERSAZIONALE | 1,6% | 40,0% | 3,0% |
| DIALOGICO | 0,3% | 21,3% | 1,8% |
| MONOLOGICO | 2,1% | 31,3% | 0,5% |

Questi dati ci mostrano come gesti e azione linguistica siano fortemente legati anche nella ripartizione all'interno del corpus seguendo la stessa tendenza distribuzionale sulla base dei generi comunicativi. Cercando di analizzare quanto siano legati gesti e azioni linguistiche, abbiamo estratto il numero di gesti che vengono prodotti insieme ad almeno un valore illocutivo; i gesti, infatti, possono a volte occorrere con più di un valore illocutivo.

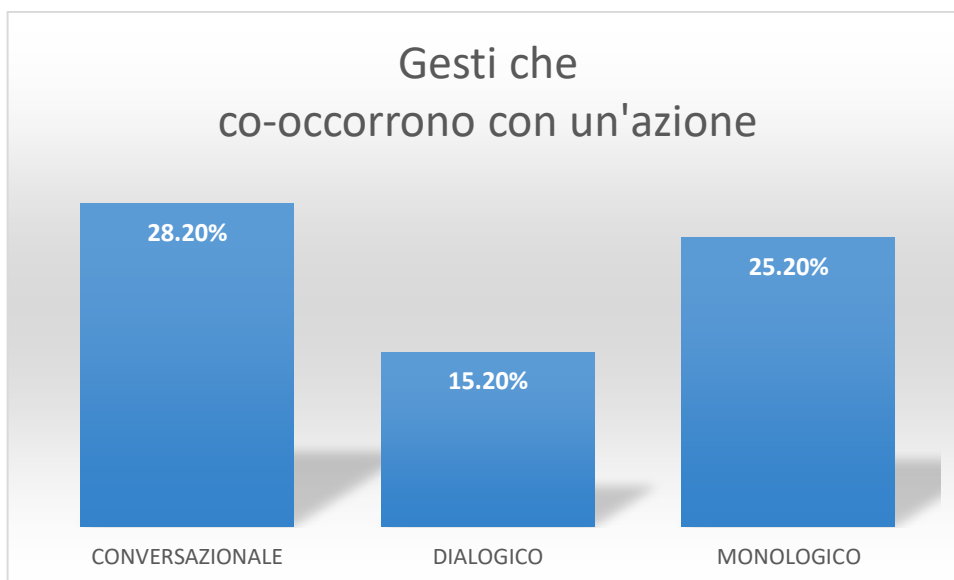


Figura 11 Percentuale di gesti che occorrono insieme a un'azione linguistica in ciascun genere interazionale

Tabella 16 Gesti che occorrono insieme a un'azione linguistica

| INTERAZIONE | OCCORRENZE | PERCENTUALE |
|-----------------|------------|-------------|
| CONVERSAZIONALE | 172 | 28,2% |
| DIALOGICO | 93 | 15,2% |
| MONOLOGICO | 154 | 25,2% |
| Totale | 419 | 68,6% |

Da questi dati risulta che quasi il 69% dei gesti viene realizzato in concomitanza con l'espressione di un valore illocutivo e la quasi totalità di questi gesti è prodotta insieme al parlato. Infatti, come è riportato nella figura 12, il 98,6% dei gesti che co-occorre con l'espressione di un'illocuzione è prodotto insieme al parlato. Questo dato se da un lato dimostra che la produzione dei gesti è fortemente legata alla produzione del parlato, dall'altro lato ci indica che i parlanti attraverso la produzione di un gesto, senza la controparte verbale, riescono a essere efficacemente comunicativi.

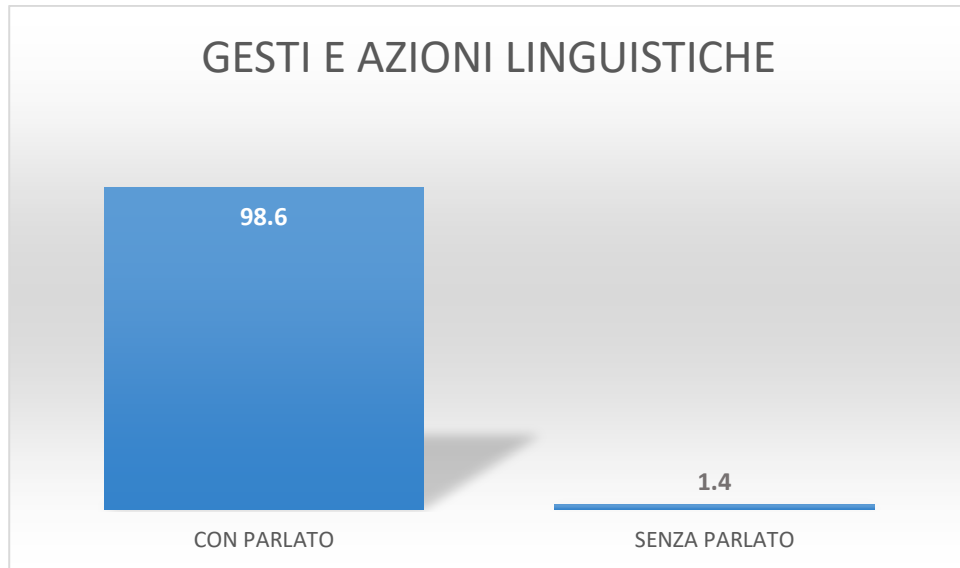


Figura 12 Gestì con valore illocutivo e parlato

Questo dato potrebbe apparire in controtendenza rispetto alla tabella 1, ma il fatto che un gesto venga realizzato in concomitanza con la produzione di almeno un'ilocuzione non implica che il gesto esprima un valore pragmatico. Abbiamo così analizzato le singole occorrenze di gesti che co-occorrono con un'ilocuzione e distinguendoli in due categorie: gesti che esprimono un valore pragmatico, e che quindi costruiscono l'azione linguistica, e gesti che non esprimono un valore pragmatico. Questi ultimi molto spesso esprimono una funzione di tipo cognitivo o rappresentativo. Nei risultati che riportiamo nella tabella 17, infatti, è possibile notare che il 36,5% dei gesti prodotti nel corpus è di tipo pragmatico: la distribuzione nei diversi generi interazionali vede una concentrazione lievemente maggiore nella conversazione con il 16,6% e una percentuale simile nell'interazione dialogica e monologica, intorno al 10%.

Tabella 17 Percentuale di gesti con valore pragmatico in CORMIP

| INTERAZIONE | PERCENTUALE |
|-----------------|-------------|
| CONVERSAZIONALE | 16,6% |
| DIALOGICO | 9,7% |
| MONOLOGICO | 10,2% |
| Totale | 36,5% |

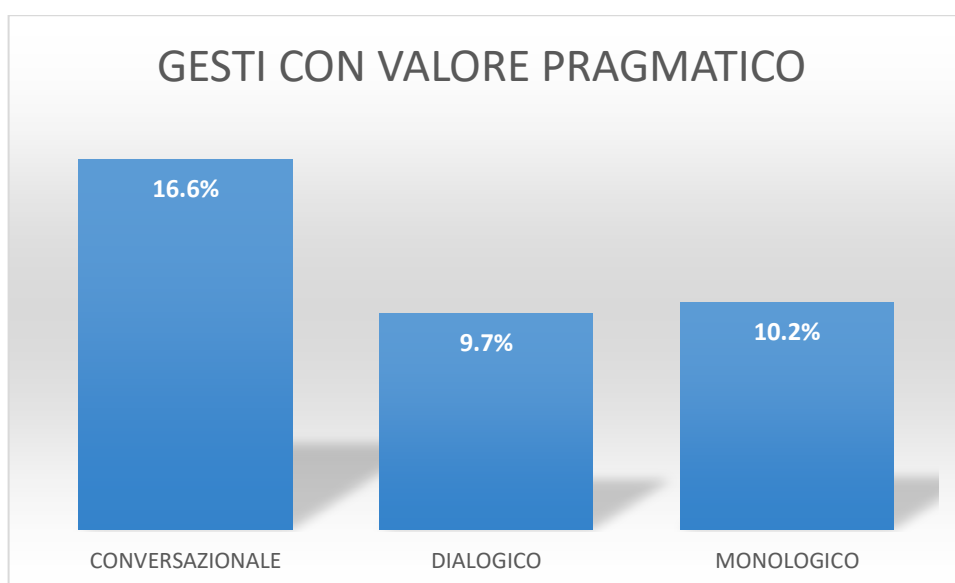


Figura 13 Gesti con valore pragmatico nei diversi tipi di interazioni

Come già dimostrato da Kendon (1995, 2004) e da Bressem e Müller (2014), i gesti che esprimono un valore pragmatico possono svolgere diverse funzioni. In questo studio li abbiamo divisi in gesti che esprimono un valore illocutivo, gesti che hanno una funzione testuale – e quindi marcano le unità informative o le prominenze del parlato, oppure svolgono una funzione di *parsing testuale* (sia a livello lessicale che a livello sintattico) – e gesti che esprimono sia un valore illocutivo che funzione testuale contemporaneamente (per esempio quei gesti che a livello ritmico marcano il parsing di ciò che esprimono, mentre forma e movimento enfatizzano o esprimono un valore illocutivo).

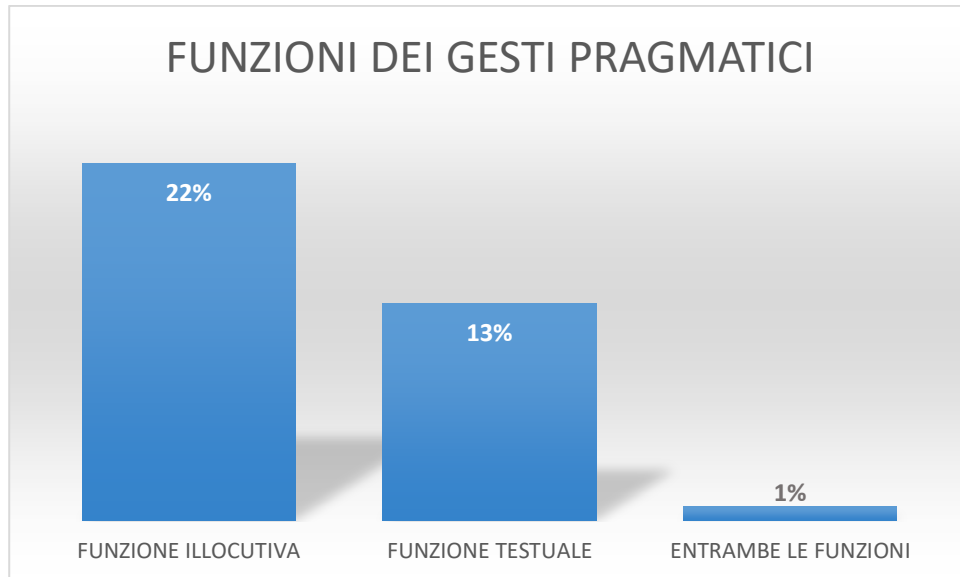


Figura 14 Gestii pragmatici per funzioni espresse

Pertanto, possiamo affermare che il 36% dei gesti nel nostro corpus svolge una funzione pragmatica. Di questi il 22% esprime una funzione illocutiva, il 13% una funzione di tipo testuale e l'1% esprime entrambe le funzioni. I dati riportati in questa analisi ci dimostrano che le funzioni pragmatiche rientrano tra le principali espresse dai gesti. Questo non esclude che il numero di movimenti gestuali compresi in queste percentuali esprima esclusivamente funzioni pragmatiche. Anzi, come vedremo nel paragrafo successivo, il gesto deve essere considerato nella sua globalità; ha infatti la capacità di esprimere più funzioni in quanto elemento linguistico che compone un enunciato. Così come in una stringa di parlato è possibile riconoscere diverse funzioni sulla base dei diversi livelli di lingua, lo stesso è possibile anche per un movimento gestuale. A ciò si aggiunge che le diverse funzioni espresse sono da noi considerate da un punto di vista multimodale. Pertanto, le diverse funzioni espresse dai gesti vanno contestualizzate per comprendere il peso che hanno nel determinare il significato dell'enunciato. Nel paragrafo successivo discuteremo alcuni esempi per cercare di comprendere come i parlanti costruiscono le azioni linguistiche multimodali.

4. Discussione dei dati

I dati estratti dal nostro corpus mostrano in modo evidente che i parlanti usano tutte le strategie a loro disposizione per esprimere le proprie esigenze comunicative. Un ruolo principale in queste strategie è giocato dal parlato nella sua interezza, dal lessico alla prosodia, dalla gesticolazione in tutte le sue forme e anche dall'espressione facciale. Questo dimostra, altresì, che i gesti e le espressioni facciali non sono modalità comunicative subordinate al parlato, ma sono strumenti che i parlanti utilizzano e organizzano in unione alla modalità orale nella costruzione delle azioni linguistiche. In questo paragrafo riporteremo alcuni esempi che ci mostrano come le diverse modalità vengono gestite dai parlanti. Inizieremo discutendo due enunciati in cui i gesti pragmatici esprimono funzioni testuali, poi vedremo tre esempi di enunciati che esprimono l'illocuzione attraverso l'uso del canale orale e gestuale e chiuderemo con due esempi in cui la costruzione dell'azione linguistica fa emergere come i gesti esprimano diverse funzioni contemporaneamente.

Partiamo col discutere i due esempi che ci dimostreranno come la struttura testuale e informativa dell'enunciato sia costruita dal parlante attraverso l'uso della prosodia e della gestualità.



Figura 15 Fotogrammi estratti dall'esempio 1

Esempio (1) Gesto e parlato costruiscono la struttura testuale e informativa
[\(DIA_scout-FI - 02':34'' - 02':46''\)](#)

il passaggio dal / idea di politica come impegno + allora / prima c'è l'informazione /

```
%PREFIX      %INTERRUPTED          %INCIPIIT    %ROOT
              %G-PHR----- | %G-PHR-----
              %Non-Pictorial          %Non-Pictorial
                                      %ASSERTION
```

poi c'è l'idea di politica come impegno politico associazioni eccetera eccetera /

```
%ROOT
----- | %G-PHR----- | %G-PHR-----
          %Non-Pictorial          %Non Pictorial
%ASSERTION
```

e poi c'è quotidianità //

```
%ROOT
----- | %G-PHR----- |
          %Non-Pictorial
%ASSERTION
```

Come possiamo vedere nell'esempio 1 il parlante, dopo aver interrotto un primo tentativo di enunciazione in cui le mani si trovano in posizione di riposo, costruisce un nuovo enunciato anche attraverso l'uso dei gesti. L'enunciato ha un'unità tonale incipitaria seguita da tre unità tonali di tipo ROOT, rispetto al canale gestuale abbiamo cinque *gesture phrase* di cui tre realizzate esclusivamente con la mano sinistra e due con entrambe le mani. Come è possibile vedere (anche nel video nel link), il parlante enuncia un elenco ordinato di priorità di impegni di natura morale. Nella costruzione dell'elenco, i gesti seguono in modo parallelo la struttura prosodica - e quindi informativa - dell'enunciato: il primo gesto è bruscamente interrotto in corrispondenza con l'interruzione del primo tentativo di enunciazione; il

secondo gesto – costituito da uno stroke che vede il parlante portare la mano (piatta con il palmo rivolto verso il basso) in alto e da uno stroke successivo in cui la mano viene abbassata – accompagna il primo elemento dell’elenco (primo stroke = *prima c’è l’informazione*) e sfocia in parte nella seconda unità ROOT abbassandosi (secondo stroke = *poi c’è*); il terzo gesto che corrisponde alla seconda unità ROOT vede la mano sinistra mantenuta nella posizione acquisita durante il gesto precedente mentre la mano destra viene sollevata parallelamente alla mano sinistra assumendone la stessa forma, in questo gesto la mano sinistra viene mossa parallelamente al pavimento per poi rimanere in fase hold in corrispondenza del secondo elemento dell’elenco (*poi c’è l’idea di politica come impegno politico associazioni eccetera eccetera*); il quarto gesto rientra in gran parte nell’ultima unità intonativa che è al suo interno scansionata, infatti il parlante introduce l’ultimo elemento dell’elenco con un gesto e una stringa introduttiva (stroke = *e poi c’è*); infine l’ultimo gesto, costituito da tre stroke, scandisce quasi sillabicamente la parola pronunciata (*quotidianità*) seguendo la stessa scansione che il parlante riproduce vocalmente. In questo esempio è riportata una costruzione complessa di un elenco e dimostra in modo chiaro come il parlante scansioni i diversi elementi dell’elenco, in ordine anche alla loro importanza, attraverso la scansione intonativa e gestuale.

Nell’esempio 2, riportiamo un brano in cui i gesti scansionano, insieme all’intonazione, due concetti espressi per rispondere a una domanda dell’interlocutore.



Figura 16 Fotogrammi estratti dall’esempio 2

Esempio (2) Gestii e intonazione svolgono insieme la funzione testuale
([DIA_students-CT - 2':13'' - 2':16''](#))

SAR: *immagina la chimica / e la fisica insieme hhh //*

```
%ROOT      %ROOT
      G-PHR-----|G-PHR|GPHR-----|
      %Non-Pic  %Non-Pic %Non-Pic
%ASSERTION  %ASSERTION
```

Nello specifico, la parlante risponde a una domanda implicita del suo interlocutore su che cosa sia la disciplina chimica fisica. La parlante nel rispondere scansiona intonativamente l'enunciato dividendolo in due unit  di tipo ROOT (*immagina la chimica / e la fisica insieme hhh //*) e accompagnando la scansione con tre gesti diversi. La particolarit  di questo esempio   data dal fatto che i gesti, oltre a scansionare i due elementi, li rappresentano esplicitando in modo gestuale che la chimica fisica   una composizione delle conoscenze chimiche e fisiche. Questo esempio dimostra che i parlanti utilizzano i gesti facendo esprimere diverse funzioni contemporaneamente, in questo caso c'  una funzione di scansione testuale, che sembra marcare informativamente le stringhe di parlato, e una rappresentativa.

Nell'esempio 3   possibile vedere come i parlanti esprimono il valore illocutivo anche attraverso l'uso dei gesti e dell'espressione facciale.

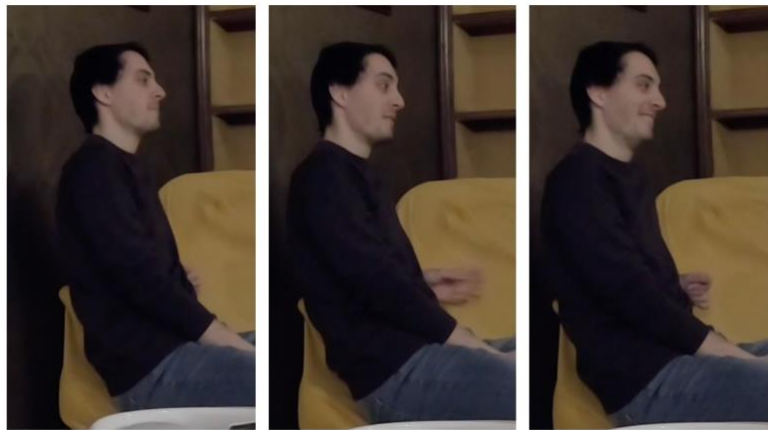


Figura 17 Fotogrammi estratti dall'esempio 3

Esempio (3) Gestì ed espressione facciale esprimono il valore illocutivo ([CONV arbitri pallamano-FI 4':10 - 4':15''](#))

MAR: *solo per fermarsi / fa tre passi //*

```

%PREFIX      %ROOT
      G-PHR-----|
      %Non-Pic
            %EXPRESSION
  
```

In questo caso, il parlante accompagna l'enunciazione con un piccolo gesto della mano sinistra e un'espressione facciale che determinano il valore dell'illocuzione. Infatti, ascoltando solamente l'audio si potrebbe classificare l'enunciato come assertivo ma il movimento della mano e l'espressione facciale rendono chiaro che l'illocuzione espressa è di tipo espressiva. Il parlante afferma in modo ironico che un determinato giocatore è solito fare tre passi per fermarsi³⁸.

Nell'esempio 4 riportiamo un enunciato in cui la parlante a destra dà conferma agli interlocutori del fatto che la miniera che andranno a visitare è ancora in attività.

³⁸ Il gioco di cui si discute è la pallamano, sport in cui il numero dei passi è strettamente regolato.



Figura 18 Fotogrammi estratti dell'esempio 4

Esempio (4) Gesto e parlato esprimono la stessa illocuzione ([CONV_viaggio-CT 3':36'' - 3':37''](#))

si perché è in attività //

%ROOT

G-PHR----- |

%Non-Pic

%ASSERTION

In questo caso abbiamo un enunciato costituito da una unità intonativa e un gesto che esprimono insieme la stessa illocuzione di conferma. L'illocuzione è costruita dalla parte gestuale attraverso un gesto delle mani aperte che assumono la forma Palm Open Hand (Bressemer e Müller, 2014) e dall'espressione facciale.

Nell'esempio 5 riportiamo un enunciato in cui vengono espresse due illocuzioni, una di tipo assertivo e l'altra di tipo espressivo, che si costruiscono con strategie diverse.



Figura 19 Fotogrammi estratti dall'esempio 5

Esempio (5) Costruzione dell'ilocuzione con diverse strategie ([DIA studenti-CT - 4':51'' - 4'54''](#))

LUI: *un punto se hai fatto l'erasmus //*

| | |
|------------|-------------|
| %ROOT | |
| G-PHR----- | G-PHR----- |
| %Conv | %Non-Pic |
| %ASSERTION | %EXPRESSION |

Il parlante sta facendo un elenco di possibilità che possano portare l'interlocutrice ad aumentare il proprio voto di laurea. Lo scambio è interamente costruito su una latente ironia, l'esempio che qui riportiamo ne è una parte. Nello specifico vediamo che il parlante, con un gesto convenzionale di enumerazione, elenca le diverse voci che compongono il voto di laurea così da poter prevedere il punteggio della sua interlocutrice. Nell'estratto riportato in esempio il parlante fa riferimento alla partecipazione al progetto erasmus, che potrebbe garantire un punto in più. Qui è interessante notare come la parlante SAR risponda esclusivamente attraverso l'espressione facciale e di conseguenza, il parlante LUI (oggetto del nostro esempio) esprime ironicamente una rassegnazione attraverso il gesto (l'ultima Phrase) che vede la mano destra indicare la mano sinistra ferma in posizione di HOLD, come se stesse a rappresentare la possibilità di un punto in più, e successivamente il parlante allarga le braccia esprimendo rassegnazione mentre con l'espressione facciale esprime l'ironia.

Negli ultimi due esempi, infine, i gesti vengono utilizzati dai parlanti per esprimere insieme le due funzioni pragmatiche. Nell'esempio 6 vediamo la parlante, al centro della scena, che esprime il valore illocutivo di tipo assertivo usando il gesto della mano, l'espressione facciale e la prosodia.



Figura 20 Fotogrammi estratti dall'esempio 6

Esempio (6) Gestì svolgono funzione testuale e funzione illocutiva
 ([CONV arbitri pallamano-FI 2':57'' - 2':58''](#))

GIU: *c'era la donna //*

%ROOT

G-PHR-----|

%Non-Pic

%ASSERTION

Andando ad analizzare in modo più specifico, vediamo la forma della mano sinistra - palmo rivolto lateralmente con indice e pollice stesi e il resto delle piegate a metà - determinare il valore illocutivo mentre il movimento sembra scansionare l'unità prosodica enfatizzando la prominenzza con il secondo stroke.

Infine, nell'esempio 7 viene riportato un enunciato costituito da due unità intonative e due gesti che scansionano la struttura intonativa e informativa.

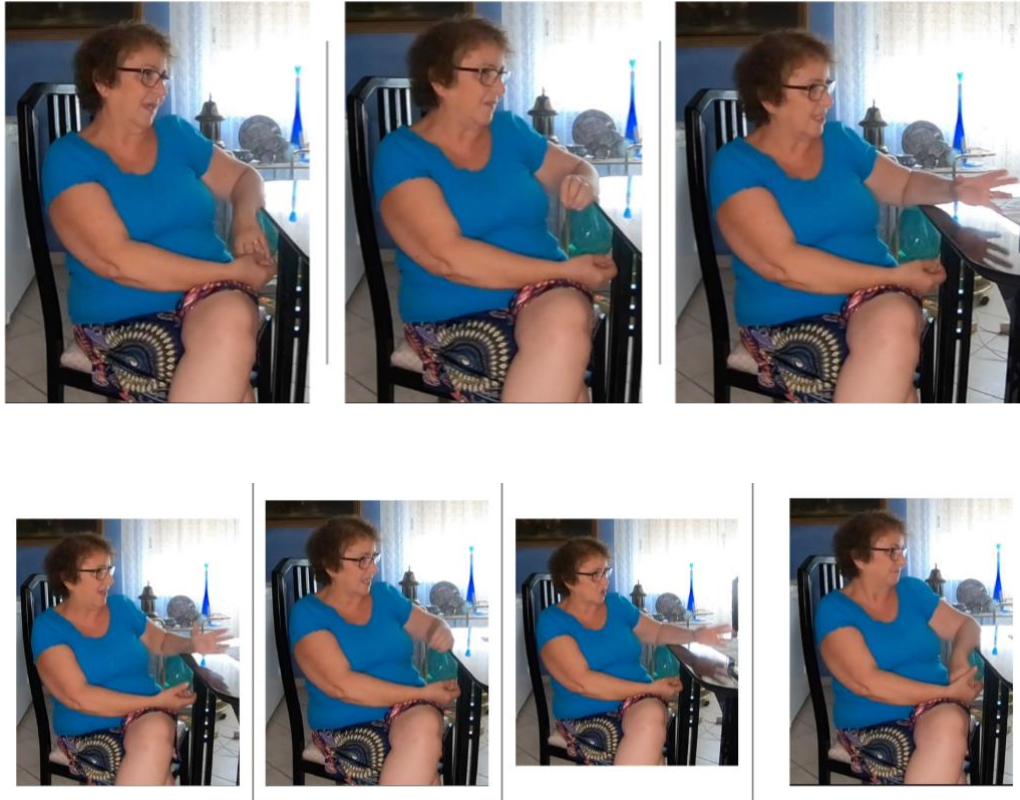


Figura 21 Fotogrammi estratti dall'esempio 7

Esempio (7) I gesti scansionano la struttura intonativa e informativa dell'enunciato
[\(MONO_racconto-CT 3':39 - 3':43''\)](#)

ELI: magari li avrei visti lo stesso / due volte l'anno //

| | |
|------------|------------|
| %ROOT | %SUFFIX |
| G-PHR----- | G-PHR----- |
| %Non-Pic | %Non-Pic |
| %ASSERTION | |

Ma l'aspetto particolarmente interessante di questo esempio è il primo gesto, co-articolato insieme all'unità ROOT. Infatti, in questa occorrenza del gesto è possibile notare come la forma collabori all'espressioni dell'illocuzione assertiva mentre il movimento scandisce il ritmo dell'enunciazione di "lo stesso /" che risulta così prominente all'interno dell'enunciato.

5. Conclusione

All'interno degli *gesture studies* l'ipotesi che i gesti abbiano un valore pragmatico e che, in particolare, esprimano i valori illocutivi è stata verificata da diversi studi (Kendon, 1995-2004; McNeill, 1992; Bressem e Müller, 2014). In particolare, è stato riconosciuto che i gesti ricorrenti e/o altamente convenzionalizzati siano preposti alla codifica dei valori illocutivi così, secondo questa prospettiva, l'azione linguistica sembra determinarsi in base alla scelta lessicale e all'uso di gesti che sembrano essere quasi del tutto lessicalizzati.

Dai dati estratti dal nostro corpus le funzioni pragmatiche sembrano essere tra le principali espresse dai gesti. In particolare, il 36% delle occorrenze di *gesture phrase* svolgono funzioni di scansione testuali o azionali. Questa percentuale, in relazione alla bassissima percentuale di gesti convenzionali, ci porta a sostenere che i gesti trasmettono l'illocuzione attraverso differenti strategie oltre all'uso delle forme convenzionali. Dal nostro punto di vista, risulterebbe più opportuno affermare che l'azione linguistica è multimodale e pertanto non è esclusivamente determinata dal gesto, tranne quando quest'ultimo occorre in assenza di parlato. Gli esempi che abbiamo riportato hanno dimostrato come l'organizzazione del flusso gestuale e del flusso parlato avvenga in concomitanza e secondo le necessità comunicative contestuali del parlante. Inoltre, abbiamo visto come i gesti esprimano diverse funzioni contemporaneamente secondo modalità che appaiono strutturate e dipendenti sia da fattori contestuali che da precise strategie comunicative realizzate dai parlanti. Tutto ciò sostiene e avvalorava la nozione di enunciato multimodale, già definita da Kendon (2004), Enfield (2009) ma al momento non ci permette di riconoscere secondo quali parametri possa essere definita e segmentata. Questo studio ci ha portato a riconoscere alcune caratteristiche e alcune strategie utilizzate dai parlanti per la realizzazione di enunciati multimodali.

Così, a conclusione di questa analisi possiamo affermare che:

- l'azione linguistica si realizza attraverso l'uso di diversi indici e canali e dunque possiamo affermare che l'azione linguistica è multimodale;
- il gesto e l'espressione facciale collaborano alla determinazione dei valori illocutivi riuscendo a modificare quanto espresso esclusivamente dal parlato;
- i valori illocutivi vengono espressi e codificati, oltre che dalla prosodia, dal tipo di movimento, dalla forma del gesto e dall'espressione facciale;
- i parlanti possono realizzare un'azione linguistica attraverso l'uso esclusivo della modalità gestuale;
- la funzione di scansione testuale è tendenzialmente espressa dal movimento del gesto.

Possiamo quindi concludere che i parlanti comunicano attraverso l'uso di diversi canali e indici rappresentati dalla voce, dalle mani, dalle braccia e dal viso. La lingua è dunque multimodale e pertanto da un lato è necessario costruire corpora multimodali di parlato sempre più grandi per comprendere al meglio le strategie messe in campo e dall'altro è indispensabile affinare i metodi di studio per individuare specifici parametri per l'analisi del movimento.

Conclusioni

Il lavoro di ricerca che abbiamo riportato in queste pagine ha avuto l'intento di realizzare un corpus multimodale pilota dell'italiano parlato in contesto spontaneo. Come abbiamo visto, questo ci ha portato a indagare e comprendere nel dettaglio la complessità del fenomeno linguistico. Abbiamo affrontato la questione partendo da una rassegna di studi sui due principali canali semiotici utilizzati da parlanti: il parlato e il gesto.

Rispetto al parlato è emersa una sfaccettatura di approcci al fenomeno che ne determinano differenze rispetto agli aspetti da indagare. L'interesse che riguarda questo lavoro ci ha portato ad approcciarci al parlato come fenomeno pragmatico che vede i parlanti realizzare le loro azioni linguistiche attraverso l'uso dell'intonazione e degli strumenti lessicali e morfo-sintattici.

Per il gesto è emersa una situazione ugualmente complessa e variegata a quella del parlato. L'interesse nei confronti del gesto è testimoniato sin dai romani e si è esteso fino a oggi con una incredibile varietà di interessi e obiettivi di indagine. Solo negli ultimi decenni la gestualità è diventata oggetto di studio della linguistica a seguito di un lungo dibattito sulla sua linguisticità, sulla sua natura verbale divenendo così elemento costitutivo dell'enunciazione e non linguaggio a sé (e di grado inferiore rispetto al parlato) o elemento aggiuntivo del parlato.

Tutto ciò ci ha portato a comprendere la complessità del fenomeno lingua e delle questioni legate alla sua natura multimodale. Infatti, la definizione di multimodalità della lingua e una letteratura sul gesto relativamente giovane e molto sfaccettata ci hanno posto davanti a tre fondamentali questioni. La prima è data dalla definizione di lingua. Infatti, ancora oggi prevale negli studi linguistici la nozione tradizionale lingua, che non vede il linguaggio come sistema multimodale. Questo fa sì che a oggi, e questo è la seconda questione, i linguisti non sono riusciti a definire un'unità

linguistica multimodale. Nelle pagine precedenti abbiamo visto alcuni tentativi di definizione che però rimangono a un livello di analisi esclusivamente teorico e non permettono di segmentare il flusso linguistico (gestuale e parlato) nel suo realizzarsi. Infine, terza questione, la mancanza di uno standard di analisi e trascrizione del gesto. Questo ha portato da un lato al proliferare di tipologie di corpora diversi e dall'altro all'impossibilità di condividere dati e risultati tra ricercatori.

Ponendoci davanti a questo quadro abbiamo deciso di comprendere al meglio come poter analizzare il gesto sulla base dei principali approcci emersi: quello cognitivo, quello linguistico, quello conversazionalista e quello kinesiological. Dal canto nostro abbiamo deciso di approcciarci alla lingua come sistema multimodale da un punto di vista pragmatico e percettivo. Così, abbiamo costruito il nostro corpus basandoci sulle unità di analisi individuate dal quadro teorico della Teoria della Lingua in Atto (Cresti 2000), per il parlato, e dall'architettura del gesto di Kendon, per il gesto. Due quadri teorici che permettono di segmentare il flusso del parlato e del gesto in unità percettivamente riconoscibili: da un lato le unità prosodiche e dall'altro le unità del movimento gestuale. L'approccio pragmatico ci ha permesso, da un lato di mantenere separati i due canali (gestuale e parlato) così da poter indagare come ciascuno di essi costruisce le informazioni e come si struttura in unità più grandi, e dall'altro di indagare come queste due modalità, pur nelle loro differenze, interagiscono nella costruzione dell'enunciato.

Pertanto, come la lingua nella sua modalità parlata articola e costruisce le informazioni e i contenuti semantici attraverso i diversi livelli segmentali e soprasegmentali, così avviene anche per la modalità gestuale. In questo modo, gesto e parlato costituiscono un unico sistema che esprime una sola azione linguistica che si costruisce su più livelli di lingua. Così, dunque, il corpo diventa strumento di espressione linguistica che usa tutte le sue potenzialità articolatorie: quelle orali, usate in modo principale e più articolate, e quelle gestuali, meno usate e meno articolate. Rispetto a ciò, nel

costruire il nostro corpus, abbiamo focalizzato la nostra attenzione al parlato come fenomeno intonativo, al gesto nei movimenti degli arti superiori e dell'espressione facciale.

Per poter indagare la lingua come sistema multimodale, e studiare le interazioni tra la modalità gestuale e orale, abbiamo deciso di annotare l'azione linguistica a livello multimodale utilizzando le classi illocutive come mezzo di indagine.

In conclusione, possiamo affermare che questo nostro approccio ci ha permesso di costruire un corpus pilota - che raccoglie sei brani di interazioni spontanee in contesti diversi e in diverse tipologie di interazioni - e di indagare le interazioni comunicative da un punto di vista multimodale senza però creare uno strumento che possa apparire centrato maggiormente su una delle due modalità. Questo è stato possibile grazie all'approccio percettivo e pragmatico al fenomeno lingua. Inoltre, in questo lavoro emerge l'esigenza di corpora sempre più grandi contenenti dati spontanei e realizzati attraverso l'uso di standard largamente condivisi. Questo può portare a una maggiore capacità di indagine dell'intera comunità scientifica affinché la lingua possa essere indagata nella sua complessità e si possa pian piano cercare di mettere ordine nell'apparente caos della comunicazione umana.

BIBLIOGRAFIA

't Hart J., Collier R., Cohen A. (1990), *A Perceptual Study of Intonation. An Experimentel-Phonetic Approach to Speech Melody*, Cambridge University Press, Cambridge.

't Hart Johan (1981), *Differential Sensitivity to Pitch Distance, Particularly in Speech*, in "Journal of Acoustical Society of America", 69, pp. 811-821.

Albano Leoni F. e Maturi P. (1992), *Per una verifica pragmatica dei modelli fonologici*, in Gobber 1992, pp. 39-49.

Argyle M. (1991), *Il corpo e il suo linguaggio*, Zanichelli, Bologna.

Austin J.L. (1962), *How to do things with words*, Oxford, Oxford University Press

Baldry, A. and Thibault, P. J. (2001), *Towards multimodal corpora*. In Aston, G. and Burnard, L. (eds), *Corpora in the Description and Teaching of English- Papers from the 5th ESSE Conference*. Bologna: Cooperativa Libreria Universitaria Editrice Bologna, pp. 87-102.

Baldry, A. and Thibault, P. J. (2006), *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Cours*. London: Equinox.

Barsalou L. (2003), *Situated simulation in the human conceptual system*. In *Language and Cognitive Processes*, 18, pp. 513-562.

Bates E., Camaioni L., Volterra V. (1975), *The Acquisition of Performatives prior to Speech*, in "Merrill-Palmer Quarterly of Behavior and Decelopment", 21, 3, pp. 205-226.

Bateson G, Mead M. (1942), *Balinese Character: A Photographic Analisis*, in Willbur G. Valentine (a cura di) *Special Publication of the New York Academy of Sciences*, Vol. II, New York Academy of Sciences, New York.

Bavelas J. B. et al. (2008), *Gesturing on the Telephone: Independent Effect of Dialogue and Visibility*, in "Journal of Memory and Languaga", 58, pp. 495-520.

Bazzanella C. (1994), *Le facce del parlare. Un approccio pragmatico all'italiano parlato*. LA NUOVA ITALIA Collana: Biblioteca di Italiano e oltre.

- Bazzanella C. (2014), *Linguistica cognitiva: un'introduzione*, Bari: Laterza.
- Beckman M. E., Pierrehumbert J. B. (1986), *Intonational Structure in Japanese and English*, in *Phonology Yearbook*, 3, pp. 255-309.
- Bernardelli A. e Pellerey R. (1999), *Il parlato e lo scritto*. Milano: Bompiani.
- Berruto G. (1985), *Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica?*, in Holtus G., Radtke E. (a cura di); *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Gunter Narr Verlag, pp. 120-153.
- Biber D. (1993), *Representativeness in corpus design*. *Literary and Linguistic Computing* 8(4): pp. 243-57.
- Boomer, D. S. (1965), *Hesitation and grammatical encoding*. *Language and Speech* (8), 145-148.
- Boutet D. (2008), *Une morphologie de la gestuelité: structuration articulaire*, *Cahiers de linguistique analogique*, 5, pp. 80-115.
- Boutet D. Morgenstern A. (2020), *Prélude et Ode à l'approche kinésiologique de la gestualité*, *Travaux interdisciplinaires sur la parole et le langage*, 36, 2020.
- Boutet, D. (2010), *Structuration physiologique de la gestuelle : Modèle et tests*, *Lidil, Revue de linguistique et de didactique des langues*, (42), 77-96.
- Boutet, D. (2018) *Pour une approche kinésiologique de la gestualité*. (Habilitation à diriger des recherches), Université de Rouen-Normandie, <https://hal.archives-ouvertes.fr/tel-02357282>
- Boutet, D., & C. Cuxac (2008) *Le signifiant gestuel : langue des signes et gestualité ; Avant- Propos*, *Cahiers de linguistique analogique*, (N 5), 2-15.
- Boutet, D., Morgenstern, A., & A. Cienki (2016) *Grammatical Aspect and Gesture in French: A kinesiological approach*, *Russian Journal of Linguistics*, 20(3), 132-151.
- Boutet, D., Morgenstern, A., & A. Cienki (2018a) *Summing up: Aspect as amodal or as modalitydependent*, in Alan Cienki & Olga K. Iriskhanova (Eds.), *Aspectuality cross Languages*, Amsterdam, Philadelphia: John Benjamins, 58-60.

Bressem J. (2013), *A linguistic perspective on the notation of form features in gestures*, in Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp.1079-1098.

Bressem J. e Müller C. (2014). *A Repertoire of German Recurrent Gestures with Pragmatic Functions*, in *Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk) 38/2*,. De Gruyter Mouton. Pp. 1575–91.

Bressem J. e Müller C. (2014)., *A Repertoire of German Recurrent Gestures with Pragmatic Functions*, in *Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk) 38/2*, 1575–91. De Gruyter Mouton.

Bressem J., Ladewig S. H. e Müller C. (2013b), *Linguistic Annotation System for Gestures*. In *Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk) 38/1*, 1098–1124. De Gruyter Mouton.

Brown G. (1990), *Listening to spoken English*, London - New York, Longman (1a ed. 1977).

Burnard, L. (2005), *Developing linguistic corpora: metadata for corpus work*. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 30–46.

Butterworth B., Hadar U. (1989), *Gesture, Speech and Computational Stages: A reply to McNeill*, in “*Psychological Review*”, 96, 1, pp. 168-174.

Campisi, E. (2018), *Che cos'è la gestualità?*, Carocci, Roma.

Chafe W. (1993), *Prosodic and functional units of language*, in Edwards J., Lampert M.D. (a cura di) *Talking data: Transcription and coding in discourse research*, Hillsdale (NJ), Lawrence Erlbaum Associates, pp. 33-43.

Cienki A. (1998), *Metaphoric gestures and some of their relations to verbal metaphoric expressions*. In J.P. Koenig (ed.), *Discourse and Cognition*:

Bridging the Gap, 189-204. Stanford, CA: Center for Study of Language and Information.

Cienki A. (2008), *Why study metaphor and gesture?*, In A. Cienki e C. Müller (eds.), *metaphor and Gesture*, 5-25. Amsterdam: John Benjamins.

Cienki A. (2017) *Semantic Analysis of Language as Dynamic and Multimodal: Simulation and Conceptualization*. In: *Ten Lectures on Spoken Language and Gesture from the Perspective of Cognitive Linguistics*. Brill, 2017. p. 129-144.

Cienki A. (2017), *From Paralinguistic to Variably Linguistic*, in *The Routledge Handbook of Pragmatics*, pp. 61-68.

Cienki, A. (2017). *Gesture and pragmatics: From paralinguistic to variably linguistic*. In A. Barron, G. Steen & Y. Gu (eds.), *The Routledge Handbook of Pragmatics*. London: Routledge, 61-68.

Cienki, A. J., 2013, *Cognitive Linguistics: Language and gestures as expressions of conceptualization*, in Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp. 182-201.

Clark H. H. (2003), *Pointing and placing*, in S. Kita (edit by), *Pointing: where language, Culture, and Cognition Meet*, 243-268. Mahwah, NJ: Lawrence Erlbaum.

Cook S. W. e Tanenhaus M. K. (2008), *Speakers communicate their perceptual-motor experience to listeners nonverbally*. In B. C. Love, K. McRae e V. M. Sloutsky (eds.), *Processing of the 30th Annual Conference of the Cognitive Science Society*, pp. 957-962.

Cook S. W., Tanenhaus Mo. K. (2009), *Embodied Communication: Speakers' Gesture Affect Listeners' Actions*, in "Cognition", 113, 1, pp. 98-104.

Corballis M. C. (2020), *La verità sul linguaggio (per quel che ne so)*. Roma: Carocci editore.

Cresti E. (2000), *Corpus di italiano parlato*. Accademia della Crusca, 2000.

Cresti E. (2000), *Per una nuova classificazione dell'ilocuzione*, in Burr E. (a cura di), *Tradizione e innovazione. Atti del VI Convegno Internazionale della SILFI - Gerhad-Mercator Universit (Duisburg, 28 giugno-2 luglio 2000)*. Firenze, Franco Cesati Editore, pp. 233-246.

Cresti E. (2017), *The empirical foundation of illocutionary classification*, in De Meo A., Dovetto F.M. (a cura di) *La comunicazione parlata. Atti del Congresso SLI-GSCP - Universit degli Studi di Napoli "L'Orientale"* (Napoli, 13-15 giugno) Aracne, Napoli, 243-264.

Cresti E. (2020), *The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory*, in Izre'el S., Mello H., Panunzi A., Raso T. (a cura di). *In search of basic units of spoken language: A corpus-driven approach*, Amsterdam, John Benjamins, 181-219.

Cresti E., Moneglia M. (2005) (a cura di.), *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*, Amsterdam, John Benjamins.

De Jorio A. (2002) [1832], *La mimica degli antichi investigata nel gestire napoletano*, Forni, Bologna, ristampa anastatica [I edizione Stamperia e Cartiera del Fibreno, Napoli, 1832].

De Jorio, A. (1979), *La mimica degli antichi investigata nel gestire napoletano*. Bologna: Arnaldo Forni. [=ristampa anastatica dall'edizione orifinale del 1832.]

De Mauro T. (2000), *Vocalità, gestualità, lingue segnate e non segnate*, in Bagnara C., Chiappini G. Conte M. P., Ott M. (a cura di), *Viaggio nella città invisibile*. Edizioni del Cerro, Pisa, pp. 17-45.

Diadori P. (1990), *Senza parole: 100 gesti degli italiani*.

Diadori P. (1990), *Senza parole: 100 gesti degli italiani*. Rome: Bonacci.

Du Bois J. W. et al. (1993), *Outline of discourse transcription*, in Edwards J.A., Lampert M.D. (a cura di) *Talking data: Transcription and coding in discourse research*, Hillsdale, NJ: Lawrence Erlbaum, pp. 45-89.

Dybkj r, L. and Ole Bernsen, N. (2004), *Recommendations for natural interactivity and multimodal annotation schemes*. Proceedings of the LREC'2004 Workshop on Multimodal Corpora, Lisbon.

Edwards, J. (1993), *Principles and contrasting systems of discourse transcription*. In Edwards, J. and Lampert, M. (eds), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 3-44.

Efron D, (1972) [1941], *Gesture, race and culture*. The Hague: Mouton.

Ekman P. Friesen W. Y. (1969), *The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding*, in "Semiotica", 1, 1, pp. 49-98.

Enfield N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Language Culture and Cognition. Cambridge University Press.

Enfield N. J. 2006, *Social Consequences of Common Ground*, in *Roots of Human Sociality. Culture, Cognition and Human Interaction*, edited by S. C. Levinson, 399-430. Wenner-Gren International Symposium Series. Oxford: Berg.

Enfield N. J. A (2013), *Composite Utterances approach to meaning*, In: *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK) 38/1*. De Gruyter, 2013.

Enfield N. J., Kita S. and De Ruiter J. P. (2007). *Primary and Secondary Pragmatic Functions of Pointing Gestures*, in *Journal of Pragmatics* 39 (10): 1722-41.

Enfield N. J., Kita S. e De Ruiter J. P. (2007), *Primary and Secondary Pragmatic Functions of Pointing Gestures*, in *Journal of Pragmatics* 39 (10), pp. 1722-41.

Feyereisen P. (1987), *Gesture and Speech, Interaction and Separations: A Repley to McNeill (1985)*, in "Psychological Review", 94, 4, pp. 493-498.

Feyereisen P., Van de Wiele M., Du Bois F., (1988), *The meaning of Gestures: What Can Be Understood without Speech?*, in "Cahiers de Psychologie Cognitive. European Bulletin of Cognitive Psychology", 8, 1, pp. 3-25.

Fontana. S. (2009), *Linguaggio e multimodalità: gestualità e oralità nelle lingue vocali e nelle lingue dei segni*. Edizioni ETS, Pisa.

Ford C. E.; Fox B. A.; Thompson S. A., *Practices in the construction of turns: The "TCU" revisited*, in *Pragmatics*, 1996, 6.3, pp. 427-454.

Gadeau P. (1983), *La civilisation incarnée. Synthèse d'une approche de la culture gestuelle italienne*, in *Geste et Image* 3, pp. 19-36.

Gallese, V. (2007). *Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1480): 659-669.

Gibbon D., Gut U., Hell B., Looks K., Thies A. and Trippel t., (2003), *A Computational Model of Arm Gestures in Conversation*. In *Eighth European Conference on Speech Communication and Technology*. Geneva, Switzerland.

Goffman, E. (1963), *Behavior in Public Places*. New York: The Free Press.

Goldin-Meadow, S., and Brentari, D. (2017). *Gesture, sign, and language: the coming of age of sign language and gesture studies*. In *Behav. Brain Sci.* 40, 1-17.

Goodwin C. (1979), *The interactive construction of a sentence in natural conversation*, in Psathas G. (a cura di) *Everyday Language*. *Studies in Ethnomethodology*, New York, Irvington, pp. 97-121.

Goodwin C. (2003), *The semiotic body in its environment*. In J. Coupland e R. Gwyn (eds), *Discourse of the body*, New York: Palgrave/Macmillan, pp. 19-42.

Goodwin C. (2007), *Participation, stance and affect in the organization of the activities*. *Discourse and society* 18(1), pp. 53-73.

Goodwin, C. (2000). *Action and embodiment within situated human interaction*. *Journal of Pragmatics* 32, 1489-1522.

Graziano M., Gullberg M. (2018), *When Speech Stops, Gesture Stops: Evidence from Developmental and Crosslinguistic Comparisons*, in *Frontiers in Psychology* 9: 879.

Grice H. P. (1967), *Logic and Conversation*, in P. Cole e J. Morgan (a cura di), *Syntax and Semantics 3: Speech Acts*, Academic Press, New York, pp. 41-58.

Gu, Y. (2006), *Multimodal text analysis: a corpus linguistic approach to situated discourse*. Text and Talk 26(2): pp. 127–67.

Halliday, M. A. (1963). *The tones of English*, in Arch. Linguist. (15), 1-28.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd ed.). London: Edward Arnold.

Hockett C. F. (1960), *The origin of speech*, in Scientific American, 203, pp. 88-96.

Hostetter B., Alibali M. W. (2008), *Visible Embodiment: Gestures as Simulated Action*, in “Psychonomic Bulletin & Review”, 15, 3, pp. 495-514.

Hostetter, A. & M. Alibali. (2008). *Visible embodiment: Gestures as simulated action*. Psychonomic Bulletin & Review 15(3): 495–514.

Jefferson G. (2004), *Glossary of transcript symbols with an introduction*. In Lerner, G. H. (ed.), *Conversation Analysis: Studies from the first generation*. Amsterdam/ Philadelphia: John Benjamins, pp. 13–31.

Jewitt C, Bezemer J., O'halloran K. (2016) *Introducing multimodality*. Routledge, 2016.

Kelly S., Barr D. J. (1999), *Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory*, in “Journal of Memory and Language”, 40, pp. 577-592.

Kendon A. (1972), *Some relationships between body motion and speech: an analysis of an example*. In Siegman A. e Pope B. (edited by), “*Studies in Dyadic Communication*”, Elmsford, New York, pp 69 – 89.

Kendon A. (1980), *Gesticulation and speech: Two aspects of the process of utterance*. In Key M. R. (edited by), *The relationship of verbal and nonverbal communication*. De Gruyter Mouton, pp. 207-228.

Kendon A. (1988), *How gestures can become like words*, in Poyatos F. (edited by), *Cross Cultural Perspectives in Nonverbal Communication*, Lewiston, New York, pp. 131-141.

Kendon A. (1995), *Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation*, in “Journal of Pragmatics”, 23, 3, pp. 247-279.

Kendon A. (1995). *Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation*, in *Journal of Pragmatics* 23 (3), pp. 247–79.

Kendon A. (2000), *Language and Gesture: Unity or Duality?*, in McNeill D. (edited by), *Language and Gesture*, Cambridge University Press, Cambridge.

Kendon A. (2004), *Gesture: Visible Action as Utterance*. Cambridge University Press.

Kendon A. (2004), *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kendon A. (2014), *Semiotic diversity in utterance production and the concept of language*, in *Philosophical Transactions of the Royal Society B*, 369: 20130293, pp. 1-13.

Kendon A. (2014). *Semiotic diversity in utterance production and the concept of 'language'*; in *Phil. Trans. R. Soc. B* 369: 20130293.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.

Kipp, M., Neff, M. and Albrecht, I. (2007), *An annotation scheme for conversational gestures: how to economically capture timing and form*. *Language Resources and Evaluation* 41(3/4): pp. 325–39.

Kita S. (2000), *How representational Gestures Help Speaking*, in D. McNeill (edited by), *Language and Gesture*, Cambridge University Press, Cambridge, pp. 162-185.

Kita, S., van Gijn, I., & van der Hulst, H. (1998). *Movement phases in signs and co-speech gestures, and their transcription by human coders*. In I. Wachsmuth, & M. Fr hlich (A cura di), *Gesture and Sign Language in Human- Computer Interaction*, International Gesture Workshop Bielefeld, Germany, September 17-19, 1997, Proceedings. *Lecture Notes in Artificial Intelligence* (Vol. 1317, p. 23-35). Berlin: Springer-Verlag.

Klima E, Bellugi U, (1979), *The Sign of Language*, Cambridge, MA: Harvard University Press.

Knight, D. (2011) *Multimodality and active listenership: A corpus approach*. A&C Black.

Knight, D., Evans, D., Carter, R. A. and Adolphs, S. (2009), *Redrafting corpus development methodologies: Blueprints for 3rd generation multimodal, multimedia corpora*. *Corpora*, p. 32.

Kok K. I., Cienki A. (2016), *Cognitive Grammar and Gesture: Points of convergence, advances and challenges*. In *Cognitive Linguistics*, 2016, 27, pp. 67-100.

Krauss, R. (1998). *Why do we gesture when we speak?* *Curr. Dir. Psychol. Sci.* 7, 54-60.

Kress, G. (2014). *What is a Mode?* In C. Jewitt (Ed.) *Routledge Handbook of Multimodal Analysis* (pp. 60-75). London: Routledge.

Kress, G. R., & Hodge, R. I. V. (1988). *Social Semiotics*. Cambridge: Polity Press.

Kress, G. R., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London; New York: Edward Arnold/Nottinghamshire; Oxford University Press.

Labov, W. (1972), *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Ladd, D. R. (1996), *Intonational Phonology*. Cambridge University Press, Cambridge.

Ladewig S. H. (2012), *Linear integration of gestures into speech*. In Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp. 1662-1677.

Ladewig, S. H. (2011). *Putting the cyclic gesture on a cognitive basis*. *CogniTextes. Revue de l'Association française de linguistique cognitive*, (Volume 6).

Langacker R. W. (1988), *A usage-based model*, in B. Rudzka-Ostyn (ed.), *Topic in Cognitive Linguistics*, 127-161. Amsterdam: John Benjamins.

Langacker R. W. (1993), *Reference-point constructions*, *Cognitive Linguistics*, 4, 1-38.

Langacker R. W. (2008), *Cognitive grammar: a Basic Introduction*. Oxford: Oxford University Press.

Langacker, Ronald W. (2012) *Interactive cognition: Toward a unified account of structure, processing, and discourse*. *International Journal of Cognitive Linguistics* 3(2). 95-125.

Lausberg H. (2013), *NEUROGES – A Coding System for the Empirical Analysis of Hand Movement Behaviour as a Reflection of Cognitive, Emotional, and Interactive Processes*. In *Body - Language - Communication*, 1:1022-37. De Gruyter Mouton.

Lausberg H., Sloetjes H. (2015), *The revised NEUROGES-ELAN: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture*, *Behavior Research Methods*, 2009, 48, pp. 973-993.

Leech G. (1997), *Introducing corpus annotation*. In Garside, R., Leech, G. and McEnery, T. (eds) *Corpus annotation: Linguistic Information from Computer Text Corpora*. London: Longman, pp. 1-18.

Levinson, S. C., (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Lieberman M., Pierrehumbert J. B. (1984), *Intonational Invariance under Changes in Pitch and Length*, in M. Aronoff, R. T. Oehrle (edito da), *Language Sound Structure*, MIT Press, Cambridge, pp. 157-233.

Lieberman Ph. (1967), *Intonation, Perception and Language*, MIT Press, Cambridge (MA).

Linell, P. (2009) *Rethinking Language, Mind and World dialogically: Interactional And Contextual Theories Of Human Sense-Making*, Charlotte, NC: Information Age Publishing.

Locke J. (1690) [traduzione italiana, 1955], *Saggio sull'intelletto umano*, Zanichelli, Bologna.

Loehr D. (2007), *Aspects of Rhythm in Gesture and Speech*, *Gesture* 7 (2): 179-214.

Loehr D. (2014) *Gesture and Prosody*. In *Body - Language - Communication*, 2:1381-91. De Gruyter Mouton.

Magno Caldognetto E., Ursini F. e Poggi I. (2004), *Italiano parlato o comunicazione multimodale? (Le ragioni di un successo)*, in *Il parlato italiano. Atti del convegno nazionale (Napoli, 13-15 febbraio 2003)*, a cura di F. Albano Leoni et al., Napoli, M. D'Auria (Cd-Rom, G08).

Martinet A. (1960), *Éléments de linguistique Générale*. Paris: Armand Colin.

McEnery, T. and Wilson, A. (1996), *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McNeil D. (1987), *So You Do Think that Gestures Are Non-Verbal! A Reply to Feyereisen (1987)*, in "Psychological Review", 94, 4, pp. 499-504.

McNeill D. (1985), *So You Think Gestures Are Nonverbal?*, in "Psychological Review", 92, 3, pp. 350-371.

McNeill D. (2005), *Gesture & Thought*. Chicago e London: University of Chicago Press.

McNeill D., Pedelty L. L.; Levy E. T. (1990). *Speech and gesture*. In "Advances in psychology"; North-Holland, p. 203-256.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, D. (2013). *The growth point hypothesis of language and gesture as a dynamic and integrated system*. In *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK) 38/1* (pp. 135-155). De Gruyter Mouton.

Mondada (2013), *Conversation analysis: Talk and bodily resources for the organization of social interaction*, In Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp.218-227.

Mondada L. (2018) *Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality*, *Research on Language and Social Interaction*, 51:1, 85-106,

Mondada, L., (2011) *The organization of concurrent courses of action in surgical demonstrations*. In: Streeck, J., Goodwin, C., LeBaron, C. (Eds.), *Embodied Interaction, Language and Body in the Material World*. CUP, Cambridge, pp. 207-226.

Mondada, L., (2016), *Challenges of multimodality: language and the body in social interaction*. *J. Sociolinguist.* 20 (2), 2-32.

Moneglia M., Raso T. (2014), *Notes on Language into Act Theory (L-Act)*, in Raso T., Mello H. (a cura di) *Spoken corpora and linguistic studies*, Amsterdam, John Benjamins, 468-495.

Morrel-Samuels, P., and Krauss, R. (1992). *Word familiarity predicts temporal asynchrony of hand gestures and speech*. In *J. Exp. Psychol.* 18, 615-622.

Morsella, E., and Krauss, R. M. (2005). *Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories*. *J. Psycholinguist. Res.* 34,415-427.

Müller C, Bressemer J., Ladewig S. H. (2013), *Towards a unified grammar of gesture and speech: A multimodal approach*. In Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp.707-733.

Müller C. (2013), *Gestural Modes of representation as techniques of depiction*, In Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp. 1687-1702.

Müller C. (2013b), *Gestures as a medium of expression: the linguistic potential of gestures*, In Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. & Teßendorf, S. (eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton, pp. 202-217.

Müller C. (2018) *Gesture and Sign: Cataclysmic Break or Dynamic Relations?*; in *Front. Psychol.* 9:1651.

Munari B. (1963), *Supplemento al dizionario italiano*. Milan: Muggiani.

Nencioni G. (1976), *Parlato-parlato, parlato-scritto, parlato-recitato*, in «Strumenti critici», 29, pp. 1-56.

Ochs, E. (1979), *Transcription as theory*. In Ochs, E. and Schieffelin, B. B. (eds) *Developmental Pragmatics*. New York: Academic Press, pp. 43-72.

Ong W. J., 2011 [1982] *Oralità e scrittura*. Bologna: Il Mulino.

Özyürek A. (2002), *Do Speakers Design Their Co-Speech Gestures for Their Addressees? The Effects of Addressee Location on Representational Gesture*, in “*Journal of Memory and Language*”, 46, 4, pp. 688-704.

Panunzi A. and Scarano A. (2009), *Parlato Spontaneo E Testo: Analisi Del Racconto Di Vita*. In *I Parlanti E Le Loro Storie: Competenze Linguistiche, Strategie Comunicative, Livelli Di Analisi*, edited by L. Amenta and G. Paternostro, 121-32.

Panunzi A. e Gregori L. (2012). *DB-IPIC. An XML Database for the Representation of Information Structure in Spoken Language*. In: H. Mello, A. Panunzi, T. Raso. *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*, Firenze: Firenze University Press. pp. 133-150

Pettorino M. e Giannini A (2005), *Analisi delle disfluenze e del ritmo del dialogo romano*, in *Italiano parlato. Analisi di un dialogo*, a cura di F. Albano Leoni & R. Giordano, Napoli, Liguori, pp. 89-104.

Pierrehumbert J. B. (1980), *The Phonology and Phonetics of English Intonational*, Ph.D. Dissertation, MIT, Cambridge (MA).

Pierrehumbert J. B., Beckman M. E. (1988), *Japanese Tone Structure*, MIT Press, Cambridge (MA).

Pietrandrea P. (2000), *Complessità dell'interazione di iconicità e arbitrarietà nel lessico della LIS*, in Bagnara C., Chiappini G. Conte M. P., Ott M. (a cura di), *Viaggio nella città invisibile*. Edizioni del Cerro, Pisa, pp. 38 - 49.

Pietrandrea P. (2012); *L'approccio socio-semiotico alle lingue dei segni*; in Thornton A. M., Voghera M. (a cura di) "Per Tullio de Mauro: studi offerti dalle allieve in occasione del suo 80 compleanno", Aracne, Roma, pp. 265-285.

Pitrè G. (1877), *Gesti ed insegne del popolo siciliano*, in *Rivista di letteratura popolare* 1: 32-43.

Pitrè G. (1887), *I gesti*. In: G. Pitrè, ed., *Biblioteca delle tradizioni popolari siciliani*. Palermo: Luigi Pedone Lauriel. Pp. 341-377.

Pouw, W. T.J.L. et al. (2014) *Toward a more embedded/extended perspective on the cognitive function of gestures*. *Frontiers in psychology*, 5: 359, pp. 1-14.

Quintiliano (2001), *Institution oratoria*, A. Pennaccini (a cura di), Einaudi, Torino.

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). *Gesture, speech, and lexical access: the role of lexical movements in speech production*, in *Psychol. Sci.* 7, 226-231.

Ricci Bitti P. E., Cortesi S. (1977), *Comportamento non verbale e comunicazione*, Il Mulino, Bologna.

Rizzolati G., Sinigaglia C. (2019), *Specchi nel cervello: come comprendiamo gli altri dall'interno*, Raffaele Cortina, Milano.

Sabatini F. (1985), *L'italiano dell'uso medio*, in Holtus G., Radtke E. (a cura di); *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Gunter Narr Verlag, pp. 154-184.

Sacks H. (1992), *Lectures on Conversation*, Oxford, Basil Blackwell.

SackS H., Schegloff E. A., Jefferson G. (1974), *A simplest Systematics for the Organization of Turn taking for Conversation*, in "Language", 50, 4, pp. 696-735.

Sbisà, M. (1994). *Per una pragmatica degli atti linguistici: quasi un bilancio*. In Orletti F. (a cura di) *Fra conversazione e discorso: L'analisi dell'interazione verbale*. Roma: Carocci, pp. 29-47.

Sbisà, M. (2013). *Locution, illocution, perlocution*. In *Pragmatics of speech actions* (pp. 25-76). De Gruyter Mouton.

Schegloff E. A (1972), *Sequencing in Conversational Openings*, in *American Anthropologist*, 70, 6, pp. 1075-95.

Schegloff E. A (1992), *In Another Context*, in A. Duranti, Ch. Goodwin (edito da), *Rethinking Context*. Cambridge University Press, Cambridge, pp. 193-227.

Schegloff E. A (1996), *Turn-organization: One intersection of grammar and interaction*, in Ochs E., Schegloff E.A., Thompson S. (a cura di) «*Interaction and Grammar*», Cambridge, Cambridge University Press, pp. 52-133.

Schegloff E. A. (1982), *Discourse as an Interactional Achievement: Some uses of 'Uh Huh' and Other Things that Come between Sentences*, in Tannen, D. (a cura di), pp- 71-93.

Schegloff E. A. (1984), *On some gestures' relation in Talk*, in J M. Atkinson e J. Heritage (eds.), *structure of Social Action*, Cambridge: Cambridge University Press, pp. 266-296.

Schegloff E. A. (2002), *Beginnings in the telephone*. In J. E. Katz e M. Aakhus (eds), *Perceptual contact: Mobile Communication, Private Tal, Public Performace*, Cambridge: Cambridge University Press, pp. 284-300.

Schegloff E.A. (2007), *Sequence Organization in Interaction: a Primer in Conversation Analysis*, Volume 1. Cambridge: Cambridge University Press.

Schmitt J. C. (1990), *Il gesto nel medioevo*, Bari: Editori Laterza.

Searle J.R. (1969), *Speech Acts. An essay in the philosophy of language*, Cambridge, Cambridge University Press, [tr. it. 1976, *Atti linguistici. Saggio di filosofia del linguaggio*, Torino, Boringhieri].

Sinclair J. (2005), *Corpus and text- basic principles*. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 1-16.

Sinclair J. (2008), *Borrowed ideas*. In Gerbig, A. and Mason, O. (eds), *In Language, People, Numbers- Corpus Linguistics and Society* . Amsterdam: Rodopi BV, pp. 21-42.

Sorianello P. (2014), *Prosodia: modelli e ricerca empirica*, Carocci Editore, Roma.

- Sornicola R. (1985), *Sul parlato*. Bologna: Il Mulino.
- Sotaro K., A. Özyürek. 2003, *What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking.* Journal of Memory and Language 48 (1), pp. 16-32.
- Streeck J. (1993), *Gesture as communication I: Its coordination with gaze and speech*. Communication Monographs 60, pp. 275-299.
- Streeck J. (2009) *Gesturecraft: The manu-facture of meaning*. John Benjamins Publishing.
- Thompson, P. (2005), *Spoken language corpora*. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 59-70.
- Trippel T., Gibbon D., Thies A., Milde J., Looks K, Hell B. e Gut U. (2004). *CoGesT: A Formal Transcription System for Conversational Gesture.* In Proceedings of the Fourth International Conference on Language Resources and Evaluation (Lrec'04). Lisbon, Portugal.
- Voghera M. (1992), *Sintassi e intonazione nell'italiano parlato*. Bologna: Il Mulino.
- Voghera M. (2010). *Lingua parlata*; in Raffaele S. (a cura di); *Enciclopedia dell'italiano*. Roma: Treccani.
- Volterra V., Roccaforte M., Di Renzo M., Fontana S. (2019), *Descrivere la lingua dei segni italiana: una prospettiva cognitiva e sociosemiotica*, Bologna: Il Mulino.
- Watzlawick P., Beavin J. H., Jackson D. (1971), *Pragmatica della comunicazione umana. Studio dei modelli interattivi, delle patologie e dei paradossi*. Astrolabio-Ubaldini, Roma.
- Wesp R, Hesse J., Keutmann D. and Wheaton (2001), *Gestures maintain spatial imagery*, in *The American Journal of Psychology*.
- Wundt W. (1973), *The Language of Gestures*. Tradotto da J, S, Thayer, C. m. Greenleaf e M. D. Silberman da *Völkerpsychologie, etc.*, Volume I, quarta

edizione, Capitolo 2. Stuttgart: Alfred Kröner VERLAG, 1921. The Hague: Mouton.

Wynne, M. (2005), *Archiving, distribution and preservation*. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 71-8.

Zipoli Caiani S. (2016), *Corporeità e cognizione. La filosofia della mente incorporata*". Milano: Le Monnier.