

Think multimodally! The GFFIntD corpus and visibility on screen: presenting a multimodal English-Italian film festival interpreting corpus

Laura Picchio



Laura Picchio
University of Macerata;
lpicchio1@unimc.it;
ORCID: [0000-0003-3596-243X](https://orcid.org/0000-0003-3596-243X)



Abstract

This paper presents GFFIntD, a multimodal corpus of interpreting practices performed at the Giffoni Film Festival, with a focus on data collection, transcription and analysis. In particular, the paper draws upon the concept of visibility on screen and discusses the two-step analytical process applied to certain excerpts in which multimodal elements co-construct meaning during guest-initiated vignettes.

Keywords: corpus-based interpreting studies; ELAN; embodiment; film festival interpreting; GFFIntD; multimodality; visibility on screen.

Resumen

Este artículo presenta el corpus multimodal GFFIntD en términos de recopilación, transcripción y análisis de las prácticas de interpretación realizadas en el Giffoni Film Festival. Concretamente, se basa en el concepto de visibilidad en pantalla y discute el proceso analítico en dos etapas de aquellos extractos en los que los elementos multimodales co-construyen el significado durante las escenas de interacción iniciadas por los asistentes.

Palabras clave: estudios de interpretación basados en corpus; ELAN; corporalidad; interpretación en festivales de cine; GFFIntD; multimodalidad; visibilidad en pantalla .

Resum

Aquest article presenta el corpus multimodal GFFIntD en termes de recopilació, transcripció i anàlisi de les pràctiques d'interpretació portades a terme al Giffoni Film Festival. Concretament, es basa en el concepte de visibilitat en pantalla i discuteix el procés analític en dues etapes d'aquells extractes en què els elements multimodals co-construeixen el significat durant les escenes d'interacció iniciades pels assistents.

Paraules clau: estudis d'interpretació basats en corpus; ELAN; corporalitat; interpretació en festivals de cine; GFFIntD; multimodalitat; visibilitat en pantalla .

1. Introduction

Over the last 20 years, scholars of monolingual communication have increasingly focused their attention on the notion of embodiment, sharing the assumption that all human communication is embodied (see, for instance, Hazel et al., 2014). So as to effectively investigate whether and how verbal and embodied resources interact to create meaning (Jewitt, 2014), such studies have drawn on the concept of multimodality.

Traditionally, scholars of interpreting studies have predominantly focused on the verbal components of oral texts. The situation appears rather different in dialogue interpreting (for a definition, see Merlini, 2020), “where calls for a multimodal approach are very recent, but innovative studies building on and developing new theoretical and methodological perspectives [...] have already been published” (Pöchhacker, 2020: 23). Dealing with such aspects as gestures, proxemics, gaze, body orientation, and the handling of artefacts, dialogue interpreting scholars have highlighted how meaning is co-constructed not only at a verbal but also at a non-verbal level, stressing the need for an embodied (Nevile, 2015), multimodal turn (Mondada, 2016: 340; Davitti, 2019: 12). Some such studies (section 2.2) have also drawn on a corpus-based approach that “could greatly contribute to describing and analysing the configuration of meanings in [dialogue interpreting]” (Gao and Wang, 2017: 19).

Against the backdrop of the relevant literature in both multimodal monolingual conversation analysis and corpus-based dialogue interpreting studies, this paper presents the multimodal corpus GFFIntD (which stands for Giffoni Dialogue Interpreting). Its data were transcribed by means of the software package ELAN (see, for instance, Wittenburg et al., 2006) and investigated in the author’s PhD dissertation (Picchio, 2023a). Covering a timespan from 2017 to 2020, this corpus includes video clips of English-Italian dialogue interpreting broadcast in live streaming on YouTube during the Giffoni Film Festival (henceforth GFF); to a lesser extent, it also contains audio-only encounters that were not live streamed, even though they share some basic characteristics with the former category of events (see section 4.1).

Acknowledging that building and analysing a multimodal corpus is not an easy task (Gao and Wang, 2017: 20), the objective of this paper is to present GFFIntD in terms of data collection, transcription and analysis, focusing in particular on the two-step transcription and analysis process applied to certain eloquent excerpts in which verbal and embodied elements are seen to co-construct meaning during hilarious, guest-initiated vignettes. In the case of the latter, the discussion revolves around the concept of “visibility on screen” put forward in Picchio (2023a): GFFIntD investigates whether, when, and how participants (interpreters included) are visible in the shot for the benefit of streaming viewers, with either positive or negative consequences as far as interaction and entertainment are concerned. In this regard, visibility on screen draws upon Katan and Straniero Sergio’s (2001) concept of “ethics of entertainment”, which refers to the playful macro-function of media communication whose primary aim is to entertain the audience and keep their interest level high, through means including ad-lib embodied sketches (see section 3). It is worth highlighting that streaming viewers’ access to GFF

events is mediated by video cameras, which can either reinforce or hamper the reception of entertaining speech-body interaction, depending on what/who is visible on screen — and, therefore, available for analysis. This is in line with the “video turn” Pöchhacker put forward in 2020: here, the term “video” is intended to mean a medium for capturing and transmitting the interpreter’s input/output in technology-mediated online scenarios (for the benefit of the remote audience), as well as for capturing and recording audio-visual data on interpreting as an object of study (*Ibid.*: 35).

2. Multimodality in communication and corpus-based interpreting studies

2.1 Defining multimodality

Multimodality “is an inter-disciplinary approach drawn from social semiotics that understands communication and representation as more than language and attends systematically to the social interpretation of a range of forms of making meaning” (Jewitt, 2013: 250).

Brought to the fore at the beginning of the 21st century (see, for instance, Kress and van Leeuwen, 2001; Jewitt and Kress, 2003), this approach has drawn from different theoretical and methodological perspectives to analyse how verbal and embodied resources interact to create meaning in the here and now (see, for instance, Kress, 2010; Jewitt, 2014). Multimodality thus emphasises situated interaction, its specific sociocultural context, and the range of resources speakers select to communicate: “With this emphasis, a key question for multimodality is how people make meaning in context to achieve specific aims” (Jewitt, 2013: 252).

Unlike writing, which is mediated by a “concrete” medium (paper or a display screen, for instance), face-to-face communication between two or more human beings could easily be conceived of as unmediated, since the medium through which it is performed is “invisible” (Pöchhacker, 2020: 16). Nonetheless, “a system of human organs and bodily functions (from vocal folds to respiratory organs) is used to produce sounds” (*Ibid.*: 17). Therefore, speech production is a mediated, ergo embodied phenomenon through which verbal and paraverbal elements are produced by the speaker’s vocal-respiratory system and perceived by the listener’s auditory apparatus in a bimodal oral-aural modality. It is only when the visual channel comes into play that speech becomes multimodal, as the speaker produces a whole array of visible actions, such as facial expressions, head movements, eye gaze, hand gestures, body posture, etc.: the combination of (para)verbal elements and visible embodied signals makes speech unquestionably multimodal (*Ibid.*: 17–18). In other words, in the visuo-gestural modality, interaction is inherently multimodal given the abundance of embodied resources that can also be used in combination to communicate something — even in the absence of speech. As a result, when interaction takes place in the auditory (spoken language) and visual (face-to-face) channels simultaneously, “communication achieves most of its multimodal potential” (*Ibid.*).

Multimodality challenges the prior dominance of speech (Scollon and Scollon, 2009), which is conceived of as just one of the many components of the so-called “multimodal ensemble” (Jewitt, 2013). Without denying that spoken language often plays a key role in interaction (Norris, 2004), all resources other than speech also have the potential to contribute equally to meaning.

Traditionally, monolingual conversation analysis scholars have focused on how co-participants organise their social encounters in and through vocal conduct and turns-at-speaking (Hazel et al., 2014: 1-2). However, early research into human interactive practices also encompassed studies on how gestures (see, for instance, Goodwin, 1986; Schegloff, 1984) are organised with or in relation to vocal conduct. Following in their footsteps, other scholars applied speech-oriented conversation analysis concepts to non-vocal modes to systematically analyse how embodied semiotic resources (see Goodwin, 2000) are recognisable social practices that co-construct ongoing interaction. The interrelation between different media is a key analytical point: multiple semiotic resources are simultaneously used by co-participants, to the extent that different media (speech and gestures, for instance) mutually elaborate each other within a situated embodied participation framework (see Goodwin, 2000); Mondada (2014, 139) describes this holistically integrated ensemble of resources in terms of a “complex multimodal Gestalt”.

One of the challenges posed by this multimodal approach is linked to the transcription stage (see section 4.2), a necessary step in analysing oral interaction. This process is influenced by research purposes, the technologies available to scholars and the concrete needs they have (e.g. publishing a paper). This opens up several questions about “how to represent in adequate and relevant ways these Gestalts, how to develop conventions and tools able to capture both holistically and analytically these complexities, including their fine-grained temporality, and how to avoid the tacit reproduction of language-centred conventions and conceptions” (Mondada, 2014: 154). To this end, Mondada (2006, 2018), among others, developed a set of *ad hoc* multimodal conventions, while also suggesting structuring different versions of the transcript in a “reflective”, recursive way, from a basic transcript to a fine-grained final version with *ad hoc* conventions and screenshots.

2.2 The multimodal turn in corpus-based dialogue interpreting studies

In face-to-face interaction, almost 65% of communicative understanding comes through embodied resources (Straniero Sergio, 2007: 359). Dialogue interpreting that takes place both in the oral-aural and in the visual-gestural modalities should thus be analysed through the lens of a multimodal approach.

Leaving aside comprehensive overviews of both corpus-based interpreting studies (see Bendazzoli, 2017) and multimodal approaches to dialogue interpreting (see Davitti, 2019), this section focuses specifically on three eloquent research projects that have drawn on a corpus-based multimodal methodology to investigate video-recorded interpreter-mediated encounters by means of ELAN.

Davitti (2013) discusses the role of gaze and body orientations in triggering, eliciting and elucidating conversational moves in English-Italian parent-teacher encounters. She observes that the interpreter used gaze to actively promote alignments between the main parties and deliberately sustained mutual gaze with the marginal participant (i.e. the mother) in order to elicit response or encourage her engagement in the interaction. Despite these efforts, Davitti notes that the integration of both interlocutors' engagement was not always successful, suggesting that the asymmetrical institutional relations between them might be the factor that limited the effect of the interpreter's active attempts. Methodologically, Davitti adopts a multimodal conversation analytical approach; more specifically, her linguistic transcription conventions draw on Jefferson (2004), while taking inspiration from Rossano (2012) to design her *ad hoc* multimodal set, which is mapped onto transcripts of interaction by means of ELAN and complemented by relevant screenshots.

Bao-Rozée's PhD dissertation (2016) focuses on multimodal analysis to investigate the role of interpreters in six simulated (non-)institutional encounters between Chinese and English native speakers. The roles played by the primary participants were deliberately matched with their real professional or social roles. Interactions involved "parent-teacher, businessman-businessman, doctor-patient, neighbour-neighbour, interviewer-interviewee and traveller-traveller" (*Ibid.*: 7) meetings. Bao-Rozée shows that interpreters coordinate interaction through gaze and body orientation, while also establishing a relationship with the co-participants and actively using gestures to facilitate the cross-cultural communication process. Methodologically, she uses both conversation and multimodal analysis: drawing on Jefferson (2004) and McNeil (2006), her transcription scheme is "rather flexible and diverse, aiming to use the best possible presentation of transcripts to support the focus of the analysis" (Bao-Rozée, 2016: 95). In particular, the ELAN vertical layout provides a basis for the analysis of gaze and body orientation, aligning non-verbal and verbal annotations.

In their exploratory paper, Gao and Wang (2017) present a corpus-based multi-layer analytic framework that is particularly pertinent to distant languages (like Chinese and English), "where non-verbal meaning making semiotics, like gaze and gestures, could compensate what is missing or shifted verbally" (*Ibid.*: 29). This framework includes the written transcript (lexicogrammar/semantics), auditory properties (phonology/phonetics), visual semiotics (gaze, gesture, body orientation, proxemics, handling of artefacts), and context (situation and culture). In their proposal, the scholars draw both on linguistic theory of stratification from systemic functional linguistics and on previous multimodal approaches to dialogue interpreting studies. The complementarity between these approaches is vital not only at a theoretical but also at an empirical level: in the case of the latter, machine-readable formal data (e.g. at the lexico-grammatical strata) are combined with ELAN-based "corpus techniques to seek more meaningful multimodal data, which provide interpretation, explanation or triangulated results to demystify what descriptive interpreting studies alone are unable to explain" (*Ibid.*: 33). Indeed, Gao and Wang argue that the multimodal data they analyse through this corpus-based framework

“could enable [...] researchers to explain shifts or non-expectancies in one layer of semiotic configuration with the answers triangulated from other layers” (*Ibid.*: 29).

3. Film festival interpreting at the Giffoni Film Festival: setting the scene

Film festival interpreting refers to an under-investigated setting of professional dialogue interpreting practice involving the use of the short consecutive mode at international film festivals (Merlini, 2017: 137–138). It typically deals with on-stage face-to-face speech, such as Q&A sessions, the awarding of prizes, and interviews. As film festivals habitually attract media coverage by radio, television and web networks, a large part of the interpreting activity performed there may arguably be subsumed under the wider category of “media interpreting” (see Dal Fovo, 2020) on at least three counts: the diversity of addressees and communication levels stemming from having an additional, remote audience; the resulting high degree of visibility and exposure experienced by interpreters; and the reference to the “ethics of entertainment” (Katan and Straniero Sergio, 2001), making interpreting behaviour consistent with the playful macro-function of TV communication. According to Katan and Straniero Sergio (2001), media talk is specifically designed to maintain the interest of the remote audience, while ensuring that all participants — both on- and off-screen — feel comfortable with what they are doing. In compliance with this overarching “comfort factor”, interpreters are called upon to contribute to the entertainment goal of the media event, through means including ad-lib embodied vignettes. More specifically, film festival interpreting can be defined as “interpreting *also for* the media” (Merlini, 2017: 139, emphasis in the original), since film festivals do attract media coverage even if they are neither an activity organised by the media themselves and performed within a studio nor a situationally displaced translation of events taking place in a distant location, staged and edited for the benefit of the remote audience.

The reference to media interpreting can also be applied to the interpreting practices performed at the GFF because, as already said, some of the festival’s events are broadcast in live streaming on its official YouTube channel, so interpreters have to cater for the communicative needs of two audiences: the flesh-and-blood jurors sitting in the cinema and the remote audience following the live streaming. The GFF is one of Italy’s most important international film festivals. It has been taking place since 1971, with the unique feature of addressing an audience of children and young people. Juries known as +13, +16 and +18 (in reference to age groups) are composed of both Italian and foreign people, which is the reason why their meetings with foreign guests are interpreted between Italian and English: these are the GFF official languages, with English being used as a *lingua franca* by foreign participants. Meetings with international guests include post-screening Q&A sessions with members of film crews and encounters with stars invited to Giffoni to receive special awards or present their film *premières*.

Given its characteristics, the GFF was chosen as a case study for the PhD dissertation the present paper is based on (Picchio, 2023a; see also Picchio, 2023b). Through GFFIntD,

the mixed-methods PhD dissertation in question aims to analyse several aspects of the interpreting practices performed at the GFF, including visibility on screen, i.e. whether, when, and how interpreters are visible in the shot for the benefit of streaming viewers, with either positive or negative consequences as far as interaction and entertainment are concerned. Leaving aside a full discussion of this topic (see Picchio, 2023a), the excerpts discussed below present the two-step transcription and analysis process opted for, thus offering an illustration of the potential of the corpus.

It is worth highlighting that in the field of media interpreting, scholars (Straniero Sergio, 2007: 361-362; Wadensjö, 2008: 192-193; Sandrelli, 2015: 94-96) have already pointed to the importance of what/who is visible on screen for a successful spectacle-oriented performance. Nonetheless, those few cursory observations have not drawn on a systematic multimodal approach to data.

4. GFFIntD: from data collection to analysis

To the author's knowledge, GFFIntD is one of a kind, since no other scholar has systematically devoted attention to film festival interpreting. It is worth mentioning that CorIT (Falbo, 2012), which focuses on TV interpreting, shares GFFIntD's overall reference to the broader category of media interpreting.

4.1 Data collection

GFFIntD contains 23 items, giving a total length of over 920 minutes (see Table 1), and covers a timespan from 2017 to 2020. The clips were either recorded onsite at the 47th GFF or retrieved from online public platforms (at the time of writing, they are still publicly available on YouTube).¹ In the case of the former, fieldnotes were also taken; as will soon become evident, they were extremely important for transcribing and analysing these clips. The majority of the data refer to streamed materials collected in the audio-visual format; a small part of the corpus is devoted to non-streamed audio-only clips. Since this paper revolves around visibility on screen, the excerpts discussed below are taken from the former category; nonetheless, the second category was included in GFFIntD because it allows for investigation of other concepts (i.e. audience design and interpreters' agency — for a discussion, see Picchio, 2023a, 2023b; see also Merlini and Picchio, 2019).

¹ The GFF YouTube channel is available at <https://www.youtube.com/@giffonifilmfestival> (Accessed: 20240912).

STREAMED (AUDIO-VISUAL FORMAT)	NON-STREAMED (AUDIO-ONLY FORMAT)
MJ_1_47	DPP_4_47
MJ_2_47	DPP_5_47
MJ_3_47	DPP_6_47
MJ_4_47	DPP_7_47
MJ_5_47	DPP_9_47
MJ_6_49	DPP_10_47
MS_1_47	
MS_2_47	
PRC_2_49	
PRC_3_49	
MJ_7_50	
MJ_8_50	
MJ_9_50	
MJ_10_50	
MJ_11_50	
DPP_11_50	
DPP_12_50	
TOT. 709.34 MIN.	tot. 215.12 min.
TOT. 924.46 MIN. = 15.4 H.	

Table 1. The structure of GFFIntD

Each clip was assigned a code consisting of:

- letters identifying the type of event: MJ (“meet the jury” — the jurors meet stars), MS (“meet the star” — a group of cinemagoers meets popular guests), PRC (“press conference” — journalists interact with celebrities), or DPP (“*dibattito post-proiezione*”, i.e. post-screening Q&A session — jurors talk with members of film crews);
- a progressive number;
- a number identifying the edition of the GFF.

Despite there being differences between them (streaming vs. non-streaming, audio-visual vs. audio-only format, type of event), all the items included in GFFIntD share the following characteristics:

- the same kinds of participants, i.e. an Italian host, an international guest, Italian and international jurors, and an interpreter;
- the dialogic format of interaction, i.e. a juror asks a question or makes a comment and the guest answers;
- dialogue interpreting in the short consecutive mode between Italian and English;

Therefore, right from the data collection stage, GFFIntD required a flexible approach to deal with these different — albeit very similar — items.

4.2 Data transcription

Despite being time-consuming, this activity is a necessary stage in dealing with oral data. Back in the early 2000s, Falbo (2005: 33) acknowledged the benefits of using a transcription software package:

Depuis quelques années, le développement de logiciels spécifiques [...] permet d'informatiser le son par la création de fichiers son et de procéder à la transcription en se servant d'un seul outil: l'ordinateur. Le texte transcrit est ainsi accompagné du texte oral.

As already mentioned, ELAN (2021) was used in the case of GFFIntD. It enabled the creation of a speech corpus (Straniero Sergio and Falbo, 2012) in which each transcript is aligned with its original clip, this being fundamental for the analysis of multimodal data — the purpose for which GFFIntD was created. Niemants (2012: 168) points out that “the alignment of audio[/video] and transcript uncovers the transformations, and

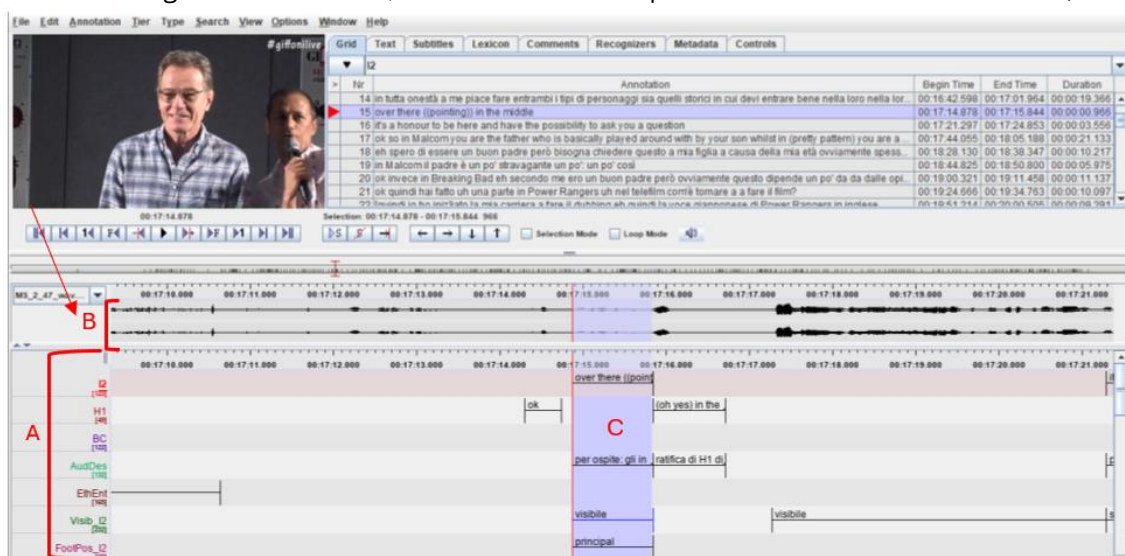


Figure 1. ELAN window in GFFIntD

the losses, perpetrated in the transcription process”, arguing that the audio/video of an event is the primary source of data, and the transcript a secondary source. Each clip is affected by the recording process, in which elements are selected from the observable

events. In GFFIntD, the fieldnotes taken onsite were of paramount importance, not only to disambiguate elements in the non-streamed audio clip, but also to reconstruct the whole audio-visual streamed event, which was affected by directorial choices to show a particular participant (usually the guest) on screen rather than another.

As a time-based tool, ELAN relates content to time and is suitable for the transcription and analysis of interactional data (see Niemants, 2012). Its intuitive vertical tiers (see letter “A” in Figure 1 above) resemble a musical score and enable scholars to clearly represent verbal and multimodal resources. In the same window, transcribers see their tiers and the corresponding audio-video segment (see letter “B”).

Before transcribing, scholars have to make several decisions that will affect their subsequent analyses (Niemants, 2018: 672). In GFFIntD, choices were made regarding tier types and attributes (i.e. where to transcribe); and units of analysis, conventions and the content of each annotation (i.e. what to transcribe).

Figure 2 shows GFFIntD’s tier types and attributes. Two types of tiers were created: the “speaker tiers” identify each speaker (e.g. in Figure 2, I4 identifies an interpreter), while the “analysis tiers” enable the researcher to carry out preliminary analyses while transcribing (e.g. in Figure 2, Visib_I4 refers to the on-screen visibility of the interpreter in question). In terms of attributes, each tier required a name (e.g. I4, Visib_I4) and was assigned to a participant (e.g. interpreter). All tiers in GFFIntD are independent, which means that the relationships between them were created exclusively at a temporal level by means of overlaps (see letter “C” in Figure 1 above). This made it possible to link together tiers related to different interlocutors: for instance, when a guest made a relevant gesture while the interpreter was talking, the guest’s analytical tier (Visib_G) was temporarily linked to the interpreter’s speaker tier (e.g. I4).

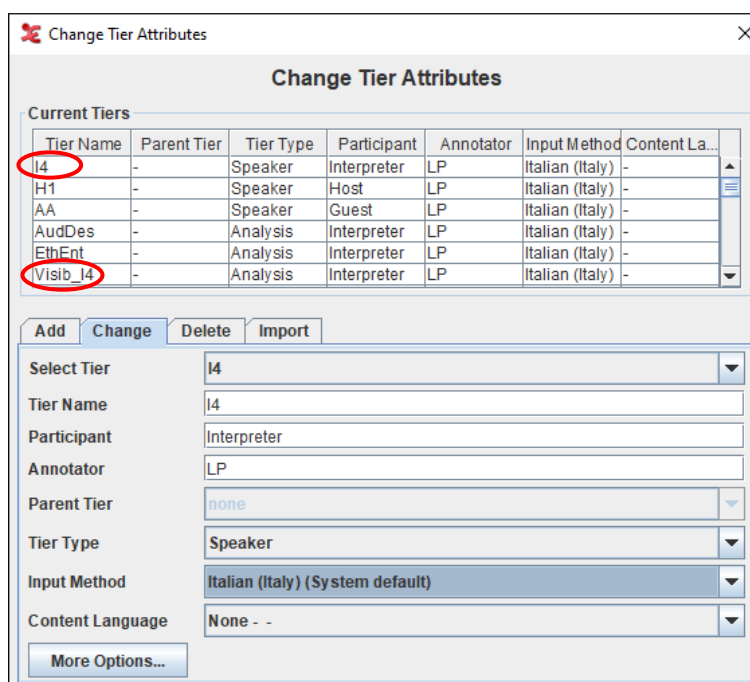


Figure 2. ELAN tier types and attributes in GFFIntD

As far as speaker tiers are concerned, the unit of analysis corresponds to a participant's turn. Canonically, in GFFIntD, the turn-taking routine is juror's question/comment (J) – interpretation (I) – guest's answer (G) – interpretation, i.e. J – I – G – I (excerpt 1 below); however, sometimes there are sequences such as J – G – I – J – G – I – J – G – I (excerpt 2 below). Excerpts 1 and 2 show that speaker tiers include the verbal and paraverbal transcription of each interactional turn using Jefferson-like conventions (Jefferson, 2004). Since language-specific features are not the object of the present analysis, sequences are shown in English, with italics signalling the original Italian utterances; idiomatic translations have been kept as close as possible to the syntactic structure of the original language to enable a sufficiently faithful reproduction of orality traits. The complete transcription key is provided in the Appendix.

Excerpt 1

- 1 J hi hi Amy I'm {name} from {country} and I heard that when you were a child you wanted to be a singer and we have also seen you singing in many movies and also in the video for the Golden Globes' opening and I was wondering if you would like to have the main role in a musical thank you
2 I *I know that when you were a child you longed to sing I'd like to know if you would like to star in a musical still today*
3 G absolutely I'm known to break into songs even on sets that aren't musicals so it would be great if uh: to get- to be in another musical uhm it's my first love doing musical theatres so I would always love returning to that
4 I *absolutely I'd like to star in a musical again it's my first love*

Excerpt 2

- 1 J I wanted to ask if you know any Italian films and if your answer is yes I want to know which one do you prefer and why
2 G my favourite film?
3 I Ita[lian film]
4 J [Italian] film
5 G Italian film
6 I (1) *I wanted to know if there are some Italian films that you like and if so which ones and why*
(3)
7 J or (.) also a director an Italian director
8 G °uh (.) uhm° (.) Matteo Garrone
9 I Matteo Garrone

As for the analysis tiers, their content is described through verbal glosses (for a comprehensive overview, see Picchio, 2023a). Focusing specifically on the multimodal tier — devoted, precisely, to visibility on screen (Visib) — semiotic embodied resources were transcribed in a two-step process. Rather than a full description of each resource, verbal glosses indicating the (in)visibility on screen of the main participants, including interpreters, were firstly included in ELAN Visib tiers. In Figures 3 and 4 (see the red circles below), the glosses “visible” and “invisible” describe whether the interpreter is visible (albeit partially) or not on screen.

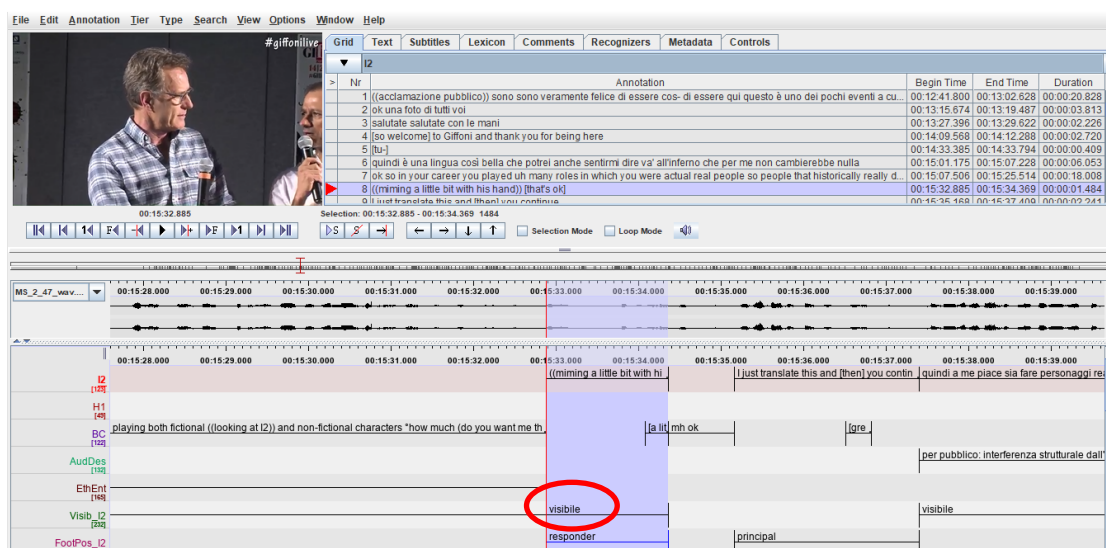


Figure 3. ELAN window with the interpreter visible on screen (middle: guest; right: interpreter)

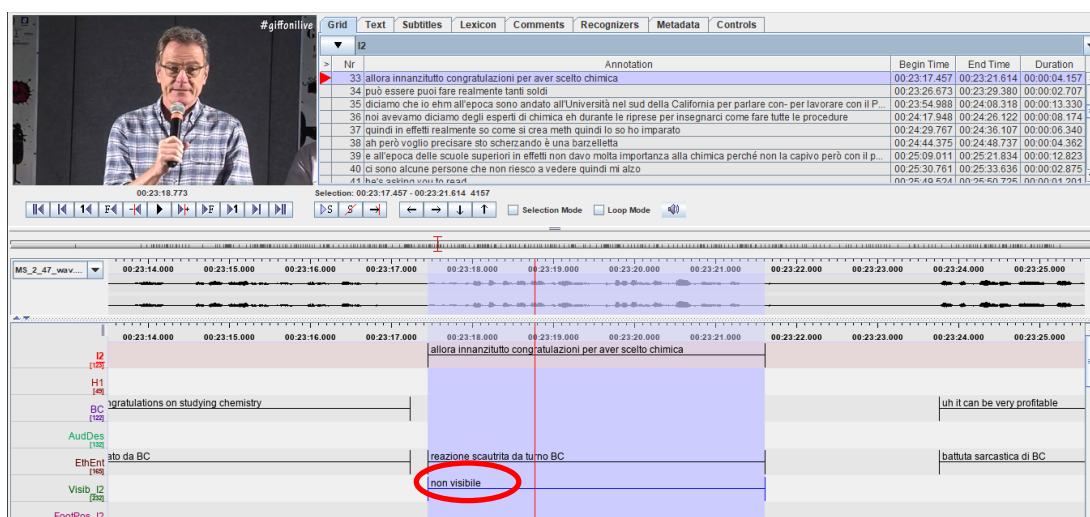


Figure 4. ELAN window with the interpreter not visible on screen

When analysing and discussing the excerpts (see also section 4.3 below), the glosses are replaced with *ad hoc* multimodal conventions (Picchio, 2023a). Excerpt 3 shows Mondada-like conventions (Mondada, 2018; see key in Appendix) and Figure 5 presents a screenshot intended to replace the generic gloss “visible”, which was included in both the Visib_I and Visib_G tiers. Multimodal conventions include the lower case for participants’ embodied behaviours; symbols that delimit the speakers’ non-verbal resources; symbols that describe the action’s length; and hashtags and screenshots to identify the points at which freeze-frames have been taken. In excerpt 3, the supplemental cam (i.e. camera shot) tier has also been added to include additional comments related to the juror’s visibility on screen. Following Mondada’s (2018) suggestions, in order to enhance the readability of the transcripts, various colours and fonts are used: black for turns-at-speaking originally uttered in English; black and italics for words originally uttered in Italian; grey for embodied behaviours; and bold for features of interest.

In excerpt 3, the guest talks with a juror and jokes about his² appearance and age (turn 1), miming his beard (turn 4). The interpreter makes the same gesture (turn 8; Figure 5 below), yet the juror is never visible on screen for the benefit of online viewers.

Excerpt 3

- 1 G you're: you're what? eleven twelve years old?
2 I *how old are you eleven twelve years old?*
3 J twenty-one
cam J not visible--->>>
4 G +°I could tell° by the beard+
g +mimes the beard--->>
5 I *uh yes I wa[s joking]*
6 J [like you] in Breaking Bad
7 G ((chuckling)) that's right
8 I *so like you I *have the #beard* like Walter White in Breaking Bad*
i *mimes the beard*
#Fig. 5



Figure 5. Excerpt 3 (left: guest; right: interpreter)³

In line with the recursive approach put forward by Mondada (2006), this flexible, two-step method enables the transcriber to annotate basic features in ELAN without preventing further *ad hoc* developments, which can be subsequently added when necessary. Three other excerpts with an “enriched” transcript are provided in section 4.3.

To keep track of relevant features, each transcript’s metadata (date, name of participants, noteworthy comments about analysis tiers) were saved in separate Microsoft Word files.

² Personal pronouns and possessive adjectives used in the following paragraphs agree in gender with the person they refer to.

³ The video clip in Figure 5 is included in GFFIntD and is also available at <https://www.youtube.com/watch?v=gYQqJJFSMy0> (Accessed: 20240219).

4.3 Multimodal data analysis: between visibility and entertainment

The tiers specifically devoted to visibility on screen enabled a wide range of analyses (Picchio, 2023a) for investigating whether, when, and how interpreters are visible in the shot for the benefit of streaming viewers, with either positive or negative consequences as far as interaction and entertainment are concerned. Here, by way of example, three eloquent excerpts in which verbal and embodied elements are seen to co-construct meaning during hilarious, guest-initiated vignettes are discussed. As already said (see sections 1 and 3), all the participants in a TV-/online broadcast are required to maintain the interest of the remote audience, ensuring they enjoy the show. In compliance with this “ethics of entertainment”, interpreters themselves should contribute to the entertainment goal of the media event, by also getting involved in funny, ad-lib embodied vignettes and showing that they are completely at ease with what they are doing. Bearing in mind that some GFF events are live streamed, the remote audience has mediated access to them: camera shots can either reinforce or hamper the reception of such sketches, depending on who/what is visible on screen. Thus, the multimodal excerpts discussed below bring together visibility on screen and the “ethics of entertainment”. As described in section 4.2, the following excerpts have transcripts automatically retrieved from ELAN by searching for the annotation “visible” in the interpreters’ Visib tiers. Subsequently, those verbal glosses were replaced with *ad hoc* specific multimodal conventions and screenshots. Excerpts 4 and 5 are taken from a traditional, onsite GFF event (2017), while excerpt 6 is from a hybrid event from the COVID-19 pandemic era (2020).

In excerpt 4, a juror asks the guest (Bryan Cranston), in English, if he has ever “brought home” one of his characters, i.e. if a role changed his personality or lifestyle. The actor answers that he loved playing Walter White in the TV series *Breaking Bad*, but “didn’t want to take him home” (turn 5) — as White, in the guise of his alter ego Heisenberg, becomes a drug manufacturer and dealer. Cranston explains that he usually took two hot, moist towels and draped one over his head like a turban and one over his face “like you’re getting a professional shave” (turn 7), to pull out “the character’s toxins” (turn 12). While describing these actions verbally, the guest uses two hand gestures that illustrate the content of his utterance (turn 7). When the interpreter gives his rendition (turns 8-9), he mimes Cranston’s gestures, and the actor himself mimes the same gestures on the interpreter’s body (Figures 6 and 7 below). It is precisely the re-enactment by the interpreter and the re-re-enactment by the guest which make jurors laugh.

Excerpt 4

- 1 J ok so my question is uhm a lot of actors always said they brought a part of their- of one character- of characters- of their characters home and my question for you is is there a character you uh that has a particular part of his personality that you brought home with you and changed your personality and changed your way to live for example?
- 2 I okay there are many actors who say that they brought home a part of this or that character and so this or that character became part of their lives so I'd like to ask whether there is a character you played that you took home and became part of you

- 3 G it might be a better question for my wife to answer ((chuckling))
4 I *maybe it might be an easier question to- for my wife to answer*
5 G no I for instance Walter White (.) I loved playing that character but I didn't want to take him home
((jurors laugh))
6 I *for instance Walter White it was great to play that character that I have never wanted to take home*
7 G so I would have a routine when I was done with the workday I would take two hot and moist towels
and I would drape +one over my head like a turban+ +and one over my entire face+ almost like you're=
g +mimes the gesture on his body+ +mimes the gesture on his body +
=you're getting a professional shave
8 I *okay so I had a kind of routine at the end of each workday I took two*
very hot towels and I put one on my head like a #turban
i **mimes the gesture on his body----->**
bc +mimes the gesture on l's body----->+
#fig. 6

((jurors laugh))

- 9 I *and [*one on my face*#*
i **mimes the gesture on his body**
10 G [+e due (2,4)+
g +mimes the gesture on l's body+
#fig. 7

((jurors laugh))

- 11 I **like if- when one goes to the barber's to get a professional shave**
i **mimes the gesture on his body----->**
12 G so it would- it would pull out the dirt and the toxins of the day and also the: the character's toxins
13 I *so that this routine pulled out all the toxins of the day along with the toxins also the negative aspects*
of the character I was playing



Figure 6. Excerpt 4/1 (left: guest; right: interpreter)⁴

⁴ The video clips in Figures 6 and 7 are included in GFFIntD and are also available at <https://www.youtube.com/watch?v=gYQqJJFSMy0> (Accessed: 20240219).



Figure 7. Excerpt 4/2 (left: guest; right: interpreter)

In excerpt 5, a juror refers to the theme of GFF 2017 (“Into the magic”) and asks the guest (Kit Harington), in Italian, if he likes magic and what he would use it for — if it existed (turn 1). The actor, who is known for playing John Snow in the TV series *Game of Thrones*, answers that he has always wanted a wand (turn 3); he is a huge Harry Potter fan (turns 5-6) and he would love to have played a wizard, Voldemort in particular, because of the peculiar way he uses his wand, an action he demonstrates using a hand gesture (turn 10). While interpreting, the interpreter tries to mime Voldemort, but his gesture is not accurate (turn 12; Figure 8 below). Harington thus corrects him, miming the right gesture a second time (Figure 9 below). The interpreter ratifies this correction with a smile and mimes Voldemort correctly (turn 13; Figure 10 below). It is precisely when the guest repeats the action and the interpreter rectifies his embodied behaviour that jurors are entertained, as their laughter shows.

Excerpt 5

- 1 J *hi Kit uhm the theme of the Giffoni this year is into the magic and we know that magic is one of the main themes in Game of Thrones so I'd like to know uh what do you think about magic and if it existed what would you use it for?*
- 2 I *uh one of- the theme of the Giffoni Film Festival this year is into the magic and there's a lot of magic in Game of Thrones and if you could ever use magic how would you use it?*
- 3 G *I've always wanted a wand*
- 4 I *I've always wanted a wand*
- 5 G *I think there's something uhm amazing I was a huge Harry Potter fan*
(applause from the jurors)
- 6 G *I'm a I'm a Gryffindor so you know*
(applause from the jurors)
- 7 G *but I I think there's- (.) +sorry do you want to go ahead+?*
g *+looking at I----->+*
- 8 I **I think it was clear ((chuckling)) Harry Potter fan**
i **looking at G----->**
- 9 I **he is a great Harry Potter fan he has always been**
i **looking at the jurors in the stalls----->**

- 10 G I uhm if I uhm I think there's something- I personally would love to have played but- to have played the wizard to +have a wand there's something so powerful about using a wand you know how Voldemort=
g +miming Voldemort--->>
G =does it when he has it like that?
((applause from the jurors))
11 G yeah like that that's what (I do)
12 I so I'd like to have played the role of a wizard and also the way one uses the wand have you seen the particular way in which Voldemort *#uses the wand? precisely that way the way in which=
i *tries to mime Voldemort--->>
#fig. 8
=+he turns it+
g +#mimes the correct gesture+
#fig. 9
((jurors laugh))
13 I *°like that° ((smiling))*
i *#mimes the correct gesture*
#fig. 10
((jurors laugh))



Figure 8. Excerpt 5/1 (left: host; middle: guest; right: interpreter)⁵



Figure 9. Excerpt 5/2 (left: host; middle: guest; right: interpreter)

⁵ The video clips in Figures 8-10 are included in GFFIntD and are also available at <https://www.youtube.com/watch?v=PNCgyNwGYV4> (Accessed: 20240219).



Figure 10. Excerpt 5/3 (left: host; middle: guest; right: interpreter)

Excerpt 6 is taken from a hybrid event in August 2020. Because of COVID-19 travel restrictions, the host, the interpreter and a few Italian jurors were able to participate onsite and in person, while the guest and many other jurors were connected to a Zoom videocall, which was also broadcast on YouTube for the benefit of the streaming audience. A juror asks the guest (Sylvester Stallone), in English, to pick the director he would most like to work with (turn 1). Despite his remote location, Stallone initiates a sketch that revolves around his achievements (turns 3 and 7): he first answers that it depends on the genre of the film, and mentions Ford Coppola, the Russo Brothers and Bay, but then argues that his favourite director is... himself (turn 3). As the interpreter is not visible on screen, he can only convey the irony of Stallone's sentence verbally (turn 4): he thus lengthens the "s" in "is" and makes a short pause before completing his turn ("but I must admit that my favourite director I prefer to work with is: (.) is myself"). A second juror later asks the guest if he still has goals to achieve. Stallone states that he has fulfilled all his dreams and shows the Academy Award he won for Rocky (turn 7; Figure 11 below). Given that the prize is no longer visible on the screen during the interpreter's rendition (Figure 12 below), he is only able to use verbal strategies and thus refers to "the statue you have just seen": multimodality is replaced with speech.

Excerpt 6

- 1 J1 uhm hello uhm it's really you know amazing to have this opportunity to talk to you uhm my question is there have been (lots of) amazingly talented directors past and present uhm and if you could choose one director to work with of all time who would that be?
- 2 I if you coul- if you could choose any director to work with uhm because you worked with many directors in the past whom would you like to work with now?
- 3 G well I think I will go back into history a little bit to see (.) there're so many fantastic directors it depends on what kind of movie (.) if I was doing a film like Rocky I would choose maybe Francis Ford Coppola if I was doing a science fiction film I'd choose someone like the Russo brothers or Michael Bay someone who really understands the camera **but my favourite director that I get along with the best is me ((laughing))**

- 4 I obviously it depends on movies on what kind of movie one wants to make obviously there have been many directors in the past it depends on what kind of movie one wants to make if for instance you want to make a Rocky-style movie it would be great to work with someone like Francis Ford Coppola or if you want to make a science fiction movie someone like Michael Bay because he knows precisely what to make with- with the camera **but I must admit that my favourite director I prefer to work with=**
cam I not visible--->>
=is: (.) is myself
- (...)
- 5 J2 okay my name is {name} and the question is the following so do you feel like you have achieved everything in your career or in your life or is there still something that you would like to do?
- 6 I do you feel like you achieved everything in your career or in your life or there is still something you'd like to achieve?
- 7 G I've achieved everything I wanted to achieve when you know (from) a young man so uhm all my goals have been achieved and beyond uhm when ((talking to his assistant)) °+actually you should bring=
g +pointing upwards--->
=over (.) the statue (xxx)° a long time ago °no right in the middle there it is+ bring them bring them=
g --->+
=here thank you° so when I was thirty years old +we won this right #for Rocky which is really amazing=
g +shows the Oscar--->
#fig. 11
=so (.) I- I can't get better than this+ so what I want to do now is produce and write and find young=
g --->+
=filmmakers and teach them and help them because I have nothing more to prove and you have so much to prove so many dreams you do have to live so never quit always keep fighting
- 8 I I must admit that all my dreams came true I achieved all my goals and even beyond since I was young for instance **#this statue that you have just seen** I won it- it is the Academy Award that=
#fig. 12
- cam the Academy Awards is no more visible--->>
=I won for Rocky and I was thirty years old so now I feel I want to produce help writers and young directors achieve their goals I'd like to help you make your dreams come true



Figure 11. Excerpt 6/1 (guest)⁶

⁶ The video clips in Figures 11 and 12 are included in GFFIntD and are also available at <https://www.youtube.com/watch?v=QnHOMLXcb-Q&list=LL&index=18&t=300s> (Accessed: 20240219).



Figure 12. Excerpt 6/2 (on stage: host and interpreter; in the background: guest)

Excerpts 4 and 5 show that the guests and the interpreters entertain the jurors sitting in the cinema through hilarious embodied vignettes, initiated by the former and co-constructed with the latter. This is both highlighted in the GFFIntD “enriched” transcripts and depicted in the screenshots accompanying the excerpts. In line with the “ethics of entertainment”, not only are the interpreters comfortable with what they are doing when the guest unexpectedly involves them in funny ad-lib sketches, but also the flesh-and-blood jurors enjoy what is going on on stage. The remote audience following the live streaming is presumably entertained as well, given that these sketches are fully visible on screen. Excerpt 6, on the other hand, demonstrates that remoteness impacts on visibility, multimodality and the “ethics of entertainment”. When the guest and the interpreter do not share the same physical location, vignettes are performed differently: personal objects that are usually and “jealously” kept on a shelf are publicly displayed, and while the interpreter translates the guest’s words and quips, he carefully monitors both what is shown by the star and what is visible on the Zoom/YouTube screen — adapting his output accordingly.

5. Conclusions

Speech production is a mediated, ergo embodied phenomenon through which verbal and paraverbal elements are produced by the speaker’s vocal-respiratory system and perceived by the listener’s auditory apparatus. When the visual channel comes into play, speech becomes truly multimodal, as the speaker produces a whole array of visible actions, such as gestures. While interpreting studies researchers have predominantly focused on the verbal components of oral texts, the multimodal approach has been increasingly adopted by dialogue interpreting scholars, along with a corpus-based methodology: dialogue interpreting takes place both in the oral-aural and in the visual-gestural modalities, and thus lends itself well to a multimodal, corpus-based approach.

Following that line of research, this paper has presented the multimodal corpus GFFIntD in terms of data collection, transcription and analysis. In particular, it has discussed the two-step transcription and analysis process applied to certain eloquent

excerpts in which verbal and embodied elements (i.e. gestures) co-construct meaning during hilarious, guest-initiated vignettes.

The corpus consists of English-Italian dialogue interpreting practices performed at the Giffoni Film Festival. It mainly includes live streamed audio-video data (discussed here), but also contains a few audio-only encounters that were not live streamed. The items in the corpus feature different types of events, such as meetings between jurors and stars or Q&A sessions with members of film crews. Despite such differences, the clips share basic characteristics related to types of participants, interaction format, and interpreting mode. Right from the data collection stage, GFFIntD required a flexible approach to deal fruitfully with different items. GFFIntD's materials were transcribed by means of the software package ELAN, which enabled the creation of a speech corpus in which each transcript is aligned with its original clip. While speaker tiers identified the interlocutors' contributions, analysis tiers allowed for preliminary investigation during transcription. The unit of analysis corresponds to a participant's turn, including the Jefferson-like (para)verbal transcription of their words. One of the analytical tiers was specifically devoted to investigation of the participants' visibility on screen, analysing whether, when, and how interpreters are visible in the shot for the benefit of streaming viewers, with either positive or negative consequences as far as interaction and entertainment are concerned. During transcription, verbal glosses describing whether or not the interpreter is visible on screen were included. Nonetheless, in line with the recursive approach put forward by Mondada (2006), such annotations can be replaced using *ad hoc* Mondada-like multimodal conventions and screenshots, when necessary — as the excerpts discussed in this paper have shown.

To illustrate one aspect of the potential of the corpus, three eloquent examples, in which speech and gestures co-construct meaning during hilarious, guest-initiated vignettes, have been discussed here. Like all participants in media events, interpreters themselves are called upon to entertain their audience by also getting involved in funny ad-lib vignettes. Camera shots can either reinforce or hamper the reception of such sketches, depending on who/what is visible on the screen. Thus, the excerpts presented bring together multimodality, visibility on screen, and the “ethics of entertainment”, in both onsite and remote scenarios. In the former, the guests and interpreters co-construct funny embodied sketches that entertain both the flesh-and-blood jurors and (presumably) the audience following the live streams — since the vignettes are fully visible on screen. On the other hand, when different locations are involved, remoteness affects both multimodality and the “ethics of entertainment”: the guest makes use of personal objects to entertain his audience and the interpreter has to use verbal resources and deictics to refer to what has been shown on the screen. This is in line with findings in monolingual conversation analysis and dialogue interpreting studies, which show that gestures are organised with or in relation to vocal conduct to co-construct the ongoing interaction: co-participants use multiple semiotic resources simultaneously, to the extent that different media, such as speech and gestures, mutually elaborate each other within an embodied participation framework.

While acknowledging that the GFFIntD corpus does not draw on a “pure” multimodal method, since it does not directly include a specific set of multimodal conventions in ELAN (cfr. Bao-Rozée, 2017), the multimodal approach introduced here enables scholars to use and benefit from the advantages of a transcription tool like ELAN without preventing any future *ad hoc* developments.

References

- Bao-Rozée, Jie (2016). Using Multimodal Analysis to Investigate the Role of the Interpreter [PhD dissertation]. University of Stirling. Stirling.
<<http://hdl.handle.net/1893/24665>>. [Accessed: 20241126]
- Bendazzoli, Claudio (2017). Editorial. A Dialogue on Dialogue Interpreting (DI) Corpora. *The Interpreters' Newsletter*, n. 22, pp. VII–XVII. <[10.13137/2421-714X/20734](https://doi.org/10.13137/2421-714X/20734)>. [Accessed: 20241126].
- Dal Fovo, Eugenia (2020). Media Interpreting. In: Mona Baker; Gabriela Saldanha (eds.). *Routledge Encyclopedia of Translation Studies*. London/New York: Routledge, pp. 315–320. <<https://doi.org/10.4324/9781315678627-67>>. [Accessed: 20241126].
- Davitti, Elena (2013). Dialogue Interpreting as Intercultural Mediation: Interpreters' Use of Upgrading Moves in Parent-Teacher Meetings. *Interpreting*, v. 15, n. 2, pp. 168–199. <<https://doi.org/10.1075/intp.15.2.02dav>>. [Accessed: 20241126].
- Davitti, Elena (2019). Methodological Explorations of Interpreter-Mediated Interaction: Novel Insights from Multimodal Analysis. *Qualitative Research*, v. 19, n. 1, pp. 7–29. <<https://doi.org/10.1177/1468794118761492>>. [Accessed: 20241126]
- ELAN (Version 5.9/6.0) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <<https://archive.mpi.nl/tla/elan>>. [Accessed: 20240221].
- Falbo, Caterina (2005). La transcription: une tâche paradoxale. *The Interpreters' Newsletter*, n. 13, pp. 25–38. <<http://hdl.handle.net/10077/2468>>. [Accessed: 20241126]
- Falbo, Caterina (2012). CorIT (Italian Television Interpreting Corpus): Classification Criteria., In: Francesco Straniero Sergio; Caterina Falbo (eds.). *Breaking Ground in Corpus-Based Interpreting Studies*. Bern: Peter Lang, pp. 155–185.
- Gao, Fei; Wang, Binhua (2017). A Multimodal Corpus Approach to Dialogue Interpreting Studies in the Chinese Context: Towards a Multi-Layer Analytical Framework. *The Interpreters' Newsletter*, n. 22, pp. 17–38. <[10.13137/2421-714X/20736](https://doi.org/10.13137/2421-714X/20736)>. [Accessed: 20241126].
- Goodwin, Charles (1986). Gestures as a Resource for the Organization of Mutual Orientation. *Semiotica*, v. 62, n. 1–2, pp. 29–49.
<<https://doi.org/10.1515/semi.1986.62.1-2.29>>. [Accessed: 20241126].

- Goodwin, Charles (2000). Action and Embodiment within Situated Human Interaction. *Journal of Pragmatics*, v. 32, n. 10, pp. 1489–1522. <[https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X)>. [Accessed: 20241126].
- Hazel, Spencer; Mortensen, Kristian; Rasmussen, Gitte (2014). Introduction. A Body of Resources – CA Studies of Social Conduct. *Journal of Pragmatics*, n. 65, pp. 1–9. <<https://doi.org/10.1016/j.pragma.2013.10.007>>. [Accessed: 20241126].
- Jefferson, Gail (2004). Glossary of Transcript Symbols with an Introduction. In: Lerner, Gene H. (ed.). *Conversation Analysis: Studies from the First Generation*. Amsterdam: John Benjamins, pp. 13–31. <<https://doi.org/10.1075/pbns.125.02jef>>. [Accessed: 20241126].
- Jewitt, Carey (2013). Multimodal Methods for Researching Digital Technologies. In: Price, Sara; Jewitt, Carey; Brown, Barry (eds.). *The SAGE Handbook of Digital Technology Research*. London: Sage, pp. 250–265. <<https://doi.org/10.4135/9781446282229>>. [Accessed: 20241126].
- Jewitt, Carey (ed.) (2014). *The Routledge Handbook of Multimodal Analysis*. 2nd ed. London/New York: Routledge.
- Jewitt, Carey; Kress, Gunther (eds.) (2003). *Multimodal Literacy*. New York: Peter Lang.
- Katan, David; Straniero Sergio, Francesco (2001). Look Who's Talking. The Ethics of Entertainment and Talk Show Interpreting. *The Return to Ethics. Special issue of The Translator*, v. 7, n. 2, pp. 213–237. <<https://doi.org/10.1080/13556509.2001.10799102>>. [Accessed: 20241126].
- Kress, Gunther (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London/New York: Routledge.
- Kress, Gunther; van Leeuwen, Theo (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- McNeil, David (2006). Gesture, Gaze and Ground. In: Renals, Steve; Bengio, Samy (eds.). *Machine Learning for Multimodal Interaction MLMI 2005*. Berlin: Springer, pp. 1–14. <https://doi.org/10.1007/11677482_1>. [Accessed: 20241126].
- Merlini, Raffaella (2017). Developing Flexibility to Meet the Challenges of Interpreting in Film Festivals. In Cirillo, Letizia; Niemants, Natacha (eds.). *Teaching Dialogue Interpreting. Research-based Proposals for Higher Education*. New York/Amsterdam: John Benjamins, pp. 137–157. <<https://doi.org/10.1075/btl.138.07mer>>. [Accessed: 20241126].
- Merlini, Raffaella (2020). Dialogue Interpreting. In: Mona Baker; Gabriela Saldanha (eds.). *Routledge Encyclopedia of Translation Studies*. London/New York: Routledge, pp. 147–152. <<https://doi.org/10.4324/9781315678627-32>>. [Accessed: 20241126].
- Merlini, Raffaella; Picchio, Laura (2019). Liminalità e interpretazione: sconfinamenti tra posizioni interazionali e piani comunicativi. In: Diego Poli (ed.). *In Limine: frontiere e integrazioni*. Roma: Il Calamo, pp. 199–233.

- Mondada, Lorenza (2006). La pertinenza del dettaglio: registrazione e trascrizione di dati video per la linguistica interazionale In: Yvette Bürki; Elwys de Stefani (eds.). *Trascrivere la lingua. Dalla filologia all'analisi conversazionale*. Bern: Lang, pp. 313–344.
- Mondada, Lorenza (2014). The Local Constitution of Multimodal Resources for Social Interaction. *Journal of Pragmatics*, n. 65, pp. 137–156.
<<https://doi.org/10.1016/j.pragma.2014.04.004>>. [Accessed: 20241126].
- Mondada, Lorenza (2016). Challenges of Multimodality: Language and the Body in Social Interaction. *Journal of Sociolinguistics*, v. 20, n. 3, pp. 336–366.
<https://doi.org/10.1111/josl.1_12177>. [Accessed: 20241126].
- Mondada, Lorenza (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, v. 51, n. 1, pp. 85–106. <<https://doi.org/10.1080/08351813.2018.1413878>>. [Accessed: 20241126].
- Neville, Maurice (2015). The Embodied Turn in Research on Language and Social Interaction. *Research on Language and Social Interaction*, n. 48, pp. 121–151.
<<https://doi.org/10.1080/08351813.2015.1025499>>. [Accessed: 20241126].
- Niemants, Natacha (2012). The Transcription of Interpreting Data. *Interpreting*, v. 14, n. 2, pp. 165–191. <<https://doi.org/10.1075/intp.14.2.03nie>>. [Accessed: 20241126].
- Niemants, Natacha (2018). Des enregistrements aux corpus: transcription et extraction de données d'interprétation en milieu médical. *Meta*, v. 63, n. 3, pp. 665–694.
<<https://doi.org/10.7202/1060168ar>>. [Accessed: 20241126].
- Norris, Sigfrid (2004). *Analyzing Multimodal Interaction*. London: Routledge Falmer.
<<https://doi.org/10.4324/9780203379493>>. [Accessed: 20241126].
- Picchio, Laura (2023a). L'interpretazione dialogica tra cinema e media digitali: il Giffoni Film Festival come caso di studio [PhD dissertation]. University of Macerata. Macerata. <<https://hdl.handle.net/11393/317130>>. [Accessed: 20241126].
- Picchio, Laura (2023b). Distance vs. (Non-) Streamed Interpreting Performances: A Focus on the Rendition of Film Scenes. *The Interpreters' Newsletter*, n. 28, pp. 171–188. <[1007/35560](https://doi.org/10.1007/35560)>. [Accessed: 20241126].
- Pöschhacker, Franz (2020). 'Going Video'. Mediality and Multimodality in Interpreting. In: Salaets, Heidi; Brône, Geert (eds.). *Linking up with Video. Perspectives on Interpreting Practice and Research*. Amsterdam/Philadelphia: John Benjamins, pp. 13–45.
<<https://doi.org/10.1075/btl.149.02poc>>. [Accessed: 20241126].
- Rossano, Federico (2012). Gaze Behavior in Face-to-Face Interaction [PhD dissertation]. Radboud University: Nijmegen.
- Sandrelli, Annalisa (2015). 'And maybe you can translate also what I say': Interpreters on Football Press Conferences. *The Interpreters' Newsletter*, n. 20, pp. 87–105.
<[http://hdl.handle.net/10077/11854](https://hdl.handle.net/10077/11854)>. [Accessed: 20241126].

- Schegloff, Emanuel A. (1984). On some Gestures' Relation to Talk. In: Atkinson, J. Maxwell; Heritage, John (eds.). *Structures of Social Action. Studies in Conversation Analysis*. Cambridge University Press: Cambridge, pp. 266–296.
<<https://doi.org/10.1017/CBO9780511665868.018>>. [Accessed: 20241126].
- Scollon, Ron; Scollon Wong, Suzie (2009). Multimodality and Language: A Retrospective and Prospective View. In: Carey Jewitt (ed.). *The Routledge Handbook of Multimodal Analysis*. London: Routledge, pp. 170–180.
- Straniero Sergio, Francesco (2007). *Talkshow interpreting: La mediazione linguistica nella conversazione spettacolo*. Trieste: EUT.
- Straniero Sergio, Francesco; Falbo, Caterina (2012). Studying Interpreting through Corpora: An Introduction. In: Francesco Straniero Sergio; Caterina Falbo (eds.). *Breaking ground in Corpus-Based Interpreting Studies*. Bern: Peter Lang, pp. 9–52.
<<https://doi.org/10.3726/978-3-0351-0377-9>>. [Accessed: 20241126].
- Wadensjö, Cecilia (2008). In and off the Show. Co-Constructing Invisibility in an Interpreter-Mediated Encounter. *Meta*, v. 53, n. 1, pp. 184–203.
<<https://doi.org/10.7202/017982ar>>. [Accessed: 20241126].
- Wittenburg, Peter; Brugman, Hennie; Russel, Albert; Klassmann, Alex; Sloetjes, Han (2006). ELAN: A Professional Framework for Multimodality Research. In: Calzolari, Nicoletta; Choukri, Khalid; Gangemi, Aldo; Maegaard, Bente; Mariani, Joseph; Odijk, Jan; Tapias, Daniel (eds.). *Proceedings of LREC 2006. Fifth International Conference on Language Resources and Evaluation*. European Language Resources Association, 1556–1559. <http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf>. [Accessed: 20240221].

Appendix. Transcription key

(Para)verbal elements – in black	
J G I H	upper case for participants as speakers
{text}	anonymised detail
(text)	transcriber's guess
[text] [text]	overlap
uh , uhm	vocalised pause
text:	lengthened sound
text-	abrupt cut-off in the flow of speech

°text°	decreased volume
=text	latching
?	rising intonation
(1)	timed pause
(.)	untimed short pause
(...)	omitted speech
((text))	paraverbal behaviour
<i>text</i>	transcriber's translation into English
text	feature of interest
Multimodal elements – in grey	
g i	lower case for participants' embodied behaviours
cam	comments about camera shots
+text+	delimitation of description of G's actions
text	delimitation of description of I's actions
--->+	action described continues until the same symbol is reached
--->>	action described continues until the turn's end
--->>>	action described continues until the extract's end
#fig. #	the exact point where a screenshot (figure) has been taken is indicated with a specific sign showing its position within a turn at speaking
text	feature of interest