

LORENZO SPREAFICO¹

Un corpus di cartoni animati come input per l'acquisizione e l'apprendimento linguistico: studio pilota su ambiente acustico e paesaggio sonoro

Abstract

In this paper, a multimodal corpus for the study of cartoons as input for language acquisition is presented. After explaining what prompted us to compile the corpus, the stages of defining, collecting, and transcribing the data are briefly discussed. Eventually, we show how the information contained in the corpus can be used as a starting point for studies of the acoustic environment and the soundscape to which viewers of the cartoons considered are exposed.

1. Introduzione

Per quanto i cartoni animati² siano stati oggetto di numerose indagini scientifiche – tipicamente da parte degli studiosi di comunicazione di massa e di educazione ai media (Cappuccio 2015) – in ambito italiano solo poche ricerche ne hanno trattato in ottica linguistica e, nel caso, privilegiando un approccio sociolinguistico monomodale tendenzialmente finalizzato a identificare i tratti caratterizzanti della specifica varietà di italiano trasmesso (Sardo & Ferro 2008; Sardo 2018).

Per questa ragione, si è ritenuto utile trattare della questione anche a partire da una base di dati più ampia e caratterizzata in ottica multimodale così da non solo contribuire ad aggiornare la caratterizzazione del tipo testuale a più livelli, ma anche da rendere disponibile una fonte di dati attendibile da cui muovere per verificare le ipotesi già in letteratura sul rapporto tra cartoni animati, acquisizione e apprendimento linguistico. Di qui la volontà di procedere ad uno studio pilota per la costruzione di un *Corpus di cartoni come input per l'acquisizione e l'apprendimento*

¹ Libera Università di Bolzano.

² “Cartoni animati è un'espressione popolare derivata dall'inglese *cartoons* e, benché meno elegante ed erronea, si è imposta su quella, più corretta in lingua italiana, di *disegni animati*. *Film d'animazione o cinema d'animazione*, invece, sono definizioni tecniche, che comprendono molto altro oltre a quella produzione seriale e industriale di lungometraggi e cortometraggi cinematografici o televisivi a cui, nel linguaggio quotidiano, diamo l'etichetta familiare di *cartoni*” (Tagliapietra 2010: 18). In questo contributo consideriamo *cartone* quella particolare specie del film di animazione (Rondolino 2003) caratterizzata dall'essere un cortometraggio seriale non fotografico trasmesso (anche) da una rete televisiva italiana e basato su una “forma di animazione che utilizza disegni o modelli bi- o tridimensionali per creare una sequenza di immagini in movimento” (ChatGPT 2023).

linguistico (COCAINAA), le cui fasi di approntamento sono delineate in §2 e la cui spendibilità è esemplificata in §3.

2. Costruzione del corpus

Per lo più le operazioni di redazione di un corpus di cartoni non differiscono da quelle per la costruzione di ogni altro corpus audiovisivo (Thompson 2022) e contemplano perlomeno tre fasi: quella di definizione dei contenuti (§2.1), quella di raccolta dei contenuti (§2.2), quella di trascrizione dei contenuti (§2.3).

2.1 Definizione dei contenuti

La fase di definizione dei contenuti è finalizzata a garantire che il corpus sia rappresentativo, ovvero dia testimonianza di quanti più possibili fenomeni rilevanti.

In tal senso, per quanto riguarda COCAINAA si è deciso di includere nella base di dati sia cartoni con dialoghi originali in lingua italiana, sia la versione in lingua italiana di cartoni con dialoghi originali in altra lingua³, ché questi ultimi sono i più trasmessi dalle emittenti televisive italiane (Pumares & Fontaine 2021) e i più visti dal pubblico italiano (Auditel c.p.). Tuttavia, poiché “despite the passing of a decade and the ever-increasing sophistication of speech-to-text technologies, a fully accurate and sophisticated automated approach to spoken corpus construction has still not been developed” e dunque “the collection and processing of spoken corpora remains largely manually driven and is typically costlier and more time-consuming to undertake than the construction of written corpora” (Knight & Adolphs 2022: 21), si è deciso di operare un’ulteriore selezione rispetto alla moltitudine di cartoni trasmessi dalle emittenti italiane e quindi di considerare solo quelli in onda dopo l’anno 2000 e, tra questi, quelli che (a) non fossero esplicitamente rivolti a un pubblico adulto stando alle note di accompagnamento dell’opera oppure alla valutazione del suo pubblico di riferimento secondo le emittenti che li trasmettono e (b) siano (stati) liberamente e gratuitamente accessibili al pubblico.

Così facendo si sono identificati quale base di dati iniziale per il corpus le serie di cartoni *La Pimpa* (lingua originale: italiano; anni di produzione: 1975-2005)⁴; *Bentornato Topo Gigio* (italiano; 1988-1989); *Winx Club* (italiano, inglese; 2004-2019); *Yo Yo* (italiano; 2017-in corso); *Cuccioli* (italiano; 2003-2012); *Mini cuccioli* (italiano; 2016-2019); *Pocoyo* (spagnolo, inglese; 2005-in corso); *Blaze e le mega machine* (inglese, 2014-in corso); *Topo Tip* (italiano, 2014-2019); *Nina&Olga* (italiano, 2021-in corso).

³ La differenza, di cui si tiene traccia nei metadati, permetterà di confrontare il parlato dei due tipi di prodotto e, in caso di differenze, di valutarne il ruolo per l’acquisizione.

⁴ I dati sono ripresi dalla pagina web ufficiale del cartone, se disponibile; oppure dalla pagina web dedicata all’opera in Wikipedia; oppure dalla pagina web dedicata all’opera nel sito web www.imdb.com.

2.2 Raccolta dei contenuti

La fase di raccolta dei dati è finalizzata a recuperare tutte e solo le informazioni che servono per costruire il corpus.

Per la fase pilota, mirata a valutare la fattibilità del lavoro, si è deciso anzitutto di focalizzarsi sulla prima e l'ultima puntata di ciascuna serie di cartoni elencata in §2.1 disponibili al momento della costruzione del corpus, anche per massimizzare sin da subito la variabilità dei dati da trattare e, dunque, far emergere quanti più possibili aspetti problematici cui dare risposta. Successivamente si sono rintracciate nel World Wide Web copie digitali liberamente accessibili delle puntate selezionate, per es. sulle pagine web www.raiplay.it, www.youtube.com, www.vimeo.com. Infine, si sono scaricate copie elettroniche delle puntate identificate ricorrendo al software *easy youtube video downloader* (versione 18.5) così da standardizzare la procedura di acquisizione dei dati tanto con riferimento al formato dei file (.mp4), quanto alla frequenza di campionamento e di impostazione dei livelli audio in modo da normalizzare eventuali successive analisi automatiche o impieghi in contesti sperimentali. Dalle copie elettroniche delle puntate individuate si è infine estratto e salvato in formato digitale .wav il sonoro.

Poiché né il rinvenimento dei cartoni in pagine web liberamente accessibili, né il loro sfruttamento per finalità non commerciali legittimano di per loro l'utilizzazione di prodotti audiovisivi tutelati dalla legge sul diritto di autore, si è reso necessario identificare delle modalità che garantissero che l'impiego delle copie dei cartoni fosse conforme alle normative. A tal fine, laddove fattibile – come nel caso di *La Pimpa* – si è sottoscritta una liberatoria tra i detentori dei diritti di utilizzazione dell'audiovisivo e il gruppo di ricercatori che lo avrebbero trattato, sempre e solo per finalità scientifiche e di ricerca (fig. 1); altrimenti si è preso a riferimento il decreto legislativo 177/2021 di recepimento della direttiva 2019/790/UE che prevede alcune limitazioni all'esercizio del diritto di autore motivandole con un superiore interesse generale e pertanto legittima “le riproduzioni compiute da organismi di ricerca [...] per scopi di ricerca scientifica, ai fini dell'estrazione di testo e di dati da opere o da altri materiali disponibili in reti o banche di dati cui essi hanno lecitamente accesso, nonché la comunicazione al pubblico degli esiti della ricerca ove espressi in nuove opere originali” (articolo 70-ter)⁵.

⁵ La normativa legittima l'uso ma non anche la distribuzione delle riproduzioni, il che impone che al momento COCAINAA non possa essere reso né pubblicamente accessibile, né fruibile da parte di chi non sia stato esplicitamente autorizzato a lavorarci da parte dei detentori dei diritti. Poiché come rilevato anche da uno dei revisori ciò renderebbe la base di dati non FAIR (Wilkinson *et al.* 2016; <https://www.go-fair.org/>), ci si sta impegnando per trovare una soluzione adeguata al problema.

Figura 1 – Copia di liberatoria anonimizzata

LIBERATORIA

Tra la _____, con sede in _____ (P.IVA _____) nella persona del suo legale rappresentante, Signora _____ (qui di seguito _____) da una parte;

e dall'altra parte i Professori _____, domiciliati ai _____, con sede in _____, fini della presente liberatoria presso la _____ (di seguito **RICERCATORI**).

PREMESO CHE:

- _____ agisce nella sua qualità di agente e rappresentante esclusivo di _____ (in arte _____) - qui di seguito semplicemente l'AUTORE, creatore del personaggio _____ e unico autore della serie di storie di cui essa è protagonista (di seguito il PERSONAGGIO).

SI CONVIENE E SI STIPULA QUANTO SEGUE:

- 1.- La _____, concede il permesso ai RICERCATORI, alle condizioni e con i limiti previsti dalla presente liberatoria, il permesso di utilizzare per fini di ricerca i video disponibili sul sito del PERSONAGGIO:
- 2.- I RICERCATORI hanno come obiettivo costruire una banca dati linguistica per indagare come la lingua italiana usata nei cartoni animati si sia evoluta nel tempo e se la lingua lì impiegata possa essere utilmente impiegata come strumento per l'insegnamento delle lingue.
- 3.- I RICERCATORI dopo aver scaricato i video dei cartoni animati del PERSONAGGIO, trascriveranno con i criteri propri della trascrizione scientifica del parlato. I dati trascritti saranno utilizzati unicamente dai componenti del gruppo di ricerca cui afferiscono i RICERCATORI presso _____ ed esclusivamente per scopi di ricerca, dunque senza fini di lucro, commerciali, artistici o creativi.
- 4.- I RICERCATORI inoltre riprodurranno parte dei dati trascritti anche in pubblicazioni scientifiche accompagnandoli, se necessario per chiarezza espositiva, con la riproduzione di fermi immagine del PERSONAGGIO ripresi dai video.
- 5.- È espressamente stabilito ed accettato che i RICERCATORI non acquisteranno alcun diritto sul PERSONAGGIO al di fuori di quanto stabilito nella presente autorizzazione e che la presente liberatoria non costituisce un diritto di commercializzazione.

INIZIALI


2.3 Trascrizione e annotazione dei contenuti

La fase di trascrizione comporta la trasposizione scritta per finalità analitiche delle informazioni contenute nei cartoni, *in primis* con riferimento alla componente parlata.

In COCAINAA questa viene trattata muovendo dalla consapevolezza che “spoken discourse is multi-modal in nature” (Knight & Adolphs 2022: 21), ovvero che il parlato è l’esito della sintesi di “verbal and embodied actions” (Mondada 2018), laddove le azioni incarnate sono per es. gesti, sguardi, posture, movimenti, manipolazioni, pratiche sensoriali.

Tuttavia, in COCAINAA il punto di partenza della trascrizione dei cartoni non è, come nel caso dei corpora di parlato, la registrazione di un evento come per es. il contesto di fruizione del cartone da parte di alcuni bambini che vi interagiscono e lo commentano verbalmente, bensì l’originale audiovisivo stesso che, a sua volta, è “a semiotic construct comprising several signifying codes that operate simultaneously in the production of meaning” (Chaume 2004: 16). Pertanto, nel caso dei cartoni la multimodalità è connaturata al fenomeno e consegue dalla compresenza di una serie di segni sempre intenzionali (perché artefatti) parti di un codice di comunicazione a più livelli: *illustrativo non fotografico* (dato da disegni, colori, montaggio, ecc. e veicolato nel canale visuale), *musicale* (dato da musiche inserite in post-produzione e veicolate nel canale aurale), *rumorale* (dato da effetti sonori artificiali, non colti in presa diretta e veicolati nel canale aurale), *paralinguistico* (messo in scena dai doppiatori come nel caso di risate previste dal copione o portato naturalmente come per la qualità vocale, e veicolato nel canale aurale) e infine *linguistico*, veicolato nel canale aurale per il parlato e il cantato e in quello visuale per lo scritto.

Sebbene questi codici possano essere concettualizzati da chi idea il cartone come parti di un tutto⁶, in realtà sono elaborati singolarmente e in quanto tali sono fruiti e percepiti dallo spettatore, come testimoniato dal fatto che alcuni livelli condizionano la percezione degli altri (Ansani *et al.* 2020) e che ogni percettore attribuisce loro un peso diverso, per es. in funzione di variabili soggettive come l’età⁷. Per tale ragione, in fase di trascrizione si è deciso di riportare in maniera discreta ciascuno dei codici di comunicazione, ovvero di adottare una trascrizione multimodale e dunque, in ultima istanza, di elaborare un corpus multimodale.

Peraltro, va esplicitato che in COCAINAA il riferimento alla multimodalità non è limitato come da accezione diffusa nelle discipline linguistiche alla resa della moltitudine di modalità comunicative – tradizionalmente le parlate e le gestuali – che concorrono alla (de)codifica del significato, ma è esteso anche alla tracciatura delle diverse modalità di saturazione di ciascun canale comunicativo, per es. con riferimento a quello acustico, che nel cartone veicola segni che rimandano a referenti che nel mondo reale assolverebbero tanto a funzioni comunicative linguistiche (ad

⁶ Prova ne è il ricorso alla tecnica di *mickey-mousing* (Leeuwen 1999).

⁷ Bruner (1966) sostiene che fino a circa 7 anni il bambino impiega soprattutto la visione e le immagini per (de)codificare la realtà.

es. le parole) e non linguistiche (ad es. il pianto), quanto a funzioni non comunicative (ad es. uno schiocco). In tal senso, l'approccio alla trascrizione di COCAINAA è prossimo a quello rintracciabile nelle ricerche di impronta cineiconica (Burn 2013), laddove “the kineikonic mode is a term for the moving image as a multimodal form. It indicates an approach to the analysis of film, video, television [...] that examines how systems of signification such as image, speech, dramatic action, music and other communicative processes work together to create meaning within the spatial and temporal frames of filming and editing” (Wikipedia 2023).

Per tale ragione, in COCAINAA la trasposizione scritta di ognuno dei livelli ipotizzati è stata effettuata impiegando l'ambiente di trascrizione multitraccia ELAN (Elan 2023) predisponendo non solo un file indipendente per ciascuna puntata trascritta, ma anche approntando per ciascun file righe di trascrizione allineate temporalmente alla traccia audio del file multimediale per più livelli (fig. 2):

- 1a. la trasposizione ortografica del *parlato* di ciascun personaggio o narratore parlanti, prendendo a riferimento per la trascrizione le forme contenute nel Vocabolario Treccani (edizione 2018)⁸;
- 1b. la trasposizione per sintagmi fonologici del parlato di ciascun personaggio o narratore parlanti, prendendo a riferimento per la resa Nespor (1993);
- 1c. la qualità vocalica dominante negli enunciati di personaggi o narratori parlanti, riferita secondo le convenzioni proposte da Ball *et al.* (2018);
2. la presenza di *cantato*, intesa come lingua parlata modulata per ritmo, melodia e armonia per finalità più estetiche che comunicative⁹;
3. la presenza di *musica*, ovvero di successioni strutturate di suoni prodotti con strumenti musicali organizzati secondo le dimensioni di ritmo, melodia e armonia;
4. la presenza di *effetti sonori*, intesi come suoni e rumori creati appositamente per imitare suoni del mondo reale, per indicizzare prototipicamente eventi del mondo reale, oppure per accompagnare o enfatizzare le animazioni;
5. la presenza di segni o gesti di *scrittura*, reali o fittizi.

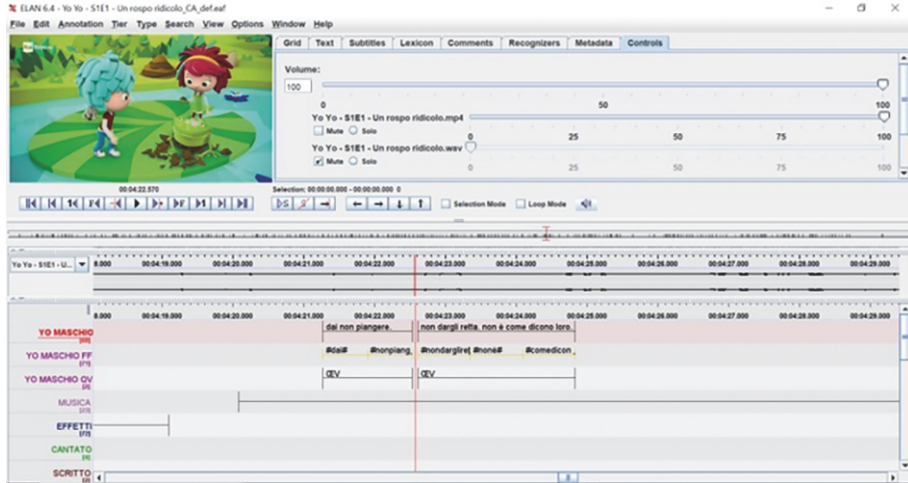
La trascrizione, operata da Aurora Cantoni, Debora Arnoldi, Iris Dall'Aglio, Matteo Largaiolli e dall'autore di questo articolo, è stata effettuata uditivamente senza ricorrere in alcuna fase del processo a sistemi per il riconoscimento o l'analisi automatica del parlato, del cantato, della musica, degli effetti, delle immagini così da tenere un approccio emico prossimo a quello dell'esperienza di fruizione tipica

⁸ La trascrizione del parlato in luogo dell'inserimento di una copia della sceneggiatura originale è legittimata dal fatto che quest'ultima, pur essendo scritta per esser letta come non fosse mai stata scritta (pena l'innaturalità), può in realtà esser resa dai doppiatori in maniera diversa da quanto previsto dagli sceneggiatori.

⁹ L'assegnazione di parlato e cantato a linee di trascrizione diverse pur a parità di parlante consegue dalla letteratura scientifica di riferimento dove si legge che “the question of whether speech and song recruit shared or distinct neural systems remains a contentious and controversial topic” (Hymers *et al.* 2015).

di un cartone e da disporre di una base di raffronto da contrastare a tentativi di trascrizione automatica già in programma per il prossimo futuro.

Figura 2 – *Esempio di trascrizione*



3. *Uso del corpus*

Numerosi sono gli ambiti di spendibilità e gli utilizzi di un corpus multimodale. A titolo di esempio nei prossimi due paragrafi si tratta della caratterizzazione dell'ambiente acustico e del paesaggio sonoro – fenomeni rilevanti per una riflessione sul rapporto tra *input*, *intake* e acquisizione – prendendo a riferimento una porzione del corpus che contiene puntate da *La Pimpa*, *Yo Yo*, *Pocoyo*, *Mini cuccioli*, *Blaze e le mega macchine*, *Topo Tip* per un totale di 55 minuti.

3.1 Ambiente acustico

Le linee di trascrizione relative ai livelli da 1 a 4 permettono di tenere traccia di ogni evento che si dia nel canale acustico e provenga dal cartone, vale a dire consentono di definirne l'ambiente acustico, ovvero l'insieme oggettivo e misurabile dei fenomeni sonori cui lo spettatore è potenzialmente esposto (Brown *et al.* 2015). In tal senso, per quanto ancora limitati, i dati sinora disponibili in COCAINAA permettono di ricavare alcune informazioni interessanti, anzitutto con riferimento alla distribuzione dei diversi fenomeni con valori o effetti comunicativi. Risulta infatti che, in termini assoluti di durata e con riferimento alle coppie di puntate considerate nel loro complesso includendo anche le sigle di testa e di coda, la musica è presente in tutti i cartoni e in termini di frequenza costituisce il primo input aurale per lo spettatore in quattro serie di cartoni su sei (tab. 1). La cosa colpisce, perché almeno impressionisticamente la prevalenza della musica non pare caratterizzare

nessun'altra esperienza di acquisizione linguistica tra quelle che si danno tanto in contesti naturali quanto in contesti formali¹⁰.

La base di dati permette di osservare poi che anche il parlato è presente in tutti i cartoni considerati, pur non costituendo il tipo di input aurale più diffuso. Anche in questo caso pare impressionisticamente a chi scrive che la situazione che si rileva per i cartoni non coincida con quella che si dà nei contesti di interazione e apprendimento reali, dove a dominare sono invece i rumori, che accompagnano ogni momento dell'esistenza. Di fatti, i rumori – che nei cartoni sono tipicamente ridotti a suoni prototipici non mimetici del reale né nella forma né nella durata ma univocamente indicizzanti un evento (movimento, sentimento, ecc.) – sono presenti in ogni puntata considerata seppur risultano essere meno frequenti di musica e parlato. Il dato è inatteso perché, come evidenziato da Kock & Louven (2018), gli spettatori dei cartoni abbisognano continuamente di informazioni sonore che accompagnino quelle visive dato che la natura illustrativa non fotografica di questo genere audiovisivo impedisce la presentazione agli spettatori di informazioni sonore registrate in sede di ripresa che forniscano informazioni ambientali di contorno. In tal senso è possibile – e andrà quindi verificato sperimentalmente – che l'assenza o la presenza¹¹ di effetti possano determinare differenti esiti di coinvolgimento degli spettatori e di interpretazione delle situazioni illustrate nel cartone, dunque eventualmente anche di processazione delle informazioni linguistiche co-occorrenti e, in ultima istanza, di acquisizione linguistica. Infine, con riferimento alla presenza dei diversi tipi di segni, emerge che il cantato non è un tratto caratterizzante dei cartoni, nemmeno nelle sigle, visto che per es. *Pocoyo* e *Mini cuccioli* non ne contengono.

Tabella 1 – *Durata totale in secondi delle annotazioni ai vari livelli*

	<i>Pimpa</i>	<i>Yo Yo</i>	<i>Pocoyo</i>	<i>Mini cuccioli</i>	<i>Blaze e le mega macchine</i>	<i>TopoTip</i>
totale	300	415	431	365	1304	434
parlato	144	494	168	259	706	286
cantato	15	77	0	0	58	29
musica	219	336	363	270	666	318
effetti	58	127	204	142	240	ND

Le trascrizioni disponibili permettono una seconda serie di osservazioni relative all'ambiente acustico¹², ovvero quelle sulle sovrapposizioni di segni linguistici e non linguistici diversi nel canale aurale (tab. 2). I dati rivelano che nelle puntate considerate il fenomeno, seppur presente, non è nel complesso frequente e diminuisce al crescere del numero di tipi di segni diversi sovrapposti. Più precisamente, la verifica empirica mostra che la saturazione del canale dovuta alla presenza simultanea di informazioni relative a quattro livelli – ovvero *parlato*, *cantato*, *musica* ed *effetti* – è

¹⁰ Ovviamente se si escludono gli approcci all'insegnamento della lingua basati sul ricorso a musica e canzoni (Kušnierek 2016).

¹¹ E, nel caso di presenza, la loro distribuzione temporale e l'associazione alle immagini.

¹² Incomplete alla data di chiusura di questo contributo e quindi non riportate per *Topo Tip*.

di fatto assente, se si esclude un'istanza della durata inferiore a due secondi in una puntata del cartone *Yo Yo*. Un poco più diffuse sono invece le sovrapposizioni di tre livelli – ovvero *parlato*, *musica* ed *effetti* – dato che le si rintracciano in quattro serie di cartoni, sebbene sempre con un'incidenza percentuale limitata sul totale della puntata, visto che nel caso più diffuso il valore massimo è 10,6%, pari a un valore assoluto di circa 46 secondi. Più diffuse ancora (ma comunque marginali) sono le sovrapposizioni tra *parlato* ed *effetti* che raggiungono al più 12,8% in *Pocoyo*. Tendenzialmente più frequenti sono infine le sovrapposizioni nello stesso canale aurale tra *musica* e uno tra *parlato* o *effetti*.

Tabella 2 – *Sovrapposizioni tra i livelli (% sul totale della durata)*.

	<i>Pimpa</i>	<i>Yo Yo</i>	<i>Pocoyo</i>	<i>Mini cuccioli</i>	<i>Blaze e le mega macchine</i>
PCME	0	0,4	0	0	0
PME	1,2	3,1	10,6	5,4	0
PM	15,2	23,2	22,2	12,9	5,1
PE	3,2	6,1	12,8	7	0
ME	8,9	16,8	38,9	23,0	0,2

Legenda: P = parlato; C = cantato; M = musica; E = effetti

I dati appena presentati sono rappresentativi di due delle molte possibili alternative per definire l'ambiente acustico dei cartoni nel corpus COCAINAA.

3.2. Paesaggio sonoro

Nell'ottica della linguistica applicata l'ambiente acustico è rilevante perché costituisce il punto di partenza del *paesaggio sonoro* ovvero, secondo l'Organizzazione Internazionale per la Normazione (ISO 12913-1:2014) che cerca standardizzarne la definizione muovendo dalla proposta fatta da Schafer (1977) nell'ambito degli studi sulla cultura sonora, quella che sia la percezione che una o più persone ne hanno in specifiche condizioni di ascolto. Il paesaggio sonoro, a sua volta, merita di essere indagato dalla linguistica applicata perché l'organizzazione dei diversi codici accessibili a livello aurale concorre a definire e a fare dei cartoni nel loro complesso (nonché delle singole serie e puntate che li costituiscono) dei contesti da cui recuperare, dopo opportuna selezione e correlazione delle parti, quello che è l'input sonoro linguisticamente rilevante, ovvero quelle che sono le componenti di significato fonico-acustico dei segni linguistici.

Con riferimento ai cartoni questa operazione di estrazione dell'input linguistico non è banale per i fruitori – tanto più per quelli con una competenza linguistica in sviluppo – perché, paradossalmente, marcatamente monomodale se messa in relazione alla definizione tradizionale di multimodalità.

In primo luogo perché, visto che nei cartoni si danno illustrazioni grafiche più o meno mimetiche del reale e più o meno decorative ma mai fotografiche, solitamen-

te¹³ non si dà presenza o perlomeno sincronizzazione veritiera tra le componenti acustiche e le componenti articolatorie facciali del parlante percepibili dall'ascoltatore. In conseguenza di ciò, l'estrazione del significante fonico-acustico grava esclusivamente sulle capacità e sull'apparato uditivo-percettivo del ricevente. È importante esplicitarlo con riferimento alla più generale riflessione che si possa fare circa la spendibilità dei cartoni per la didattica della lingua visto che numerose ricerche confermano che l'accesso visivo al volto del parlante accresce significativamente la percezione del parlato (Vatikiotis-Bateson & Munhall 2015) e che la mancata sincronizzazione o il disaccoppiamento tra l'informazione visiva circa il volto del parlante e il suo parlato danno luogo a percezioni erronee o falsate del detto (McGurk & MacDonald 1976). In tal senso, il cartone rappresenta un ambiente di ascolto specifico che presenta all'ascoltatore un input meno multimodale – dunque più difficile – rispetto sia alla conversazione faccia a faccia, che alle esperienze di fruizione di audiovisivi fotografici.

In secondo luogo, l'operazione di estrazione di input linguistico fonico dai cartoni non è banale per lo spettatore perché vi si danno sovrapposizioni di informazioni aurali che possono essere considerate disturbanti a livello uditivo e percettivo. A livello uditivo perché gli effetti sonori aggiunti per fornire verosimiglianza acustica alla situazione simulata o per orientare e stimolare lo spettatore possono mascherare i suoni linguistici. A livello percettivo perché la presenza di effetti rallenta sempre la processazione delle informazioni linguistiche, come messo in evidenza da ricerche sugli *irrelevant sound effects* che, a seconda del loro essere presenti o assenti e, se presenti, a seconda del loro essere sovrapposti o meno al parlato, possono determinare esiti di coinvolgimento differenti rispetto alla visione del cartone, dunque idealmente anche di processazione delle informazioni linguistiche lì contenute, per es. in termini di segmentazione (Alqattan & Turner 2021).

Da ultimo, gli ostacoli a livello percettivo possono divenire ancora più marcati quando gli *effetti sonori* si sovrappongano o a *parlato* e *musiche*, o a *cantato*, di per loro già causa di *mondegreens* (Fraser 2014).

4. Conclusioni

Pare a chi scrive che un'indagine linguistica multimodale dei cartoni del tipo abbozzato nei paragrafi precedenti possa rivelarsi utile, anzitutto per la linguistica applicata e con riferimento alla descrizione e migliore comprensione del ruolo di questo genere audiovisivo per l'acquisizione, l'apprendimento e in ultima istanza anche l'insegnamento delle lingue, soprattutto di individui in età prescolare e scolare.

Infatti, in letteratura è ben documentato quanto i primi anni di vita di un bambino siano importanti per il processo di acquisizione linguistica tutto, in primis con riferimento all'input ricevuto (Friedmann & Rusou 2015). Tuttavia, se fino ad alcuni decenni fa que-

¹³ Costituiscono parziale eccezione i casi di cartoni animati che contengono forme di sincronizzazione labiale.

sto input era sostanzialmente limitato a quanto il bambino poteva ricavare dall'ambiente sociale circostante in funzione della sua interazione su base linguistica con le figure deputate alle cure primarie e secondarie a lui dirette, oggi giorno la situazione è significativamente cambiata perché, in conseguenza della maggiore accessibilità e diffusione di alcuni mezzi di comunicazione di massa cui anche infanti e bambini sono esposti precocemente e massicciamente, l'input comprende ormai anche quanto ricavabile dalla visione di programmi televisivi.

A tal proposito, una recente ricerca ISTAT (Bologna & Savioli 2022) riporta che i bambini residenti in Italia guardano la televisione e la cosiddetta televisione fuori dalla televisione – cioè i programmi televisivi visionati ricorrendo a dispositivi diversi dal televisore come per es. smartphone, tablet, computer – con alta frequenza. Più precisamente, la ricerca riferisce che – anche in conseguenza di alcuni cambiamenti sociali affermatasi durante il periodo delle restrizioni emanate per scongiurare la diffusione della malattia da virus SARS-CoV-2 – a tutto il 2022 il 94% dei bambini di età compresa tra i 3 e i 5 anni guardava la televisione quotidianamente. Peraltro, poiché da una ricerca di Ciccotti e Sabbadini (2007) risulta che il 98% dei maschi e il 77% di femmine di età compresa tra 3 e 5 anni che guardano la televisione fruiscono di cartoni, non pare azzardato affermare che la maggioranza dei bambini in età prescolare sia esposta quasi quotidianamente a cartoni.

Indipendentemente da ogni considerazione sociologica, psicologica, pedagogica, morale o moraleggiante, chi si occupa di linguistica applicata non può ignorare questi dati perché, se è vero o se si condivide quel che sostengono per es. Piske e Young-Scholten (2008: 1) secondo cui “as long there’s input, acquisition will occur”, allora si deve ritenere che anche la sola visione di cartoni possa in qualche misura condizionare il processo di acquisizione¹⁴ linguistica.

In realtà, l'effettivo ruolo della televisione sull'acquisizione linguistica è da tempo oggetto di un articolato dibattito che fa registrare posizioni anche molto distanti tra loro. Per es., da un lato si danno autori come Kuhl *et al.* (2003) che ritengono che questo tipo di input, non essendo negoziato e non essendo accompagnato da alcuna forma di interazione sociale tra l'apprendente umano e l'audiovisivo, non abbia effetti rilevanti sull'acquisizione. D'altro canto, autori come Rice (1983) contestano “the conventional view among developmental psychologists [...] that television viewing does not contribute to a young viewer’s language acquisition” e sostengono che “children draw upon television as a source of verbal routines for their own play interactions”.

Attualmente la posizione di Rice (1983) – e più in generale quella per cui la televisione condiziona l'acquisizione – è la più diffusa e viene ritenuta valida con riferimento tanto alle prime lingue, quanto alle lingue dopo le prime; tanto ad apprendenti precoci, quanto adulti; tanto ad apprendenti iniziali, quanto intermedi. Nel quadro di questa visione, l'influenza positiva dei cartoni viene solitamente spiegata richiamandosi anzitutto all'alto grado di motivazione e gratificazione esperita da chi i cartoni guardi che, per tutta risposta, determinerebbe un abbassamento del filtro affettivo nonché di tutte le altre barriere che possano frenare l'acquisizione, in un'ottica quasi suggestopedica (Poštič 2015).

¹⁴ Da qui in avanti si utilizzerà solo il termine *acquisizione* considerandolo iperonimo di tutte le possibili esperienze che determinano un accrescimento della competenza linguistica e comunicativa.

In secondo luogo, il ruolo positivo dei cartoni viene motivato con riferimento al fatto che i cartoni rendono disponibile un input non adattato, tendenzialmente naturale, caratterizzato da frequenti discorsi contestualizzati e dalla frequente ripetizione di strutture testuali, discorsive, sintattiche e lessicali, incluse quelle a bassa frequenza che diventerebbero comprensibili grazie al supporto semiotico delle immagini che consentono di identificare e risolvere eventuali difficoltà di assegnazione di eventi e referenti (Rodgers 2013), ovvero richiamandosi implicitamente a teorie dell'acquisizione linguistica multisensoriale o multimodale (Dressmann 2020).

In tal senso, quanto riportato in §3 documenta che il corpus multimodale quale COCAINAA permette di identificare passaggi che presentino ai fruitori potenziali sfide uditive e percettive che potrebbero condizionare la comprensione dell'input e quindi l'acquisizione linguistica tutta, tanto più che “selective attention to speech crucially determines the course of spoken language development. From birth on, typically-developing humans pay more attention to speech than to other auditory input, and selectively attend to relevant linguistic information useful for language development” (Saksida *et al.* 2022).

In tal senso, la pur ridotta base di dati attualmente disponibile in COCAINAA consente già di osservare che l'ambiente acustico – e dunque anche il paesaggio sonoro – cambiano di cartone in cartone rendendo probabilmente disponibili a chi guardi i cartoni animati serie ed episodi con proprietà e complessità variabili rispetto allo specifico parametro, sebbene tra le serie si diano alcuni tratti condivisi come la frequente presenza di musica e la non trascurabile sovrapposizione di musica e parlato che caratterizzano il genere.

Oltre a ciò, la costruzione di un corpus multimodale di cartoni animati pare permetterebbe di elaborare osservazioni relativamente anche ad altri piani tra loro tanto diversi come, per es., quello del rapporto tra esposizione ad immagini animate e acquisizione incidentale del lessico; delle sequenze di pedagogia naturale inscenata in cartoni animati educativi (cfr. *Blaze e le mega macchine*); della definizione di paesaggi linguistici inusuali, improbabili o irreali e della loro influenza sugli apprendenti. Pertanto, si ritiene utile procedere nella redazione del corpus arricchendolo delle registrazioni, trascrizioni e annotazioni ancora mancanti.

Ringraziamenti

L'autore ringrazia Maria Elena Favilla e Matteo Largaiolli per la loro disponibilità a discutere di molti degli aspetti riassunti in questo contributo. L'autore ringrazia inoltre Aurora Cantoni, Debora Arnoldi (cfr. Arnoldi 2023), e Iris Dall'Aglio per il loro aiuto con le trascrizioni. L'autore ringrazia infine i due revisori anonimi e le curatrici per i loro utili commenti. Ricerca finanziata dall'Unione europea - Next Generation EU nell'ambito del progetto PRIN 2022 2022NPXYHH CIP - *Corpus of Italian language for Preschoolers*.

Bibliografia

- Alqattan, Danah & Turner, Paul. 2021. The effect of background noise on speech perception in monolingual and bilingual adults with normal hearing. *Noise Health* 23. 67-74.
- Ansani, Alessandro & Marini, Marco & Poggi, Isabella. 2020. How soundtracks shape what we see. *Frontiers in Psychology* 11. 2242.
- Arnoldi, Debora. 2023. *Il lessico dei cartoni animati*. Tesi di laurea non pubblicata. Libera Università di Bolzano.
- Ball, Martin & Esling, John & Dickson, Craig. 2018. Revisions to the VoQS system for the transcription of voice quality. *Journal of the International Phonetic Association* 48. 165-171.
- Bologna, Emanuela & Savioli, Miria. 2022. *Tempo libero e partecipazione culturale: tra vecchie e nuove pratiche*. Roma: Istat.
- Burn, Andrew. 2013. *The kineikonic mode: Towards a multimodal approach to moving image media*. <https://eprints.ncrm.ac.uk/id/eprint/3085/>
- Brown, Lex & Gjestland, Truls & Dubois, Danièle. 2015. Acoustic environments and soundscapes. In Kang, Jian & Schulte-Fortkamp, Brigitte (eds.), *Soundscape and the built environment*, 16-25. Boca Raton: CRC Press.
- Bruner, Jerome. 1966. *Toward a theory of instruction*. Harvard: Harvard University Press.
- Cappuccio, Giuseppa. 2015 *Cartoons di qualità nella prima infanzia*. Roma: Aracne.
- Chaume, Frederic. 2004. Film studies and translation studies. *Meta* 49. 12-24.
- ChatGPT. 2023. *Definizione di cartone animato*. [Response to user question]. <https://chat.openai.com>.
- Ciccotti, Ermenegildo & Sabbadini, Linda. 2007. *Come cambia la vita dei bambini*. Firenze: Istituto degli Innocenti.
- Dressmann, Mark. 2020. Multimodality and language learning. In Dressman, Mark & Sadler, William (eds.), *The handbook of informal language learning*, 39-55. Hoboken: Wiley Blackwell.
- ELAN (Version 6.6). 2023. Nijmegen: Max Planck Institute for Psycholinguistics. <https://archive.mpi.nl/tla/elan>.
- Fraser, Helen. 2014. When teaching phonology isn't enough. *Speak Out!* 50. 29-33.
- Friedmann, Naama & Rusou, Dana. 2015. Critical period for first language: the crucial role of language input during the first year of life. *Current Opinion in Neurobiology* 35. 27-34.
- Hymers, Mark & Prendergast, Garreth & Liu, Can & Schulze, Anja & Young, Michellie & Wastling, Stephen & Barker, Gareth & Millman, Rebecca. 2015. Neural mechanisms underlying song and speech perception can be differentiated using an illusory percept. *NeuroImage* 108. 225-233.
- Knight, Dawn & Adolphs, Svenja. 2022. Building a spoken corpus: what are the basics? In O'Keeffe, Anne & McCarthy, Michael (eds.), *The Routledge handbook of corpus linguistics. Second edition*, 21-34. Abingdon: Routledge.
- Kock, Maximilian & Louven, Christoph. 2018. The power of sound design in a moving picture. *Empirical Musicology Review* 13. 132-148.
- Kuhl, Patricia & Tsao, Feng-Ming & Liu, Hwei-Mei. 2003. *Foreign-language experience in infancy*. Proceedings of the National Academy of Sciences 100. 9096-9101.

- Kuśnierek, Anna. 2016. The role of music and songs in teaching English vocabulary to students. *World Scientific News* 43. 1-55.
- Leeuwen, Theo. 1999. *Speech, music, sound*. London: Red Globe Press.
- MacDonald, John. 1976. Hearing lips and seeing voices. *Nature* 264. 746-748.
- Mondada, Lorenza. 2018. Multiple temporalities of language and body in interaction. *Research on Language and Social Interaction* 51. 85-106.
- Nespor, Marina. 1993. *Fonologia*. Bologna: il Mulino.
- Piske, Thorsten & Young-Scholten, Martha. 2008. *Input matters in SLA*. Bristol: Multilingual Matters.
- Poštič, Svetozar. 2015. Influence of cartoon network on the acquisition of American English during childhood. *Verbum* 60. 188-195.
- Pumares, Marta Jiménez & Fontaine, Gilles. 2021. *Animation films and TV series in Europe. Key figures*. European Audiovisual Observatory. <https://rm.coe.int/animation-films-and-tv-series-in-europe-key-figures/1680a4a717>.
- Rice, Mabel. 1983. The role of television in language acquisition. *Developmental Review* 3. 211-224.
- Rodgers, Michael. 2013. *English language learning through viewing television*. Unpublished doctoral dissertation. Wellington: Victoria University.
- Rondolino, Gianni. 2003. *Storia del cinema di animazione*. Torino: Utet libreria.
- Saksida, Amanda & Ghiselli, Sara & Picinali, Lorenzo & Pintonello, Sara & Battelino, Saba & Orzan, Eva. 2022. Attention to speech and music in young children with bilateral cochlear implants. *Journal of Clinical Medicine* 11. 1745.
- Sardo, Rosaria. 2018. C'era una volta la tv per ragazzi... *Lingue e culture dei media* 1. 1-57.
- Sardo, Rosaria & Ferro, Patrizia. 2008. La TV per bambini e per ragazzi. In Alfieri, Gabriella & Bonomi, Ilaria (a cura di), *Gli italiani del piccolo schermo*, 379-432. Roma: Cesati.
- Schafer, Murray. 1977. *The soundscape*. New York: Knopf.
- Tagliapietra, Andrea. 2010. *Filosofia dei cartoni animati*. Torino: Bollati Boringhieri.
- Thompson, Paul. 2022. Building a specialised audiovisual corpus. In O'Keeffe, Anne & McCarthy, Michael (eds.), *The Routledge handbook of corpus linguistics. Second edition*, 75-88. Abingdon, Routledge.
- [TRECCANI]. 2018. *Vocabolario Treccani online*. Istituto della Enciclopedia Italiana. <https://www.treccani.it/vocabolario>
- Vatikiotis-Bateson, Eric & Munhall, Kevin. 2015 Auditory-visual speech processing. In Redford, Melissa (ed.), *The handbook of speech production*, 178-199. Chichester: John Wiley & Sons.
- Wikipedia contributors, 2023, *Kineikonic mode*. https://en.wikipedia.org/wiki/Kineikonic_Mode.
- Wilkinson, Mark & Dumontier, Michel & Aalbersberg, IJsbrand Jan *et al.*. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, 160018. <https://www.nature.com/articles/sdata201618#citeas>.