

TRAIETTORIE DELLA MULTIMODALITÀ: GLI SNODI TEORICI E I MODELLI APPLICATIVI

*Maria Grazia Sindoni*¹

1. PREMESSA

Gli studi sulla semiosi della comunicazione, o multimodali, godono di una certa circolazione in Italia già da circa un ventennio, principalmente in ambito anglistico (Baldry, 2000; Baldry, 2005; Garzone, Poncini, Catenaccio, 2007), ma hanno guadagnato una popolarità crescente con l'avvento del digitale, dal momento che forniscono una serie di strumenti e metodologie utili per la comprensione della complessa, e sempre più pervasiva, testualità della rete (Sindoni, 2013; Prada, 2015). Tuttavia, la multimodalità emerge da premesse teoriche ben distanti dai fenomeni del mondo digitale, sebbene un equivoco lessicale ancora radicato porti a un uso interscambiabile dei termini “multimodale” e “multimediale”. Anzi, è sempre più frequente la combinazione “video multimodale” rispetto a “video multimediale”, con un rapporto di 2:1, secondo una ricerca delle due occorrenze in lingua italiana tramite il motore di ricerca Google².

Questo contributo nasce con l'intento di tracciare i confini teorici e metodologici dell'ambito di studi multimodali, per sgombrare il campo da ogni possibile fraintendimento lessicale, che, purtroppo, tende a trascinare in un fraintendimento semantico. Con questo obiettivo, si tenterà di ripercorrere i primi passi di questa area disciplinare, ormai ampiamente frequentata in Italia, per poi rintracciarne alcune delle direzioni epistemologiche principali, per orientare il dibattito attuale tramite la messa a fuoco degli ambiti preferenziali di studio e delle metodologie specifiche della multimodalità. Identificare i passaggi tramite cui la multimodalità si è sviluppata come disciplina autonoma, attraverso l'individuazione delle sue matrici teoriche, servirà dunque a inquadrare gli scenari in cui gli studi multimodali trovano la loro maggiore utilità applicativa. Ma al di là degli scenari applicativi, l'obiettivo principale intende definire la specificità epistemologica della multimodalità, distinta da discipline affini, quali la linguistica o la semiotica. Nei paragrafi successivi, si ricostruirà una breve storia di questo ambito attraverso un breve *excursus* dei maggiori snodi teorici, da cui discenderanno alcuni scenari applicativi, con suggerimenti su come la multimodalità possa orientare le domande di ricerca, fornire i metodi e prestarsi all'interpretazione di tipi di testualità e di eventi performativi che vanno ben al di là del digitale e della rete. Nella Sezione finale, si presenterà un breve esempio di analisi multimodale di un video TikTok molto popolare di Khaby Lame tramite il metodo classico della trascrizione e annotazione manuale, allo scopo di illustrare l'applicazione di principi e metodologie discusse nel saggio e di proporre un modello di analisi multimodale riproducibile in altri contesti, ad esempio scolastici, professionali o universitari.

¹ Università degli Studi di Messina.

² Quasi 30 milioni per la prima contro poco meno di 15 milioni per la seconda occorrenza, con una ricerca su Google svolta il 26 giugno 2022, con esito di entrambi a circa mezzo secondo (0,43 sec).

2. COME E PERCHÉ NASCE LA MULTIMODALITÀ

I primi studi che gettano le basi per un ripensamento radicale della teorizzazione della comunicazione sono *The Language of Displayed Art* di Michael O’Toole (1994) e *Reading Images: The Grammar of Visual Design* di Gunther Kress e Theo van Leeuwen (1996), quest’ultimo arrivato nel 2020 alla sua terza edizione. Sebbene Michael O’Toole sia stato fra i primi a ipotizzare che altri sistemi, oltre a quello linguistico, fossero dotati di vere e proprie grammatiche, il testo fondante è da molti riconosciuto nella Grammatica appena citata di Kress e van Leeuwen. Si tratta di un testo che ha riscritto il modo in cui i fenomeni della semiosi possono essere letti e interpretati e a cui ne sono seguiti molti altri che, analogamente, hanno adottato teorie e metodi per scrivere grammatiche di altri sistemi di semiosi non verbali, quali la musica (van Leeuwen, 1999), la danza (Maiorani, 2021), la matematica (O’Halloran, 2005), l’opera lirica (Rossi, Sindoni, 2016) e molti altri ancora.

La composizione di una grammatica in grado di spiegare il funzionamento sotteso ai meccanismi della comunicazione delle immagini prende le mosse da una riproposizione e reinterpretazione della grammatica sistemico-funzionale di Michael A. K. Halliday. Essa postula, in maniera simile a quanto aveva già fatto Saussure, che la comunicazione sia un sistema di produzione del significato (*meaning making*), al quale la lingua contribuisce solo in parte e non necessariamente in modo determinante o dominante (Halliday, 1978). Il sistema linguaggio è inizialmente uno stadio astratto, un potenziale di significato (*meaning potential*), che, successivamente, dopo essersi realizzato nel dominio intermedio della semantica, si concretizza in atti comunicativi che si esplicitano attraverso la lessicogrammatica. Quest’ultima si traduce in atti verbali tramite la risorsa semiotica della lingua, articolata nelle sue componenti materiali di scritto e parlato (Halliday, 1989a). Le grammatiche delle lingue descrivono lo stadio finale, più concreto, della lessicogrammatica, nella presentazione dei principi del funzionamento della lingua tramite sistemi di regolarità (i cosiddetti enunciati) in contesti sociali precisi (Halliday, Matthiessen, 2014). Si vedano, ad esempio, le grammatiche che tengono conto della varietà formale *vs.* quella informale, in una categorizzazione di registro, che si riferisce al contesto di produzione linguistica e non a un sistema lingua astratto (*inter alia*, Rossi, Ruggiano, 2022).

Tuttavia, i processi di significazione non sono vincolati solo al sistema linguistico, ma si possono attivare tramite diversi sistemi di segni organizzati e regolari che, a loro volta, si servono di risorse semiotiche specifiche di quel sistema. La regolarità con cui queste risorse “altre” si dispongono, per generare reti di senso interpretabili in modo coerente da una stessa comunità di parlanti, permette che questi sistemi possano essere descritti per mezzo di *grammatiche*. Coerentemente con questo approccio, alcuni multimodalisti hanno prodotto grammatiche riferite ad altri sistemi di semiosi, quali la musica (van Leeuwen, 1999) o la matematica (O’Halloran, 2005). Il concetto teorico di risorsa semiotica incorpora una componente *fissa*, legata alle proprietà intrinseche o alla materialità del segno e una componente *mobile*, determinata dalle convenzioni, abitudini, preferenze e pratiche sociali relative all’uso della risorsa stessa in un tempo e in un luogo specifico (Kress, 2009). Ad esempio, il colore è una risorsa semiotica nel momento in cui genera un senso sociale condiviso (Kress, van Leeuwen, 2002): il rosso del semaforo nel codice della strada significa obbligo di fermata, quindi è provvisto di uno schema di funzionamento regolare in quel contesto. La semiosi della materialità riguarda invece la differenza nella realizzazione di un artefatto multimodale con tratti di colore a olio rosso prodotte da un pennello (quali un dipinto), in luogo di una o più linee rosse prodotte da un’app digitale. Il tempo e il luogo determineranno la scelta di un certo uso o addirittura la possibilità della scelta stessa. Il concetto di risorsa semiotica, o modo (*mode*, Kress,

2009), è cruciale all'interno degli studi multimodali, dunque ritorneremo più avanti a parlarne.

Halliday, già a partire dalla fine degli anni Settanta, aveva sottolineato che uno studio diacronico della strutturazione e organizzazione delle risorse semiotiche è in grado di svelare i meccanismi usati dai parlanti per *modificare* la rappresentazione del reale (1978), indicando dunque una strada attraverso cui sfuggire al senso di determinismo linguistico che è possibile cogliere in alcuni approcci strutturalisti e post-strutturalisti. Michael Halliday crea le basi per lo sviluppo dell'ambito di indagine denominato *social semiotics*³, ovvero *semiotica sociale*, che riconosceva la lingua come un sistema di significazione di natura sociale (rifiutando l'innatismo e gli universali grammaticali di matrice chomskiana) e, allo stesso tempo, apriva lo sguardo a fenomeni semiotici, dunque non esclusivamente verbali, nei processi di produzione e interpretazione del significato, come si evince dalla seguente celebre affermazione:

È possibile dare la definizione forse più semplice di testo come lingua funzionale. Con funzionale si intende una lingua che svolge un certo compito in un certo contesto, al contrario di parole isolate o frasi scritte alla lavagna [...]. Dunque, ogni esempio di lingua viva che gioca un ruolo in un contesto situazionale può essere chiamato testo. *Può essere un testo parlato o scritto, e non solo, può essere prodotto con qualsiasi altro mezzo di espressione che ci può venire in mente.*⁴

La possibilità che il testo non debba necessariamente essere orale o scritto crea una via maestra per l'incorporazione della semiotica all'interno della linguistica. Questa idea non era affatto nuova, essendo già stata formulata da Saussure, che concepiva la lingua come parte di un più ampio sistema dei segni, da studiarsi tramite la disciplina "futuristica" denominata *semiologia*, di cui Saussure stesso preconizzò la centralità nel XX secolo. Questa disciplina, a suo dire, sarebbe stata determinata da uno sviluppo dei modelli teorici forniti proprio dalla linguistica ([1916] 1957). Si inizia a intravedere a questo punto un primo importante snodo che i primi multimodalisti intendevano far emergere, ovvero una presa di distanza da tutte quelle teorie che vedevano la centralità della lingua, il sistema prioritario sotteso a ogni fenomeno della semiosi. Questi studiosi volevano sfuggire alla morsa logocentrica che aveva vincolato ogni analisi di testi o interazioni a un paradigma esclusivamente linguistico. Non è però possibile rendere conto in modo esaustivo delle origini e delle motivazioni che spingevano a un tale deciso distanziamento, ma è certamente plausibile affermare che questi studiosi, *in primis* Michael Halliday, che getta le basi teoriche per lo sviluppo dell'intera disciplina, erano mossi da un forte impulso militante. Erano infatti alla ricerca di modelli in grado di scalzare le strutture pedagogiche imperanti nell'Inghilterra del secondo dopoguerra per potersi liberare dalle maglie di un sistema scolastico che faticava ad affrancarsi dalle ideologie linguistiche colonialistiche, secondo cui la misura del rendimento scolastico era direttamente proporzionale alla capacità di adeguarsi a uno "standard" monolitico e dunque escludente per definizione (lo Standard English, per l'appunto)⁵. In questo scenario, la lingua inglese era percepita, a

³ Espressione tratta dal titolo del volume di Michael Halliday, *Language as Social Semiotic* (1978), in cui la nozione di "semiotica sociale" fu introdotta, per la prima volta, in linguistica. Bob Hodge e Gunther Kress svilupparono le idee di Michael Halliday nel loro volume *Social Semiotics* (1988), creando un modello di riferimento teorico che da quel momento in poi accelerò la diffusione della disciplina, che guadagnò sempre più circolazione grazie alla costituzione della rivista *Social Semiotics* nel 1991, tutt'ora attiva ed edita da Routledge (Taylor and Francis).

⁴ Halliday (1989b: 10, corsivo e traduzione miei).

⁵ Si veda, fra gli altri, il lavoro di Raymond Quirk, che presentava l'insegnamento della lingua inglese, nella sua versione Standard, come l'unica possibile, invitando i discenti a non porsi come obiettivo «nulla che fosse inferiore al meglio» (Quirk, 1990, traduzione mia).

ragione, come un'arma integrante dell'arsenale imperialista (Kachru, 1991), atta a perpetuare la dominazione britannica, mentre lo studio della lingua, e la linguistica inglese tutta, erano articolate secondo logiche metriche di "competenza", che non solo non tutelavano i soggetti postcoloniali più vulnerabili, che l'avevano appresa come "lingua seconda", ma creavano le condizioni per una perpetua sottomissione all'eterna e ineludibile superiorità del madrelingua⁶. In questo *humus* culturale, si svilupparono premesse decisive per un ridimensionamento della linguistica intesa come potenziale strumento di perpetuazione dell'oppressione culturale colonialista e imperialistica (Atkinson, Delamont, Davies, 1995). Accanto a queste spinte ideologiche linguistico-disciplinari, la cultura dell'immagine iniziava a farsi sempre più forte, con la proliferazione di riviste con immagini a colori, pubblicità televisive e un progressivo interesse per il visuale, esploso definitivamente e capillarmente alla fine degli anni Settanta, per poi consolidarsi in Gran Bretagna all'inizio degli anni Ottanta. Infine, questi studi raccolgono, seppur implicitamente, l'invito epistemologico dei Cultural Studies (Hoggart, 1957; Williams, 1958), sviluppatosi negli anni Sessanta a Birmingham, che intrecciavano strumenti critici e teorici del marxismo, femminismo, etnografia e semiotica per studiare i fenomeni della comunicazione di massa come pratica sociale, eliminando la tradizionale divisione fra cultura cosiddetta "alta" (romanzi, poesia) e "bassa" (pubblicità, graffiti), per allargare lo sguardo verso la testualità quotidiana, ordinaria e prodotta da concreti attori sociali, che non erano mai stati studiati ma giocavano un grande ruolo nell'arena (mass)mediale. Coerentemente con questo focus di ricerca, anche gli approcci multimodali non escludono aprioristicamente alcun fenomeno dal loro radar e non eleggono oggetti di studio preferenziali, non creando distinzioni fra testi scientifici, interazioni spontanee, istruzioni di pronto soccorso, costruzioni Lego, monumenti, performance teatrali, fotogiornalismo o fumetti, come si vedrà in maggior dettaglio nella Sezione 5.

2.1. *Leggere le immagini: la grammatica visuale di Gunther Kress e Theo van Leeuwen*

La grammatica delle immagini di Gunther Kress e Theo van Leeuwen non è il primo lavoro che si propone di tradurre la grammatica della lingua inglese sistemico-funzionale di Halliday in una grammatica per altri sistemi di semiosi, ma probabilmente è il più noto come "*academic bestseller*", se si valuta l'impatto citazionale e la frequenza con cui compare come testo di studio primario in corsi di multimodalità in paesi anglofoni e non (come riferito da van Leeuwen, in una comunicazione personale del luglio 2022). Si prenderà quindi questo lavoro a modello per estrapolare dei concetti che sono stati trattati da numerosi altri autori, non perché debba necessariamente considerarsi come il precursore, bensì perché si presta a illustrare le strategie di significazione delle immagini secondo un sistema di regolarità che gli autori stessi definiscono *grammatica*.

Tutte le categorie sviluppate nel modello sistemico-funzionale della lingua di Halliday sono pertanto adottate e adattate per spiegare la significazione di ogni tipo di immagine statica, quali foto, dipinti, cartoline, pubblicità, schizzi, ma anche di ogni genere di rappresentazione grafica astratta, quali infografiche e tabelle, oltre che per suggerire alcune metodiche di funzionamento delle immagini in movimento, come video e film. Le macrocategorie a cui la scuola multimodale classica fa riferimento sono le metafunzioni hallidayiane. Per Halliday, la lingua ha lo scopo di attivare processi di significazione

⁶ Si veda ad esempio il lavoro di Basil Bernstein sulla sociologia della pedagogia e tutte le implicazioni legate alla sociologia dell'apprendimento della lingua madre (1971, 1973 e 1975). Per una trattazione del ruolo giocato da Bernstein nella concezione dello sviluppo di pedagogie alternative nel Regno Unito, si vedano Atkinson, Delamont, Davies (1995).

nell'interazione, reciproca e contemporanea, tra tre “funzioni delle funzioni”, o metafunzioni, ovvero l'*ideazionale* (che include l'*esperienziale* e la *logica*), l'*interpersonale* e la *testuale*. Il concetto di metafunzione è centrale nella concettualizzazione hallidayana di lingua come *sistema funzionale*, perché queste rappresentano gli strumenti per la significazione e strutturazione del pensiero e della realtà. Queste metafunzioni permettono la realizzazione di espressioni potenziali al livello astratto del sistema sociosemantico in veri e propri enunciati, resi possibili dalla lessicogrammatica; sono per questo motivo collegate alla transizione del linguaggio del bambino, definito come un sistema a due livelli (il contenuto e l'espressione) a quello dell'adulto, che invece è a tre livelli (il contenuto, la forma e l'espressione)⁷. Secondo Halliday, le metafunzioni sono componenti funzionali nella disposizione del sistema semantico, laddove nel linguaggio dell'adulto il concetto di *funzione* si dissocia da quello di *uso* e le funzioni si evolvono in componenti del sistema semantico e gli usi nei contesti sociali o tipologie situazionali. Questo passaggio, come dicevamo, riguarda la transizione al linguaggio degli adulti, che:

è ancora funzionale in origine ma dove il concetto di “funzione” ha subito un cambiamento importante: non è più semplicemente sinonimo di “uso”, ma è diventato molto più astratto, una sorta di “metafunzione” attraverso cui tutti gli innumerevoli usi concreti della lingua di un adulto trovano espressione simbolica in una forma sistematica e finita.⁸

Pertanto, le metafunzioni possono definirsi come degli aggregati del sistema, ovvero funzioni astratte con potenziale semantico, non ancora realizzate in enunciato, ma in grado di generare certi tipi di espressioni, che vengono, a loro volta, proiettate all'interno dell'enunciato, che può infine essere scomposto in tre diversi punti di vista (*ideazionale*, *interpersonale* e *testuale*). Queste funzioni delle funzioni possono essere osservate e studiate separatamente in modo euristico dal linguista, ma, come suggerisce Halliday stesso, operando simultaneamente all'interno della lingua in veri contesti d'uso. Dal momento che questi concetti sono stati, e continuano ad essere, ampiamente utilizzati in ambito multimodale, soprattutto nel vasto campo dell'approccio denominato “Multimodal Discourse Analysis” (MDA, si veda, fra gli altri O'Halloran, Lim, 2014), se ne riassumeranno le principali caratteristiche.

La metafunzione *esperienziale* riguarda le rappresentazioni del mondo esterno e interno e si collega alla nozione di campo (*field*) in ambito di registro, quindi, sintetizzando, si concentra su ciò che viene raccontato, vissuto, esperito dal parlante. Gli avvenimenti esterni al parlante, o interni alla sua coscienza, sono espressi tramite i *processi* (i verbi), i *partecipanti* (gli argomenti del verbo) e le *circostanze* (i complementi indiretti). La metafunzione ideazionale è tradotta in *rappresentazionale* nella grammatica multimodale di Kress e van Leeuwen, perciò aiuta a comprendere e a descrivere in modo sistematico ciò che l'immagine racconta in termini contenutistici o, come si potrebbe riformulare nelle domande: cosa mostra questa immagine? Che storia racconta (se la racconta)? Le categorie di interesse di questa metafunzione, fra le altre, riguardano quanti soggetti sono presenti nell'immagine, le relazioni che intercorrono fra loro in termini di direzionalità (*vectoriality*), ovvero le azioni che vengono rappresentate, mimeticamente o meno. L'immagine considerata fra le più mimetiche è la fotografia, mentre un esempio di immagine non mimetica è un'infografica o una tabella.

⁷ Halliday (1978: 56).

⁸ Halliday (1978: 22).

Tramite la metafunzione rappresentazionale, gli eventi visivi mostrati si estrinsecano tramite i *processi* (il verbo è sostituito dalla direzionalità indicata da un vettore)⁹, i *partecipanti rappresentati* (possono essere umani, antropomorfici o astratti, come le regioni in una carta geografica) e le *circostanze* (elementi di contorno che danno indicazioni sul dove o quando è avvenuta l'azione rappresentata). La vettorialità, in questo contesto, è fondamentale, perché riproduce ciò che nella lingua si realizza con il verbo; può essere espressa da rappresentazioni visive di movimento nelle azioni di natura fisica o materiale, ma anche dalla direzione dello sguardo di chi è messo in scena, ad esempio inscenando una madre che guarda la sua bambina. Come capisco cosa avviene nella scena? Seguendo la direzione (immaginaria) dello sguardo della madre. La metafunzione *logica*, invece, permette di collegare le rappresentazioni del mondo, ad esempio stabilendo relazioni semantiche fra le proposizioni (per espandere o riferire un'informazione) e spiegando la natura stessa di tali relazioni (ad esempio di dipendenza o indipendenza, ovvero ipotassi e paratassi visiva). Quest'ultima, integrata alla metafunzione esperienziale, dà luogo alla metafunzione ideazionale, come detto precedentemente.

La metafunzione *interpersonale*, invece, istanzia le relazioni fra le diverse componenti dell'enunciato per come sono realizzate nell'enunciato stesso. Non si guarda infatti a ciò che viene detto, ma a come viene presentata l'informazione dal punto di vista dell'enunciatario, prestando un'attenzione particolare al cosiddetto "tenor", ovvero alla relazione che l'enunciatario stabilisce fra il soggetto dell'enunciato e un 1) *operatore temporale* svincolato dalla componente lessicale (ovvero un marcatore che permette di stabilire in che posizione temporale si colloca l'enunciato rispetto al tempo dell'enunciazione o alla percezione del tempo trasmesso dall'enunciatario) o un 2) *operatore modale*, che rende conto della posizione dell'enunciatario rispetto all'enunciato, ad esempio se è presente una modalizzazione epistemica, ovvero una vicinanza o presa di distanza rispetto all'enunciato. Esempi classici che spiegano il funzionamento dell'operatore temporale potrebbero essere spiegati in lingua italiana con la seguente formulazione: *mi ha detto che era tornata*. Questo implica che l'enunciatario colloca sia il dire sia l'azione del tornare da parte di una persona di genere femminile nel passato, non sappiamo quanto distante, mentre *dice che è tornata*, riferisce un'azione passata che è stata riportata all'enunciatario in un momento vicino o coincidente con il presente. In lingua italiana, l'esempio *dice che è tornata* potrebbe essere persino letto pragmaticamente, come a indicare un punto di vista esterno (*si dice*) o epistemicamente, quasi a mettere in dubbio quello che è stato detto (*lo ha detto, ma sarà vero?*). Questi due esempi non sono tanto focalizzati su quando è avvenuta l'azione del *tornare*, ma piuttosto quanto sia vicina nel tempo il *dire* dal punto di vista di chi ha prodotto l'enunciato. Invece, la funzione dell'operatore modale è di natura epistemica o deontica e serve, anche in questo caso, non tanto a trasmettere un'informazione sull'azione avvenuta, bensì a spiegare come l'azione o l'evento siano percepiti da chi ne parla, dando infine informazioni sul punto di vista di chi giudica l'azione o l'evento

⁹ Ad esempio, una donna che è rappresentata in una foto in movimento è interpretata come tale perché è possibile cogliere visualmente un senso di direzionalità attraverso l'immaginazione di un vettore che porta verso una certa direzione (chi guarda potrà mentalmente visualizzare la direzione della camminata o della corsa). L'ampiezza dei movimenti compiuti (ma rappresentati) variano a seconda che la donna sia rappresentata durante una camminata (la microvettorialità sarà inferiore, indicando movimenti più lenti) o una corsa (la microvettorialità ci farà "vedere" dei movimenti più ampi, da noi interpretati come più veloci). Traducendo questa immagine in linguaggio verbale, si potrebbe dunque dire: "La donna cammina" o "La donna corre". Questo dimostra empiricamente che le azioni, siano esse enunciate attraverso il linguaggio verbale ("la donna cammina" o "corre"), o siano esse rappresentate visualmente, possono essere descritte come *processi*, ovvero *verbi* nel linguaggio verbale (ciò che mette in azione l'enunciato, allo stesso tempo dandogli un senso semanticamente) e *vettori*, o indicazioni di direzione, movimento, azione nel linguaggio visivo.

predicati, ad esempio come probabili, improbabili, sicuri, frequenti, obbligatori, consigliabili, e così via.

Nella grammatica multimodale, questa metafunzione è chiamata *interattiva* e, a sua volta, adotta una serie di criteri e categorie atte a comprendere la rappresentazione delle relazioni, ad esempio considerando il punto di vista rispetto alla posizione di chi guarda l'immagine, così come il colore, la luce, la distanza, la prospettiva e altro. Da notare, a questo punto, che, similmente a quanto descritto per la vettorialità, lo sguardo segue una sua grammatica nella sfera interattiva, ovvero può stabilire relazioni sia fra i partecipanti rappresentati e chi li guarda (se i partecipanti guardano verso lo spettatore), sia fra i partecipanti stessi (se i loro sguardi creano delle relazioni di senso all'interno del mondo rappresentato, per esempio a suggerire un dialogo o un altro tipo di interazione). Queste relazioni sono stabilite, rispettivamente, tra i *partecipanti interattivi*, dunque proiettando la relazione al di fuori della scena, o tra i *partecipanti rappresentati*, ovvero rimanendo all'interno della scena visiva ed escludendo il mondo di chi la scena la guarda nel e dal mondo reale. La differenza fondamentale nell'analisi dello sguardo in una prospettiva rappresentazionale e interattiva dipende da due punti di vista: se lo sguardo rimane solo interno alla scena, pertanto in una dimensione esclusivamente interna, o intradiegetica, oppure se lo sguardo è in grado di proiettare delle rappresentazioni di relazioni sia interne sia esterne, o extradiegetiche, ovvero coinvolgendo anche chi guarda.

La metafunzione *testuale* ha un ruolo sintattico nella lingua e serve a dare un ordine di priorità informativa all'enunciato. Questo ordine riflette una gerarchia rispetto al modo in cui le informazioni sono distribuite all'interno dell'enunciato, con l'enunciario che decide come organizzare questo sistema di priorità informativa. Esistono nella lingua delle preferenze non marcate, ovvero SVO nelle lingue SVO, ma le opzioni marcate, che alterano l'ordine non marcato, rispondono sempre a preferenze di distribuzione dell'informazione, che si riflettono, a grandi linee, nella ripartizione fra Tema e Rema. Questa funzione della lingua si traduce nella grammatica visuale di Kress e van Leeuwen in metafunzione *composizionale*, che però, chiaramente, non si può distribuire in maniera sintagmatica e lineare come per la lingua verbale (come nel caso del Tema e del Rema), ma si può disporre all'interno dello spazio in modo più libero, ad esempio con elementi in alto e in basso in una pagina (rispettivamente in posizione di *Ideale* e *Reale*) o al centro e ai lati (*Centro* e *Margini*), e così via.

2.2. *Altre teorie, altri percorsi*

Quanto discusso nella precedente Sezione è un esempio di un'applicazione dei principi stabiliti dalla linguistica sistemico-funzionale a fenomeni di semiosi non verbali. Tuttavia, in quel clima culturale, la grammatica visuale non fu l'unico tentativo di descrizione e sistematizzazione delle risorse semiotiche non verbali. Altri esempi di notevole impatto e circolazione sono certamente il lavoro di Michael O'Toole, che, come anticipato all'inizio della Sezione 2.1, aveva pubblicato, già due anni prima rispetto a Gunther Kress e Theo van Leeuwen, un primo tentativo di sistematizzazione di quell'approccio denominato *Systemic-Functional Multimodal Discourse Analysis* o, in breve, SF-MDA, ovvero analisi sistemico-funzionale del discorso multimodale. Nel suo studio, O'Toole descrive il funzionamento di diverse forme di arte, visive e verbali, in chiave sistemico-funzionale. Coerentemente con quanto è già stato illustrato, la nozione di metafunzione è applicata alla comprensione dei processi di significazione, che operano in sistemi semiotici non verbali, quali la pittura o il disegno, ma anche ad apparati tridimensionali, quali la scultura e l'architettura. O'Toole usa gli stessi tre criteri metafunzionali, già descritti a proposito

della grammatica visuale di Kress e van Leeuwen, per spiegare, rispettivamente: 1) *cosa è rappresentato* (focus rappresentazionale), 2) *come questa rappresentazione interagisce con il pubblico* (focus modale), e 3) *come questa rappresentazione è composta* (focus compositivo). L'ipotesi, nuovamente, è che lo spettatore colga i tre messaggi, intrecciati tra loro, simultaneamente, proprio come avviene nell'interpretazione del linguaggio verbale. L'aver proiettato la nozione di testo ben al di là del verbale e del visivo resta uno dei fondamentali contributi teorici di O'Toole, che amplia i confini epistemologici della testualità oltre ciò che pertiene alla verbalità, come dimostrato nel celebre saggio "Opera Ludentes: the Sydney Opera House at work and play", pubblicato a cura di un'altra figura molto nota nel panorama multimodale, ovvero Kay L. O'Halloran nel 2004.

Proprio come nella lingua, il potenziale Collocazionale degli elementi architettonici – la loro Congiunzione in stanze, piani ed edifici, il Riferimento sia reciproco che tra questi stessi e sia in riferimento all'ambiente in cui sono collocati – è ciò che li rende dei "testi" coerenti e usabili (O'Toole in O'Halloran, 2004: 11).

In questo saggio, O'Toole considera l'edificio del teatro dell'Opera di Sidney come un testo, che può essere letto nelle componenti *esperienziali* (o rappresentazionali), *interpersonali* (o modali) e *testuali* (o compositivi), osservando sia le relazioni interne fra le varie parti che costituiscono l'edificio, sia le relazioni esterne fra l'edificio e l'ambiente circostante, fisico, sociale e semiotico. Se l'edificio stesso costituisce la funzione esperienziale, con i suoi scopi pratici intessuti nel contenuto "lessicale" (come definito da O'Halloran, 2004) dei suoi ambienti, come il teatro, il proscenio, le poltrone, il sistema di luci e così via, le componenti interpersonali si estrinsecano nel modo in cui sono costruite (o immaginate) le relazioni fra i vari partecipanti all'evento performativo, quali gli spettatori, gli attori, i cantanti, l'orchestra. Non si tratta ovviamente di relazioni fra persone fisiche, ma piuttosto di relazioni immaginate e proiettate nella distribuzione degli spazi, che possono spiegare le ideologie costruttive di chi ha pianificato, ad esempio, come dovrebbero essere accolti gli abitanti provvisori dello spazio teatrale (cioè gli spettatori), cioè secondo dinamiche di potere o di solidarietà, di affiliazione o distanziamento. Infine, la funzione testuale si realizza tramite la relazione fra le varie parti, sia interne sia rispetto all'ambiente circostante, con cui il teatro in questione presenta elementi di continuità e di discontinuità. Questo modello riflette quanto già discusso a proposito della grammatica visuale di Kress e van Leeuwen, in quanto accoglie l'idea della tripartizione della semiosi in linea con la linguistica sistemico-funzionale elaborato da Michael Halliday.

Altri modelli che godono di circolazione sono quelli elaborati da Kay O'Halloran nel suo tentativo di integrare lo studio della lingua e dell'immagine come visualizzazione del simbolismo in matematica. Partendo dal riconoscimento che il linguaggio della matematica è talmente ricco di simbolismo da risultare inaccessibile ai più, O'Halloran ne traccia la sua evoluzione a partire dal linguaggio, fino ad arrivare a forme di rappresentazione simbolica che permette di incapsulare il senso globale in formule precise e non ambigue. L'integrazione della lingua, delle immagini e del simbolismo matematico mostra come ogni risorsa abbia una sua precisa funzione in matematica. Se la funzione della lingua serve al ragionamento e all'argomentazione euristica, il simbolismo serve piuttosto a catturare le dinamiche delle relazioni tra le entità e i processi matematici, che permettono infine di sistematizzare i risultati a cui si addiunge. Infine, le relazioni matematiche sono visualizzate nell'ottica di una semiotica visiva, che O'Halloran definisce un "circuitto semiotico", dove il senso potenziale di ogni risorsa è reso disponibile per la comprensione e risemantizzazione del significato matematico stesso, in una continua estensione del potenziale di rappresentazione e simbolizzazione della realtà fisica.

Un'altra teoria fondamentale all'interno della scuola SF-MDA proposto da un altro noto e prolifico multimodalista, John A. Bateman, è stata sviluppata per comprendere il rapporto testo-immagine (2014b). Il punto di partenza di questa riflessione mutua ancora una volta le intuizioni e le sistematizzazioni delle categorie del verbale di matrice hallidayiana, con particolare riferimento alle relazioni logiche di ipotassi e paratassi e a come queste possano essere utilizzate per comprendere le relazioni fra testo e immagine. Questo approccio è utile per comprendere la costruzione del messaggio anche attraverso diversi tipi di rappresentazione grafica, ad esempio paragonando le raffigurazioni pittoriche ai diagrammi scientifici, oltre a rivestire un valore predittivo, dal momento che consente di fare riferimento a griglie classificatorie dei generi multimodali (brochure turistica, ricetta interattiva di cucina, infografiche e statistiche), al pari dei generi testuali tradizionalmente intesi (romanzo, poesia, lettera commerciale, ecc.).

Di altra derivazione, invece, è la cosiddetta scuola *sociosemiotica*, anch'essa di discendenza hallidayiana, che si concentra sulla dimensione sociale della significazione nelle sue varie articolazioni, ad esempio in riferimento alla produzione, interpretazione e diffusione dei messaggi e dei testi in un dato contesto socioculturale con l'obiettivo di identificare quali siano gli attori sociali coinvolti e quali interessi essi possano riflettere o resistere. Anche in questo approccio, la nozione di risorsa semiotica è centrale, ma la risorsa è vista come uno strumento che risponde agli interessi di alcune classi sociali, osservabili in un'ottica spazio-temporale definita e il cui orientamento trova una sua collocazione non solo nella parte più militante della linguistica hallidayiana, ma anche nella semiotica classica e nella linguistica critica (*Critical linguistics*), che oggi è meglio conosciuta come Analisi Critica del Discorso (*Critical Discourse Analysis*). I maggiori fautori di questa scuola sono stati gli stessi Gunther Kress (insieme a Robert Hodge, 1979) e Theo van Leeuwen, che hanno però, da questo punto di osservazione, elaborato teorie atte a decostruire le dinamiche del potere, che si struttura nel e con il discorso in senso foucaultiano. Gli studi multimodali sociosemiotici, più che indagini del discorso da un punto di vista strettamente mutuato dalla linguistica, mettono in conversazione la semiotica, la filosofia e la sociologia. Uno degli assunti cardine di questa teoria è che il legame fra significato e significante non è arbitrario, come postulato da Saussure, bensì *motivato dall'interesse del produttore del segno*. Questo è un cambio paradigmatico che discende dall'ipotesi più ampia del sistema di segni come un apparato fluido e in perpetuo movimento, al contrario dell'ipotesi della lingua come struttura. Il segno, nell'approccio sociosemiotico, non è dunque il risultato arbitrario di una certa combinazione, ma scaturisce dal conglomerato di pratiche e consuetudini sociali, con la possibilità di uno scarto, o innovazione, inverabile grazie all'intervento del singolo, ovvero il produttore del segno.

Nella prossima Sezione, si passeranno in rassegna alcuni assunti teorici fondamentali, per rendere poi conto delle principali aree applicative e delle domande di ricerca che possono essere affrontate in queste prospettive di studio.

3. ALCUNI PRINCIPI FONDAMENTALI

Un principio che sottende a tutta la teoria e pratica legata al mondo della semiosi multimodale consiste nel non considerare la lingua come la risorsa più importante, o addirittura l'unica, per la comunicazione. Sebbene la linguistica come disciplina lasci presupporre che la lingua sia la risorsa per eccellenza per la comunicazione umana, epitome di quella scintilla che rende gli esseri umani diversi da qualsiasi altro essere vivente, la multimodalità si distanzia da questo assunto logocentrico, spostando la propria

attenzione sulla produzione del significato, risultante dalla combinazione di più risorse. Sotto questa luce, la comunicazione umana non avviene solo tramite la lingua accompagnata da risorse ancillari, liquidate come “paraverbali” o “extralinguistiche” (si veda, *inter alia*, Sindoni, 2011 e 2019). Ad esempio, l’analisi della conversazione non è considerata come uno scambio di enunciati a cui fanno da sottofondo elementi di contorno, come lo scambio di sguardi, il tono della voce, la gestualità e la distanza che si pone fra gli interlocutori. Nessuno di questi elementi è considerato aprioristicamente come secondario, ma ne viene esaminato il contributo specifico in ogni singolo momento dell’enunciato o dell’interazione, dando per assodato che una risorsa singola, come la lingua, non può avere una precedenza sulle altre in modo continuativo durante una conversazione, ma può invece diventare saliente, e dunque indispensabile per l’interpretazione stessa di un tratto o parte della conversazione stessa, in alcuni momenti specifici. Gunther Kress parlava, a tale proposito, dell’*orchestrazione del repertorio* delle risorse da parte dei partecipanti all’interazione o da parte degli autori di un testo (Kress, 2010). L’espressione “orchestrazione del repertorio” cattura alcuni elementi cardine della semiosi multimodale: proprio come in un’orchestra, la bellezza (e dunque il senso) del pezzo suonato non risiede nel contributo di uno strumento unico, ma, appunto, da quanto efficacemente tutti gli strumenti apportano un contributo al suono generato dall’insieme, che, tuttavia, non è semplicemente la *somma* di tutti i singoli suoni, ma un insieme, un profluvio di armonie che produce il piacere estetico dell’ascolto. Inoltre, non è detto che si debba necessariamente raccogliere il contributo di tutte le risorse a disposizione in un repertorio, ma che si possa (o si voglia) selettivamente trarre da alcune di esse per produrre messaggi nella forma di un testo, di una musica, di una foto o di una coreografia. Questo concetto di selezione, ancora una volta, è di derivazione squisitamente hallidayiana, ovvero l’idea che il linguaggio costituisca la codifica di un potenziale di azione (*behaviour potential*) in un potenziale di significato (*meaning potential*), cioè un mezzo per trasformare ciò che l’essere umano può *fare* in ciò che l’essere umano può *significare* (Halliday, 1975). Il passaggio dal sistema sociosemantico alla lessicogrammatica è un processo di riduzione: dal numero potenzialmente infinito di scelte per rappresentare la realtà, la lingua come sistema ne produce un numero definito di tipologie di processi (azioni, eventi, cioè i verbi), partecipanti (classi di oggetti e fenomeni, cioè i sostantivi), etc., che il parlante può usare per esprimere ciò che desidera, selezionando, pertanto, tra le opzioni messe a disposizione dal sistema lingua di riferimento.

Nel caso della testualità digitale, la descrizione delle scelte disponibili per produrre testi nati dall’orchestrazione delle risorse è parimenti ricca rispetto ad altre forme di testualità statica (come le immagini) o tridimensionale (come l’architettura). Basti pensare a una pagina web, a un blog o a un video TikTok, che si compongono di una serie variegata di risorse, ovvero immagini, layout della pagina (una forma di configurazione grafica diversa dall’immagine), musica, lingua scritta e parlata e altro. Queste risorse non saranno organizzate sempre allo stesso modo, ma al contempo adotteranno una serie di criteri facilmente riconoscibili per l’osservatore, ovvero si caratterizzeranno per una serie di regolarità, e alcune forme di variazione, che daranno l’indicazione di quanto il produttore della pagina web o il tiktokker sappia (o possa) essere creativo. Questa orchestrazione, quindi, rappresenta sia la stabilità che ha anche un carattere predittivo del modo in cui le regolarità si ergono a sistema, sia la rottura creativa e variazionale.

Abbiamo dunque potuto osservare che la lingua non è una risorsa gerarchicamente più importante in un modello di interpretazione multimodale, e che ogni risorsa, in combinazione con altre, apporta un contributo alla semiosi, sia esso un’interazione dal vivo fra parlanti, sia un artefatto multimodale come un oggetto 3D o una scultura, o, ancora, un testo scritto o una pagina web.

Da una prospettiva non dissimile, altri due multimodalisti, Anthony Baldry e Paul Thibault (2006), elaborano il *principio di integrazione delle risorse (resource integration principle)*, secondo il quale le risorse collaborano, integrandosi fra loro, per la produzione del significato. Secondo questo principio, nessun testo può essere definito monomodale, perché anche un libro di testo densamente scritto e privo di qualsiasi immagine, è, a ben guardare, composto visivamente con risorse che non hanno nulla di linguistico *stricto sensu*. Un libro di testo non contiene solo parole portatrici di senso, ma risorse grafiche, quali i margini della pagina, la suddivisione in capitoli e paragrafi, come anche degli accorgimenti tipografici, quali il corsivo e il grassetto, che possono enfatizzare talune parole, andando oltre il dato linguistico secco e riportando al lettore messaggi che rispondono alle intenzioni di chi le ha usate (si veda, come esempio, l'uso del corsivo come stilema di un autore). Possiamo addirittura ricordare a questo punto che la stessa spaziatura fra le parole non è altro che un'ulteriore convenzione grafica, a cui siamo, da lettori, talmente abituati da far sì che passi totalmente inosservata. In realtà, la risorsa semiotica della spaziatura pare sia stata gradualmente introdotta in vari contesti nell'antichità, secondo alcuni, da scribi irlandesi intorno al VII secolo per creare quelli che sono stati definiti *testi areati*, creando le condizioni per la nascita embrionale della lettura silenziosa (Saenger, 1997). Baldry e Thibault definiscono gli elementi di continuità, *generici*, e gli elementi di innovazione o unicità, *specifici-del-testo* (i.e., *text-specific*). A questo punto è collegata l'elaborazione di un secondo principio, detto di *compressione del significato (meaning compression principle)*, che si riferisce agli effetti prodotti dall'interazione fra le risorse semiotiche al micro-livello, cioè per come operano all'interno di un testo o di una interazione, e quelle ad un più alto livello di astrazione, livello in cui avviene l'osservazione e l'interpretazione del messaggio complessivo da parte dell'osservatore, che è in grado di leggere (*reading pathway*) il modo in cui le varie risorse si intrecciano fra loro. Secondo Baldry e Thibault, questo è un principio di economia, laddove le combinazioni multimodali, ad esempio configurate visualmente e verbalmente al micro-livello, come in una pagina web, forniscono modelli semiotici su una più vasta scala, ovvero il contesto situazionale e culturale di cui il testo in questione è parte. Un certo amalgama di risorse, dunque, "comprime" i significati, ovvero produce una sensazione di unitarietà all'occhio di chi osserva, ma per capirne a fondo la motivazione, la costruzione e gli interessi degli autori del testo (visti in questo contesto come *i produttori del significato*) occorre scioglierne l'intreccio. Questo amalgama di risorse può essere dunque sciolto, o, con un termine di origine multimodalista tanto caro a Baldry, *spacchettato (unpacked)*, al fine di comprenderne il preciso funzionamento in ogni sua componente, che può variare se il testo si sviluppa nello spazio (come una pagina, o, nel caso di un testo tridimensionale, di una scultura, un giocattolo o un edificio, si veda Stenglin, 2009) o nel tempo (come nel caso di una sinfonia, di un balletto o di un'opera teatrale).

Un altro importante concetto che verrà infine menzionato in questo contesto è l'*affordance*, un termine ripreso dallo psicologo della percezione James Gibson già a partire dagli anni Cinquanta e successivamente ripreso nel lavoro *The Ecological Approach to Visual Perception* del 1979, ma già estesamente trattato nel 1977. L'iniziale concezione gibsoniana vedeva nell'*affordance* un potenziale d'uso di un qualsiasi oggetto e per questo tradotto in italiano come "invito all'uso". Invece, Kress amplia il concetto e si riferisce a tutto ciò che un oggetto, o, in maniera traslata, una risorsa o un *mode* possono (*potentialities*, i.e. potenzialità) o non possono (*constraints*, i.e. vincoli) compiere, in un insieme di potenzialità e limiti, più recentemente definiti, in un contesto più ampio, guadagni e perdite (*gains and losses*, 2005). Le *affordance* di un mezzo, altra componente basilare entro cui si realizza un atto comunicativo, contribuiranno al modo in cui esso si possa esplicitare. Ad esempio, le *affordance* della piattaforma WhatsApp privilegiano l'uso di brevi testi scritti che possono

essere tipicamente utilizzati per la messaggistica istantanea. Tuttavia, le potenzialità del mezzo permettono anche l'invio di messaggi vocali, che l'utente potrà scegliere se e quando usare. Un'altra *affordance* del mezzo è di poter cancellare un messaggio già inviato: quello che invece non è possibile fare è eliminare traccia della cancellazione del messaggio. L'utente vedrà una scritta informativa autogenerata da WhatsApp nella chat: "messaggio cancellato". Questa impossibilità di cancellazione totale, anche delle tracce di cancellazione, è tecnicamente chiamato un *constraint*. Non è un testo prodotto dall'utente, bensì un messaggio automatico che informa della cancellazione di un contenuto. Una conseguenza comunicativa pragmatica potrà essere la richiesta, da parte dell'interlocutore, di chiarimento sul contenuto cancellato (ad esempio: "cosa hai cancellato?"), che genererà un testo non legato allo scambio in sé, bensì al modo in cui la piattaforma interviene e modifica l'interazione, in senso sia esplicito (nella forma di una domanda) o implicito (in una curiosità, fuggevole o intensa, a seconda del contesto in cui si svolge l'interazione e a seconda del momento in cui il "messaggio cancellato" interviene nella catena sintagmatica: qualcosa indica che il messaggio cancellato fosse particolarmente importante?). Altra *affordance*, sempre usando WhatsApp come piattaforma di esempio, è il segno di spunta singolo grigio (inviato ma non ricevuto), doppio grigio (consegnato ma non letto) o doppio azzurro (ricevuto e letto), che hanno portato a una serie di dinamiche conversazionali e interazionali tali da spingere gli sviluppatori dell'app a elaborare un'ulteriore *affordance* di privacy, ovvero l'opzione di non rendere nota l'avvenuta lettura del messaggio. Questi esempi potranno forse rafforzare nel lettore l'idea che tali fenomeni, seppur non linguistici, abbiano un impatto, in misura maggiore o minore a seconda dei casi, nella comunicazione.

John Bateman, sebbene da una prospettiva diversa, parla di *materialità*, per definire tutto ciò che serve a creare uno scenario o una griglia semiotica per l'interpretazione dei meccanismi di semiosi (2017). La pagina di un documento scritto impaginato a colonne preparerà il lettore a un certo genere scientifico, giornalistico o saggistico. *Affordances* e *materialità* sono concetti che rimandano a tutto ciò che circonda testi e interazioni, in un'ottica che spinge a riflettere sulle molteplici componenti che contribuiscono alla semiosi.

4. METODI E AMBITI

Ci si chiederà a questo punto che strumenti metodologici abbia sviluppato la disciplina della multimodalità per affrontare tutta l'ampia casistica di testualità che è abbracciata dall'interesse per una così corrispondentemente ampia fenomenologia. Prima di procedere con una rassegna dei principali metodi di cui si avvale un analista multimodale, è utile riflettere ancora per un momento su quali siano gli oggetti di studio di questa disciplina e chiedersi se e perché la testualità digitale sia un ambito di interesse preferenziale in campo multimodale o meno. Bisognerà anzitutto sgombrare il campo da una certa sovrapposizione che spesso è determinata da una ambiguità terminologica, come già accennato nella prima Sezione: la multimodalità e la multimedialità sono termini interscambiabili? Come già menzionato, i due ambiti sono completamente distinti e separati, in quanto la multimedialità riguarda la compresenza di strumenti e di tecnologie, spesso in campo massmediale. Esempi di multimedialità possono essere un video, un film o un podcast, perché rappresentano casi in cui più di un medium è usato per la produzione di un prodotto. In ottica multimodale, invece, la molteplicità a cui fa riferimento il prefisso multi- *riguarda le risorse e non i canali*. Ad esempio, una conversazione spontanea è un esempio di evento multimodale, perché riunisce le risorse della lingua parlata, la gestualità,

lo sguardo, la distanza sociale (prossemica) e tutto ciò che contribuisce alla semiosi. Qualsiasi testo o evento che abbia uno scopo comunicativo, visto sotto questa luce, è dunque multimodale. In tal senso, la multimodalità supera come ampiezza di interessi la multimedialità, che è solitamente riferita a testi massmediali, come dicevamo, pertanto di natura molto più recente. La multimodalità, invece, si può applicare, in questa concettualizzazione ampia di testo, ai graffiti preistorici, a un codice miniato o a una performance di un cantastorie, remota o attuale.

Tuttavia, la testualità digitale può certamente considerarsi come terreno ideale di applicazione delle teorie e dei metodi della multimodalità, perché presenta tipicamente lo schieramento di una serie di risorse semiotiche complesse, interattive e spesso modificabili dall'utente che può, almeno su certi livelli, operare, cambiare e ricombinare il testo. Questo è testimoniato dalla ricchezza di studi di matrice multimodale e sociosemiotica su testi digitali quali pagine web, social media, video su piattaforme di condivisione di media, audio e video clip, podcast. In breve, la multimedialità riguarda i canali, gli strumenti e le tecnologie, mentre la multimodalità si muove entro un orizzonte sociosemiotico, più interessato alla comprensione dei meccanismi che sottendono alla semiosi in un contesto sociale e culturale che costruisce "spazi di affinità" (traduzione letterale di *affinity spaces*, si veda, *inter alia*, Gee, 2004).

Come già brevemente illustrato, le domande di ricerca a cui si può rispondere utilizzando, in senso molto esteso, le teorie e i metodi multimodali sono: come funziona la produzione di senso in questo testo? Come attivare i processi di significazione, quali risorse entrano in gioco? In che modo queste si combinano e/o si alternano per scambiare messaggi all'interno di una comunità specifica, come un gruppo ristretto, o una più ampia, come i lettori di una testata giornalistica internazionale? Quali sono gli interessi dei produttori del testo o dei messaggi e come i produttori usano gli strumenti a disposizione (risorse, modi, *affordances*) per attivare dei percorsi di lettura e interpretazione del testo in questione?

Chiaramente, questo tipo di domande di ricerca non solo include una casistica estremamente vasta che oltrepassa i limiti della linguistica, ma richiede l'uso di strumenti e metodi molto variegati per lo studio e per la descrizione. Rimandando ad altre sedi la discussione dettagliata di tutto il repertorio metodologico maturato in questo campo, qui si menzionerà, a titolo esemplificativo, solo la grande distinzione fra metodi di tipo quantitativo e qualitativo. Sottolineando preliminarmente che difficilmente uno studio di matrice multimodale potrà essere di natura solo quantitativa, è bene ricordare che gli studi su corpora multimodali sono ben diffusi e continuano a produrre strumenti utili per le sfide della contemporaneità, per esempio legate ai big data, all'Intelligenza Artificiale e al *machine learning*. In tal senso, strumenti molto utili per la ricerca multimodale guidata dai dati (*data-driven*) sono stati prodotti da Anthony Baldry (Open MWS, si veda Baldry *et al.*, 2020a e 2020b), da Kay L. O'Halloran (MAP, si veda O'Halloran, 2008; O'Halloran, Pal e Jin, 2021) e da John Bateman (Bateman, 2014a). Al di là di software computazionali multimodali, i metodi qualitativi tradizionali includono la raccolta, anche a campionatura o selettiva, di uno o di una serie di testi utili per rispondere alla domanda di ricerca in questione. Il procedimento classico qualitativo è l'annotazione e trascrizione multimodale. Il multimodalista può sia servirsi di software per la trascrizione e annotazione (esempi sono OpenMWS, Elan, Transana, Multimodal Lab), che richiedono l'intervento umano, oppure di semplici tabelle, che saranno riempite con le annotazioni pertinenti alle risorse o i modi di interesse specifico per quel testo o evento interattivo che si intende studiare. La trascrizione manuale tramite tabelle è ancora ampiamente utilizzata, sebbene sia ancora fortemente criticata all'interno della comunità dei multimodalisti stessi, sia per la presentazione dei dati lineare e basata sui limiti grafici (*constraints*) della pagina scritta

tradizionale, che quindi si appiattisce secondo il sistema di scrittura usata e non rende conto della complessità, simultaneità e diversità semiotica delle risorse e dei modi in campo (si vedano, fra gli altri, Norris, 2002 e Flewitt *et al.*, 2009). Tuttavia, resta a tutt'oggi una metodica classica e mai abbandonata, anche per gli indubbi meriti didattici che riveste (Sindoni, 2019). Lo spaccettamento delle risorse, diversamente definito e concettualizzato dai vari teorici, è un punto di accesso utile alla comprensione dei fenomeni in chiave multimodale, perché l'interesse della disciplina è di provare a comprendere le modalità di significazione di uno scambio, un testo o di un evento attraverso l'uso combinato di diverse risorse in un contesto sociale e culturale, come si è provato a illustrare selettivamente con alcuni concetti teorici chiave corredati da esempi. Nella prossima Sezione, si presenterà un breve esempio di analisi di un testo della rete, per indicare alcune delle possibili traiettorie di indagine sostenute dalle teorie multimodali e dagli strumenti ad esse connesse.

5. APPLICAZIONI E ESEMPIO DI ANALISI

In questa Sezione si illustrerà un fenomeno della testualità digitale piuttosto recente per indicare alcune delle possibili traiettorie di applicazione multimodale tramite la trascrizione e annotazione manuale, replicabile anche in contesti didattici universitari per l'applicabilità e la funzionalità all'interno di diversi corsi e discipline, quali l'insegnamento dell'italiano come L2, la linguistica italiana o la didattica delle lingue. Come già anticipato nella Sezione 2, i possibili oggetti di studio verbali e non verbali sono potenzialmente infiniti e non si limitano alla testualità digitale. Agli occhi di un multimodalista, l'interessamento a un fenomeno non è legato al valore artistico, didattico o etico di un qualsiasi testo o evento. Nella scelta di un testo, nessun rilievo è dato a un pregiudizio sulla qualità vera o presunta del materiale, seguendo l'idea che ciò che motiva la produzione del segno sia l'*interesse del suo produttore* (Kress, van Leeuwen, 2001). Nel caso della testualità digitale l'interesse è, piuttosto, legato all'innovazione che un certo tipo di testo o evento apporta alla comunicazione o alla generazione di nuove tipologie, generi testuali o pratiche interazionali, soprattutto se precedentemente assenti o riconfigurati nel panorama della rete. Dunque, si escluderanno esempi di testi digitali da piattaforme studiate già da molti anni e ormai largamente presenti nella letteratura più consolidata nel settore, quali Facebook (Moschini, 2022; Tagg, Sergeant, Brown 2017), Twitter (Squires, 2015; Zappavigna, 2012) e YouTube (Benson, 2016; Androutsopoulos, Tereick, 2015; Sindoni, 2013), a favore di TikTok, meno studiato, ma estremamente popolare e in fase di crescita esponenziale (Kantrowitz, 2022).

5.1. *TikTok: il caso di Khaby Lame*

TikTok è una piattaforma cinese di condivisione di contenuti video generati dagli utenti, detti "creator", dalla durata breve, a partire da 15 secondi, impostati per default, fino a 3 minuti¹⁰. Sebbene le piattaforme di condivisione video fossero già abbondantemente diffuse, basti pensare al caso di YouTube, TikTok, sin dal suo lancio nel mercato extra-cinese alla fine del 2017, ha guadagnato un'immensa popolarità. È stata dichiarata il sito più usato al mondo a fine 2021, superando persino il colosso Google

¹⁰ La durata del video può anche essere estesa fino a 10 minuti con gli aggiornamenti disponibili a partire da febbraio 2022.

(Rosenblatt, 2021). TikTok è una piattaforma principalmente utilizzata dai giovanissimi, con il 41% degli utenti in una fascia di età compresa fra i 16 e i 24 anni, con un uso giornaliero di 1 ora e 25 minuti per il 91% di utenti (Meola, 2020). Nel 2022, TikTok è la terza piattaforma più diffusa globalmente, con numeri di utenti raddoppiati dal 2019 al 2022, da 294 a 656 milioni in meno di tre anni¹¹. Va da sé che un fenomeno così pervasivo nelle vite di ragazzi e studenti di ogni ordine e grado non possa essere semplicisticamente liquidato come propagatore e esaltatore di banalità e insulsaggini, nel migliore dei casi, o come strumento di cattura di dati sensibili, ideato per creare dipendenza psicologica e potenzialmente a rischio di diffusione di contenuti pericolosi o dannosi, nel peggiore dei casi. Dati e statistiche recenti affermano che il fenomeno è destinato a crescere, anche grazie a politiche meno severe sul controllo dei dati forniti da parte della casa madre e a un uso sempre più diffuso fra i giovanissimi a livello globale. Per questi motivi, agli educatori, docenti e ricercatori, non sfuggirà l'utilità di conoscere alcune delle dinamiche insite in questa piattaforma social, o, per dirla in termini multimodalisti, di comprendere il funzionamento della semiosi nell'orchestrazione del repertorio delle risorse disponibili in un certo spazio-tempo.

L'esempio scelto in questa sede riguarda quello che è al momento della scrittura di questo saggio (luglio 2022), il tiktokker più seguito al mondo, Khaby Lame, all'anagrafe Khabane Lame, classe 2000, ragazzo nato a Dakar, in Senegal, e trasferitosi a un anno di età con la famiglia in provincia di Torino, a Chivasso (Pasqui, 2021). Dopo essere stato licenziato da una ditta dove lavorava come operatore di macchine CMC a seguito della pandemia da Covid-19, Khaby Lame ha iniziato a pubblicare contenuti a partire dal marzo 2020 (Vinci, 2021). Attualmente vive a Milano e ha ottenuto questo record in una platea di tiktokker sempre più agguerriti e informati sulle strategie di marketing digitale, che permette una crescita di followers e consensi sotto forma di like e condivisioni. Tale primato è stato raggiunto il 23 giugno 2022 con l'impressionante cifra di quasi 150 milioni di followers e circa 2,5 miliardi di like. Per dare un contesto a queste cifre, la ben più famosa Chiara Ferragni, blogger giudicata dalla rivista Forbes nel 2017 «l'influencer più importante al mondo» (Forbes, 2017), totalizza, solo sulla piattaforma Instagram, l'assai più misera cifra di 28,4 milioni di followers, con una scalata social molto aggressiva, sostenuta da un immenso apparato finanziario, e, per di più, partita undici anni prima. Sebbene viva in Italia da 21 anni, Khaby Lame ha ricevuto l'annuncio tramite tweet del Sottosegretario all'Interno Carlo Sibilìa della concessione della cittadinanza italiana all'indomani della consacrazione a «tiktokker più famoso al mondo» (ANSA, 2022). Si veda sotto (Figura 1) la riproduzione del tweet originale del sottosegretario Sibilìa, risalente al 24 giugno 2022:

Figura 1. *Il tweet del sottosegretario Sibilìa che informa Khaby Lame della concessione della cittadinanza italiana*



¹¹ Dati riportati da Insider Intelligence al 1 giugno 2022 e consultati il 1 luglio 2022. Si veda: <https://www.insiderintelligence.com/charts/global-tiktok-user-stats>.

Questo annuncio segue l'ampia risonanza mediatica data all'impossibilità di Lama di raggiungere gli Stati Uniti per partecipare alla cerimonia di ViCon per l'incoronazione a re di TikTok, essendo ancora in attesa di visto, poiché privo di passaporto italiano. Ironicamente, come Lama ha fatto mestamente notare in più occasioni, lo stesso è conosciuto a livello globale come «tiktokker italiano» (Luna, 2021).

I tiktok di Lama sono brevissimi video senza voce, commento vocale, né testo in sovraimpressione, pertanto del tutto sprovvisti di contributo linguistico (Karimi, Mezzofiore, 2021). Spesso accompagnati da musica o rumori ambientali, sono tipicamente organizzati in modo bipartito, con un primo segmento che mostra un video preesistente che presenta un complicatissimo “*life hack*” (metodi per risolvere piccoli problemi quotidiani, come riparare un danno con metodi fai-da-te), riproposto da Khaby nel secondo segmento, che, silenziosamente, risolve il presunto problema (che in realtà non è tale), nel modo più ovvio possibile. «Se vuoi ridere, sei nel posto giusto» recita la presentazione del profilo.

In questa sede, la scelta per una breve trascrizione, annotazione e analisi multimodale è ricaduta sul video in Appendice, intitolato “Specchietto retrovisore”, perché ha totalizzato il maggior numero di visualizzazioni, ergo quasi 356 milioni a un anno esatto dalla pubblicazione¹². È diviso in due brevissimi segmenti narrativi: nel primo (Scene 1-5), si vede una ragazza che, per evitare di essere investita dalle auto mentre esce dall'abitacolo di una vettura, mostra uno specchietto adesivo da applicare nel lato posteriore dell'auto, utile a controllare se la strada sia sgombra e se sia possibile uscire in sicurezza. Nel secondo segmento, Khaby mostra la sua soluzione al problema: abbassa il finestrino, guarda se ci siano veicoli nei dintorni e infine chiude il video con il suo gesto che è ormai riconoscibile globalmente: un movimento delle mani con i palmi aperti rivolti verso l'altro, come a dire “Voilà”, che segna la degna conclusione del suo messaggio che potremmo “tradurre” in termini verbali come: “Questo non è un vero problema, ecco come fare, ridete insieme a me.”

Per brevità, si illustra questo caso con una trascrizione semplificata, riportata in Appendice, e con solo alcuni riferimenti generali alle metafunzioni discusse nelle precedenti Sezioni. La prima colonna è numerica e serve usualmente a indicare il numero del turno, che però, in questo caso, è assente, perché Khaby Lama non fa alcun ricorso alla lingua. Una funzione importante del numero di turno è la possibilità di discutere il turno o scena, menzionando il numero della riga in questione. Una colonna successiva, non impiegata in questo esempio, indica invece il minutaggio, altro elemento imprescindibile in un'ottica multimodale, soprattutto quando ci sono turni di parola e/o di testo scritto. In questa occasione, la colonna temporale è stata omessa, perché la brevità del video ha permesso una quasi integrale riproposizione degli screenshot singoli, ovviamente senza la replicazione di uno stesso screenshot quando permaneva a video per più di un secondo. Nella seconda colonna, invece, è stata inserito lo screenshot, ovvero un'immagine che mostra un momento saliente dell'evento. In trascrizioni molto più lunghe, l'analista multimodale di solito inserisce screenshot selettivi, solo per indicare i cambi di scena in senso lato, ma in questo caso, come già detto, è stato possibile restituire una trascrizione praticamente integrale. L'elemento visivo è molto importante e guida la lettura delle colonne seguenti, che fanno esplicito riferimento, sebbene in modo generico, alle tre metafunzioni discusse: la *Rappresentazione*, che riporta i contenuti e le esperienze, la storia – o una brevissima porzione di storia – che viene narrata attraverso varie strategie, fra cui la vettorialità dello sguardo che, come già detto, gioca un ruolo simile al verbo; l'*Interazione*, che allude a come i rapporti fra partecipanti interni e esterni alla storia (ovvero

¹² Si veda la versione YouTube, probabilmente più facilmente accessibile, al seguente indirizzo: https://www.youtube.com/watch?v=WFUj_0Txxxw.

il pubblico, l'osservatore) siano costruiti tramite i linguaggi usati; la *Composizione*, che richiama i concetti grammaticali di Dato e Nuovo, nella distribuzione spaziale degli elementi visivi e non presenti nell'immagine. Da notare, ancora una volta, come il linguaggio visivo traduca o risemiotizzi il linguaggio verbale, con strategie differenti ma, cionondimeno, produttrici di significato. Queste tre categorie qui impiegate per l'esemplificazione richiamano esplicitamente le metafunzioni rappresentazionali, interattive e composizionali della grammatica visiva di Kress e van Leeuwen, sebbene molti altri abbiano lavorato su altre forme di spaccettamento delle funzioni hallidayiane per la lingua verbale, non perché queste siano mutate dalla lingua, bensì per trasmettere il potenziale semiotico grammaticale (inteso come sistema di regolarità con caratteristiche di predittività) delle altre risorse, quali immagini statiche o in movimento, come in questo caso. L'ultima colonna, infine, commenta la componente sonora, che in questo esempio si suddivide in musica rap stereotipata nel primo segmento narrativo (Scene 1-5) e in rumori ambientali (Scene 6-10), integrati da effetti sonori in post-editing nelle Scene 11-13 per enfatizzare la gestualità iconica di Khaby Lame, che, con il suo gesto delle mani con i palmi aperti rivolti verso l'altro, come a dire "Voilà", mostra gli effetti della sua semplificazione di un'operazione presentata nel primo segmento narrativo come un "life back". Il trucco salvavita è demistificato nella macro-narrazione generale digitale di Khaby, mostrato come un'inutile complicazione e il trionfo della vanità, spesso legata a un acquisto superfluo, che viene quindi ridicolizzato dal gesto enfatico, a sua volta sottolineato dall'effetto sonoro da cartone animato, come se le mani fendessero l'aria rumorosamente.

L'annotazione e trascrizione riportata in Appendice permette di comprendere come le tre metafunzioni possano essere isolate e dissezionate singolarmente e come ciascuna apporti un contributo unico, ma rafforzato dalla presenza contestuale degli altri. La storia che viene raccontata è stata poco prima riassunta e vede la presentazione di un oggetto inutile (lo specchietto retrovisore per i passeggeri del lato posteriore dell'auto) nelle Scene 1-5 e la risoluzione di Khaby tramite la semplice apertura del finestrino posteriore con il capo voltato a controllare l'eventuale arrivo di macchine nelle Scene 6-13. Da un punto di vista interazionale, invece, se nella prima parte del video il messaggio riflette il punto di vista di marketing commerciale classico (questo è il problema, ecco la soluzione), la seconda parte del video mostra che la sezione precedente presenta in realtà un falso problema, tramite un costante incrocio di sguardi fra il protagonista/risolutore/demistificatore (Khaby) e il pubblico che osserva. La dimensione interazionale cambia in termini di attivo coinvolgimento del pubblico all'interno della scena, ora con uno sguardo in macchina, come si dice in gergo cinematografico, ora con ammiccamenti che fanno scattare meccanismi di identificazione fra Khaby e l'osservatore. Infine, alcune osservazioni sulla distribuzione spaziale delle immagini mostrate a schermo possono aiutare a comprendere come l'organizzazione delle informazioni sia sempre funzionale ai bisogni comunicativi specifici, perché allo spaccettamento, o divisione immaginaria dei vari spazi, segue una compressione degli elementi visivi, secondo il principio di compressione del significato elaborato da Baldry e Thibault. L'interesse del produttore di significato, Khaby, è di mostrare una piccola scena in cui l'esperienza ordinaria viene complicata senza ragione, a cui segue un *bathos* che lo vede protagonista, un *everyman* che provoca un sorriso o una risata. Gli elementi di novità, quale l'obiettivo della momentanea dissacrazione, e la ripetitività della struttura del video, insieme a quei gesti e sguardi che segnano la sua indiscussa celebrità digitale, ci riportano alla discontinuità e continuità del segno quando si fa testo, di cui abbiamo già discusso. La stabilità è data dalle *affordance* di TikTok (quello che è materialmente possibile fare con la piattaforma, tipicamente un video che in 15 secondi sia capace di suscitare curiosità e successiva fidelizzazione) e dalla

ripetitività di certi micro-generi e stilemi, mentre la discontinuità è ciò che riesce a far distinguere *questo* video da milioni di altri simili, ma ignoti.

Per riprendere altri principi illustrati precedentemente, Khaby Lame orchestra le risorse del suo repertorio, ovvero le sue risorse personali, come lo sguardo, l'espressione del viso che crea un'immediata sintonia con il pubblico di riferimento, la gestualità universale e le *affordance* del mezzo TikTok, che con il suo format breve ed editabile in maniera semplice e intuitiva, consente di parodiare contenuti già esistenti e intrappolare chi guarda per i 15 preziosi secondi, che costituiscono l'unità semiotica del testo. Per una lettura integrale di questi sistemi di significazione secondo per secondo, o meglio, screenshot per screenshot, si rimanda all'Appendice.

6. CONCLUSIONI

“Molto rumore per nulla” è la reazione classica di chi si avvicina per la prima volta a questo tipo di analisi, ovvero: “Perché sprecare tempo ed energia per studiare dettagliatamente un video così insignificante? Queste sono mode passeggere indegne dell'attenzione di un docente universitario”. Ci si chiederà, legittimamente peraltro, cosa abbia spinto 356 milioni di persone a visualizzare questo video, a riderci sopra, a commentarlo e apprezzarlo tanto da trasformare un operaio disoccupato ventenne di origini senegalesi dell'hinterland torinese, nel breve volgere di due anni, in una celebrità indiscussa che il mondo corteggia e ci invidia, tanto da spingere il governo italiano a conferire una frettolosa cittadinanza “compensatoria”. La stupidità umana crescente? La mancanza di passioni, il tedioso susseguirsi dei giorni, la pandemia o, più in generale, la glorificazione del nulla assurda a sistema all'interno del mondo del digitale, che predispone le menti a dilettersi con idiozie insulse nel tentativo di anestetizzarsi dalla noia o dalla disperazione? Non credo stia a noi trovare delle risposte che forse i sociologi, gli antropologi, i filosofi o gli studiosi di comunicazione di massa potrebbero affrontare da vari punti di vista. La reazione dei linguisti, che *non siano anche* analisti multimodali, è invece: “Questo video non contiene lingua, non ha alcun dialogo o elemento che lo renda linguistico. Perché dovrei occuparmene? Non rientra nell'alveo della mia disciplina, che se ne occupino altri, se proprio ne hanno voglia”. Questa affermazione può essere vera e anche questa posizione può essere legittimamente sostenuta con il supporto di declaratorie ufficiali o di teorie e prassi disciplinari. Tuttavia, il punto di vista che si intende sostenere in questa sede è che chi si occupa di comunicazione umana, e ancora più della cosiddetta lingua della rete (posto che la lingua della rete non può e non esiste al di fuori di un contesto intensamente multimodale), non può fare a meno di interrogarsi su questi fenomeni (Prada, 2022). Non solo perché, banalmente, sono “popolari”. Non solo perché hanno un numero di visualizzazioni elevato. Nemmeno perché conoscere questi fenomeni, e saperli interpretare criticamente, può aprire un canale di dialogo vero e profondo con i nostri figli o studenti sui mutevoli bisogni espressivi degli esseri umani. La motivazione che rende necessario confrontarsi con questa testualità non linguistica ma ipermodale (Lemke, 2002) è la dilagante propensione del mondo dei giovani a dedicare una sempre più preponderante parte delle loro interazioni e dei loro scambi a una testualità che è per lo più sconosciuta, e istintivamente allontanante, proprio per i citati motivi di inconsistenza, per il mondo della scuola e dell'università. La lingua italiana, come la lingua inglese e tutte le lingue che si trovano in rete, è ormai immersa in contesti spesso ignoti e si intreccia con pratiche comunicative che non è più possibile ignorare (Prada, 2022). Fenomeni quali il cyberbullismo, il body shaming, il revenge porn, i gruppi social che si incontrano dal vivo solo per creare delle risse o le sfide, talvolta a rischio della vita, lanciate

su TikTok, che sempre più spesso balzano agli onori della cronaca per la diffusione e la propagazione di messaggi di violenza, automutilazione, umiliazione di sé e degli altri non possono essere liquidati come esclusivo appannaggio del sociologo o del pedagogo, ma spingono a un profondo coinvolgimento di tutte le scienze sociali. Allo stesso modo, svelare i meccanismi dell'ironia e delle dinamiche che scatenano consenso e motivazione fra i più giovani può guidare la costruzione di percorsi didattici per il potenziamento delle abilità digitali e della comunicazione interculturale in contesti di italiano L2 o di inglese come lingua franca, come ad esempio illustrato nel *Common Framework of Reference for Intercultural Digital Literacies*, esito di un progetto europeo coordinato dall'autrice di questo saggio (Sindoni *et al.*, 2019) e disponibile nella sua traduzione in lingua italiana (Sindoni *et al.*, 2021). Sulla sperimentazione pluriennale che ha condotto allo sviluppo di questo strumento, si rimanda ad altra sede per una trattazione più ampia e più dettagliata, che esula dagli scopi di questo saggio (Sindoni *et al.*, 2021; Sindoni *et al.*, 2022).

Si è tentato di dimostrare che la multimodalità, sebbene abbia degli interessi in comune con gli studi massmediali e multimediali, è una disciplina che può essere fruttuosamente combinata con la linguistica, con cui condivide l'interesse per la lingua, sebbene non si tratti, come discusso, di un interesse semiotico esclusivo (Prada, 2022). Gli studi che intreccino linguistica e multimodalità possono trovare risposte alle domande e ai bisogni educativi delle giovanissime generazioni, che spesso lamentano un disallineamento totale fra le loro esperienze con la testualità digitale e i percorsi di apprendimento formale a scuola e all'università. Nel peggiore dei casi, le nuove generazioni non hanno neanche gli strumenti per formulare le domande e per esprimere i propri bisogni, per poter sfuggire agli innumerevoli pericoli e alle insidie che una testualità così strabordante, mutevole e talvolta incomprensibile mette in campo.

La storia di Khaby Lame, ormai prossimo a ottenere la tanto agognata cittadinanza italiana per i suoi meriti digitali, paradossalmente, non produce contenuti in lingua italiana, ma sceglie la comunicazione muta per raggiungere più persone possibili, come da sua stessa ammissione (Onukwe, 2021). Restano da esplorare grandi zone grigie, o zone diversamente colorate, come la comunicazione muta e non verbale, le immagini, la grammatica degli hashtag, le challenges (sfide) spesso indecifrabili agli occhi di chi non abita questi spazi di affinità, la viralità di contenuti pericolosi, il successo straordinario, e altrettanto oscuro, di testi o interazioni digitali, e molti altri fenomeni ancora, che nascono e si diffondono senza che un adulto medio ne sappia o ne possa tenere traccia. Questa è una delle sfide educative che la contemporaneità ci pone e la multimodalità può senza dubbio affiancare altre discipline con una lunga e importante tradizione in Italia, quali la linguistica italiana, per rispondere alle domande più urgenti per i giovani in formazione. Quelle che non vengono poste.

RIFERIMENTI BIBLIOGRAFICI

- Androutsopoulos J., Tereick J. (2015), "YouTube: language and discourse practices in participatory culture", in Georgakopoulou A., Spilioti T. (eds.), *The Routledge Handbook of Language and Digital Communication*, Routledge, London e New York, pp. 354-370.
- ANSA (2002), "World's most followed TikToker to get Italy citizenship", *ansa.it*, 24 giugno 2022: <https://www.ansa.it/english/news/2022/06/24/worlds-most>

[followed-tiktok-to-get-italy-citizenship_4421e607-aebc-477a-ae5c-97d0edd06f61.html](https://www.researchgate.net/publication/360107010-followed-tiktok-to-get-italy-citizenship_4421e607-aebc-477a-ae5c-97d0edd06f61.html) 01/07/2022.

- Atkinson P., Delamont S., Davies B., (1995) (eds.), *Discourse and Reproduction: Essays in honour of Basil Bernstein*, Hampton Press, Cresskill, NJ.
- Baldry A. (2000), *Multimodality and Multimediality in the Distance Learning Age*, Palladino, Campobasso.
- Baldry A. (2005), *A Multimodal Approach to Text Studies in English*, Palladino, Campobasso.
- Baldry A., Thibault P. J. (2006), *Multimodal Transcription and Text Analysis. A Multimedia Toolkit and Coursebook*, Equinox, London-Oakville.
- Baldry A., Kantz D., Loiacono A., Marenzi I., Taibi D., Tursi F. (2020a) “Accessing Specialised Medical Discourse in the Online Age”, in Bianchi F., Manca E. (eds.), *Specialised Discourse and Multimedia. Linguistic Features and Translation Issues*, ESE: Salento University Publishing, pp. 433-472.
- Baldry A., Thibault P.J., Coccetta F., Kants D., Taibi D. (2020b), “Multimodal Ecological Literacy: Animal and Human Interactions in the Animal Rescue Genre”, in Vasta N., Baldry A. (eds.), *Multiliteracy Advances and Multimodal Challenges in ELT Environments*, Forum, Udine, pp. 155-218.
- Bateman J. A. (2014a), “Using multimodal corpora for empirical research”, in Jewitt C. (ed.), *The Routledge handbook of multimodal analysis*. Second ed., Routledge, London-New York, pp. 238-252.
- Bateman J. A. (2014b), *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. Routledge, London-New York.
- Bateman J. A. (2017), “Triangulating transmediality: A multimodal semiotic framework relating media, modes and genres”, in *Discourse, Context & Media*, 20, pp. 160-174.
- Bateman J. A., Schmidt K.H. (2012), *Multimodal Film Analysis: How Films Mean*, Routledge, London-New York.
- Benson P. (2016), *The Discourse of YouTube: Multimodal Text in a Global Context*, Routledge, London-New York.
- Bernstein B. (1971), *Class, Codes and Control: Volume 1 – Theoretical Studies Towards A Sociology of Language*, Routledge e Kegan Paul, London.
- Bernstein B. (1973), *Class, Codes and Control: Volume 2 – Applied Studies Towards A Sociology of Language*, Routledge e Kegan Paul, London.
- Bernstein B. (1975), *Class, Codes and Control: Volume 3 – Towards A Theory of Educational Transmission*, Routledge e Kegan Paul, London.
- De Saussure F. (1967[1916]), *Corso di linguistica generale*. Introd., trad. e commento di T. De Mauro, Editori Laterza, Bari.
- Flewitt R., Hampel R., Hauck M., Lancaster R. (2009), “What are multimodal data and transcription?”, in Jewitt C. (ed.), *The Routledge Handbook of Multimodal Analysis*, Routledge, London-New York, pp. 40-53.
- Forbes (2017), “2017 Top influencers. Chiara Ferragni”, *Forbes.com*:
<https://www.forbes.com/profile/chiara-ferragni/?sh=53a2ab5d5a54>.
- Garzone G., Poncini G., Catenaccio P. (eds.) (2007), *Multimodality in Corporate Communication. Web Genres and Discursive Identity*, FrancoAngeli, Milano.
- Gee J. P. (2004), *Situated Language and Learning: A Critique of Traditional Schooling*, Routledge, London-New York.
- Gibson J. J. (1977), “The theory of affordances”, in Shaw R., Bransford J. (eds.), *Perceiving, Acting, and Knowing. Toward and Ecological Psychology*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 67-82.
- Gibson J. J. (1979), *The Ecological Approach to Visual Perception*, Lawrence Erlbaum, Hillsdale, NJ.

- Halliday M. A. K. (1975), *Learning How to Mean: Explorations into the Development of Language*, Edward Arnold, London.
- Halliday M. A. K. (1978), *Language as Social Semiotics*, Arnold, London.
- Halliday M. A. K. (1989a), *Spoken and Written Language*, Oxford University Press, London e Oxford.
- Halliday M. A. K. (1989b), “Part A”, in Halliday M. A. K., Hasan R. (eds.), *Language, Context, and Text: Aspects of Language in a Social Semiotic Perspective*, Oxford University Press, Oxford, pp. 1-49.
- Halliday M. A. K., Matthiessen C. M. I. M. (2014⁴), *Halliday’s Introduction to Functional Grammar*, Routledge, London e New York.
- Hodge B., Kress G. (1988), *Social Semiotics*, Polity Press, Cambridge.
- Hoggart R. (1957), *The Uses of Literacy: Aspects of Working-Class Life with Special References to Publications and Entertainments*, Chatto and Windus, London.
- Jewitt C. (2013), “Multimodal methods for researching digital technologies”, in Price S., Jewitt C., Brown B. (eds.), *The SAGE Handbook of Digital Technology Research*, SAGE, London, pp. 250-265.
- Kachru B.B. (1991), “Liberation linguistics and the quirk concern”, in *English Today*, 7, 1, pp. 3-13.
- Kantowitz A. (2022), “Inside TikTok’s explosive growth”, *CMSWire*, 16 maggio 2022: <https://www.cmswire.com/digital-experience/inside-tiktoks-explosive-growth/>
- Karimi F., Mezzofiore G. (2021), “He’s the most popular man on TikTok. And he doesn’t say a word”, in *CNN*: <https://edition.cnn.com/2021/10/02/world/tiktok-khaby-lame-cec/index.html>
- Kress G. (2005), “Gains and losses: New forms of texts, knowledge, and learning”, in *Computers and Composition*, 22, 1, pp. 5-22.
- Kress G. (2009), “What is mode?”, in Jewitt C. (ed.), *The Routledge Handbook of Multimodal Analysis*, Routledge, London-New York, pp. 54-66.
- Kress G. (2010), *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge, London-New York.
- Kress G., Hodge B. (1979), *Language as Ideology*, Routledge and Kegan, London.
- Kress G., van Leeuwen T. (2001), *Multimodal Discourse – The Modes and Media of Contemporary Communication*, Arnold, London.
- Kress G., van Leeuwen T. (2002), “Colour as a semiotic mode: notes for a grammar of colour”, in *Visual Communication*, 1, 3, pp. 343-368.
- Kress G., van Leeuwen T. (2006²), *Reading Images. The Grammar of Visual Design*, Routledge, London-New York.
- Lemke J. L. (2002). “Travels in Hypermodality”, in *Visual Communication*, 1, 3, pp. 299-325.
- Luna R. (2022). “Ministra Lamorgese, le presento Khaby Lame”, in *La Repubblica*, 23 giugno 2022: https://www.repubblica.it/tecnologia/blog/stazionefuturo/2022/06/23/news/ministro_lamorgese_le_presento_khaby_lame-355122560/.
- Maiorani A. (2021), *Kinesemiotics: Modelling How Choreographed Movement Means in Space*. Routledge, London-New York.
- Meola A. (2020), “Analyzing Tik Tok user growth and usage patterns in 2020”, in *Business Insider*, 25 febbraio 2020.
- Moschini I. (2022), *The World as a Global Community? A Critical Multimodal Discourse Analysis of Facebook’s Institutional Communication and Technical Documentation*, Effigi Edizioni, Grosseto.

- Norris S. (2002), “The implication of visual research for discourse analysis: Transcription beyond language”, in *Visual Communication*, 1, 1, pp. 97-121.
- O’Halloran K. L. (ed.), (2004), *Multimodal Discourse Analysis*, Continuum, London.
- O’Halloran K. L. (2005), *Mathematical Discourse: Language, Symbolism and Visual Images*. Continuum, London-New York.
- O’Halloran K. L. (2008), “Systemic Functional-Multimodal Discourse Analysis (SF-MDA): constructing ideational meaning using language and visual imagery”, in *Visual Communication*, 7, 4, pp. 443-475.
- O’Halloran K. L., Tan S., Smith B., Podlasov A. (2010), “Challenges in designing digital interfaces for the study of multimodal phenomena”, in *Information Design Journal*, 18, pp. 2-21.
- O’Halloran K. L., Lim F. V. (2014), “Systemic Functional Multimodal Discourse Analysis”, in Norris S., Maier C. D. (eds.), *Interactions, Images and Texts: A Reader in Multimodality*, Mouton de Gruyter, Berlin, pp. 137-154.
- O’Halloran K.L., Pal G., Jin M. (2021), “Multimodal approach to analysing big social and news media data”, in *Discourse, Context & Media*. Numero Speciale a cura di Sindoni M. G., Moschini I. *What’s Past is Prologue*.
<https://doi.org/10.1016/j.dcm.2021.100467>.
- Onukwe A. (2021), “A Senegal-born factory worker is Europe’s first TikTok megastar”, in *Quartz Africa*, 17 agosto 2021. <https://qz.com/africa/2048483/senegal-born-khaby-lame-is-as-famous-on-tiktok-as-charli-damelio/>.
- O’Toole M. (1994), *The Language of Displayed Art*, Leicester University Press, London.
- O’Toole M. (2004), “Opera Ludentes: the Sydney Opera House at work and play”, in O’Halloran K. L. (ed.), *Multimodal Discourse Analysis*, Continuum, London, pp. 11-28.
- Pasqui G. (2021), “Khaby Lame, ecco chi è l’italiano che ha più follower di Mark Zuckerberg ed è diventato una star dei social senza mai dire una parola”, in *Il Fatto Quotidiano*, 13 maggio 2021. <https://www.ilfattoquotidiano.it/2021/05/13/khaby-lame-ecco-chi-e-litaliano-che-ha-piu-follower-di-mark-zuckerberg-ed-e-diventato-una-star-dei-social-senza-mai-dire-una-parola/6196830/>.
- Prada M. (2015), *L’italiano in rete: usi e generi della comunicazione*, FrancoAngeli, Milano.
- Prada M. (2022), *Non solo parole. Percorsi di didattica della scrittura. Dai testi funzionali a quelli multimodali*, FrancoAngeli, Milano.
- Quirk R. (1990), “Language varieties and standard language”, in *English Today*, 6, 1, pp. 3-10.
- Rosenblatt K. (2021), “TikTok surpasses Google as most popular website of the year, new data suggests”, in *NBC News*, 22 dicembre 2021:
<https://www.nbcnews.com/tech/tech-news/tiktok-surpasses-google-popular-website-year-new-data-suggests-rcna9648>.
- Rossi F., Sindoni M. G. (2016), “The phantoms of the opera. Towards a multidimensional interpretative framework of analysis”, in Sindoni M. G., Wildfeuer J., O’Halloran K. L. (eds.), *Mapping Multimodal Performance Studies*, Routledge, London-New York, pp. 61-84.
- Rossi F., Ruggiano F. (2022), *La Grammatica Treccani per la scuola secondaria di primo grado*, Treccani Giunti T.V.P., Firenze, 4 voll.
- Saenger P. (1997), *Space Between Words: The Origins of Silent Reading*, Stanford University Press, Stanford.
- Sindoni M. G. (2011), *Systemic-Functional Grammar and Multimodal Studies. An Introduction with Text Analysis*, Ibis, Como-Pavia.

- Sindoni M. G. (2013), *Spoken and Written Discourse in Online Interactions. A Multimodal Approach*, Routledge, London-New York.
- Sindoni M. G. (2019), “Mode-switching in video-mediated interaction: integrating linguistic phenomena into multimodal transcription tasks”, in *Linguistics and Education*, 62, pp. 1-11.
- Sindoni M. G., Adami E., Karatza S., Marenzi I., Moschini I., Petroni S., Rocca M. (2019), *Common Framework of Reference for Intercultural Digital Literacies*, pp. 1-255.
<https://www.eumade4ll.eu/wp-content/uploads/2020/02/cfridil-framework-linked-fin1.pdf>.
- Sindoni M. G., Calarese S., Garito G., Restuccia G., Runcio V. (2021a), *Quadro Comune di Riferimento per l'Alfabetizzazione Digitale Interculturale*. Traduzione del *Common Framework of Reference for Intercultural Digital Literacies*:
https://www.eumade4ll.eu/wpcontent/uploads/2022/03/QCoRADI_definitivo.pdf.
- Sindoni M. G., Adami E., Karatza S., Moschini I., Petroni S. (2021b), “The theory and practice of the Common Framework of Reference for Intercultural Digital Literacies (CFRIDiL)”, in Sindoni M. G., Moschini I. (eds.), *Multimodal Literacies Across Digital Learning Contexts*, Routledge, London-New York, pp. 166-184.
- Sindoni M. G., Moschini I., Adami E., Karatza, S. (2022), “An introduction to the Common Framework of Reference for Intercultural Digital Literacies: learning as meaning-making and assessment as recognition”, in Ørevik S., Diamantopoulou S. (eds.), *Multimodality in English Language Learning*, Routledge, London-New York, pp. 221-237.
- Squires L. (2015), “Twitter: design, discourse, and the implications of public text”, in Georgakopoulou A., Spilioti T. (eds.), *The Routledge Handbook of Language and Digital Communication*, Routledge, London-New York, pp. 239-255.
- Stenglin M. K. (2009), “Space odyssey: towards a social semiotic model of three-dimensional space”, in *Visual Communication*, 8, pp. 35-64.
- Tagg C., Sergeant P., Brown A. (2017), *Taking Offence on Social Media. Conviviality and Communication on Facebook*, Palgrave Macmillan, Basingstoke.
- van Leeuwen T. (1999), *Speech, Music, Sound*, Palgrave Macmillan, Basingstoke.
- Vinci A. (2021), “Khaby Lama è l'italiano più seguito su TikTok: dalle case popolari al sorpasso su Gianluca Vacchi”, in *Corriere.it*, 23 aprile 2021:
<https://www.corriere.it/tecnologia/scuola-dad-didattica-in-presenza-futuro/notizie/khaby-lame-l-italiano-piu-seguito-tiktok-case-popolari-sorpasso-gianluca-vacchi-682a32ca-a38a-11eb-8b99-a42a4f90039f.shtml>.
- Williams R. (1958), *Culture and Society, 1780-1950*, Chatto and Windus, London.
- Zappavigna M. (ed.) (2012), *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web*, Continuum, New York.

APPENDICE

Annotazione e trascrizione multimodale del video di Khaby Lame

Scena	Screenshot	Rappresentazione	Interazione	Composizione	Suoni ambientali/Musica
1		<p>Una portiera di auto si apre, mentre un'altra auto sfreccia vicina.</p> <p>La mano che ha aperto la portiera anticipa l'uscita di un passeggero. La vicinanza dei due veicoli è molto stretta.</p>	<p>L'inquadratura suggerisce la possibilità che chi sta per uscire dal veicolo possa essere investito.</p>	<p>La macchina sulla sinistra è il Dato del video: il pericolo che sta in agguato e prende la forma del Nuovo nella parte destra del video.</p>	<p>Musica rap in sottofondo.</p>
2		<p>La persona rappresentata, presumibilmente la stessa che aveva rischiato prima di essere investita, non è identificabile se non per dei tratti femminili. Siede nel lato guida. La sua mano mostra uno specchietto adesivo, facendolo flettere avanti e indietro tramite veloci movimenti del polso.</p>	<p>Lo sguardo, o punto di vista, è interno all'auto. L'inquadratura è obliqua e restringe di molto il campo visivo dell'osservatore. Il focus è sull'oggetto mostrato, lo specchietto.</p>	<p>La composizione dell'immagine è statica, se non per il movimento che mostra le proprietà riflettenti dello specchietto, che è in posizione intermedia fra Dato e Nuovo.</p>	<p>Musica rap in sottofondo.</p>

3		La mano appoggia lo specchietto adesivo nella portiera posteriore dell'auto. Lo specchietto ha un orientamento verticale, rispetto a quello tradizionale che si intravede nella posizione usuale, ovvero accanto alla portiera anteriore.	Lo sguardo si sposta fuori dall'abitacolo. Il posizionamento dello specchietto invita l'osservatore a identificare il metodo d'uso: una semplice applicazione adesiva.	La composizione dell'immagine vede il Dato ormai stabilizzato visivamente, la sicurezza dell'auto, resa possibile dal Nuovo, il prodotto pubblicizzato.	Musica rap in sottofondo.
4		Lo specchietto è in primo piano e serve ad avvisare la passeggera sul momento giusto per scendere, perché riflette le immagini di auto che passano in velocità.	Lo sguardo si sposta dall'interno verso l'esterno. La doppia inquadratura, quella parallela all'abitacolo, e quella che proietta immagini da dietro, riproduce le proprietà dello specchietto a beneficio della passeggera, di cui l'osservatore si presume debba adottare il punto di vista.	Il Nuovo è diventato Dato a beneficio di chi lo sta provando, ma anche dello spettatore.	Musica rap in sottofondo.
5		La ragazza, dopo avere controllato lo specchietto e aver verificato l'assenza di pericolo, aziona la portiera ed esce dall'abitacolo. Il volto non viene mai mostrato e le fattezze generali sono solo accennate, per dare così rilievo alla funzione dello specchietto e non a tratti fisici o etnici riconoscibili della passeggera.	Il punto di vista è orizzontale e obliquo, come se l'osservatore fosse seduto a fianco della ragazza inquadrata. L'implicazione è che anche chi guarda il video possa verificare facilmente l'utilità di questo strumento.	Il Dato e il Nuovo sono congiunti e composizionalmente la focalizzazione è sulla figura che scende dal veicolo.	Musica rap in sottofondo.

7		Cambio di scena. Khaby è inquadrato frontalmente, con la sua tipica espressione fra lo stupito, il sardonico e il rassegnato, dentro l'abitacolo di una macchina. È seduto e tiene le mani intrecciate in grembo.	La prospettiva è rovesciata, non solo perché la nuova figura umana è ripresa frontalmente, ma anche perché la direzione di uscita del passeggero è opposta rispetto a quella presentata nel primo segmento del video. Khaby guarda direttamente negli occhi degli spettatori e sembra interrogarli su quanto visto fino a ora.	Il Dato è Khaby, una figura altamente riconoscibile nel contesto dello “spazio di affinità” del video. Il Nuovo è la portiera, sulla destra, che aspetta di realizzarsi nell'azione che sta per compiersi.	Suoni ambientali di sottofondo.
8		Khaby si sposta leggermente verso la portiera dell'auto e il suo sguardo si sposta su un dispositivo ben noto: il pulsante che aziona il movimento del finestrino. La vettorialità dello sguardo di Khaby ci fa capire come sarà risolto il falso problema presentato nella premessa del video.	L'osservatore vede ciò che l'unica figura in scena sta a sua volta osservando. Questo punto di vista crea una sensazione di identificazione con il semplice gesto che chiunque sia stato in auto come passeggero nell'abitacolo posteriore ha compiuto, ovvero usare il pulsante per gestire l'apertura del finestrino.	Il Dato rimane costante, mentre il Nuovo è introdotto dalla riproposizione, sintatticamente desueta, di un altro Dato: il pulsante del finestrino.	Suoni ambientali di sottofondo.
9		Khaby, di nuovo ripreso frontalmente, guarda gli spettatori. Mentre il suo sguardo sardonico, marchio di fabbrica, indugia proiettandosi al di fuori della scena rappresentata per raggiungere l'osservatore, l'unico gesto visibile è il dito che usa il pulsante per abbassare il finestrino.	Lo sguardo di Khaby non è mimetico, non segue l'azione che sta compiendo, è uno sguardo “in macchina”, come si direbbe nel linguaggio di semiosi del cinema, invita lo spettatore a aderire all'implicita derisione dell'oggetto presentato nella prima sequenza del video.	Anche in questo caso, il Dato è Khaby, mentre il Nuovo è il gesto che viene simbolicamente reso tale tramite la separazione, o dislocazione a sinistra, per adottare una descrizione grammaticale, fra il gesto e lo sguardo.	Suoni ambientali di sottofondo, rumore del finestrino che si abbassa.

10		Khaby guarda fuori dal finestrino con un movimento rapido e improvviso. Ecco svelato come fare a meno dello specchietto per passeggeri, presentato come così utile nella prima sezione del video: basta semplicemente aprire il finestrino e controllare che non passi nessuno.	Nessuno sguardo crea più un contatto fra il rappresentato e chi guarda, così la scena torna a essere mimetica e suggestiva di un'azione con cui tutti possono identificarsi facilmente.	Il Dato e il Nuovo sono congiunti in un unico enunciato visivo da un punto di vista compositazionale.	Suoni ambientali di sottofondo.
11		Khaby apre la portiera dell'auto. Oltre a presentare il suo sguardo come marchio di fabbrica, introduce a questo punto il gesto che lo ha reso riconoscibile globalmente, ovvero un movimento enfatico delle braccia e delle mani con i palmi aperti e rivolti verso l'alto, con direzionalità dall'interno verso l'esterno, che, senza far alcun uso delle parole, sembra voler dire: Ecco a voi: così si fa!	Lo sguardo è rivolto verso lo spettatore e continua a ingaggiare un dialogo muto, che a tratti può sfidare e a tratti sbeffeggiare chi guarda.	Il Dato e il Nuovo sono congiunti in un unico enunciato visivo da un punto di vista compositazionale.	Suoni ambientali di sottofondo. Effetto di suono di post-editing che enfatizza il movimento delle braccia e delle mani, come se producessero un suono "fumettistico" fendendo l'aria.
12		Il movimento delle mani, ritmicamente, passa dall'esterno verso l'interno, per mostrare la facilità dell'azione di guardare fuori dal finestrino prima di scendere dalla macchina, per verificare che non ci siano veicoli vicini in movimento. La coloratura e l'intenzionalità dello sguardo rimangono sempre invariate, altro tratto distintivo di questa comicità di natura auto-evidente.	Ancora sguardo rivolto verso chi osserva, a reiterare la banalità della situazione, in un invito a cogliere il ridicolo di chi vuole vendere un oggetto di cui in realtà nessuno ha veramente bisogno. La staticità dell'intenzione dello sguardo invita ad aderire al messaggio muto che si prende gioco dell'inutilità dello specchietto per passeggeri.	Il Dato e il Nuovo sono congiunti in un unico enunciato visivo da un punto di vista compositazionale.	Suoni ambientali di sottofondo. Effetto di suono di post-editing che enfatizza il movimento delle braccia e delle mani, come se producessero un suono "fumettistico" fendendo l'aria.

13		<p>La sequenza descritta nelle scene 11 e 12 si ripete, ma questa volta in modo molto più enfatico, con movimenti più ampi che mostrano il passaggio dall'interno verso l'esterno. La gamba sinistra è già fuori dal veicolo.</p>	<p>Lo sguardo, sebbene il corpo sia notevolmente sbilanciato verso il lato destro di chi osserva, rimane ancorato fortemente al punto di vista di chi osserva, come a intrecciare un gioco di sguardi con gli spettatori.</p>	<p>Il Dato rimane sempre lo stesso, mentre il Nuovo si proietta al di fuori dell'abitacolo, grazie alle linee disegnate dalla gamba sinistra di Khaby.</p>	<p>Effetto di suono di post-editing che enfatizza il movimento delle braccia e delle mani, come se producessero un suono "fumettistico" fendendo l'aria.</p>
----	---	---	---	--	--