Revisiting multimodal analysis methods for multimodal interactions

Dondon Parohinog, Wannapa Trakulkasemsuk and Sompatu Vungthong
King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Abstract
Technological affordances resulting from the prevalence of portable devices have given qualitative researchers the convenience of gathering data using video recorders. Analyses, however, only occur when these recordings are converted into transcripts. Multimodal transcription and analysis methods have been developed to analyze interactions in a recorded video. These methods are anchored in different philosophical principles depending on the variety of audience, research purposes and the modes to be investigated. However, some issues contiguous with transcription lead to problematic re-presentation of interactions using transvisuals (transcripts). Qualitative researchers face issues related to what and how of transcription. In this study, we summarized articles by analyzing multimodal analysis methods including their approaches, transcription and analysis and their application in video recorded data. Further, the analyses paved the way for a deeper understanding on what effects transcriptions have on analysis and interpretation and how modes are organized in the transcripts to highlight the research purposes. Transcription conventions suitable for different research purposes will be suggested.

1. Introduction

The growing interests to multimodal studies have attracted scholars from other disciplines such as sociology, business, linguistics and health who have also seen the exponential growth (Bezemer & Jewitt, 2018) of this novel field of inquiry. The rationale points back to the need to understand academic and social phenomena from a multimodal perspective. Pursuing a multimodal study comes with a handful of challenges for novice researchers like ourselves. These challenges include employing specific approaches to analyze video-recorded interactions and transcription methods. Leading scholars in this field have authored articles that shed light on issues concerning concepts and steps in undertaking a study in multimodality. This paper aims at providing a comprehensive map on the approaches to multimodal studies by looking closely at the methodological processes of previous research and drawing insights which may serve as springboard for researchers coming into this field of inquiry.

Approaches to multimodal studies have distinct methods of addressing questions, collecting and analyzing data. Data sets like video-recorded interactions must be in naturalistic settings. It gives emphasis on a contextualized data collection method considering the social and cultural influences to making sense of the world. Along with these approaches come the development of technology which supports the expedited data collection and analysis. Some of these include technology for data collection (use of camcorders, mobile phones for video recording) and software for analyzing the data at hand (e.g. ELAN).

As mentioned, researchers coming from different fields also brought with them diverse terms and concepts that have created boundaries among disciplines. Yet, scholars highlighted the common interest in analyzing 'meaning-making'. Echoing Bezemer and Jewitt (2018) suggestions, researchers need consistent and clear stance as to what categories and terms to use when engaged in multimodal research. In this article, we analyze previously published articles on multimodality outlining the approaches, transcription methods and analysis as well as highlighting concepts to guide other researchers who are planning a multimodal research project.

1.1 Research questions

This study aims to provide a comprehensive account of processes to multimodal studies. Specifically, this study aims to answer these questions:

1. What are the approaches to multimodal methods and analysis?
2. What are the gains from using transcription methods to specific sets of data?

2. Literature review

Multimodal studies attract scholars from different fields ranging from sociology to psychology and health all coming together to understand the phenomena underpinning multimodal interactions. The increased interest in multimodal studies have also uncovered different theoretical, methodological and domain-related issues (O'Halloran, 2011) especially in theorizing meaning-making processes and understanding the contexts where modes are chosen. To address such issues, there is a need for a clear distinction among approaches to multimodal studies. In the succeeding section, we discuss three main approaches along with their distinct principles and concepts, research focus, and method of analysis.

2.1 Approaches to multimodal studies

In her article, Jewitt (2009) explicitly discussed the three main perspectives within multimodality including social semiotic multimodal analysis, systemic functional approach (multimodal discourse analysis) and multimodal interactional analysis. The discussion acknowledges the emergence of a wide range of interests and distinct theoretical concepts and frameworks (O'Halloran & Smith, 2012). Although theories and analytical framework might differ from one approach to the other, these approaches are all anchored on the same interest- to understand how modes or semiotic resources realize their potential in meaning-making in a specific context.

2.1.1 Social semiotic approach

In 1978, Halliday introduced the term social semiotics and later developed by scholars Hodge and Kress (1988) who argued against the separate system of language and its social context. The work of Kress and van Leeuwen (2001) can be a point of reference for multimodal study which focus on considering the potential of semiotic resources not only in linguistics but also in other social theories with interest in meaning-making. In their book Reading Images, Kress and van Leeuwen figured how meaning is realized visually through composition, modality and framing (Jewitt, 2009, 2014; Jewitt, Bezemer & O'Halloran,2016). These concepts including visual semiotic resources manifest potential in communicating ideologies and discourses. Available semiotic resources are chosen from a range of resources which belong to a system. Kress and van Leeuwen, however, emphasized character of meaning-making which values the situated use of semiotic resources. It foregrounds the context where communication occurs, and the modes being used. For example, a museum visitor's interpretation of a painting or an image may depend on his awareness of how different modes of communication are integrated in art form. These art forms are called 'artefacts' which is the focus of social semiotics research. Lately, this approach also included the study of video recorded interactions.

Central to social semiotic theory is the concept of mode which is defined as socially and culturally shaped resources for meaning making. Social semiotic approach aims to analyze the interplay of modes and how are they articulated in various context of social interactions. It also foregrounds the question as to which modes people choose among those which are made available to them.

2.1.2 Systemic functional multimodal discourse analysis (SF-MDA)

Investigating interactions from multimodal perspectives stemmed back from Halliday's (1985) Systemic Functional Linguistics (SFL) which views language as a social semiotic system, a resource for meaning making (Jewitt, et al, 2016). SFL aims at developing functional grammar to describe the potential of language in meaning making. However, language or speech is not the sole resource for meaning making. Halliday acknowledged this claim especially that cultural forms such as art, sculpture, and music, which are not bounded by language, also bear meaning. Systemic Functional Theory (SFT) was conceptualized to explore the meaning-making potentials not only by language but also non-linguistic resources including images, gestures, space and the likes. In their book Reading Images: The Grammar of Visual Design, Kress & van Leeuwen (2006) first applied SFT in multimodal studies which provides a comprehensive and systematic account of grammar in visual design (Thuy, 2017). Under the umbrella theory of SFT is Systemic Functional Multimodal Discourse Analysis (SF-MDA) which analyzes the relationships of semiotic resources and the meanings they create when afforded in a social context. It is in the inter-relationship of these resources that the term 'systemic' is realized. Furthermore, meanings can also be realized using the analysis of three meta functions (discussed in detail by Halliday and Matthiessen (1985) in their book An Introduction to Functional Grammar). Eggins (2004) wrote her simplified version of SFL in her book An Introduction to Systemic Functional Linguistics. These meta functions include ideational, interpersonal and textual. Ideational metafunction realizes meaning through representation of reality through linguistic text and create the same experience through various lexico-grammatical options (Haratyan, 2011). The core concept of this metafunction is the system of transitivity and is comprised of experiential and logical meanings. Interpersonal metafunction puts its focus on social role and relationships through formality degree, pronouns and clausal mood. This also allows the exploration of degree of intimacy and type of relationship between the writer and the reader through the type of modality. Lastly, the textual metafunction focuses on the message with is realized by a theme. Thematic structure may point towards theme and rheme or the old and new information structure. Halliday (1982) explained that theme includes message in a text that indicates identity of text relations.

This approach highlights the concept of 'discourse' in the combined approaches of Halliday's Systemic Functional Grammar (SFG) and O'Halloran's Multimodal Discourse Analysis (MDA). The term now becomes Systemic Functional -Multimodal Discourse Analysis (SF-MDA) with these aims: 1) model the meaning potential of resources as interrelated sets of system and 2) analyze the meaning arising from the interactions in multimodal processes and texts according to context (Jewitt, et al, 2014). Like social semiotic approach, context is given prominence when analyzing modes in context. This approach agrees with social semiotics which considers how modes are shaped by social practices and culture.

2.1.3 Conversation analysis (CA)

The mid-twentieth century origin of CA can be traced back to Blumer, Goffman and Garfinkel's sociological approach called interactionism (Jewitt, et al, 2016). The approach proposed to investigate social interaction in fine-grained manner for purposes of observing, examining and describing instantiations of social interactions. Later, it became a theoretical and methodological approach in understanding interactions. The former scholars' influence on Sacks, Schegloff and Sudnow lead to the latter's work in understanding the structure and logics of interactions theorized how callers of Suicide Prevention Centre in Los Angeles opened and closed their conversations. Instances of talk in various naturalistic settings became the focus of CA research. To maintain the naturalistic character of interactions, CA utilizes fieldwork in collecting data and ascertains the minimal participation of the researchers.

Ethnography of communication shares the same traditions as that of CA especially the integration of fieldwork as means of data collection.

CA focuses on the analysis of speech or talk-in-interaction. Characteristics of speech and other linguistic elements become more salient when transcripts are produced from audio and video recordings. These transcripts produce fine-grained and moment-by-moment account of language which may co-occur with other semiotic resources. One notable work in CA-based multimodal study was conducted by Mondada (2008) who studied the video recordings from a call center setting. The call center setting typically produces audio recordings with potentialities for speech or talk analysis but Mondada went to analyze the bigger picture of situated complex activities in the workplace. The results provided evidence for temporal and sequential activity where call center agents continuity in the workflow showing coordination and simultaneity of activities. This study proves the widened scope of CA-oriented studies from speech or talk to a multimodal research focus.

2.1.4 Other approaches to multimodal studies

Interactions can be analyzed from perspectives dependent on the researchers' orientation. Some scholars turn to other fields such as ethnography of communication and corpus-based approaches to deal with data collection processes and analysis which is beyond the scope of previously discussed approaches. Ethnographic-based approach to multimodal communication includes collection of data through fieldwork and video recordings (Jewitt, et al, 2014) in naturalistic setting. Material artefacts, objects and the environment are also considered during the data collection process. It emphasizes the contextualized process of meaning-making where social practices and culture shape how modes are chosen from all the available resources. The complexity of analyzing interactions is made easier by putting them in social, cultural and historical contexts (Flewitt, 2011). Flewitt also sees the importance of staying 'faithful' to the phenomena under investigation. It requires constant cross-checking with the participants instead of relying entirely on 'outsider' perspective.

Corpus-based approach, on the other hand, brings its principle of systematic analysis of large corpora to multimodal field. Most approaches to multimodal studies deal with smaller number of texts and it delimits the parameter for hypothesis-testing, generalizing and building theories. The main proponent of corpus approach, Bateman (2014), echoed the importance of evaluating hypotheses and theories given the attention multimodality receives as it progresses. He further argued that insufficient quantities of data may not result to generalizable findings and generalizability can be achieved if facilities to analyze large scale of data exist. As technology advances, so as data collection and availability of tools to prepare, annotate and analyze data in large quantities. Data which can be analyzed using a corpus approach include printed and digital artefacts and video-recorded interactions. Corpus-based multimodal analysis includes two stages: construction of corpus and annotation (Bateman, Delin & Henschel, 2004).

2.2 Videos as multimodal data

Videos have shown immense potential among qualitative researchers to examine interactions in social and academic settings. Aside from its practical contribution in recording 'naturally occurring events', videos also account temporal and sequential structure (Knoblauch, et. al, 2009) of any interactions. Although videos project to mirror reality or a replica of what happened (Jewitt 2012), a belief that it also distorts reality was floated. Ethnographers and visual anthropologists refused to concur instead they paid attention to how videos can be exploited to understand aspects of social interactions. As discussed in the preceding section, all three approaches analyze small fragments of videos. As argued earlier, transcribing videos justifies the deeper examination of resources and how social actors use the available modes to make sense of things around them. Transcription, however, is not straightforward as it is a part of the analytic process (Pirini, 2014). The decisions to select specific parts of a video must be informed by the theoretical and methodological principles avoiding opinionated and biased accounts from initial viewing of the videos. Minimizing biases in video selection, Goldman et al. (2007) offered three ways to approach video selection: inductive, deductive and narrative evolving. Qualitative research including those directed towards analyzing interactions may follow an inductive method by ethnographically viewing the data to identify multimodal instantiations.

Among these transcriptions, there are three prominent systems that emerged from multimodal studies. First, the Jefferson Transcription System (Jefferson, 1974) which is adopted by conversation analysts to study speech patterns and the style of a participants to a conversation. Jefferson transcription system produces transcript not only of speech but also the manner of speaking. Paralinguistic elements such pauses, fillers and intonations become evident when transcribed. We find this system particularly useful when the study is focused on speech and its characteristics. On the other hand, multilevel transcription (Heath, et al, 2010) adopted by Bezemer (2014) in his article to account a step-by-step process of multimodal transcription. Multilevel transcription finds its appropriacy when the study aims to account for multiple modes in one transcript. Using a transcription software, modes can be presented in different tiers including image captures. Further, fine-grained analysis of the temporal and sequential actions by the participants can be read in this transcript. Finally, Pirini (2014) used the data through text and image transcription which was developed by Norris (2002, 2004, 2011). From the perspective of multimodal interaction analysis, the transcript includes an image that chosen based on when the action occurs and not based on the temporal element. It is composed of a chain of lower-level actions that comprise higher-level actions (Norris, 2004 and Norris & Jones, 2005). This method is useful especially in recognizing how individual modes are used by social actors to construct higher-level actions.

Since this paper maps out the general approaches to multimodal studies and their ontological and epistemological configurations, we may consider the theoretical and methodological stance of the researcher. These concepts are crucial elements in analyzing naturally occurring interactions. Many multimodal researchers who have investigated interactions have certain beliefs in transcription and analysis methods. Conversation analysts do not use the term 'mode' instead resource in multimodal interaction while social semioticians refer to them as modes, resources or semiotic resources. In reviewing published articles, we create clear distinction about the terms, theories, approaches and methods in doing multimodal studies.

3. Methodology

The main purpose of this paper is to streamline basic concepts, approaches and transcription and analysis methods of multimodal studies for varied audiences. Given the attention this field receives from scholars in different disciplines and domains, we respond to Bezemer & Mavers' (2011) call to analyze a wide range of articles for purposes it is deemed significant. In the succeeding section, we discuss the criteria for the selection of the articles as data sources and the method of analysis.

3.1 Criteria for selecting articles

The articles that we analyzed purpose for this study were carefully selected based on some general and specific criteria. The general criterion in selecting articles is the methodological framework. It is a must that the articles work within the framework of multimodality. The central concern must revolve around the investigation of meaning-making resources including pedagogical space (Lim & O'Halloran, 2011), gesture that co-occurs with speech (Martinec, 2004), gaze that co-occurs with speech (Brône, et al, 2017), creative use of modes other than language (Taylor, 2014), children's interaction in multiple modes (Cowan, 2013), speech and body movements in surgical theatre (Bezemer, 2011), roles of multiple modes in science classroom (Danielson, 2016), gestures in mediated interaction (Lee, et al, 2019) language, gaze and posture in creative classroom interaction (Taylor, 2016), and the role of body in interaction (Price & Jewitt, 2013). These articles were published in high quality journals (e.g. Language and Education, Classroom Discourse, Cambridge Journal of Education, System) by leading scholars in the field. The citation index of these journals is remarkably high indicating the journals' classification and ranking. Finally, we examined the articles' transcription methods being an integral stage preceding analysis. The justifications on the use of transcription method must be traceable to the articles research purposes.

3.2 Methods of analysis

The analysis of the articles included several stages content analysis. The purpose of doing so is to follow the authors' methods in conducting research in multimodality. Krippendorff (2018) explained that content analysis as the analysis of the theories underpinning the material. First, each article was summarized giving prominence to research purposes, sources and the methods of analysis. The methods of analysis part provided more insight into the selection of methodological framework which is highly influenced by the research purposes. The modes to be highlighted are also determined in this section before the next stage of defining the purpose of transcription. As we argued elsewhere, transcription plays a huge role in the analysis. It produces the transcript that allows deeper examination of the relationship of modes that emerged from the data set. The design of the transcript is determined by the theoretical and methodological stance of the researcher. The second stage in the content analysis is coding the data. It assures consistent categorization of content. From the coding process, we identified the approaches of multimodal studies with the corresponding transcription methods. These approaches and transcription methods are discussed in detail in the results section.

4. Results

RQ1: *What are the approaches to multimodal methods and analysis?*

The nature of interactions is built from complex, interwoven ensembles of communicative modes or meaning-making resources. Multimodal studies, therefore, must seek to understand the complexity of these interactions by identifying the approaches, classifying possible research foci, rationalizing the use of transcription methods based on modes of interest and theorizing how these modes are afforded in specific contexts of interactions. It should be noted as well that all the articles deal with video recorded interactions as data. Out of the ten articles we analyzed, five approaches emerged: conversation analysis, systemic functional-multimodal discourse analysis, social semiotics, multimodal-ethnography and corpus-based approaches. In the succeeding paragraphs, we briefly discuss how these approaches were employed in each article highlighting their concepts and principles.

4.1 Conversation analysis

Conversation analysis (CA) was employed by Cowan's (2013) study of children in a play-enhanced learning environment and Bezemer's, et al (2011) analysis of the interaction between a surgeon and scrub nurse during a surgical operation. Although situated in different contexts, these two articles investigated features of language including speed and latency unraveling some paralinguistic features invisible from its spoken form. CA usually situates its analysis in turn-taking, sequence and repair which are visibly prominent when adopting Jeffersonian's transcription system. After the transcription of a selected segment of the surgical operation, a 'repair work' was done when communication was about to fail. The surgeon did a repair of his request by clarifying the kind scissors he needed. Although CA focuses on speech or talk-in-interaction, it has not set its boundaries in dealing with resources other than language. CA's research foci clearly point toward speech and talk, it should theorize how resources other than language shape the interaction and yet, these resources are not given emphasis in the analysis.

4.2 Systemic functional multimodal discourse analysis

The systemic functional-multimodal discourse analysis sets its goal towards the development of functional grammar to account for the meaning-making potential of language (Jewitt, 2014) in artefacts such as digital and printed texts, videos and other three-dimensional objects. Although SFL was developed to focus only on language, systemic functional theory caters to the study of semiotic resources other than language encompassing the application of systemic functional- multimodal discourse analysis (SF-MDA). The articles written by Taylor (2014, 2016) and Danielson (2016) use SF-MDA approach to analyze interactions through the following metafunctions: experiential, logical, interpersonal and textual. Danielson analyzed resources through the ideational metafunction and only presented results regarding processes. These processes were originally proposed by Halliday & Matthiessen (2004): material, behavioural, mental, verbal, relational and existential. On the other hand, Taylor's (2016) paper drew upon SF-MDA perspective to the importance of how uncovering the temporal structure in classroom interaction and understanding the texture of engagement. Using the metafunctions as framework of analysis, she endeavors to uncover how meaning is made modes such as language, gaze and posture. SF-MDA allows an in-depth analysis of language as meaning-making resources through metafunctions and processes.

4.3 Social semiotics

Another approach that focuses on the analysis of artefacts especially in print media is the social semiotics. Aside from artefacts, this approach also deals with the social interactions recorded in a video (Jewitt, 2014). Social semioticians define modes as socially and culturally shaped resources for meaning-making. From the results of our analysis, social semiotic approach was used in exploring pedagogical space (Lim, F.V, 2012), affordances of semiotic resources in chemistry class (Danielson, 2016), examining embodied modes in tangible learning environment (Price & Jewitt, 2013). Particularly interesting is Lim's and his colleagues who investigated the use of space by mapping the positioning and movement of the teacher. Based on the argument that teachers use different resources in different spaces and each space is always reconfigured. Since the focus is on studying 'pedagogical space', the researchers used a coding technique to represent movement. The challenge now is how to represent movements and use of space inside the classroom. A software used to visualize molecular interaction networks, Cytoscape (Shannon, et al, 2003) which has been adopted to analyze network graphs in other social science research. Effective and purposeful as the software was, it also poses challenges for a more accessible, innovative and user-friendly application for the purpose of analyzing space.

4.4 Multimodal ethnography

Ethnography as a methodological approach describes the daily routine of a population in their most natural environment where the researcher becomes a participant who observes, interprets (etic) and cross-checks (emic) the meaning of data being gathered. As an inductive method, ethnography drives researchers to collect available data that may shed light on social issues (Hammersley & Atkinson, 1995) let alone providing accounts of how multiple social factors shape actions. Ethnographic approach to multimodality addresses inquiry relevant to daily activities, practices and context and how they influence meaning making processes. In observing a specific group or community, data can be collected from using fieldnotes and video recordings to document the participants' lived experiences as influenced by their society and culture. Among the articles analyzed in this study, Taylor's (2014, 2016) follow ethnographic principles with one article focusing on Linguistic Ethnography (LE) and the other following an ethnographic approach to data collection. LE investigates how language use in the classrooms mirrors social norms. As proposed by Mavin & Tusting (2011) LE employs ethnographic principles and linguistic methodologies to study language across settings including interactions with language as used as one mode. Taylor argued that understanding the context in which communication takes place is one job that a researcher should do instead of making assumptions about meaning making. Her data collection involved months of observing classroom communication and creativity which involves recontextualizing and re-presenting a specific phenomenon.

4.5 Corpus-based multimodal analysis

Finally, one approach that was utilized to study multimodality is corpus-based approach. It emphasizes the use of corpora that instantiates a phenomenon leading towards a warranted bases for hypotheses formation, generalizing and building multimodal theory (Jewitt, et al, (2016). Corpus is not a stand-alone approach to studying multimodal interactions. Its principles and analytical tools are combined with other approaches such as social semiotics for multimodal study of gaze in relation to other modes. The possible data for corpus-multimodal study include artefacts and video-recorded interactions. Similar with other approaches, analysis is carried out by fine-grained analysis of fragments of a video which instantiate the mode in focus.

Brône and his colleagues (2017) repurposed a multimodal video corpus of face-to-face interactions from Insight Interaction Corpus (Brône & Oben, 2015). The conversations in Dutch consist of dyads and three-party interactions for 30 minutes and 15 minutes which are taken from different contexts with different purposes ranging from storytelling to casual conversation without a predefined topic. The study aims to provide details of gaze patterns in relation to co-occurring speech and gesture to co-participant behavior. The results claimed that gaze aversion means turn-holding and shared gaze indicate urgency to take turn during the interactions.

Unlike other approaches to multimodal studies, the corpus-based approach analyzes data which are pre-annotated for modes of interest, in this study-- gaze and gesture. As novice researchers to the field of multimodality, we understand the potential of corpus-based tools and analysis in multimodal study of interactions. Based on corpus principles, data must be in naturalistic setting. However, collecting videos of naturally occurring interactions poses challenges and ethical issues. We find this necessary to emphasize to inform future researchers of the possible challenges they might encounter.

Table 1. Five of the approaches to multimodal studies and the brief description of their aims, focus and methodology.

Approach	Approach
■ Conversation analysis	■ **Aim:** To recognize social order in interactions ■ **Focus:** Video recordings of naturally occurring social interactions ■ **Methodology:** Transcribes and analyzes fragments of interactions but with focus on speech and talk-in-interaction
■ Systemic functional multimodal discourse analysis	■ **Aim:** To understand how the system of language is used to fulfill social functions ■ **Focus:** Collection of 'artefacts' including print and digital texts, forms of media (e.g. advertisements), educational media (e.g. textbooks) and arts and crafts (e.g. sculptures) ■ **Methodology:** Transcribes and analyzes in detail fragments of texts and corpora
■ Social semiotics	■ **Aim:** To investigate meaning-making in a situated context and what modes are chosen based on available modes ■ **Focus:** 'artefacts' (e.g. print media, film and games) and later video-recorded interactions ■ **Methodology:** Analyzes snippets of videos from a bigger set of data
■ Ethnographic multimodal approach	■ **Aim:** Investigates how social and cultural contexts shape meanings through multimodal affordances ■ **Focus:** Cultural and social practices informed by prolong observations ■ **Methodology:** Combines with social semiotics to micro-analyze artefacts' design and social and cultural norms

Approach	Approach
■ Corpus-based approach	■ **Aim:** To evaluate, critique and validate multimodal hypotheses and theories of meaning-making. ■ **Focus:** co-occurrence of modes on corpora of multimodal artefacts and interactions ■ **Methodology:** Systematically analyzes larger corpora of texts to support multimodal theory

RQ2: *What are the gains from using transcription methods to specific sets of data?*

All the articles that we analyzed for this study dealt with video-recorded interactions which are situated in naturally occurring contexts except the corpus-based study which dealt with stimulated conversations in various topics. Since transcription methods allow us to see data from a different perspective, they also create an impact on how see the data. Outlined in Table 1 are five of the approaches to multimodal studies that emerged from the articles that we summarized, their possible transcription methods. It is important to note, nevertheless, that these transcriptions methods may vary depending on the research purposes and the modes of interest. Some of these transcription methods were not utilized in the articles that we analyzed but putting them here may give additional information to future researchers. On the third column, we included the summarized description of the potential gains from using these approaches and transcription methods.

Conversation analysis approach to multimodal interactions has the greater likelihood in investigating linguistic elements that point towards turn-taking, sequence and repair. Repair is evident in Bezemer, et al (2011) who studied interaction in the operation theatre where a previously vague utterance was repaired by clarifying which type of scissors is needed. Such repair is evident in the transcript as shown in Figure 1. Consequently, CA's adaptation of Jeffersonian System of transcription produces transcripts that allow the analysis of less salient elements in spoken data.

Episode 2: "Scissors, please" (CS: consultant surgeon; SN: scrub nurse)

1. CS S . . . scissors please (slightly raised tone)
2. SN (3.0)( )
3. CS No I'm gonna cut in the ar:tree. Lo::ng, (.) th the
4. surgical dissecting scissors

Figure 1. CA-based transcription of interaction with 'repair' by Bezemer, et al. (2011)

A multimodal transcription (a transcription grid produced using ELAN software where modes are presented in different tiers) might be the most appropriate. Cowan (2013) deviced two types of multimodal transcription for different purposes. First, the tabular layout where each mode is separated and presented in different columns to show specificity but at the same time creating a picture that shows the relationship of the modes. Second, the timeline layout presents video stills and spatial modes. Unlike the tabular layout which presents each mode in written form, timeline layout allows the examination of other modes (e.g. gaze and space) through the visual elements. In Bezemer, et al. (2011), multimodal transcription method was utilized to account the temporal and sequential representation of modes including speech to body movements. The transcript with captured images highlights a segment of an interaction where other modes may be slightly put in the background.

Interestingly, the corpus-based approach to multimodality as exemplified by the study of Brône, et al, 2017 utilized multiple methods related to transcription: 1) orthographic transcription method to account for accent and intonational contour, 2) multimodal annotation using ELAN for gaze fixations and 3) MUMIN Coding Scheme for annotating intonation units and turn management features. These three methods seem impractical, they might be necessary to realize the purpose of measuring gaze events and their synchronicity with verbal markers to signal turns in the interaction.

Table 2. Five of the approaches to multimodal studies, possible transcription methods and their gains.

Approach	Transcription methods	Gains
■ Conversation analysis	■ Jefferson's transcription system	■ Provides ways to analyze linguistic patterns and paralinguistic features such as pauses, latency and fillers.
■ Systemic functional multimodal discourse analysis	■ Multimodal analysis video software	■ Describes semiotic choices that realize meaning through meta-functions by annotating, analyzing and visualizing combinations of multimodal choices.
■ Social semiotic approach	■ Multilevel transcription grid	■ Allows a fine-grained analysis of temporal and sequential elements of modes and its relationship to other modes and contexts.
■ Ethnographic multimodal approach	■ Data through text and image	■ Recognizes how social actors use individual modes to construct higher-level actions.
■ Corpus-based approach	■ Orthographic transcription for speech ■ ELAN multimodal annotation environment ■ MUMIN coding scheme	■ Provides prosodic information such as main accent and intonation contour per intonational unit ■ Segments information based on gaze fixations and relevant areas of interests. ■ Annotates intonation units for turn management features

On the final note, transcription is anchored in the research purposes in addition to the modes of interests. These interests, nevertheless, are not constant for they may also be dependent on the data sets at hand. Furthermore, the researcher's orientation and stance must be considered for they serve as benchmark for the development of research questions. We suggest that careful formulation of the research aims as well as initial analysis of the data at hand would be helpful in deciding which transcription method to adopt.

5. Discussion

This article primarily aims to provide a comprehensive account of multimodal studies which encompasses the analysis of formulating research questions, selection of segments of videos to be analyzed and choosing methodological framework. The discussion also details the researcher's theoretical and methodological stance. Our goal is to provide an easy-to-read report for audience inside and outside the field of multimodality.

When a researcher foregrounds language as a dominant mode in the interaction, it discounts the meaning-making potential of other modes which may also play an integral role in the communicative act. The approach is leaning towards CA approach and the use of the Jeffersonian Transcription System that codes speech including paralinguistic features. Other approaches (e.g. Social Semiotics) aim to understand the relationships of multiple modes in a specific context as shaped by social and cultural practices. A multilevel transcription is adopted to produce transcripts that allow the examination of salient modes or even modes that are not foregrounded. The grid highlights the sequential and temporal unfolding of modes (Bezemer, 2014) including the overlapping use of these modes in the interaction. If the research purpose aims to uncover what mode is afforded in a specific time, multilevel transcription proved to be useful. Through this, we see how research purposes impact the employment of a specific transcription method.

Theoretical and methodological orientation emphasizes the researcher's stance when pursuing multimodal studies. As mentioned in the preceding section, it is important to be consistent with our stance in approaching multimodal studies. Each approach carries distinct principles along with the terms to refer to specific concepts (e.g. modes or semiotic resources and interaction or communication) and mixing them up may not sit well with experts from these fields. As novice researchers in the field, the results of this paper provided room to familiarize ourselves with existing approaches and their methodological processes. The experience of doing so shed light on our quest for understanding multimodal phenomena.

6. Future directions

Going back to the main purpose of this study which is to provide a comprehensive account of approaches to analyzing multimodal interactions, the results of this paper become the springboard for future research. The approaches outlined in this study as well as the multimodal method and analysis is useful for researchers in multimodal interactions. Further, a research project which employs various approaches to a specific data set may be conducted to see the commonalities and differences of the results. The idea may not sound convincing but in doing so, we contribute to the development of theories that govern studies in multimodality.

References

Bateman, J., Delin, J., & Henschel, R. (2004). Multimodality and empiricism. *Perspectives on multimodality, 6*, 65-87.

Bateman, J. A. (2014). Using multimodal corpora for empirical research. *The Routledge handbook of multimodal analysis*, 238-252.

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: A social semiotic perspective. *International Journal of Social Research Methodology, 14*(3), 191-206.

Bezemer, J., Murtagh, G., Cope, A., Kress, G., & Kneebone, R. (2011). "Scissors, please": the practical accomplishment of surgical work in the operating theater. *Symbolic Interaction, 34*(3), 398-414.

Bezemer, J. (2014). 14 Multimodal transcription: A case study. *Interactions, images and texts: A reader in multimodality, 11*, 155.

Bezemer, J., & Jewitt, C. (2018). Multimodality: A guide for linguists. *Research methods in linguistics, 28*, 1-3.

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative inquiry, 11*(2), 226-248.

Brône, G., & Oben, B. (2015). InSight Interaction: A multimodal and multifocal dialogue corpus. *Language resources and evaluation, 49*(1), 195-214.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., & Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics, 28*(3), 449-483.

Cowan, K. (2013). Multimodal transcription of video: Examining interaction in Early Years classrooms. *Classroom Discourse, 5*(1), 6-21.

Danielsson, K. (2016). Modes and meaning in the classroom–the role of different semiotic resources to convey meaning in science classrooms. *Linguistics and Education, 35*, 88-99.

Dicks, B., Flewitt, R., Lancaster, L., & Pahl, K. (2011). Multimodality and ethnography: Working at the intersection.

Eggins, S. (2004). *Introduction to systemic functional linguistics*. A&c Black.

Erickson, F. (1992). Ethnographic microanalysis of interaction. *The handbook of qualitative research in education*, 201-225.

Erickson, F. (2011). Uses of video in social research: A brief history. *International Journal of Social Research Methodology, 14*(3): 179–189.

Flewitt, R. (2006). Using video to investigate preschool classroom interaction: Education research assumptions and methodological practices. *Visual Communication, 5*(1): 25–50.

Flewitt, R. (2011). Bringing ethnography to a multimodal investigation of early literacy in a digital age. *Qualitative Research, 11*(3), 293-310.

Goldman, R. (2007). *Video representations and the perspectivity framework: Epistemology, ethnography, evaluation, and ethics*. NA.

Halliday, M. A. (1982). Linguistics in teacher education. *Linguistics and the Teacher*, 10-15.

Halliday, M. A. K., & Matthiessen, C. M. (1985). An introduction to functional grammar. Edward Arnold, London. *Australian Rev. Appl. Linguist, 10*(2), 163-181.

Halliday, M. A. K., & Matthiessen, C. M. (2013). *Halliday's introduction to functional grammar*. Routledge.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice. 2nd ed. Routledge*.

Haratyan, F. (2011). Halliday's SFL and social meaning. In *2nd International Conference on Humanities, Historical and Social Sciences* (Vol. 17, pp. 260-264).

ten Have, P. (2007). Doing conversation analysis. 2nd ed. Sage.
Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage Publications.
Hepburn, A., & Bolden, G. B. (2012). *The conversation analytic approach to transcription. The handbook of conversation analysis,* 57–76.
Hodge, R., & Kress, G. R. (1988). *Social semiotics*. Polity Press.
Jewitt, C. 2008. Technology, literacy and learning: A multimodal perspective. Routledge.
Jewitt, C. (2009). Different approaches to multimodality. Routledge.
Jewitt, C. (2012). An introduction to using video for research.
Jewitt, C. (2014). 12. Multimodal approaches. In *Interactions, images and texts* (pp. 127-136). De Gruyter Mouton.
Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge.
Knoblauch, H., Schnettler, B., Raab, J., & Soeffner, H. G. (Eds.). (2012). *Video analysis: Methodology and methods*. Peter Lang.
Kress, G. & van Leeuwen, T. (2001) Multimodal Discourse: The modes and media of contemporary communication. Arnold.
Kress, G., Jewitt, C., Ogborn, J. & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. Continuum.
Kress, G. (2003). *Learning to write*. Routledge.
Kress, G., Jewitt, C., Bourne, J., Franks, A., Hardcastle, J., & Jones, K. (2005). Urban English classrooms: Multimodal perspectives.
Kress, G. & van Leeuwen, T. (2006) Reading images: The grammar of visual design. 2nd ed, Routledge.
Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System, 81*, 26-38.
Lim, F. V., O'Halloran, K. L., & Podlasov, A. (2012). Spatial pedagogy: Mapping meanings in the use of classroom space. *Cambridge journal of education, 42*(2), 235-251.
Martinec, R. (2004). Gestures that co-occur with speech as a systematic resource: The realization of experiential meanings in indexes. *Social Semiotics, 14*(2), 193-213.
Maybin, J., & Tusting, K. (2011). Linguistic ethnography. In *The Routledge handbook of applied linguistics* (pp. 535-548). Routledge.
Mondada, L. (2008, September). Using video for a sequential and multimodal analysis of social interaction: Videotaping institutional telephone calls. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 9, No. 3).
Norris, S. (2002). The implication of visual research for discourse analysis: Transcription beyond language. *Visual communication, 1*(1), 97-121.
Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge.
Norris, S. (2011). *Identity in (inter) action: Introducing multimodal (inter) action analysis* (Vol. 4). Walter de Gruyter.
Norris, S., & Jones, R. H. (2005). Introducing mediated action. In *Discourse in Action* (pp. 29-31). Routledge.
O'Halloran, K. L. (2011). Multimodal discourse analysis. *Companion to Discourse. London and New York: Continuum*.
O'Halloran, K., & Smith, B. A. (Eds.). (2012). *Multimodal studies: Exploring issues and domains* (Vol. 2). Routledge.

Pahl, K. 1999. *Transformations: Meaning making in nursery education*. Trentham Books.

Pirini, J. (2014). 9. Introduction to multimodal (inter) action analysis. In *Interactions, images and texts* (pp. 77-92). De Gruyter Mouton.

Price, S., & Jewitt, C. (2013, February). A multimodal approach to examining'embodiment'in tangible learning environments. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 43-50).

Schegloff, E., Jefferson, G., & Sacks, H. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome research, 13*(11), 2498-2504.

Taylor, R. (2014). *Meaning between, in and around words, gestures and postures–multimodal meaning-making in children's classroom discourse*. Language and Education, 28(5), 401-420.

Taylor, R. (2016). *The multimodal texture of engagement: Prosodic language, gaze and posture in engaged, creative classroom interaction*. Thinking Skills and Creativity, 20, 83-96.

Thuy, T. T. H. (2017). Reading images-the grammar of visual design. *VNU Journal of Foreign Studies, 33*(6).

Van Leeuwen, T. (2005). *Introducing social semiotics*. Psychology Press.

Zacks, R., & Tversky, D. (2001). Event structure in perception and conception. *Psychological Bulletin,127*, 3-21.