

Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME

Abstract

During the CHROME project (Cultural Heritage Resources Orienting Multimodal Experience) digital resources were collected and used in technological applications to be adopted in cultural heritage linguistical and technological applications. The project outcomes have been obtained through interdisciplinary work carried out by computer scientists, psychologists, computational and general linguists, architects, museologists, and art historians. A set of texts taken from a wide range of scientific and non-scientific literature have been collected, describing the architectural environments of Carthusian monasteries and their artistic heritages. The collection ranges from scientific texts written by experts addressed to other experts to museum brochures addressed to children and speech corpora, architectural surveys of buildings and environments, video corpora filming tour guides describing points of artistic interest to visitors, from which gestural expression data have been extracted. Preliminary attempts were also made to integrate all digital data sources into a multidimensional platform for which an integrated multi-level annotation system was designed, oriented towards computer applications enriched by knowledge representation systems. The three magnificent Carthusian monasteries in Campania, namely San Martino (Naples), San Lorenzo (Padula, Salerno) and San Giacomo (Capri), were chosen as a case study for testing the technologies, methods, and data collected. The fruit of this interdisciplinary commitment has resulted in a large collection of multidimensional data articulated in heterogeneous sub-sectors, ranging from the collection of texts, through collections of audio-video materials, to three-dimensional reconstructions of valuable architectural environments which are virtually available. The project provides full access to all collected data, analysis tools and software demos for those who request it.

This contribution aims to highlight the main concepts and results achieved in terms of the multimodal analysis of videos in which professional operators illustrate the major aspects of the Charterhouse of San Martino.

1. *Il progetto CHROME*

Nell'attuale panorama della ricerca scientifica internazionale, sui temi dell'innovazione tecnologica iniettata nelle applicazioni di *Digital Culture* si possono trovare moltissimi progetti di grande rilevanza e una letteratura vastissima. In Italia, in particolare, il contributo in questo settore portato dalla Fondazione Bruno Kessler

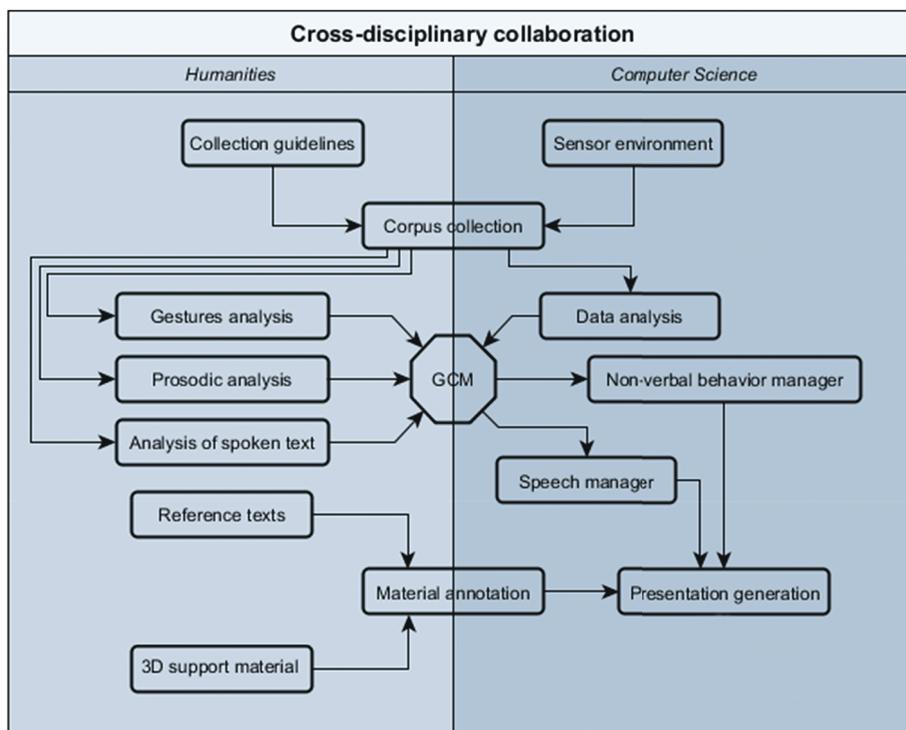
¹ Centro Interdipartimentale di Ricerche Urban/Eco – Università degli Studi di Napoli Federico II.

² Dipartimento di Scienze Umanistiche – Università degli studi di Catania.

in generale e da Oliviero Stock in particolare, ha generato attività di enorme interesse scientifico (Stock & Zancanaro 2007, Kuflik *et al.* 2015). Tuttavia, sebbene questi progetti si siano sempre mossi in un quadro esplicitamente interdisciplinare, gli aspetti legati al linguaggio ed alla comunicazione (verbale e non verbale), al rapporto fra spazi reali, esseri umani ed opere d'arte e monumenti sono stati gli ambiti meno investigati. Il progetto che presentiamo in questo lavoro punta a colmare questa lacuna, proponendo raccolte di dati, formalizzazione di metodi di ricerca e progettazione di dimostratori linguistici, architettonici e di applicazioni informatiche.

Più specificamente, l'obiettivo finale, sintesi di tutte le attività che sono state portate a termine durante il progetto CHROME, è la creazione di un agente umanoide progettato per assumere il ruolo di guida turistica virtuale. Tutte le risorse raccolte in CHROME sono utili a questo scopo: i testi che descrivono i luoghi e le opere ivi presenti, classificati per livello di articolazione e specializzazione nonché per la definizione del tipo di pubblico a cui sono rivolti, sono impiegati per definire il lessico specialistico i principali concetti semantici, il legame con gli oggetti fisici che realmente compongono il posto e la loro nomenclatura. In quest'ottica, i rilievi architettonici digitali, che sono un ulteriore risorsa informatica raccolta in questo progetto, forniscono rappresentazioni codificate dello spazio reale e degli oggetti ivi contenuti. Il continuum spaziale viene segmentato ed etichettato da esperti storici dell'arte e della architettura, creandosi, in questo modo, un collegamento fra entità presenti nell'ambiente e i loro nomi. La definizione del lessico specialistico si avvale anche del contributo di risorse esterne quali l'AAT della Getty Foundation, dai nomi e dai termini specialistici che sono messi in relazione alle posizioni e alle coordinate geometriche che questi ultimi assumono negli spazi investigati. Il materiale audiovideo permette la generazione di una definizione del modello di operatore turistico virtuale; i suoi testi parlati, trasformati in materiale etichettato consentono la creazione di uno stile di parlato che dà alla voce sintetica un timbro e delle caratteristiche estremamente naturali, i contenuti che il sistema automatico "inietta" nell'agente derivano dalla migliore sintesi fra quanto effettivamente prodotto dalle guide durante le loro presentazioni, sia nelle caratteristiche testuali che in quelle di pronuncia, e quanto riscontrato nei testi bibliografici di riferimento. I movimenti delle mani e del corpo delle guide vengono analizzati e ri-sintetizzati per rendere "naturale" il comportamento dell'agente virtuale che, inoltre, collegandosi alla mappatura degli oggetti nello spazio tridimensionale è "consapevole" della posizione di questi ultimi e può imparare ad indicarli e a puntarli con gesti deittici. L'aspetto più coinvolgente della complessa struttura interdisciplinare del progetto è la pianificazione di attività che vedono a stretto contatto il mondo delle scienze "mollì", almeno in principio, in cui gli aspetti umanistici sono dominanti, con il mondo, sempre più estremizzato, delle nuove tecnologie e della intelligenza artificiale. Nella figura 1 questa pianificazione viene schematizzata e sono illustrati i principali moduli di intervento e le relazioni fra questi.

Figura 1 - *La pianificazione delle parti del progetto e la non separazione fra ruolo degli umanisti e ruolo della innovazione tecnologica (fonte: documenti originali del progetto sottoposto al MIUR, la sigla GCM sta per "Gatekeeper Computational Model")*



Questo impianto progettuale trova la sua giustificazione nella seguente citazione:

it has often been said that museums are the places where you can see the real things. It is this presentation of reality, sampled and structured according to arbitrary principles of reality, that makes museums different from any other communication system. Museums are not simple communication systems that we can describe as: a sender, a medium, more than one receiver. It is a complex system with several senders, multiples media and different categories of receivers. The primary system of museum communication includes a curator who is the sender, real things that are the media and visitors who are the receivers. Curators describe nonverbal media (real things) with subsidiary media. (Cameron 1968, ma si leggano anche altre riflessioni sul tema in Chiacchio 2004)

La tendenza della scienza che si occupa dell'iniezione di nuove tecnologie per potenziare l'esperienza museale si pone esattamente l'obiettivo di trovare nuovi paradigmi di comunicazione per rendere materiale ciò che, artisticamente parlando, non lo è. Linguistica ed Interazione Uomo-Macchina, nella sua declinazione legata alla multimodalità, restano centrali in questo discorso e devono trovare nuovi modi di dialogare e di indirizzare la raccolta di risorse digitali che aiutino, nell'ambito delle moderne tecniche della intelligenza artificiale, ad addestrare sistemi innovativi. Si

raccogliono dunque testi, materiale parlato, audio e video, si effettuano rilievi digitali degli ambienti reali dove normalmente gli operatori umani agiscono, caratterizzati dalla loro struttura architettonica e dalle opere d'arte che contengono. Questo patrimonio di dati ricchissimo viene poi ripartito fra gli specialisti: la linguistica si occuperà di individuare ed estrarre corpora di testo e di parlato, di individuare ed estrarre da questi le risorse lessicali, di fare emergere stili di retorica e di oratoria, formalizzandone le caratteristiche; lo studio delle componenti non verbali, condotto dagli esperti di comunicazione non verbale, porta alla definizione di nuovi schemi di annotazione per il sistema espressivo gestuale delle guide.

Insieme agli architetti, che hanno effettuato i rilievi, si studia il rapporto fra il corpo, le parti dello spazio indicate dalle guide. Gli informatici, i linguisti e gli esperti di comunicazione non verbale definiscono mappe di comunicazione articolate, dando nomi agli oggetti nello spazio, calcolando i movimenti che portano alla deissi gestuale, rendendo virtuale finanche il rapporto fra spazio, esseri umani e testo prodotto. Proprio quest'ultimo, il testo generato automaticamente e sintetizzato per essere pronunciato dalla voce dell'agente artificiale, costituisce l'ultima, ma non meno complessa, sfida. Si parte da testi di differente complessità e destinazione d'uso, alcuni derivati dalle trascrizioni del parlato delle guide professionali, altri estratti dalla collezione di testi del corpus degli scritti raccolti; per questi, per ora attraverso un processo semiautomatico, vengono generate delle descrizioni di elementi artistici che possano essere impiegate sia per dare voce alla guida virtuale, sia per fungere da risposte a domande fatte dagli utenti quando, utilizzando il nostro sistema di dialogo, interrogano l'avatar. Il testo generato deve, da un lato avere le caratteristiche stilistiche e di contenuti coerente con quanto ci si aspetta in questo ambito nelle condizioni che simulano la situazione reale, e, dall'altro, deve essere prodotto con una voce che sia credibile in tutti gli aspetti prosodici, ivi incluse le naturali disfluenze.

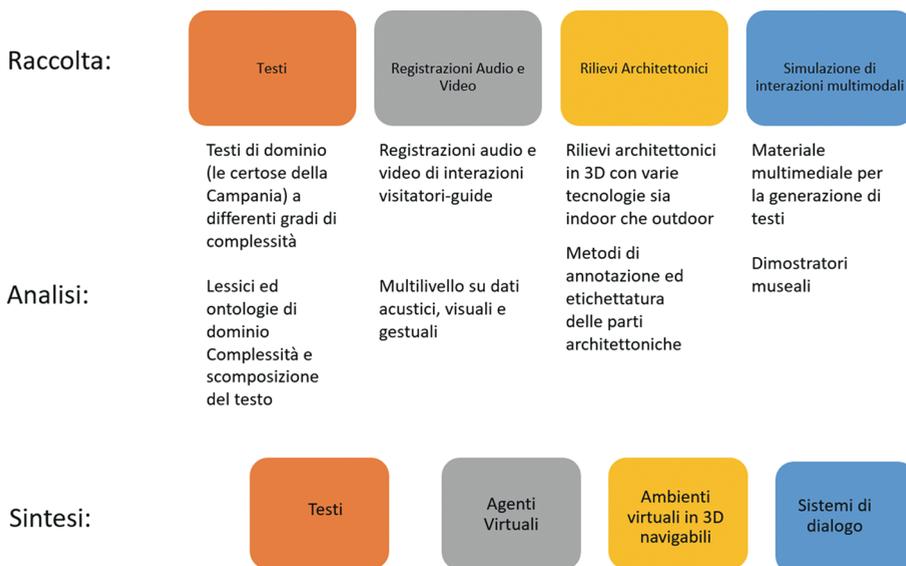
Sulla base di quanto finora esposto, dunque, dunque il progetto prevede delle fasi ben precise:

- raccolta dataset di riferimento per tutti i settori di ricerca da fornire alla comunità;
- studi su: a) le strutture linguistiche (sintattiche, lessicali, delle parti del discorso) per quanto riguarda i testi scritti, per i quali si tiene conto anche del differente stile in funzione della loro destinazione e fruizione, b) delle strutture del parlato, per le quali, oltre a molti degli aspetti pure legati alle strutture linguistiche, si considerano anche aspetti della qualità fonica sia a livello segmentale che soprasegmentale c) dei video, d) dei rilievi architettonici. Gli studi disciplinari e interdisciplinari introdotti in questo progetto, dunque, mirano a generare un sistema di analisi e riproduzione sintetica di tutte le dimensioni comunicative che si possono incontrare negli scenari evocati nel caso di studio scelto;

- processi di integrazione scientifica interdisciplinare e tecnologica per la generazione automatica di contenuti multimodali e per introdurre un modello computazionale sotto forma di agenti artificiali immersi in ambienti virtuali;
- valutazione sperimentale dei risultati scientifici conseguiti nei vari rami del progetto, da condursi attraverso la produzione di appositi dimostratori.

La figura 2 riassume e schematizza queste fasi.

Figura 2 - *Le fasi del progetto*



Il progetto offre, a chi ne faccia richiesta, il pieno accesso a tutti i dati raccolti, a tutti gli strumenti di analisi e ai dimostratori software progettati. Le modalità per le richieste si possono trovare all'indirizzo <http://www.chrome.unina.it>

2. *Le risorse raccolte in CHROME*

Alla fine del progetto sono state raccolte ingenti (per i dettagli si consulti il sito del progetto) quantità di dati che, come indicato nel paragrafo precedente, sono disponibili su richiesta all'intera comunità scientifica.

2.1 Testi sulle tre certose della Campania

Sono stati raccolti e sono disponibili in forma di consultazione, con diritti da assolvere per le opere coperte da diritto d'autore, 102 testi scritti che sono riconducibili a diverse categorie: a) testi scientifici, b) cataloghi destinati alla disseminazione, c) guide specializzate, d) materiale preso dal web e opportunamente selezionato con la supervisione di un esperto di dominio, e) brochure di sala per visitatori adulti e no.

2.2 Analisi testuale

Su parte dei testi raccolti sono state condotte analisi automatiche e manuali (Brunato *et al.* 2020). Le analisi principali sono state: a) estrazione di lessici generali e di termini specifici del dominio, b) Individuazione ed etichettatura degli eventi temporali indicati nei testi, c) analisi sintattica basata su concetti di treebanking, d) riconoscimento dei concetti dell'architettura, basato sul confronto fra termini estratti al punto a) e quelli presenti nell'Art and Architecture Thesaurus (AAT 2018)

2.3 Rilievi architettonici digitali

Gli architetti presenti nella compagine hanno eseguito diversi interventi di rilievo digitale degli edifici, degli ambienti e delle componenti strutturali ivi presenti. Sono state utilizzate diverse tecniche di rilievo, fra cui laser scanner, voli di droni e fotogrammetria aerea. I tre edifici certosini e le parti maggiormente rilevanti dal punto di vista artistico e storico di ognuno di questi hanno generato una considerevole mole di dati di digitalizzazione di beni architettonici, che saranno resi disponibili su richiesta all'interno del *repository* del progetto.

2.4 Video registrazioni delle guide museali

Tre delle quattro guide turistiche ingaggiate nel corso del progetto sono state videoregistrate durante il loro lavoro all'interno del museo ospitato nella certosa di S. Martino in Napoli mentre descrivevano, ad un pubblico di quattro persone per turno, le principali caratteristiche artistiche del sito. La registrazione è stata condotta con due telecamere, una puntata sulla guida e l'altra puntata sul pubblico, e il parlato della guida è stato registrato su tre canali audio aggiuntivi: il primo era collegato ad un microfono *headset* indossato dalla guida e tenuto a distanza fissa dalle labbra, gli altri due a microfoni simmetrici, distanti circa un metro dalla guida e puntati verso il pubblico, che mediamente stazionava di fronte alla guida.

Sulla componente linguistica di questo materiale audiovisivo sono state condotte delle analisi su vari livelli: a) livello ortografico, con trascrizioni del testo prodotto arricchita dalla annotazione delle pause silenti, delle pause piene, delle principali disfluenze (ripartenze, esitazioni, eccetera, si vedano Cataldo *et al.* 2019, Schettino *et al.* in stampa), b) livello sillabico, segmentazione in sillabe, con valutazione *a posteriori* della fluenza e della velocità di eloquio, c) livello intonativo, allineamento fra movimenti tonali della curva dell'intonazione e testo corrispondente, individuazione dei pattern di enfasi, indicazione di stili prodotti, d) livello pragmatico informativo delle unità informative maggiori di topic e comment (Gundel 1988) e di inciso (Cresti 2000), livello macrosintattico (per un approfondimento si veda anche Alfano *et al.* in stampa), e) analisi multimodale/gestuale, annotazione, secondo uno schema di annotazione dei gesti (si veda la seconda parte di questo lavoro) delle espressioni del volto, inclusa una classificazione in termini di funzionalità testuali, emotive ed interazionali. Lo strumento scelto per l'annotazione audio-video è ELAN (Sloetjes & Wittenburg 2008), che consente una architettura di annotazioni

multilivello, fondendo livelli di analisi video con livelli di analisi audio, permettendo al contempo sia la massima indipendenza possibile dei livelli di annotazione, sia di effettuare i collegamenti e le interrogazioni inter-livello fra i dati disponibili.

Di tutti gli aspetti trattati in CHROME, per questo articolo, si è scelto di procedere con la descrizione della analisi multimodale del corpus audio-video con specifico riferimento alla analisi dei gesti di una guida museale.

3. *Introduzione all'analisi multimodale*

È un dato ormai consolidato che i gesti che facciamo mentre parliamo non sono movimenti casuali ma interagiscono col parlato corrispondente a livello prosodico, sintattico, semantico e pragmatico (McNeill 1992; Kendon 2004). In particolare, numerosi studi confermano che non solo i gesti sono parte integrante dell'intenzione comunicativa del parlante, ma anche che sono percepiti come tali dai destinatari, che integrano l'informazione che viene dai gesti con quella proveniente dal parlato e, di conseguenza, consentono di raggiungere risultati migliori in compiti di comprensione (per una rassegna più completa cfr. Campisi 2018). Questo è vero non soltanto tra esseri umani ma, come un recente filone di ricerca sembra suggerire, anche nell'interazione con robot umanoidi e con agenti virtuali, che vengono apprezzati maggiormente quando gesticolano in maniera "appropriata" (Bergmann & Macedonia 2013).

Sebbene i *gesture studies* moderni siano una disciplina ormai affermata e in continua espansione, grazie anche agli sviluppi importanti delle tecnologie a supporto della ricerca sul parlato multimodale, la nostra conoscenza sulla gestualità che si accompagna al parlato (*co-speech gestures*) risente ancora di sostanziali limitazioni; tra queste, ricordiamo la scarsità di studi sistematici in lingue diverse dall'inglese americano da un lato e la frequente separazione tra chi si occupa di gesti, diventando esperto nella loro identificazione e codifica, e i linguisti che studiano il parlato, che a volte riconoscono l'importanza della componente gestuale ma non hanno gli strumenti conoscitivi per includerla nelle loro analisi.

Il progetto CHROME ha le potenzialità per contribuire a entrambe le limitazioni. Infatti, il corpus di parlato semi spontaneo e semi-monologico (Cataldo *et al.* 2019) è perfetto per un'analisi dei gesti che permetta di integrare ciò che già sappiamo sui gesti dei parlanti italiani – e nello specifico napoletani, documentati per lo più dal padre degli studi moderni sul gesto, Adam Kendon (cfr. Kendon 2004) – con informazioni nuove, provenienti dal perfezionamento delle nozioni che ha caratterizzato i *gesture studies* negli ultimi decenni e dal già menzionato sviluppo delle tecnologie a nostra disposizione.

4. L'analisi dei gesti

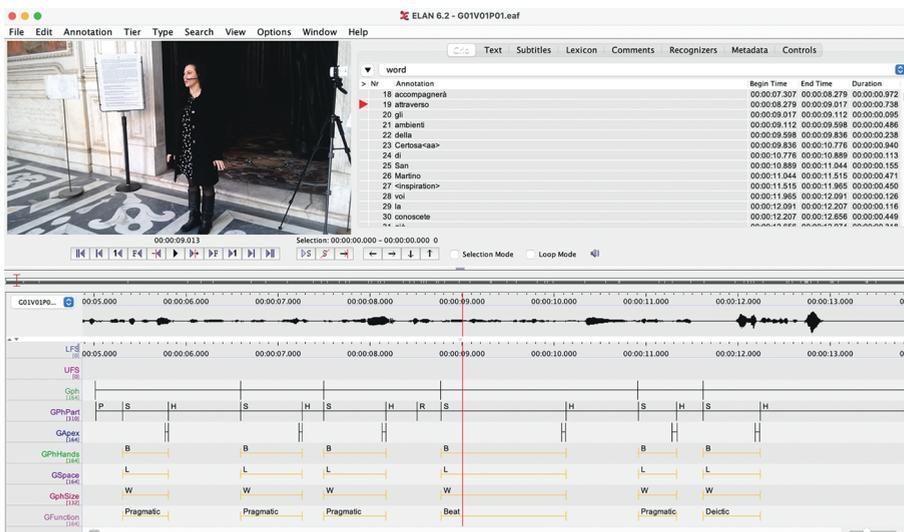
4.1 Criteri iniziali e successivi sviluppi

Fin dalle prime fasi del progetto, si auspicava un'analisi del comportamento multimodale delle guide turistiche coinvolte che prevedesse l'annotazione di gesti, le espressioni facciali e lo sguardo. Tale analisi doveva non soltanto includere la descrizione fisica di tali componenti multimodali, ma anche l'analisi semantica e la classificazione in termini di funzioni testuali, emotive e interazionali (Origlia *et al.* 2019a). In questa sede ci limitiamo a presentare la codifica, ancora in corso, dei gesti delle mani e delle braccia.

4.2 Schema di codifica attuale

Lo schema di codifica attuale dei gesti si presenta come un insieme di livelli organizzati gerarchicamente, tipici dell'interfaccia di ELAN.

Figura 3 - Schermata Elan con i livelli (tiers) di codifica dei gesti



I livelli attualmente individuati sono i seguenti:

Segmentazione dell'unità gestuale:

- identificazione delle *gesture phrases*
- identificazione delle fasi gestuali
- identificazione dell'apice del gesto

Per ogni *stroke*, identificazione delle caratteristiche cinetiche e semantiche del gesto:

- Mani
- Spazio del gesto
- Dimensioni del gesto
- Configurazione della mano
- Funzione del gesto

Nel resto del paragrafo definiremo brevemente ciascuno dei livelli.

Seguendo Kendon (1980), consideriamo un'unità gestuale (*gesture unit*) come l'intera escursione dei movimenti, da quando le mani iniziano a muoversi a quando tornano a una posizione di riposo. Un'unità gestuale può essere più o meno lunga, e può corrispondere o meno all'intera unità discorsiva. L'unità gestuale può essere a sua volta scomposta in unità più piccole, dette *gesture phrases*³, che si caratterizzano per un'escursione in cui le mani e le braccia eseguono un particolare movimento posizionandosi in un punto preciso dello spazio intorno al parlante e le mani assumono una configurazione specifica. Ogni *gesture phrase* inizia con un movimento di "preparazione", in cui gli arti superiori si preparano a raggiungere tale posizione e configurazione, e può concludersi con una "retroazione", cioè col ritorno a una posizione di riposo. Inoltre, ogni *gesture phrase* ha un suo apice, detto *stroke*, in cui il gesto raggiunge il momento massimo di sforzo e chiarezza. La complessità e la lunghezza delle *gesture phrases* possono variare notevolmente, a seconda che contengano o meno delle pause (sia *pre-stroke* sia *post-stroke*), la cui funzione principale sembra essere quella di allineare i gesti con il ritmo del parlato.

4.2.1 Mani (sinistra/destra/entrambe)

Un gesto può essere eseguito con una mano o con due mani. Usare due mani anziché una può essere considerato uno sforzo comunicativo aggiuntivo e quindi un tentativo di rendere più ricco il proprio messaggio. Questo è vero sia nel caso in cui le due mani eseguono lo stesso movimento sia nel caso di configurazioni più complesse, in cui invece i movimenti sono diversi (vedi la descrizione delle funzioni più in basso).

4.2.2 Spazio del gesto

Ogni gesto viene eseguito in un luogo specifico, cioè in una porzione di quello che si chiama *gesture space*, che possiamo immaginare come una semisfera davanti al parlante, la cui grandezza è delimitata dall'estensione massima delle braccia in avanti, di lato, in alto e in basso (McNeill 1992: 86). Secondo McNeill, la maggior parte dei gesti ha luogo al centro dello spazio gestuale, che corrisponde più o meno allo spazio davanti al torace del parlante, oppure nell'immediata periferia. Ciò comporterebbe che, più un gesto è periferico, più dovrebbe diventare saliente. Per la codifica attuale si è deciso, a causa della libertà di movimento delle guide, di non usare una griglia, ma di codificare lo spazio in base al luogo del corpo in cui il gesto si realizza (gambe, torso, testa).

4.2.3 Dimensione del gesto

Anche la dimensione del gesto può essere interpretata in termini di sforzo comunicativo, con i gesti più grandi considerati un modo per amplificare e potenziare il

³ A causa dell'ambiguità del termine *phrase*, che a seconda dell'ambito di studi linguistici può intendersi sia come *intonational phrase* sia come 'sintagma', decidiamo in questa sede di non tradurre l'espressione con un equivalente in italiano.

loro contributo (Mol *et al.* 2011). La codifica delle dimensioni può essere eseguita o tramite una scala Likert o, tramite il punto di articolazione del gesto (spalla/gomito/polso; Mol *et al.* 2011). Per la presente codifica abbiamo scelto di usare quest'ultimo sistema.

4.2.4 Configurazione della mano

Codificare i gesti in base alla configurazione che le mani assumono nell'eseguirlo può essere molto utile, a patto però di essere consapevoli del fatto che mentre nelle lingue dei segni la configurazione ha un valore lessicale abbastanza stabile (almeno nelle forme 'da dizionario'), le configurazioni dei gesti intanto hanno una natura altamente idiosincratca, e per di più non sono sempre chiaramente articolate come nelle lingue dei segni a causa di una generale maggiore rilassatezza dell'articolazione. La codifica della configurazione della mano è fondamentale quando ci si propone di analizzare famiglie di gesti (cfr. § 4).

4.2.5 Funzione del gesto

Gesti con funzione deittica: gesti che indicano "un oggetto, una posizione, o una direzione, che è scoperta proiettando una linea dritta dal punto estremo della parte del corpo che è stata estesa verso l'esterno sullo spazio che si estende oltre il parlante" (Kendon 2004: 200, trad. nostra). Per il presente progetto, ci interessa in particolare il modo in cui i gesti deittici vengono inseriti all'interno della gesticolazione. Il corpus CHROME è particolarmente ricco in questa direzione: il discorso delle guide turistiche è tipicamente accompagnato da deittici che indicano l'oggetto della spiegazione, presente nell'ambiente. Inoltre, anche in questo corpus i gesti assumono anche la cosiddetta funzione 'deittica astratta' quando rimandano a un referente assente o si riferiscono a un referente fisicamente presente, ma hanno una funzione metonimica.

Gesti con funzione rappresentativa: gesti che rappresentano un referente attraverso un'immagine che, in qualche modo, lo raffigura (McNeill 1992). Nella terminologia di McNeill e dei suoi allievi, i gesti con funzione rappresentativa includono i gesti iconici e i gesti metaforici. Un gesto iconico è correlato al contenuto semantico del discorso e lo esprime con un'immagine corrispondente. Un gesto metaforico, invece, crea un'immagine concreta per un concetto astratto. Anche in questo caso, il corpus offre numerosi esempi di entrambe le categorie: oltre a discutere di referenti presenti nell'ambiente circostante, le guide offrono narrazioni sulla storia dei luoghi di interesse o su altre opere e altri siti, accompagnando queste narrazioni con numerosi gesti rappresentativi.

Gesti con funzione pragmatica: Per gesto con funzione pragmatica si intende un gesto "che esprime aspetti della struttura dell'enunciato, inclusi lo status delle porzioni di discorso e della loro relazione e il carattere dell'atto linguistico o della mossa interazione dell'enunciato" (Kendon 1995: 247, trad. nostra). Secondo Kendon, i gesti pragmatici possono essere ulteriormente suddivisi in:

- gesti con una funzione “modale”, cioè gesti che indicano in che modo ciò che è detto deve essere interpretato;
- gesti “performativi”, che indicano il tipo di atto linguistico rappresentato dall’enunciato;
- gesti con una funzione di analisi (*parsing*), che marcano le componenti logiche del discorso ed evidenziano i punti più salienti.

Sembra quindi che, in generale, mentre i gesti rappresentativi sono correlati al contenuto della frase, i gesti pragmatici hanno a che fare con l’atto linguistico in sé e con lo scopo che il parlante si prefigge di raggiungere.

Gesti batonici (beats): gesti che non sembrano contribuire al significato del parlato corrispondente né a livello semantico né a livello pragmatico, ma hanno la funzione di scandire il ritmo del discorso.

Sono esclusi dall’analisi i cosiddetti gesti adattatori, tutti quei movimenti, cioè, che non sono legati al parlato ma piuttosto alla dimensione emotiva o ad altri aspetti dell’interazione, come ad esempio movimenti che coinvolgono parti del corpo (ad es. toccarsi i capelli) o oggetti circostanti (ad es. giocare con una penna).

5. Risultati quantitativi preliminari

L’analisi dei dati è ancora in corso, quindi i seguenti dati svolgono la sola funzione di indicare possibili tendenze. I risultati qui riportati riguardano i primi 4 minuti di spiegazione di una delle guide, per un totale di 128 *stroke*. La frequenza dei gesti (27,7 gesti ogni 100 parole) è abbastanza simile a quella ottenuta in altri studi con parlanti italiani che si rivolgono a interlocutori reali (ad es. Campisi & Özyürek 2013).

Riguardo alle caratteristiche cinetiche, la maggior parte dei gesti finora codificati è a due mani (121 su 128). Oltre ai gesti che si svolgono nello spazio centrale, ci sono molti gesti in uno spazio periferico, in particolare all’altezza delle gambe (forse a causa della posizione eretta, poco frequente nei *setting* sperimentali dove invece i partecipanti sono seduti). Per quanto riguarda la grandezza, quasi la metà dei gesti codificati (57 su 128) hanno il gomito come punto di articolazione, e quindi possono essere considerato di media grandezza, mentre solo 13 sono caratterizzati da un movimento che coinvolge la spalla.

Riguardo alla funzione, la maggior parte dei gesti codificati ha una funzione pragmatica. Per questo motivo, dedichiamo l’ultima parte di questo contributo ad analizzare alcuni esempi provenienti da una delle famiglie di gesti più diffusa, la mano aperta supina, detta anche famiglia del *palm-up*.

6. Un esempio di gesti con funzione pragmatica nel corpus CHROME: *l'open hand supine family*

Nei suoi lavori sui gesti pragmatici, Kendon introduce la nozione di famiglia di gesti (*gesture families*). L'intuizione alla base è che alcune configurazioni sono usate con una funzione pragmatica più spesso di altre, almeno nelle lingue in cui sono state studiate. Questo indicherebbe che in alcuni luoghi, come ad esempio il sud Italia, molti gesti pragmatici sono convenzionalizzati, forse a causa del loro uso pervasivo nell'interazione, e meno soggetti a variazioni individuali, come invece accade per i gesti rappresentativi. Una famiglia di gesti, quindi, è un gruppo di gesti che condividono alcune caratteristiche cinetiche e, allo stesso tempo, un tema semantico. Dato che gli studi sui gesti pragmatici sono pochi, specialmente se confrontati con quelli sui gesti rappresentativi o sui deittici, e per lo più limitati all'italiano (Kendon 2004; Graziano 2014), allo spagnolo e al tedesco (Müller 2004), risulta ancora poco chiaro quali siano i parametri che permettano di attribuire a un gesto una funzione pragmatica. Di conseguenza, durante la codifica, si corre il rischio di procedere per esclusione, considerando pragmatici tutti i gesti per cui non è possibile identificare una relazione con il contenuto del parlato. La nozione di famiglia di gesti si rivela quindi fondamentale nell'identificazione dei gesti pragmatici, almeno per quelli più altamente convenzionalizzati.

Tra le famiglie di gesti individuate da Kendon (2004), la più studiata è senz'altro quella detta *Open Hand Supine* (OHS), che include tutti quei gesti con funzione pragmatica eseguiti con la mano aperta e l'avambraccio supino, in modo che il palmo della mano sia rivolto verso l'alto, sia in verticale sia in orizzontale. Tale gesto viene di solito associato all'immagine metaforica della mano aperta che regge o offre qualcosa. Le numerose osservazioni indicano, infatti, che si associa a segmenti di parlato in cui si introduce qualcosa che sta per essere detto o che commenta, spiega o chiarisce quanto detto. Secondo Kendon, inoltre, la mano può essere o estesa nello spazio di fronte al parlante (*palm presentation gesture*, PP) oppure diretta verso l'interlocutore (in quel caso si parla di *palm addressed gesture*, PA). Infine, la mano aperta è spesso usata per indicare, sia concretamente sia in astratto. Per descrivere meglio le possibili funzioni del *palm-up*, ricorriamo ad alcuni esempi del corpus.

6.1 Palm presentation gesture

Tra le funzioni identificate da Kendon, una delle più frequenti è quella di espandere le premesse o le condizioni per comprendere qualcosa, di solito introdotto nel parlato poco prima. Vediamo in che senso. Si consideri il seguente contesto: la guida sta descrivendo le varie fasi dei lavori di costruzioni della Certosa di San Martino, che raggiungono il loro *clou* nel Settecento. Ella afferma quindi:

- (1) Agli inizi del milleottocento gli ordini vengono soppressi

A partire da una posizione di riposo, in corrispondenza di 'vengono soppressi' la guida esegue un PP (vedi figura 4). La soppressione dell'ordine è la causa dell'abbandono della Certosa da parte dei certosini: come ipotizza Kendon, quindi, il gesto pragmatico in questione presenta il parlato corrispondente come una spiegazione di quanto detto in precedenza.

Figura 4 - Palm presentation gesture



Un altro contesto tipico in cui il PP si presenta è quello in cui il parlante sta facendo un elenco o sta contrapponendo due concetti. Ad esempio, si consideri il seguente enunciato:

- (2) Oggi giorno non è solo certosa ma anche museo nazionale.

In corrispondenza di “certosa” e di “museo nazionale” la guida esegue due PP a due mani, il primo caratterizzato da un movimento verso sinistra (figura 5) e il secondo da un movimento verso destra (figura 6).

Figura 5 - PP in corrispondenza di “certosa”



Figura 6 - PP in corrispondenza di "museo nazionale"



6.2 Palm addressed gesture

Tra le funzioni più diffuse del *palm up* c'è anche quella di rivolgersi verso qualcuno (di solito l'interlocutore) o verso qualcosa. Nel secondo caso, il gesto ha una vera e propria funzione deittica, che per Kendon assume una connotazione leggermente diversa rispetto all'indicare col dito indice, e cioè quella di presentare l'oggetto all'interlocutore come qualcosa verso cui si deve guardare.

Nonostante negli anni ci siano stati studi di queste famiglie semantiche, ci sono poche analisi di corpora grandi, e pochi tentativi di estrarre regolarità statistiche a partire dai casi sopra descritti. È chiaro, ad esempio, che le funzioni indicate da Kendon per ciascuna famiglia di gesti non sono esaustive, e nemmeno vincolanti. Ad esempio, non sono rari i casi in cui il parlante adoperi un PP con la funzione che Kendon associa invece al PL (*Palm lateral*), quella cioè di identificare qualcosa come ovvio. In questo senso, il corpus CHROME permetterà un'analisi sistematica delle famiglie di gesti pragmatici presenti, anche grazie alla dettagliata analisi del parlato corrispondente.

7. Conclusioni

Il progetto CHROME ha prodotto una vasta serie di risorse digitali da applicare sia nell'ambito della linguistica applicata che in altri ambiti scientifici interdisciplinarmente collegati a questa. Come già accennato in precedenza, tutte queste risorse sono online e disponibili a fronte di una richiesta che ha solo fini di censimento dei richiedenti. Anche a progetto formalmente concluso, le ricerche attive sui vari fronti descritti in questo lavoro continuano, il gruppo di ricerca evolve, i dimostratori continuano ad essere progettati. Fra i prodotti più importanti rilasciati dopo la sca-

denza del programma PRIN 2015, vale la pena segnalare la suite Fantasia (Origlia *et al.* 2019b) che, basata su tecnologie innovative, colleziona tutto il lavoro di virtualizzazione di ambienti e agenti umanoidi, vi aggiunge un sistema di dialogo e si apre al mondo della intelligenza artificiale e delle forme più avanzate di sistemi di rappresentazione della conoscenza semantica di dominio.

Data la sede di questa pubblicazione, fra le varie tematiche legate alla linguistica applicata sviluppate in CHROME, si è deciso di porre un focus sugli aspetti della comunicazione non verbale. In questa direzione, il primo passo sarà completare la codifica dei gesti secondo lo schema di codifica presentato in questa sede. In questo modo, sarà possibile avere un quadro più preciso sulla distribuzione delle caratteristiche cinetiche e semantiche dei gesti delle guide, e poter comparare la produzione gestuale di guide diverse nei diversi punti di interesse all'interno del percorso.

In questo senso, numerosi studi dimostrano che, normalmente, i partecipanti a uno scambio comunicativo non solo tengono conto dei gesti di chi parla, ma ne traggono beneficio. Ad esempio, Kushch e Prieto (2016) dimostrano che i soggetti ricordano più parole quando l'informazione più importante della frase è accompagnata da un gesto. Inoltre, secondo Breckinridge Church *et al.* (2007), non solo chi è esposto a gesti e parole insieme tiene conto dell'informazione presente nel gesto, ma ricorda tale informazione più a lungo di quella veicolata solo dalle parole. Tuttavia, tali benefici non si verificano automaticamente solo perché il parlante muove le mani, ma si manifestano solo in alcune condizioni, ad esempio, scompaiono se il parlante produce in prevalenza gesti adattatori (Obermeier *et al.* 2015). In altre parole, i destinatari tengono conto dei gesti quando il contesto e il parlante ci dicono che è appropriato farlo. Per questo motivo, è importante che la formazione di figure professionali che fanno del parlato la loro risorsa principale tenga conto anche della gestualità.

Riconoscimenti

Questo lavoro è stato reso possibile dal finanziamento del Ministero Italiano dell'Università e della Ricerca attraverso il programma PRIN 2015, progetto *Cultural Heritage Resources Orienting Multimodal Experience* (CHROME) #B52F15000450001.

Bibliografia

AAT. *Art & Architecture Thesaurus Online*. 2018. <http://www.getty.edu/research/tools/vocabularies/aat/>.

Alfano, Iolanda & Violetta, Cataldo & Schettino, Loredana & Ansani, Alessandro & Chiera, Alessandra & Poggi, Isabella & Isora, Sessa. in stampa. Corpus CHROME: la codifica multilivello di un corpus audiovisivo. In *Corpora e studi linguistici. Atti del LIV Congresso SLI*.

Bergmann, Kirsten & Macedonia, Manuela. 2013. A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In Aylett, Ruth & Krenn,

- Brigitte & Pelachaud, Catherine & Shimodaira, Hiroshi (eds.). *Intelligent virtual Agents*, 139-148. Berlino: Springer.
- Breckinridge Church, Ruth & Garber, Philip & Rogalski, Kathryn. 2007. The role of gesture in memory and social communication. *Gesture* 7(2). 137-158.
- Brunato Dominique & Cimino Andrea & Dell'Orletta Felice & Montemagni Simonetta & Venturi Giulia. 2020. Profiling-UD: A tool for linguistic profiling of texts. In Calzolari, Nicoletta & Béchet, Frédéric & Blache, Philippe & Choukri Khalid & Cieri Christopher & Declerck, Thierry & Goggi, Sara & Isahara, Hitoshi & Maegaard, Bente & Mariani, Joseph & Mazo, Hélène & Moreno, Asuncion & Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of 12th Edition of International Conference on Language Resources and Evaluation (LREC 2020)*, 7145-7151. Marsiglia, European Language Resources Association.
- Cameron F. Duncan. 1968. The museum as a communication system and implications for museum education. *Curator* 11(1). 33-40.
- Campisi, Emanuela. 2018. *Che cos'è la gestualità*. Roma: Carocci.
- Campisi, Emanuela & Özyürek, Asli. 2013. Iconicity as a communicative strategy. Recipient design in multimodal communication for adults and children. *Journal of Pragmatics* 47, 14-27.
- Cataldo, Violetta & Schettino, Loredana & Savy, Renata & Poggi, Isabella & Origlia, Antonio & Ansani, Alessandro & Sessa, Isora & Chiera, Alessandro. 2019. Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech. In Piccardi, Duccio & Ardolino, Fabio & Calamai, Silvia (a cura di), *Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale*, 205-231. Milano: Officinaventuno.
- Chiacchio, Maria Laura. 2004. *Gli appartamenti storici: una tipologia museale e le sue strategie di comunicazione*. Tesi di dottorato in Metodologie Conoscitive per la Conservazione e la Valorizzazione dei Beni Culturali. Seconda Università di Napoli.
- Cresti, Emanuela. 2000. *Corpus di italiano parlato. Volume 1: Introduzione*. Firenze: Accademia della Crusca.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Hammond, Michael & Moravcsik, Edith & Wirth, Jessica (eds.), *Studies in syntactic typology*, 209-239. Amsterdam: Benjamins.
- Graziano, Maria. (2014). The development of two pragmatic gestures of the so-called Open Hand Supine family in Italian children. In Seyfeddinipur, Mandana & Gullberg, Marianne (eds.), *From gesture in conversation to visible action as utterance: essays in honor of Adam Kendon*, 311-330. Amsterdam: Benjamins.
- Kendon, Adam. 1980. Features of the structural analysis of human communicational behavior. In von Raffler-Engel, Walburga (ed.), 29-43. *Aspects of nonverbal communication*. Lisse: Swets & Zeitlinger.
- Kendon, Adam. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics* 23, 247-279.
- Kendon, Adam. 2004. *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kuflik, Tsvika & Wecker, Alan. J. & Lanir, Jael. & Stock, Oliviero. 2015. An integrative framework for extending the boundaries of the museum visit experience: Linking the pre, during and post visit phases. *Information Technology & Tourism* 15(1). 17-47.

- Kushch, Olga & Prieto, Pilar. 2016. The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. *Proceedings of Speech Prosody 2016*, 922-925. https://www.isca-speech.org/archive/speechprosody_2016/kushch16_speechprosody.html.
- McNeill, David. 1992. *Hand and mind*. Chicago: The University of Chicago Press.
- Mol, Lisette & Kramer, Emiel & Maes, Alfons & Swerts, Marc. 2011. Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication* 17(1). 77-100.
- Müller, Cornelia. (2004). Forms and uses of the Palm up open hand: A case of a gesture family? In Müller, Cornelia & Posner, Roland (eds.), *The semantics and pragmatics of everyday gestures. Proceedings of the Berlin conference, April 1998*, 233-256. Berlino: Weidler.
- Obermeier, Christian & Kelly, Spencer & Gunter, Thomas. 2015. A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Social Cognitive and Affective Neuroscience* 10(9). 1236-1243.
- Origlia, Antonio & Savy, Renata & Cataldo, Violetta & Schettino, Loredana & Ansani, Alessandro & Sessa, Isora & Chiera, Alessandra & Poggi, Isabella. 2019a. Human, all too human. Towards a disfluent virtual tourist guide. *Proceedings of the 10th International Workshop on Personalized Access to Cultural Heritage (PATCH 2019)*, 393-399. New York: Association for Computing Machinery.
- Origlia, Antonio & Cutugno, Francesco & Rodà, Antonio & Cosi, Piero & Zmarich, Claudio. 2019b. FANTASIA: A framework for advanced natural tools and applications in social, interactive approaches. *Multimedia Tools and Applications*, 78(10). 13613-13648.
- Schettino, Loredana & Betz, Simon & Cutugno, Francesco & Wagner, Petra. In stampa. Hesitations and individual variability in Italian tourist guides' speech. In *Atti del XVII Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*.
- Sloetjes, Han & Wittenburg, Peter. 2008. Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Stock, Oliviero & Zancanaro Massimo 2007. *PEACH-Intelligent interfaces for museum visits*. Berlino: Springer.