Linguistics and Education xxx (2019) xxx–xxx

Contents lists available at ScienceDirect

Linguistics and Education

journal homepage: www.elsevier.com/locate/linged

Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks

Maria Grazia Sindoni *

University of Messina, Italy

ARTICLE INFO

Article history:
Received 11 November 2018
Received in revised form 25 February 2019
Accepted 22 May 2019
Available online xxx

Keywords:
multimodality
video-mediated interaction
multimodal transcription and annotation
language ideologies
trancription and annotation pedagogy
mode-switching

ABSTRACT

Digital environments have shaped unprecedented configurations of spontaneous interactions (Herring, 2013), thus giving rise to emergent patterns of language variation. The borders between spoken and written language have been blurred by the interplay of semiotic resources and medium affordances (Kress & van Leeuwen, 2001), and to the point that traditional ideas about spoken/written variation – with related implications in language teaching/learning – have been challenged.

An example of language variation has been labelled *mode-switching* (which paraphrases *code-switching*, cf. Heller (1988), and draws on Halliday's register component of *mode*, 1978) that describes the alternation of speech and writing in the same communicative event in video-mediated communication (Sindoni, 2013, 2014a, 2019b). Previous findings show that mode-switching is mainly used to manage the flow of conversation and for self-repair in multi-party interactions, whereas patterns and use of mode-switching are still unmapped in one-to-one interactions.

Drawing on a corpus of multimodal data (i.e. 48 Skype video-calls involving peer interaction between postgraduate students and friends and/or relatives), examples of *mode-switching* will be presented in one-to-one interactions by expanding on conversation analysis theories (Sacks, Schegloff, & Jefferson, 1974). Assuming that mode-switching is not used to hold the floor in one-to-one contexts, other creative uses will be documented, for example with reference to medium affordances, speakers' interpersonal attitudes, and interplay with other semiotic resources (e.g. layout, visuals, etc.). Discussion of data will include excerpts of multimodal transcription of video-mediated interactions produced by students to highlight how transcription can be pedagogical and helpful in understanding how students heuristically make sense of these video-mediated events and establish priorities in their reconstructions of video data.

In conclusion, the productivity of these patterns in spontaneous video-mediated communication calls for a systematic theoretical reflection and subsequent incorporation in teacher training, syllabus design, and institutional learning contexts.

1. VMI and multimodality: background and research questions

Research on video-mediated communication (VMC) has been exponentially expanding in the last two decades. Paulus, Warren, and Lester (2016) suggest that research on VMC has been mainly enucleated around three broad theoretical stances. The first is informed by language-based methodologies (such as linguistics, see Georgakopoulou, 2011 and Zappavigna, 2012; sociolinguistics, see Androutsopoulos, 2006 and Thurlow & Mroczek, 2011; pragmatics, see Herring, Stein, & Virtanen, 2013 and Yus, 2011; and conversation analysis, see Meredith & Potter, 2013). The second strand adopts social-ethno-anthropological methodologies, such as interactional sociology (Goffman, 1981), micro-ethnography (Erickson, 2004), discursive psychology and health studies (Antaki, Ardévol, Núñez, & Vayreda, 2005; Flinkfeldt, 2011; Flinkfeldt, 2014; Sneijder & te Molder, 2005). The third "mixed" strand is informed, among others, by interactional sociolinguistics (Gumpertz 1999), linguistic anthropology (Duranti, 1997), and linguistic ethnography (Creese, 2008),

Early research on VMC focused on task-oriented interactions in the workplace (Heath & Luff, 1992) and in educational contexts (Swan et al., 2008). More recently, research started to explicitly focus on educational environments from a multimodal perspective, for example analysing the influence of task type on the use of

* Correspondence to: Dipartimento di Civiltà Antiche e Moderne, Polo Annunziata, Via Palatucci, 98168, Messina, Italy.
E-mail address: mgsindoni@unime.it

https://doi.org/10.1016/j.linged.2019.05.004
0898-5898/

communication strategies in a 3D virtual environment with English as a Lingua Franca (Shih, 2014), but also comparing face-to-face interactions with video-mediated interactions, starting in the early 1990s (O'Malley, Langton, Anderson, Doherty-Sneddon, & Bruce, 1996), and focusing on the use of spoken language (O'Conaill, Whittaker, & Wilbur, 1993), on temporal differences in interactions in team work (van der Kleij, Paashuis, & Schraagen, 2005), on the different form of mutual elaboration, coordinated action in multimodal environments (Gardner & Levy, 2010) and orchestration (Groen, Ursu, Michalakopoulos, Falelakis, & Gasparis, 2012), and on the role of video materials in EFL classrooms (Bajramia & Ismailia, 2016). Other studies have explored media affordances by investigating online live streaming of video games (Recktenwald, 2017) and multiple engagements in VMC contexts (Rosenbaun, Rafaeli, & Kurzon, 2016a, 2016b), with the aim, among others, to investigate how cross-modal communication between broadcaster and audience can be reproduced, i.e. transcribed, accounting for the unfolding of activities over time.

Most methods prioritise language as the fundamental – or in some cases *the* – resource for communication. The tradition of studying interaction, and overall communication, as equated with language is apparent in several epistemological disciplinary practices, for example the habit of using language-derived terminology to describe non-linguistic phenomena and systems (e.g. *non*-verbal, *para*-language, *body* language, etc.), thus making them explicitly derivative or secondary.

The first multimodal approaches to communication were informed by linguistic theories, such as systemic-functional grammar (Halliday, 1978), as Kress and van Leeuwen (2001 and 2006) designed their model by turning Halliday's ideational, interpersonal and textual metafunctions of language into representational, interactive and compositional metafunctions to "read images". From this initial mapping, many other theories and methodologies have been developed within multimodal studies, mainly drawing on socio-semiotics (Hodge & Kress, 1993; van Leeuwen, 2005), on SFL-informed theories, such as multimodal discourse analysis (O'Toole, 2004; O'Halloran, 2004), interactional studies, such as multimodal interaction analysis (Norris, 2004), or integrating corpus studies, with multimodal corpus linguistics approaches (Baldry & Thibault, 2008; Bateman, 2014).

Multimodal studies initially originated from the attempt to redress the balance between language and *other* semiotic systems and resources (e.g. images, music, design of objects or buildings, body movements, gaze, social distance, etc.) in the analysis and interpretation of meaning-making processes. Much of the research in multimodality thus strived to get rid of rigid language-derived epistemologies, in the form of a plethora of *other grammars*, for example grammar of music and voice (van Leeuwen, 1999, 2005), grammar of displayed art (O'Toole 1994 [2011]), grammar of the built environment (O'Toole, 2004), grammar of performing arts (Sindoni et al., 2017), etc. The new developed grammars had the ambition of unearthing new, possibly formerly ignored, kinds of data, and that was obviously not the case for language.

However, multimodality can be applied to language-based disciplines, such as conversation analysis Sacks et al., 1974, to enhance and fill the gaps left open by "pure" linguistic analyses. This paper therefore adopts an approach that combines multimodality with basic principles from conversation analysis, with the aim of showing the practical benefits of adopting transcription and annotation task in teacher training syllabi. The development of a metalanguage is here considered as a central topic for linguistic research, as it "reflects metapragmatic awareness, a crucial force behind the meaning-generating capacity of language in use" (Verschueren, 2000: 439). Metalanguage is a necessary dimension rather than an object of language use, thus, very broadly, metapragmatic awareness is also crucial to understand other people's viewpoints and help working collaboratively with others (Tomasello, 1999). From a different theoretical standpoint, critical multimodal skills are the ensemble of critical and socio-semiotic abilities that allow any individual to understand how meaning-making works in texts (be they verbal, visual, aural, etc.). The reflexive properties of language itself therefore allow individuals to position themselves and the activities they are engaged with, both indexically and meaningfully. But how can these skills be thought of and then taught?

The specific contribution that multimodal studies and socio-semiotic approaches to communication have brought about in VMC research can thus be thought of as a twofold epistemological process: (1) de-emphasising the prominence of verbal language and consequently (2) addressing the issues of transcription and annotation in video analysis.

This paper is an attempt to frame the significance of linguistic phenomena within a multimodal perspective that is not covered in the approaches mentioned beforehand, with the aim of filling the gap between epistemologies and actual data sampling and analysis in VMI. To this end, some case studies will be presented and discussed by showing that a "purely" linguistic phenomenon, that is spoken and written variation in video-mediated communication, can be studied within broader socio-semiotic and multimodal approaches by focusing on a training-oriented perspective. Even though students were trained in recognising and describing all resources in the interaction (e.g. language, gaze, kinesics, proxemics, etc.), this paper will exclusively focus on language patterns as they emerge from multimodal interactions to show how multimodal transcription tasks (i.e. transcription for language and annotation for other resources) can help students to put language in perspective. For students' analyses of resources other than language, previous research has been carried out (Sindoni 2019a and 2019b) and goes beyond the scope of this paper.

In this case study, the selection and analysis of examples of *mode-switching* (alternation between speech and writing in video communication) will document: (1) theoretical implications from a linguistic perspective, for example in terms of this pattern's productivity and in comparison with code-switching; (2) theoretical implications from a multimodal perspective, and, more specifically, in how language can illuminate overall video interactions, and finally (3) educational implications by analysing students' multimodal transcriptions, annotations and analysis of their own video conversations.

The pattern under investigation is what I have defined in previous studies as *mode-switching* (Sindoni, 2012a, 2013, 2014, 2019b), that is a change from speech to text-based chat in VMC, even though we are here referring to the *materiality* of speech and writing and not to spoken-like and written-like features of language, as research literature shows that digital languages and registers tend to blur the traditional boundaries of the two language modes, i.e. speech and writing, along a cline. Hence, we will not consider whether writing in VMI exhibits features of speech in these environments, but will exclusively focus on the material transition from "vocal speech" to written text input in the chat box, as material affordances of speech and writing are here considered to be fully meaning-making.

Mode-switching (MS henceforth) paraphrases the linguistic pattern of code-switching and describes alternation from speech to writing and *vice versa* in the same communicative event, with a general prevalence of the spoken mode with written insertions for specific communicative purposes. Initially, this pattern was context-specific to some media platforms, such as Skype or MSN, but, as technological affordances have shifted over time, a wealth of other environments, including mobile communication (Facetime, WhatsApp) have allowed the proliferation of this language pattern (e.g. text-based messages plus voice memos in WhatsApp).

In multiparty-video interactions, MS is mainly used in three communicative areas: (1) management of the conversation flow

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

(e.g. repair, self-correction, etc.), (2) specific communicative purposes (e.g. secrecy, information value, etc.), and (3) address technical issues (e.g. communication breakdown), see Sindoni (2011, 2013, 2014, 2019b). With reference to turn taking and management of conversation flows, quantitative and qualitative evidence show a consistent use of MS to ask for, or manage, parallel and/or multiple floors of conversation, repair, signal transition relevance places when other non-verbal cues were unavailable or absent. In area 2, MS is used for interpersonal communicative purposes, such as secrecy (e.g. participants bypassing overhearers) or information value (e.g. to provide addresses, phone numbers, web links, etc.). In area 3, speakers usually resort to writing when technical issues arise (e.g. reduced bandwidth and consequent video stream slow down).

Multiparty video environments (for example Camfrog) are usually chaotic and communication breakdown is a highly likely occasion, as a typical configuration of these environments incorporate multiple participants with high competition to hold the speaking floor, so that they need to use all the resources at their disposal to successfully interact. Since participants frequently log in and out from these platforms, participatory configuration is frequently changing, thus modifying at the same time how single participants interact with both the environment and the others. In other words, the difference between ratified (official) and unratified (unofficial) participants that follows Goffman's participation framework (1981) is less relevant on these platforms, as the classic subdivision between *addressed* and *unaddressed* (i.e. ratified) and *eavesdroppers* and *overhearers* (i.e. unratified) is looser and more easily compromised. This is so because participants may take up several roles with different sets of participants at the same time (for example talking to a group and writing to another group, see Sindoni, 2013).

Conversely, the typical configuration of one-to-one private video-meditated interactions only involves two ratified participants with the random engagement of unratified participants (e.g. siblings, roommates, friends) that may occasionally appear onscreen. As a consequence, the conversation flow is generally smoother and easier to manage, and speakers do not need to co-deploy speech and writing at the same time for mere conversation management.

Assuming that in one-to-one video-mediated contexts the conversation flow is smoother and taking turns is less problematic than in multiparty video-mediated platforms, we may likewise assume that different uses of mode-switching will be detected. In this paper, some examples of MS in one-to-one Skype video calls will be analysed with the aim of showing how these can be described and interpreted by both myself as a multimodal analyst and by students, ultimately suggesting MS productivity and the need to describe it, especially in educational contexts and for training-oriented courses. Multimodal transcriptions and related comments by the involved participants will be commented upon to understand the participants' own perception of their own interactions.

The following research questions will be thus addressed:

1. How can MS be described and interpreted in one-to-one VMI? When MS is not necessary to manage conversation flow and/or handle technical issues, which are its communicative functions?
2. By discussing the students' multimodal transcriptions, annotations and reflections of some excerpts, what kind of data is it possible to obtain, and how can these tasks be useful in educational contexts, with specific reference to increasing students' awareness in terms of language use and metapragmatic skills in digital environments?

A qualitative case study approach is here adopted to explore the addressed research questions to highlight the reasons why MS is used by participants and how this process is interpreted by students in their analyses. In Section 2, methodology and data will be presented, in particular discussing how video data have been gathered and analysed, and the limits of the adopted approach. Section 3 will illustrate and discuss three examples from the video datasets, while Section 4 will draw some conclusions by pointing at future lines of research in this field.

2. Method and video-datasets

The video datasets used for this study include 24 Skype video calls for a research project, "MoM: Multimodality on the Move", which experimented a joint syllabus in four University degree courses.[1] The joint syllabus was grounded on a core theoretical module in multimodal studies and taught by four researchers in their home universities. In the secondary module, workshops and tutorials were devoted to different digital text types, i.e. video-mediated interaction, blogs, "about us" webpages, fanvids and mash-up videos. All students were involved in learning the same contents and practising the production and analysis of all these texts in their home universities.

The aim of the syllabus was to help students achieve advanced skills in multimodal digital literacy on contemporary texts in English from socio-semiotic and intercultural standpoints. As part of their assignments, students had to produce their own digital texts among the different taught genres, transcribe their text and discuss text and transcription in a paper. Students produced and collected materials that covered different areas for the development of wide-ranging abilities to produce, critically understand and interpret digital texts.

The datasets presented here are limited to students' production and analysis of VMI. In the module, students were taught how semiotic resources come into play in VMC (such as language, gaze, kinesic and proxemics patterns, etc.) and were trained in basic transcription methods combining conversation analysis and multimodality. The students' final assignment for this text type included:

1) a recorded video conversation with an English native speaker, relative or friend;
2) a manual multimodal transcription and annotation of a short chunk of the video-conversation by using a pre-structured transcription grid;
3) a c. 2000-word academic paper;
4) a peer-assessment grid to anonymously evaluate another student.[2]

The assignment for video-mediated interaction included the recording of a "semi-spontaneous" video-conversation (as students already knew the category they had been asked to scrutinise), the multimodal annotation and transcription task for a fine-grained analysis of the interactional event, a piece of academic writing showing their metapragmatic reflection on the video interaction, and finally, a peer-assessment grid in which they anonymously evaluated another student with the same assignments.

[1] The MoM project involved four Italian universities, namely Messina (MA in Foreign Languages and Literatures), Chieti-Pescara (MA in Foreign Languages and Literatures), Rome Tor Vergata (BA in Languages in the Information Society), Firenze (MA in Public and Political Communication Strategies) in 2014–2015. The project has now been turned into an international project, funded within an Erasmus + "EU-MADE4LL, European Multimodal and Digital Education for Language" Project website: www.eumade4ll.eu.

[2] Not analysed in this paper.

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

For this paper, the multimodal digital video datasets include a *selection* from:

1. Twenty-four recorded Skype one-to-one video calls involving forty-eight primary ratified participants (not including unratified participants);
2. Twenty-four multimodal transcription grids where students manually annotated use of resources, such as spoken and written turns, MS, gaze, kinesics and proxemics;
3. Twenty-four academic papers.

In this paper, a qualitative case study methodology has been adopted. Qualitative research recognises the unavoidable subjectivity of the researcher and foregrounds plurality of perspectives (here represented by the dual perspectives of researcher and students on the same materials), without rejecting some notion of objectivity (Baxter & Jack, 2008). A case study approach is useful when exploring research questions dealing with *how* and *why* some phenomena occur and the research focus of this study is devoted to (1) an exploration of the involved interactional processes by considering contextual conditions, where the boundaries between context and phenomena are blurry, along the lines discussed by Yin (2003). As our case study approach is based on the collaboration between the researcher and the participants (Crabtree & Miller, 1999), the dual perspective of researcher and students will be provided to qualitatively explore the students' reflexive and critical perspectives on their learning experiences, by taking into account their matapragmatic skills shown in both transcription tasks and their analyses.

With the aim of avoiding the common pitfall associated with a case study approach, that is to try to answer too broad research questions, Yin (2003) and Stake (1995) have suggested that researchers bind a case study by placing boundaries, for example (a) by time and place (Creswell, 1998); (b) time and activity (Stake, 1995); and (c) by definition and context (Miles & Huberman, 1994). The selection was thus made by singling out the cases on the grounds of definition and context, and, more specifically, in which (1) mode-switching was present in the interaction, directly annotated by the student in the transcription grid and explicitly commented on in their written analysis; (2) students decided to edit the provided materials, such as the transcription grid, to accommodate the transcription task to the interaction; (3) the presence of patterns of use for MS other than those already explored in previous research and (4) different degrees of metapragmatic awareness shown in their analyses.

All participants were given an information sheet about the research purposes of the task and project's references. They all signed a consent form allowing us to use all the recorded materials and were asked to inform and have all the participants sign the consent form before recording the conversation. Students were given instructions in class suggesting the download of free technical tools to record the video conversation and listing possible topics of informal conversation and some classes were devoted to the ethical implications of recording and studying human subjects. They were free to record an interaction of any length, but to select and transcribe a clip of maximum 3 min. Submission included the overall conversation with a clear indication of the transcribed segments.

The transcription grid used by students has been adapted from previous research (Sindoni, 2013, 2014) and is a manual time-based (first column) model that can be used in educational contexts for training purposes.[3] The columns respectively give participant's name or nickname, spoken and written turns, MS (+ if present). The other columns are used to annotate gaze directionality, kinesics patterns, posture, proxemics patterns – that are produced not by the distance participants put between themselves and others, but by the distance projected by the webcam's position and themselves, and finally a visual screenshot, usually representing secondary participants (SP henceforth, i.e. relatives or friend recruited in the project by SCs), as student collaborators (SC henceforth) were recording the interaction.

For the following discussion, a caveat is needed: as all students producing a video-recorded conversation were aware of the research purposes, their behaviour cannot be considered as spontaneous (Yin, 2003) and will be backgrounded, for example they were heavy self-initiators of MS, as they knew that this language pattern was being investigated in the project. Building on previous research, these students are here considered as-> SCs, as they were not completely spontaneous in their interactions. However,-> SCs had not been informed about the specific research contents explored.[4]

Finally, the dual perspective, one of the author of this study, and the other of the students participating in the project will be given in this paper by following point 4 above, hence starting with a case study with a student who gives evidence of significant metapragmatic and reflexive skills in her language use and then following with two case studies with participants who show less advanced skills, but whose interactions and assignments nonetheless fall in the categories 1–3 discussed above.

3. Mode-switching in VMI and students' transcriptions and reflections

In this Section, some examples of MS will be presented by following a case study approach as motivated in the previous Section, and commented to show how this language pattern exhibits productivity in digital environments, by drawing on VMI produced by students and including a meta-analysis of the students' transcriptions grids as well as commenting their reflections about their own work. The three selected cases include some patterns not discussed in previous research, as they do not fall in any of the categories described.

The video dataset in this study is considerably smaller than other datasets for general explorations of MS (Sindoni, 2013), as it includes only 7:23 h of recorded material, so that a manual counting and close analysis of all occurrences has been carried out, considering that each MS counts as one occurrence, because MS is not always taken up by the other participant (i.e. self-initiated MS counts as one MS, self-initiated plus other-initiated MS count as two MS). 50% of participants mode-switched and this rate is considerably higher than what was detected before (Sindoni, 2011, 2013, respectively 11% and 15%, with a dataset including more than 700 h of recorded material). This can be explained by the fact that SCs were ready and willing to prompt MS so as to have the chance to describe a relevant category for their assignment. This observation is supported by the fact that 33% of occurrences have been initiated by SCs, who are by and large the highest MS self-initiators in this dataset.

Narrowing down our focus only to video calls that included at least one example of MS, 83% of prompts by changing mode resulted in a response via another MS, whereas in the remaining 17% SPs did

[3] For a full description of this model and its educational implications, see Sindoni (2013).

[4] All participants signed a consent form allowing recording and future use and publication of the materials, including screenshots showing what appears onscreen, so that all participants were aware of being recorded thus influencing to some extent their behaviour. However, research has shown that self-awareness tends to loosen at some points in conversation, and that independently from context, self-monitoring can be assessed in individuals using scales of measurement (Snyder and Gangestad, 2000).

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

not react to the prompt. This means that, when prompted, participants mostly tended to respond by accommodating to the other participant's channel of communication. Furthermore, the predominant channel used in VMI tends to be spoken, with short insertions of written turns, and only when specific conditions occur.

3.1. Case study 1: Can speech and writing overlap?

In the first example (extract length: min 12:23), Susan is a SC enrolled in MA in "Foreign Languages and Literatures" at the University of Messina, and majoring in linguistics. She involved a SP, Chris, a relative from Chicago. Susan self-assesses as a bilingual, even though she was born in Italy, and did not speak English with the same frequency as Italian, whereas Chris speaks American as a native speaker. Fig. 1 presents an unedited excerpt taken from Susan's manual transcription.

In Fig. 1, two occurrences of MS can be detected by the presence of two +, respectively in rows 5 and 7. Both are self-initiated by SP Chris. When the conversation is interrupted by Chris' dogs barking, Susan asks about their names and Chris feels the need to answer by keeping the spoken floor while at the same time writing in the text chat the dogs' names, with a self-initiated MS. It emerges from the transcription that SC Susan perceives Chris' two turns (spoken and written) as perfectly overlapping (see line 6, Fig. 1), where she transcribes in the same row both the spoken and the written turn despite the fact that she was not specifically taught about how to transcribe this specific turns' overlap.

When Susan asks the question, Chris first starts writing, then pauses for less than 1/3 of a second, then says "Brutus-Brutus and" and goes on writing "Bella", and finally looks up at the screen and hits the send button, so that the message appears onscreen after ½ of a second.

A transcription by the analyst without the grid format would have the following:

Chris:	Those are my (turns on the left).. those are my dogs.. barking
Susan:	yeah.. hello!
Chris:	(grins)
Susan:	what are..(grins) what are their names?
Chris:	(starts typing) Brutus-Brutus and (stops and types in) Bella (hits the send button)
Susan:	.. of your dogs?

(Susan looks rightward down the screen as the message appears onscreen).

Considering that the transcription assignment's goal was not to train students in CA, but to make them aware of the orchestration of the different semiotic resources in VMC, and that they were all novice multimodal analysts, Susan's transcription is a reasonable approximation and, as such, a good representation of the exchange. Her transcriptional solution consists in using the same row to report on spoken and written turns and signalling the occurrence of MS.

This is an example of the most frequent use in the video datasets here presented, that is *precision*, or to avoid misunderstanding or possible trouble in conversation. However, we may well wonder if using speech and writing at approximately the same time can be considered as a typical example of MS, which most frequently signals discrete turns uttered in spoken or written mode.

Susan herself problematises this specific occurrence and comments as reported below, unedited and with my italics:

> This chunk of conversation shows different interesting points. First and foremost, it is an example of self-initiated mode-switching. As we can notice from the repetition of many words and continuous overlapping, interaction between the two interactants is interrupted by Chris' dogs. Thus, when Susan asks her what were her dogs' names, Chris decides to mode-switch in order to give a precise information and, at the same time, be sure that her partner understands her words, in spite of her dogs' barking. *Actually, this is a controversial matter, as she says simultaneously what she writes, hence, we might claim that she does not change her channel of communication but she adds another one*. In this case, the alternation of speech and writing is used as a strategy both to avoid misunderstandings and repair a trouble in conversation. Furthermore, at the minute 02:53,12 Chris returns to converse orally by performing another self-initiated mode-switching, since she realises that Susan has correctly grasped her words.

Interestingly, Susan is able to recognise the atypical occurrence of MS: on the one hand, she categorises it by following the taxonomy she had been taught, and she omits (or backgrounds) the fact that Chris may have used MS to help Susan's understanding. Chris' MS is instead explained by Susan as an attempt at repairing problems due to the noise of barking dogs. On the other hand, Susan correctly understands and motivates her decision to represent Chris' spoken and written turns as overlapping in the transcription grid, by using the same row instead of two discrete rows. Her explanation is revealing in that she finds problematic to equate overlap (what she calls "simultaneity") with MS, so that she thinks that she needs to distinguish the overlap of speech and writing from the "orthodox" MS, which describes "alternations".

This may be explained as MS is a descriptive label modelled on the linguistic phenomenon of code-switching, which is the use of more than one language or variety within a single conversation, consistently with phonology and syntax of the varieties used. Without going into details about the different theories that explain and describe different types of code-switching,[5] we may argue that the nature itself of code-switching cannot, in theory, produce *simultaneity* or *overlap* of occurrence between two or more language varieties. The most common types of code-switching, such as inter-sentential, intra-sentential, and intra-word (Li Wei, 2000; Myers-Scotton, 1989) describe patterns of alternations, for example outside or within the sentence, at clause or tag level, or within a single lexical item, respectively. MS itself has been described as "the alternation between speech and writing in video conversations" (Sindoni, 2013: 66), making it clear that it refers to a *synchronic* use, that is within the same communicative event.

It comes as no surprise that the SC Susan realises that the detected pattern is "controversial" and cannot be interpreted as an "alternation" *stricto sensu*, but, in her own words, as an *addition* of another channel of communication.

In this first example, even though motivation/reason for use falls quite fittingly within the taxonomy of "precision" that the SC was taught and then able to identify, the observed pattern of occurrence at the intra-sentential level deviates from an unmarked use, thus resulting in an interlaced pattern of MS use. Therefore if an explanation for use is rather straightforward, the same cannot be said about a pattern of use that shows productivity and challenges the student's analytical categories. Susan's background in linguistics may have helped her in a fine-grained analysis focused on interpreting data with a view to identifying discrete linguistic phenomena, possibly other than those she had been already taught.

3.2. Case study 2: Mode-switching vs. code-switching

In the second example (extract length min 4:24), a different use of MS will be illustrated, coupled with uses of code-switching. In the excerpt reproduced in Fig. 3, the transcription grid is produced by Irene, an Italian student, enrolled in MA at the University of Firenze in Public and Political Communication Strategies. SC Irene

[5] For the interpretation of code-switching within a Markedness Model, see Myers-Scotton (1993), for sequential analysis, see Auer (1998).

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

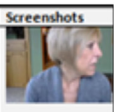
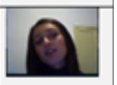
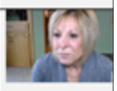
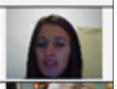
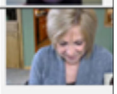
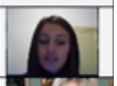
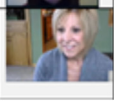
	Time	Participant	Spoken turns	Written turns	Mode-switching	Gaze	Kinesics	Proxemics	Screenshots
1	02:34,29	Chris	those are my – those are my dogs... barking		-	Looks on the right and then at screen	Turns head on the right Smiles	Medium shot Colored wall, piece of furniture, part of door, part of chair and other objects in the background	
2	02:37,08	Susan	yes, hello. What are – what – haha, what are their names?		-	Looks at screen	Small head movements, moves backwards and forwards, smiles	Close shot White wall and part of door in the background	
3	02:42,06	Chris	Bru..		-	Looks at screen	Stands still	Medium shot Colored wall, piece of furniture, part of door, part of chair and other objects in the background	
4	02:45,26	Susan	of your dogs...oh ok		-	Looks at screen	Stands still Small head movements	Very close shot White wall and part of door in the background	
5	02:46,07	Chris	brutus – brutus – brutus and Bella	brutus and bella	+	Looks at keyboard then at screen	Writes	Medium shot Colored wall, piece of furniture, part of door, part of chair and other objects in the background	
6	02:50,02	Susan	oh, brutus and bella		-	Looks at screen	Small head movements, up and down	Very close shot White wall and part of door in the background	
7	02:53,12	Chris	yes		+	Looks at screen	Smiling Small head movements	Medium shot Colored wall, piece of furniture, part of door, part of chair and other objects in the background	

Fig. 1. Susan's manual transcription and annotation (*unedited*).

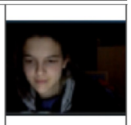
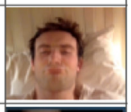
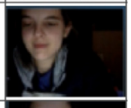
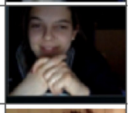
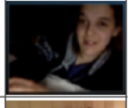
	Time	Participant	Speech	Writing	Mode-switching	Code-switching	Gaze	Kinesic action	Proxemics patterns	Visual units
46	18:19 - 4s	Irene		povero	+	+	Looks at the screen and then looks down to the keyboard.	Waits for the participant to talk with her left hand touching her head. Writes on the keyboard.	Only head visible.	
47	18:23 - 5s	Fred	//zee//		-	-	Looks at the screen.	Reads what's written. Sad grimaces.	Only head visible.	
48	18:28 - 5s	Irene		ill be there so soooooooon	-	+	Looks down at the keyboard.	Moves backwards, writes something.	Goes out of the camera for a second. Head with no hair visible.	
49	18:33 - 1s	Irene	//ehm.. did you like that?//		+	-	Looks at the screen.	Smiles and touches her hands.	Half head and both hands visible.	
50	18:33 - 1s	Fred		zee	+	-	Looks at the keyboard.	Writes something.	Only head and hair visible.	
51	18:34 - 5s	Fred	//zee//		+	-	Oblique gaze.	Still.	Only head and hair visible.	
52	18:39 - 2s	Fred		zooon	+		Oblique gaze	Writes something.	Only head and hair visible.	
53	18:41 - 5s	Irene	//zoooon…zee//	zoon	+	-	Looks down and back to the screen.	Writes, laughs and moves to her right.	Head and part of the arms visible. Very dark.	
54	18:46 - 2s	Fred	//zee//		+	-	Looks at the screen.	Smiles.	Only head visible.	

Fig. 2. Transcription grid by SC Irene.

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

involved Fred, her British boyfriend living in London, in the video conversation. The whole interaction is characterised by a number of MS and code-switching, and, quite strikingly, Irene feels the need to update the transcription grid by adding a further column, which she labels "code-switching" (see Fig. 2). Eleven occurrences of code-switching, out of 71 transcribed turns, may have prompted her to categorise the phenomenon. Both participants in the conversation code-switch (11) and mode-switch (7) several times, for reasons that can be discussed by taking as an example what is shown in Fig. 2.

SC Irene and Fred are talking indirectly about their long distance relationship. Fred is more explicit in expressing his wish to see her soon, but Irene is clearly embarrassed at his indirect request to reciprocate his feelings and decides to mode-switch (self-initiated), writing "Ill be there so soooooooon". Additionally, Irene herself describes Fred's facial expressions accompanying his words as "sad grimances".

SC Irene systematically resorts to writing when she is embarrassed and at a loss for words, as she was answering Fred's request with the Italian *povero*, i.e. *poor thing*, by a self-initiated MS, in response to Fred's kinesics patterns, humorously (thus minimising his wilful self-exposure) communicating loneliness and longing. Other occurrences of MS indicate a personal and familiar play with words or maybe sounds, as *zee* may be interpreted as a way Fred uses to imitate the Italian pronunciation of *sì*, affirmative particle for *yes* and *zoon*, possible imitation of *soon*. They amusingly play with sounds, but using a language that is available only to them, as Irene herself explains in her written assignment (unedited): "In turn 47, 50, 51, till turn 54, they both use the words 'zee' and 'zoon' that do not belong to any language and just represents their particular use of language" (see Jørgensen, 2008 for polylanguaging).

In the excerpt discussed above, occurrences of code-switching and MS can be interpreted along similar lines: both participants do not need to switch to another language or mode (spoken to written and vice versa) for a specific and referential reason. There is no need to manage the conversation flow or repair, no technical issue. Their code-switching and mode-switching are interlaced and display affective functions similar to those described in research literature for code-switching (Holmes, 2001). Code-switching in fact refers to *choice*: it does not occur when the speaker needs to change because she does not possess linguistic resources for communication nor when there is no word available in the language. Applying the same interpretative rationale to MS, we could likewise claim that it can be considered as an *autonomous linguistic phenomenon per se* only when it originates from reasons other than technical issues, such as when participants in a video conversation cannot hear each other or when the line is down.

In this example, affective functions imply the subtle and co-negotiated constructions of turns that help Irene and Fred build a solidarity network contributing to a sense of intimacy, therefore overcoming the issues of a long distance relationship and of cultural barriers, such as speaking different native languages. Building intimacy maybe be, in such conditions, particularly challenging for participants, who actively cooperate to establish a common code of communication.

Irene, however, seems to capture only a part of this complex relationship dynamics in her analysis (unedited, but with emphasis added):

> Mode-switching is less frequent than code-switching and the participants only change mode at the end of the conversation. [. . .] This means that they are not mode-switching for a specific reason, such as bad connection, or because there are other people in the room, *but simply to bring the other participant's attention to the screen and to their conversation by making them read instead of only listening.*

This means that while Irene is able to understand that there is no "specific reason" to mode-switch, she does not go beyond a mere descriptive explanation, specifically that they use MS to bring the other participant's attention to the screen by having them reading. While she correctly understands that the detected occurrences of MS cannot be clarified in technical or referential terms, as they cannot be easily categorised, she seems to lack the analytical skills to embark on a possible alternative or subtler explanation.

On another level, looking at row 53 in Fig. 2, the same phenomenon found in the example discussed beforehand by Susan can be detected: simultaneity of occurrence of the spoken and written turn is likewise recognised and identified by Irene, who utters *zoon* (another example of their "particular use of language"), while writing it at the same time. However, unlike Susan, Irene does not mention this language use in her analysis, and this could be explained, among many possible reasons, to the different academic background and consequent little training in linguistics (even though Irene's BA was in Foreign Languages and Literatures).

3.3. Case study 3: Writing is multimodal per se

A third example shows other patterns of MS (extract length, min. 5:40). The SC Raymond is a self-assessed bilingual speaker (Canadian-American and Italian) majoring in linguistics at the University of Messina. SP Amanda, Raymond's cousin, was born and raised in Canada. During their conversation, they discuss the possibility for Amanda to move in with her boyfriend Matt in Toronto. However, as Raymond immediately understands from her hesitations, her parents' cultural background has had an impact on her choice to postpone living with Matt.

Relevant exchanges are reproduced by the my transcription in Fig. 3, even though the sequence has been summarised and some turns omitted.

In the chunk of conversation selected for analysis, Raymond asks several questions to understand the real reasons why Amanda is not moving in with Matt. Amanda seems to be unwittingly admitting that there are other reasons behind lack of money and her boyfriend still studying. Raymond asks about her mother's reaction at the possibility of Amanda moving in with Matt, so that she finally reveals that her family is not backing her up, in fact her mother's reaction is a "No" that Amanda exclaims and then writes down as "No.", by a self-initiated MS. She reinforces the message and mimics the act of writing "No" and then significantly adds "With a period!" to highlight the definitive and unquestionable nature of her mother's response. When enquired about her father's opinion on the matter, Amanda admits that his reaction was even worse and says "No!", then, mode-switching once again to emphasise the illocutive power of the message, types in "No.!". Her father's reaction is even made graphically more vivid, as the exclamation point is preceded by a period, so that Amanda is pushing forward the conventional use of punctuation to creatively add the prosodic functions of exclaiming (!) and of definitiveness (.). Drawing on Lemke's definition (1998), meanings are not "additive" but "multiplicative".

Her mother's "No." and her father's "No.!" are contrasted via the use of written punctuation, that is here used to playfully represents her parents' refusal to let her move in with Matt. The affective function is here displayed to humorously represent – thus adding extra illocutionary force to – the story-telling of her parents' response. Amanda's hesitations are also evident in her kinesic shifts and uneasy movements at Raymond's questioning, and humour comes at hand to release tension around a sensitive topic.

Unlike the other examples discussed in this section, Raymond does not even recognise the occurrence of these two MSs self-initiated by SP Amanda. Even though he had been instructed to

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

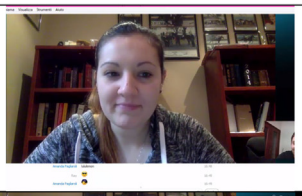
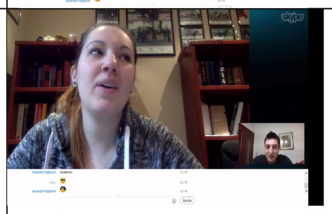
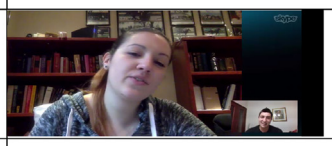
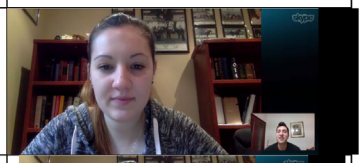
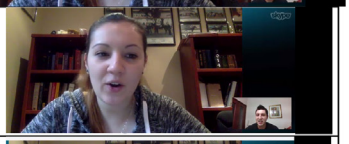
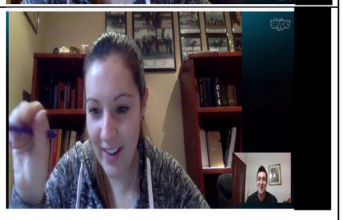
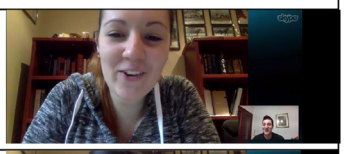
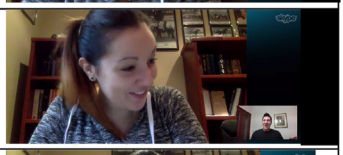
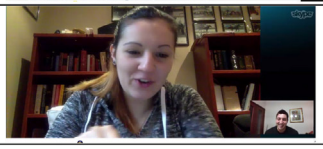
Ray	I remember that you said that you wanted to stay in a house just with Matt..			Ray grins, Amanda Is silent, nods and with a tight, forced smile	Ray looks upwards, Amanda looks sideways	
Amanda Ray	[hesitant] Yeah.. but.. it's not /possible … Because/ number one/ I have no money because of university/ Yeah			Ray raises his eyebrows	Amanda looks upwards while speaking Grins Smiles, nods	
Amanda	And number two/ he's still finishing and we don' know/ So/we're waitin'					
				ABRIDGED		
Ray	But your .. but what your mom said about this..? she/she..					
Amanda Ray	No! No?					
Amanda Ray	No! N-O Like this With a period! But why? Because she wants that you/ that you will marry before?	No.	+		Mimics writing	
Amanda Ray	Yes.. she's very old school// No, she's very I-Italian					
Amanda Ray	Yeah. Very Italian old school And what about your father? Like her.. no?					
Amanda	He's the worst. No. No. No like this So there you go.	No.!	+		Starts typing	

Fig. 3. Examples of MS for playfully use written language to emphasise expressively speech.

identify and even prompt MS, he writes the following (here reproduced unedited):

> Mode switching, is one of the most important elements to analyse in video chat. It raises a number of research questions, for example, the reasons why users decide to switch from speech to writing and *vice versa*, the range and extent of mode switches, and the reasons why users may be high or low mode-switchers. So switching may be occasioned by both internal and external factors.
>
> For example, in my video chat, I wrote and sent a smile to the other participant to see her reaction (self-initiated mode-switching). My interlocutor read what I wrote, but she continued to talk.
> This means that mode switching can produce many different reactions.

SC Raymond correctly mentions the fact that MS can be occasioned by internal and external factors, but only reports on his attempt at prompting SP's "reaction" to his self-initiated MS (i.e.

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

he sent an emoticon). He nonetheless fails to recognise the two occurrences commented above, even though he had similar training to that received by Susan, being both SCs majoring in linguistics and being taught the same course contents. Different performances are possible explanations for this different understanding shown by students, but, on a more abstract level, suggest the need for the adoption of specific training modules in linguistics, and communication in general, that may help expanding critical skills in the production and interpretation of digital textuality.

4. Discussion and concluding remarks

The examples presented from our video-datasets of 24 videocalls and materials produced by students in a research project investigating digital textuality have three different sets of interrelated implications that will be discussed in this Section, along with some final remarks indicating possible future lines of research in this field.

These three strands involve: (1) theoretical implications from a linguistic perspective; (2) educational implications, as anticipated in the premise of this paper; and (3) theoretical implications from a multimodal perspective.

The first strand of implications deals with how research boundaries in the context of VMC need to be constantly redefined as new interactional patterns emerge and evolve. The observation of the pattern of mode-switching calls for some refinement of initial descriptions, with new configurations of occurrence. MS has been shown as occurring in one-to-one semi-spontaneous videocalls. When managing the conversation flow is not necessary, and environments are easily handled by participants, due to the restricted access to interaction, new patterns of MS emerge, thus showing its productivity. Since it was initially described as a linguistic phenomenon similar to some extent to code-switching, some new distinctions involve the material impossibility of overlap when code-switching is involved, whereas in our video-datasets some overlap, or interlaced spoken and written turns, have been documented. The main modality that is used by participants is speech, with some interspersed written comments, so that it may be assumed that the interplay between spoken and written modes is not balanced in these interactions. Participants in videocalls can in fact "mode-mix", for example starting a turn in spoken mode and then type in the same or other words. We may likewise imagine contexts in which users speak and write, or speak and send multimodal materials, at exactly the same time, not necessarily repeating as in the example of Susan and Chris, but disjointing spoken and written communication, partly as has been observed in multi-party video interactions, but having here one interlocutor as the only addressee. Such cases cannot be merely described as "alternations" as is the typical case of code-switching, which, even though may exhibit different forms of interweaving of two or more language varieties, cannot be defined as "overlapping" given the syntagmatic nature of natural languages. MS is in this sense productive in that has been developing new and emergent patterns of occurrence. On another level, these new patterns call for methodological tools apt to describe them, be it in transcription and/or annotation.

Other uses discussed in this paper, however, are consistent with previous research in the context of multi-party video interactions and with code-switching uses, for example in the context of pragmatic uses that contribute to build solidarity or add to intimacy between participants. Additionally, pictorial elements, that have been previously equated with images, pictures, or any other visual elements, have here been reshaped, as the role of verbal written language can in some circumstances be associated to visual language. This was the case with SP Amanda writing down "No." and "No.!", not so much to add, explain, or specify other referential contents, but to iconically and graphically represent speech and playfully "depict" its distinctive prosodic – and emotional – features.

In broader terms, written conventions, originally used to represent speech *a posteriori*, are here resemiotised to embody speech *a priori*. Significantly, written language was originally developed to represent speech, whereas some current digital practices seem to suggest a form of "secondary writing", to paraphrase Ong's conceptualisation of "secondary orality" (Ong, 2002). In other words, these patterns may imply that writing is in some digital contexts used to "explain" speech as is common practice in digital textuality. Actually, this process is nothing new, as in face-to-face contexts writing features are commonly used for similar communicative purposes, for example when a speaker mimes inverted commas with forefingers and middle fingers to express the concept of "so-called", or similar. However, digital environments have opened new terrains for the full deployment of multimodal, sensorial and resemiotised meaning-making, in what may appear as purely "linguistic" at first sight. These reflections lead us to one of the main arguments that I wish to make, that is the *epistemologically multimodal nature of language*.

Turning to discuss the educational implications of this argument, it comes as no surprise the need to encourage students' critical digital and multimodal skills. Recognising these patterns and their metapragmatic impact on learning processes is fundamental to develop students' skills in understanding and interpreting digital textuality. The fact that students (and teachers) are able to technically use video communication does not imply that they are likewise able to semiotically make sense of all the nuances attached to each communicative choice. For example, the *technical possibility* to type in a message in the text box in a videocall does not equate with the *semiotic choice* of doing so.

When it comes to understanding and making sense of web-mediated communication, and especially so in educational environments, several layers of meaning making entail the orchestration of different semiotic resources that are co-deployed by users to adapt to these environments. Recognising these different levels and their interplay is crucial, considering the growing amount of social interaction on the web, and especially so in educational contexts. Critical awareness of video-specific patterns in syllabus design and teacher training should be further encouraged with the ultimate goal of promoting understanding of what can hinder or facilitate interaction. Failures may be due to the lack of metapragmatic and intercultural skills. The bottom line of my argument is that the much heralded adaptation to these environments is ambiguous, and precisely because, as I argued beforehand, technical possibility does not automatically correspond to semiotic action and/or critical awareness.

The students' assignments discussed in this paper include manual transcription and annotation tasks on a common grid. Even though these tasks are time-consuming and may appear of limited applicability to learning environments outside restricted and specialised contexts (for example MAs in linguistics), I argue the case for wider applications of multimodal transcription and annotation, and specifically in teacher-training courses. Their pedagogical value needs further charting, and detailed observations and transcription of VMC may be useful in developing critical skills that go beyond activities and tasks that are typically used to train linguists, and indeed should be practised in wider educational contexts.

A final remark involves the questions dealing with the epistemology of multimodal studies touched upon in the introductory section. The transcription and annotation grid adopted in this study blends basic principles of conversation analysis and multimodal foundational theories, as it prioritises language description and representation and, at the same time, incorporates the description of resources *other* than language, by using *time* as the guiding

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

principle and not *turns* as in classic CA. A caveat to this approach is that students used *language* for their transcription and annotation, even when describing *other* resources. The model rests on language for meaning making, but nonetheless points at its epistemological limits in the realm of multimodal representations. Additionally, mode-switching is placed at the centre of the grid, implicitly prioritising it by giving it centrality in the model: students in fact took pain in prompting and accounting for it. The columns devoted to language analysis are three out of nine in the transcription grid, thus openly disclosing the agenda and background of the researcher who originally developed it, that is the author of this study. However, class discussions that recognised the constraints, limits, and bias of the overall project were helpful in unpacking and de-assembling these and other questions addressed by the students themselves, who, for example, rightly pointed out the issues of capturing data with themselves involved as students collaborators, being subject and object of study at the same time, and knowing in advance the project's research agenda.

All these examples have been here selected not for the sole purpose of illustrating the productivity of mode-switching in VMC or for debating different readings and interpretation of data by different students, but with the more abstract epistemological goal introduced in the premise. I have here argued the case for the usefulness of linguistic analysis integrated in multimodal approaches to communication, and especially so in educational contexts, from an epistemological standpoint: not as an abstract principle, but to claim that studying language from a multimodal perspective can be highly productive for a trans-disciplinary cross-fertilisation of linguistics and multimodality.

An understanding of the multiple and multi-layered contribution of each resource in VMC is key to multimodal studies, but testing the boundaries of its underlying theories of knowledge, especially when it comes to probe into the relationships that multimodality exhibits with neighbour disciplines, in linguistics as well as in other social sciences, is today more than needed as the expanding realm of multiple channels of communication makes the interactional arena more complex than ever.

References

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, *10*(4), 419–438. http://dx.doi.org/10.1111/j.1467-9841.2006.00286.x

Antaki, C., Ardévol, E., Núñez, F., & Vayreda, A. (2005). "For she who knows who she is:" Managing accountability in online forum messages. *Journal of Computer-Mediated Communication*, *11*(1), 114–132. http://dx.doi.org/10.1111/j.1083-6101.2006.tb00306.x

Auer, P. (1998). *Code-switching in conversation*. London: Routledge.

Baldry, A., & Thibault, P. (2008). Applications of multimodal concordances. *Hermes*, *41*, 11–41.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*(4), 544–559.

Bajramia, L., & Ismailia, M. (2016). The role of video materials in EFL classrooms. *Procedia – Social and Behavioral Sciences*, *232*, 502–506.

Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. London and New York: Routledge.

Crabtree, B. F., & Miller, L. M. (Eds.). (1999). *Doing qualitative research*. London: Sage.

Creswell, J. (1998). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Creese, A. (2008). Linguistic ethnography. In K. A. King, & N. H. Hornberger (Eds.), *Encyclopedia of language and education. 2nd edition, Vol. 10: Research methods in language and education* (pp. 229–241). New York: Springer Science and Business Media.

Duranti, A. (1997). *Linguistic anthropology*. Cambridge: Cambridge University Press.

Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life*. Cambridge: Polity Press.

Flinkfeldt, M. (2011). 'Filling one's days': Managing sick leave legitimacy in an online forum. *Sociology Health Illness*, *33*(5), 761–776. http://dx.doi.org/10.1111/j.1467-9566.2011.01330.x

Flinkfeldt, M. (2014). Making equality relevant: Gender, housework, and sick leave legitimacy in online interaction. *Female Psychology*, *24*(3), 295–313. http://dx.doi.org/10.1177/0959353513515295

Gardner, R., & Levy, M. (2010). The coordination of talk and action in the collaborative construction of a multimodal text. *Journal of Pragmatics*, *42*, 2189–2203.

Georgakopoulou, A. (2011). Computer-mediated communication. In J. Ostman, & J. Verschueren (Eds.), *Pragmatics in practice* (pp. 93–110). Philadelphia & Amsterdam: John Benjamins.

Goffman, E. (1981). *Forms of talk*. Oxford: Blackwell.

Groen, M., Ursu, M., Michalakopoulos, S., Falelakis, M., & Gasparis, E. (2012). Improving video-mediated communication with orchestration. *Computers in Human Behavior*, *28*, 1575–1579.

Gumperz, J. J. (1999). On interactional sociolinguistic method. In S. Sarangi, & C. Roberts (Eds.), *Talk, talk, work and institutional order: Discourse in medical, mediation and management setting* (pp. 453–472). Berlin & New York: Mouton de Gruyter.

Halliday, M. A. K. (1978). *Language as social semiotic. The social interpretation of language and meaning*. London: Arnold.

Heath, C., & Luff, P. (1992). Media space and communicative asymmetries: Preliminary observations of video-mediated interaction. *Journal of Human-Computer Interaction*, *7*(3), 315–346.

Heller, M. (1988). *Code-switching: Anthropological and sociolinguistic perspectives*. Berlin & New York: Mouton de Gruyter.

Herring, S. C. (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In D. Tannen, & A. M. Trester (Eds.), *Georgetown University Round Table on languages and linguistics 2011: Discourse 2.0: Language and new media* (pp. 1–25). Washington, DC: Georgetown University Press.

Herring, S., Stein, D., & Virtanen, T. (2013). *Pragmatics of computer-mediated communication*. Boston: De Gruyter Mouton.

Hodge, R., & Kress, G. (1993). *Social semiotics*. Ithaca: Cornell University Press.

Holmes, J. (2001). *An introduction to sociolinguistics* (2nd ed.). London: Pearson Education Limited.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London: Routledge [1996].

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin, & R. Veel (Eds.), *Reading science* (pp. 87–113). London: Routledge.

Li Wei (Ed.). (2000). *The bilingualism reader*. London: Routledge.

Jørgensen, J. N. (2008). Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism*, *5*, 161–176.

Meredith, J., & Potter, J. (2013). Conversation analysis and electronic interactions: Methodological, analytic and technical considerations. In H. Lim, & F. Sudweeks (Eds.), *Innovative methods and technologies for electronic discourse analysis* (pp. 370–393). Hershey, USA: IGI Global.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Myers-Scotton, C. (1989). Codeswitching with English: Types of switching, types of communities. *World Englishes*, *8*(3), 333–346. http://dx.doi.org/10.1111/j.1467-971X.1989.tb00673.x

Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon.

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. New York: Routledge.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human–Computer Interaction*, *8*(4), 389–428.

O'Halloran, K. L. (Ed.). (2004). *Multimodal discourse analysis: Systemic functional perspectives*. London: Continuum.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, *8*(2), 177–192.

Ong, W. (2002). *Orality and literacy: The technologizing of the word* (2nd ed.). New York and London: Routledge.

O'Toole, M. (2011). *The language of displayed art* (2nd ed.). London & New York: Routledge [1994].

O'Toole, M. (2004). Opera Ludentes: The Sydney Opera House at work and play. In K. L. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives* (pp. 11–27). London: Continuum.

Paulus, T., Warren, A., & Lester, J. N. (2016). Applying conversation analysis methods to online talk: A literature review. *Discourse, Context and Media*, http://dx.doi.org/10.1016/j.dcm.2016.04.001i

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, *115*, 68–81.

Rosenbaun, L., Rafaeli, S., & Kurzon, D. (2016a). Blurring the boundaries between domestic and digital spheres: Competing engagements in public Google Hangouts. *Pragmatics*, *26*(2), 291–314.

Rosenbaun, L., Rafaeli, S., & Kurzon, D. (2016b). Participation frameworks in multiparty video chats cross-modal exchanges in public Google Hangouts. *Journal of Pragmatics*, *94*, 29–46.

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

Shih, Y.-C. (2014). Communication strategies in a multimodal virtual communication context. *System*, *42*, 34–47.

Sindoni, M. G. (2011). Online conversations. A sociolinguistic investigation into young adults' use of video chats. *Classroom Discourse*, *2*, 219–235.

Sindoni, M. G. Speech and writing in English: A cultural and linguistic debate, 2012a, Le Lettere; Firenze.

Sindoni, M. G. Spoken and written discourse in online interactions. A multimodal approach, 2013, Routledge; London and New York.

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004

Sindoni, M. G. (2014). Through the looking glass: A social semiotic and linguistic perspective on the study of video chats. *Text and Talk*, *34*, 325–347.
Sindoni M.G. and Petroni S. Multimodality and digital literacy in the English language syllabi across Europe, In: F. Cavaliere (ed.), Euromosaic: A still open challenge, 2017, Edicampus; Roma, 188-211.
Sindoni M.G. 'Of course I'm married!' Communicative strategies and transcription-related issues in video-mediated interactions, In: Garcés-Conejos Blitvich P. and Bou, P., (Eds.), Analyzing digital discourse: New insights and future directions, 2019a, Palgrave Macmillan; Basingstoke, 71-103.
Sindoni M.G. 'Tell me tell me'.Addressing cultural and linguistic diversity in cultural and linguistic video interactions, In: Carreiras i Goicocechea M., Russo G., Venuti M. (Eds.) Lingua, identità e alterità, 2019b, ETS; Pisa, 109-123.
Sneijder, P., & te Molder, H. F. M. (2005). Moral logic and logical morality: Attributions of responsibility and blame in online discourse on veganism. *Discourse Society*, *16*(5), 675–696. http://dx.doi.org/10.1177/0957926505054941
Snyder, M., & Gangestad, S. (2000). Self-monitoring: Appraisal and reappraisal. *Psychological Bulletin*, *126*(4), 530–555.
Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
Swan, P. K., Richardson, J. C., Ice, P., Garrison, D. R., Cleveland-Innes, M., & Arbaugh, J. B. (2008). Validating a measurement tool of presence in online communities of inquiry. *E-Mentor*, *2*(24), 1–12.
Thurlow, C., & Mroczek, K. (Eds.). (2011). *Digital discourse: Language in the new media*. Oxford: Oxford University Press.
Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
van der Kleij, R., Paashuis, R., & Schraagen, J. M. (2005). On the passage of time: Temporal differences in video-mediated and face-to-face interaction. *International Journal of Human–Computer Studies*, *62*, 521–542.
van Leeuwen, T. (1999). *Speech, music, sound*. London: Macmillan.
van Leeuwen, T. (2005). *Introducing social semiotics*. New York and London: Routledge.
Verschueren, J. (2000). Notes on the role of metapragmatic awareness in language use. *Pragmatics*, *4*(10), 439–456.
Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.
Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam & Philadelphia: John Benjamins.
Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the Web*. London: Continuum.

Please cite this article in press as: Sindoni, M.G. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. *Linguistics and Education* (2019), https://doi.org/10.1016/j.linged.2019.05.004