Literature Review of Multimodal Transcription and Coding Methods for Transcription of Communication Cues of Arrogance

1. Introduction

The project Changing Attitudes in Public Discourse aims to reduce arrogance in public debate through the development of practical intervention strategies. This will require an understanding of how arrogance is expressed within public debate. Communication cues for arrogance, therefore, must be identified and recorded at an early stage of the project. This literature review is concerned with the transcription of arrogance in interpersonal communication, and seeks to represent existing writing on multimodal transcription methods. This is supported by a discussion of literature regarding communication cues for arrogance themselves, so as to identify what is required from a comprehensive transcription method for arrogance in public debate.

Multimodal transcription is a relatively new and expanding field by comparison to more established and conventionalised transcription styles, such as conversation analysis. This review, therefore, has endeavoured to encapsulate the most current and effective methods for multimodal transcription. That said, this review will not extensively explore the transcription strategies used within the wider methodologies established, such as methods of transcribing speech or video shot framing within a multimodal transcription framework. It is apparent that writing discussion of multimodal transcription has flourished within the last decade, thus much of the most relevant material represented here comes within this time frame.

It must be noted, however, that many of the multimodal transcription methods exemplified in recent literature have been designed for specific purposes. Recktenwald's (2017) transcription, for example, presents a multimodal transcription method designed to encapsulate the information produced through the videogame streaming software Twitch. Similarly, Cowan's (2014) transcription is designed for early-years classrooms. Thus, where some of the multimodal transcription methods have been designed for specific situations or software yet still offer significant merits, they may require modification for the purpose of transcribing arrogance in public debate.

It has become clear that multimodal transcription methods, even within the past few years, have begun to adopt certain styles. The main body of this review will consider each of these styles in turn, though it should be taken into account that not only are these styles not completely discrete and separate. Many texts exemplify multiple methods of different styles and consequently cannot be considered dedicated to one style of multimodal transcription. This paper will then go on to discuss software available for aiding transcription, and finally a selection of writing on nonverbal communication to identify what data we are likely to expect from participants exhibiting arrogance.

2. Multimodal Transcription and Coding

i. Play-script transcription

The first style of multimodal transcription method this review will consider is play-script transcription. It is a style that most linguists will be familiar with, as it is developed from pre-existing and conventionalised transcriptions. For the most part, play-script transcriptions follow a left to right, top to bottom reading approach with different speech turns attributed to each participant; like a play script itself. Play-script style transcriptions have become standard in conversation analysis, thus most transcriptions in this style rely at least partially on conversation analysis conventions. There are extensive rules for the transcription of speech within this style, and nonverbal communication is often transcribed minimally, which is insufficient for multimodal transcription.

Mondada (2001) tackles this issue in her text 'Conventions for multimodal transcription'. She aims to expand the conventions of play-script transcriptions to incorporate nonverbal behaviour. Mondada described nonverbal communication textually, and uses a number of symbols to demonstrate the initiation, continuation and termination of these actions. Whilst there is no timestamp afforded to the various turns, temporal information of nonverbal modalities such as gaze and gesture can be inferred by comparison to the speech acts occurring simultaneously. In her own examples, Mondada also chooses to differentiate speech and additional nonverbal communication using different fonts. Using the key provided, one can deconstruct the transcription as desired, but the result is a rather complex transcription.

Further examples of Mondada's transcription method can be seen reproduced by Ayaß (2015) and again by Mondada herself in her 2016 paper 'Challenges of multimodality: Language and the body in social interaction'. It is apparent that whilst Mondada's 2001 transcription conventions are being recognised outside her own work, her conventions are not being adopted for wider use. The reasons for this seem to be the complexity of the new conventions and the logistical restraints they entail. The extensive key Mondada provided in 2001 has thirteen entries, making it difficult to follow in places. Furthermore, whilst the various modalities of a few participants can be traced using these symbols, a debate between more than a few potentials is likely to be rich with nonverbal communication, and difficult to handle using Mondada's method.

Ayaß's (2015) text is unique in its focus on adapting conversation analysis conventions to multimodal transcription requirements. Alongside her consideration of Mondada's work, she considers some other hybrid transcription styles based on play-script transcription but incorporating still-frame images. The other transcription methods exemplified within the text are not so thorough as Mondada's, but hold some similarities. There seems to be a general lack of specific temporal information attached to nonverbal modalities. Instead, duration of gaze and gesture actions is given in relation to the speech acts they occur alongside. Ayaß provides a useful collection of play-script based transcriptions, and it becomes apparent that it is difficult to incorporate significant amounts of nonverbal information into this style of transcription without saturating the transcription.

In one of the most valuable texts on multimodal transcription, Bezemer and Mavers (2011) also demonstrate an example of play-script transcription. Much like other play-script transcription, nonverbal information is described textually, differentiated from speech through italicisation. Thus, whilst gesture, gaze and information are all represented in the transcription, it is very much focused on speech. Their example contributes to a growing collection of play-script transcriptions which, whilst effective at transcribing speech, are not adaptive enough for multimodal transcriptions rich in nonverbal communication. Further material in Bezemer and Mavers' text, however, can help us understand these issues. The text considers how multimodal texts much be deconstructed into various modalities, and these modalities ranked for salience within a transcript. This consideration of the logical composition of a transcription is very useful, and Bezemer and Mavers also discuss other transcription styles which will we will return to later.

ii. Tabular transcriptions

Baldry and Thibault's (2006) seminal text *Multimodal transcription and text analysis: a multimedia toolkit and coursebook* enacted significant advancement in multimodal transcription. The text presented two major transcriptions arranged within tables, which they titled as macro-analytical and integrated. The text is also invaluable for its extensive breakdown of the two methods, discussing how the elements are arranged within the table. Unlike Mondada's 2001 text, Baldry and Thibault's model has been adopted widely within the linguistic community. Tabular transcriptions in various formats have been used regularly since the production of this text.

Indeed, in Bezemer and Mavers' (2011) aforementioned text, a tabular transcription is presented within their collection of transcription methods, albeit a less complex tabular transcription. Instead of still-frame images as used by Baldry and Thibault, this transcription uses drawn images. The transcription is useful, however, in showing the flexibility of the tabular format. Further development can be seen in Cowan's (2014) study of early-years classrooms. Cowan chooses not to include still-frame images in her tabular transcription. It does, however, highlight a drawback in the tabular transcription method. Where a timestamp is not included, such as in Cowan's transcription, it can be difficult to measure the duration of modalities such as gesture and gaze. It is also easy to recognise how a tabular transcription must be arranged carefully, as separately recording the various modalities of each participant can become difficult to manage.

Further multimodal transcriptions presented in a tabular format can be seen in Flewitt's (2009) paper collecting multimodal transcription methods. The transcriptions that Flewitt displays, however, are not so useful as they have been designed for the specific purposes. The first is transcribes audioconferencing, meaning it has no resources for transcribing gaze or gesture. The second transcribes a drawing activity between an adult and child, and whilst it is not extensive it does show how drawing progression can be included through still-frame images.

Moving from audioconferencing to videoconferencing, Helm and Dooly (2017) provide transcriptions for the software 'Soliya'. This transcription is useful as it shows how a text can be split into time fragments, then time-coded again within these fragments. Encoding temporal information seems to be a recurring issue within tabular transcription, so although this transcription offers little innovation in terms of transcribing modalities such as gaze and gesture, it does attempt to rectify non-specificity in time coding.

Another tabular transcription built around specific requirements is Recktenwald's (2017) transcription of Twitch videogame streaming. Recktenwald opts for a simplistic layout, although there is data from a number of different participants in the text. Recktenwald's text further consolidates the developing conventions of tabular transcription, but otherwise provides little ingenuity. At this point it is relevant to suggest that tabular transcriptions are becoming conventionalised; although they are modular in the sense that different columns or rows may be added or removed relative to the modalities required.

In a series of texts leading up to a more comprehensive approach in 2016, Taylor dedicates more time to transcribing image composition than most other linguists. Taylor's text focuses on subtitling Italian films, but his exemplified transcriptions are some of the best laid out examples of tabular transcription. Modalities such as gaze, movement and gesture and combined into one column, however. This may be sufficient for transcribing the overall action of a movie scene, but the finer details we will observe in public debate will require separate transcription.

In his 2013 paper 'Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze', Satar focuses expressly on transcribing gaze. Thus, whilst his tabular transcription is incomplete in terms of modalities such as gesture, Satar develops new strategies and terminology for transcribing gaze. Tabular transcriptions do not tend to suffer the same emphasis on spoken word that play-script transcriptions do, but it is uncommon for gaze to be transcribing in more detail than brief orthographic comments. Thus, this text may be invaluable for providing more extensively transcribed information on gaze activity between participants.

iii. Timeline transcriptions

Timeline transcriptions are a relatively new method, which has not seen extensive use yet. It is, however, a thoroughly flexible and effective method. In their 2011 text, Bezemer and Mavers record the usage of a timeline transcription. The method involves arranging each modality on a horizontal axis much like a timeline. Specific temporal information can be given, but regardless the arrangement of each modality (such as the speech, gaze, gesture or others as required for each participant) above one another means that each action within the transcript can be compared against any other. Moreover, the timeline arrangement means that the duration of events can be represented visually. Still-frame images are often included at the top or bottom of the transcript to contextualise the actions.

In her 2014 paper on early-years classrooms, Cowan also transcribes some of her data in a timeline style transcription. Cowan chooses to use a key as well as labelling to differentiate her modalities, and uses an innovative system to show how actions - in this case computer mouse - usage progress and terminate. This example demonstrates effectively how timeline transcription can manage a complex situation, and it is undoubtedly the most useful section of Cowan's paper in terms of this study.

Guichon and Wigham's 2015 study into webconference supported language teaching used ELAN transcription software. Transcription softwares will be discussed more thoroughly later in this review, but it is relevant to mention here as the data configured on screen is arranged similarly to a timeline transcription. This is not the final exported data, but it is interesting to see softwares adopting this arrangement of data.

Although there are relatively few examples of timeline transcriptions currently published in real-world study, it is vital that this transcription method is not dismissed. This is a new method, and will likely see much more usage in coming years.

iv. Image-based transcription

The final major transcription method is one based around still-frame images. Transcriptions following this method have been published fairly extensively and thus merit inclusion in this review, but it is largely an outdated methodology.

In Ayaß's (2015) collection of transcription methods, she demonstrates an extremely minimalist image-based transcription consisting only of three still-frame images and a caption to demonstrate what the images attempt to communicate. It is easy to see how a transcription method such as this might only be useful for transcribing recognisable gestures that need little supplementary information. Ayaß also shows a more dynamic image-based transcription which is something of a merger with a play-script transcription. Not only does this example use still-frame images, but also drawn reproductions. Directional arrows are used to show gaze, and a play-script transcription supplements the images. The images are an attempt to contextualise the verbal element of the transcript, but on a functional level little more information is added, as gaze and gesture are not clearly represented by the images.

Bezemer and Mavers (2011) text, proving to be one of the most thorough analysis of multimodal transcription, also considers image-based transcription. The transcription they exemplify is more extensive, with still-frame images showing the progression of the text, and other modalities overlayed. The speech of the two participants is differentiated by shade, and layout of the text shows intonation and stress in an interpretive fashion. Once again, gaze is shown through directional arrows and gesture is implied through the images themselves. Some of these resources are innovative attempts to represent multimodal data in an accessible and recognisable sense, but the transcription quickly becomes cluttered over the images. Certainly, transcribing a public debate through a style like this would be difficult, especially considering the subtle communication cues discussed later in this review.

Similar transcriptions can be seen in Flewitt's (2009) paper, where a both a classic image-based transcription and a hybrid form. Nevertheless, these transcriptions suffer the same restrictions as other image-based transcriptions. Whilst they can be effective for simple multimodal data, it is difficult to manage once there are multiple participants in a rich communicative situation expressing multiple modalities.

v. Conclusion

Two of the transcription methods above stand out as more likely to be useful for the Changing Attitudes in Public Discourse project. Both the tabular transcription and timeline transcription styles bear convincing arguments. Play-script transcriptions and image-based transcriptions on the other hand, hold a number of limiting factors for transcribing multimodal data.

Play-script transcriptions are very effective at transcribing largely monomodal speech discourse where there is little nonverbal communication, especially within established conventions such as conversation analysis. Due to the strictures of traditional left to right, top to bottom ordering, adding information for further modalities without disrupting the transcript can be challenging. When multiple participants are involved, the case is even more difficult, and in terms of public debate this will almost certainly be the case. Where each participant will display speech data, gesture, gaze, intonation and perhaps others not considered thus far, encoding within a play-script transcription becomes unfeasible. That said, there may be some merit to play-script transcriptions. Within a tabular transcription, there will be a column for verbal modality. Depending on the table layout, this could be a column for the speech of each participant, or a column for all the speech captured within the transcript. If the second were true, the column for speech could be transcribed as a conversation analysis play-script transcription within the larger tabular transcription.

Image-based transcriptions face similar drawbacks. Encoding all the various modalities of each participant over still-frame images incurs issues over space constraints. It is likely that multiple participants would be talking and acting within the confines of one image, thus transcribing all of the data without obscuring the image becomes challenging. Moreover, whilst still-frame images may accurately depict large, recognisable gestures such as pointing, there are subtler elements of gesture such as facial expression which will be difficult to make out from images alone. There is some merit to image-based transcription for communicative intentions such as instructions, where the data must be simple and accessible by nature. For the complex and dynamic data that will be produced from public debate, however, image-based transcriptions will not suffice.

Tabular transcriptions are one of the transcription methods that may be viable. Tabular transcriptions have the added benefit of becoming increasingly conventionalised in recent years, so they are likely to be accessible to those with a background in multimodal transcription. Both tabular transcriptions and timeline transcriptions are modular in that additional data (for different modalities) can be added or removed as necessary. In a tabular review, this is done by adding or removing certain rows or columns, depending on the way the table has been constructed. This flexibility is a huge benefit, as complex multimodal data does not make the table any more confusing, just more extensive. A tabular transcription allows two main methods of reading. Assuming time is on the vertical axis, one can read down the columns to see how one modality progresses through the text, or across the row for each time frame to gain a full picture of action within that frame. In all, tabular transcriptions hold enough flexibility and clarity to be a useful transcription method, and a table can be designed for the specific purposes of this project.

A timeline transcription could be equally effective. Time transcriptions are also inherently flexible, as each modality exists on a separate horizontal axis, thus there is no effective limit on how many may be encoded. The layout of a timeline transcription follows a logical format, so even though it is a relatively new structure, it should pose few problems in terms of accessibility. The visual representation of event duration alongside specific temporal information means that co-occurrence of events is easy to measure. Timeline transcriptions rely more on the visual element than tabular transcriptions, although much of the information must still be described orthographically, such as actions of gaze and gesture.

Conclusively, it is apparent that the two most appropriate transcription styles are tabular transcription and timeline transcription. Both offer a flexible and uncluttered approach to representing a range of multimodal data. There is little difference between the two in terms of utility, hence choosing between them may come down to personal preference on the part of the researcher. Regarding the Changing Attitudes in Public Discourse project, both styles should be suitable for transcribing all the data a public debate might produce.

3. Transcription Software

Transcription software is no longer a new phenomenon within linguistics, and users have the choice of many different programs, many open-source and some premium payment-required. This review aims to introduce some of the leading contenders in transcription software.

ANVIL (Kipp, 2014) describes itself as a video annotation tool, and focuses on information coding. ANVIL uses a progressive system by which different information, such as for different modalities, can be arranged along a timeline and colour coded. It is, in fact, similar in layout to a timeline transcription. ANVIL has the functionality to play videos within the software, represent data as waveform or pitch contour, and allows for manual speech transcription. Furthermore, ANVIL can import and export to and from most other major transcription software such as Praat and ELAN if initial work has been conducted there, or to export to a colleague using different software. The layout of ANVIL is effective and conducive to further analysis.

Chronoviz (Fouse et al, 2011) excels at presenting multiple different pieces of multimodal data on screen simultaneously, whether they be different videos and pictures, or even more specific data types such as maps and digitally penned notes. Regrettably, ChronoViz's export options are limited. Parts of the information displayed within the software can be exported, such as data annotations, parts of video clips and timeline screenshots, but information managed within ChronoViz cannot be exported as a complete transcript. In this sense, ChronoViz is less of a transcription software and more of an aid for manual transcription. The program certainly excels in managing various multimodal data sets in an accessible manner on-screen.

CLAN (MacWhinney, 2000) is one of the longest-standing pieces of transcription software available, first developed to aid the CHILDES project in 1984. CLAN allows for the simultaneous portrayal of the original multimodal data, usually a video, and the developing transcript. Transcripts created within CLAN can be temporally linked with the original media so that the relevant part of the transcript for the video is highlighted as the video plays, or alternately one can find the relevant time within the video by selecting part of the transcript. CLAN specialises in data collection and management in terms of corpus creation, as the CHILDES project aimed to create a corpus for first language acquisition data. Thus, data managed through CLAN can be tagged and organised extensively. The software also has the standard resources for displaying audio data as waveform or pitch contour. Moreover, CLAN can work directly with Praat for advanced audio analysis.

ELAN (Sloetjes and Wittenberg, 2008) is a multimodal analysis software which presents information in an accessible timeline format, thus coincidence of actions is easy to recognise. ELAN's interface is arguable among the most functional of all the software discussed in this section, as it is conducive to a final transcription. ELAN aims to be compatible with most other major multimodal analysis software. That aside, ELAN offers much of the conventional functionality of a transcription software, displaying a range of multimodal data in various formats. ELAN also offers the capacity to annotate different data sets separately.

EXMARaLDA (Schmidt and Wörner, 2009) presents a unique interface as it is a combination of timeline and tabular. Whilst the data runs on a horizontal axis much like a timeline, events can be split into rows and columns as desired. This makes EXMARaLDA one of the most flexible programs available. Moreover, EXMARaLDA is also flexible in its output. It can output in various layouts including line for line and some specific requirements such as musical score, and various file formats including HTML and Microsoft Office Word. EXMARaLDA also offers a number of integrated transcription conventions, such as CHAT and IPA alongside some German language only formats. Finally, like most of the other software available, it seeks to ease data exchange with other major software like Praat and ELAN.

Praat (Boersma and Weenink, 2017) has a slightly different focus to the other software discussed here. Praat is an advanced audio analysis tool most often used for intense speech analysis. Alongside the standard waveform and pitch contour functionality, but also more specialised forms like energy spectrograms. Praat has resources for displaying energy formants, pitch tracking, nasality ,amplitude and others. In all, Praat offers a comprehensive approach to audio analysis, and data from Praat can be imported into most other transcription software which offer limited audio analysis tools. This review does not aim to give a complete overview of Praat's ranging functionality, but to impress its potential utility in analysing audio within a multimodal text.

Transana (Woods and Fassnacht, 2017) is another transcription software that maintains functionality by displaying various data on screen simultaneously in a useful manner. Transana focuses on data coding and categorisation, whilst allowing manual transcription within its editor. A unique feature of Transana is its ability to share a project between multiple contributing researchers. Different users can work on a project from different locations at the same time, with changes occurring in real time. Thus, where it is important that multiple members of a team be working at the same time, Transana can facilitate this, especially with the aid of a conference call. It is important to note, however, that Transana is a paid software, thus this premium accessibility does come at a cost. That aside, Transana has little to set it apart from other transcription software.

One challenge in researching these programs is that it has proven difficult to find completed, exported transcripts compiled through the software. Thus, this review can only comment on the software on a functional level, rather than any completed transcripts they produce. That said, most of these programs offer largely similar functionality (with the notable exception of Praat), and furthermore there is a general aim for compatibility between the various programs. CLAN and ELAN seem to be the most prominent and widely documented programs, especially considering CLAN's longevity. Thus, whilst there may be more support on a community level for CLAN and ELAN, it seems that a researcher would not be far amiss to use most of the software discussed here, so long as they remain aware that they may need to export to a different format when sharing with colleagues using different software. That said, ChronoViz is inherently limited by its lack of cohesive export utility, and the payment requirement for Transana may be a factor to choosing different software. Praat is slightly different being a specialised audio analysis software, so it would likely prove useful in conjunction with the selected transcription software.

4. Communication Cues for Arrogance

The smaller part of this review will now consider some of the potential communication cues we may be looking for in public debate, and what actions may precede arrogance. That said, most of the writing around nonverbal communication and communication cues does not concern arrogance specifically, but the broader emotions. Consequently, this review has focused on communication cues for anger, as many of these will likely be transferrable to arrogance.

In their 2016 text *Nonverbal communication*, Burgoon et al provide one of the most thorough discussions of communication cues to various emotions. Their work on anger is particularly illuminating for this project. Where other texts focus largely on facial expressions, Burgoon et al include this information but also consider additional cues including vocal characteristics, posture, upper bodily gestures and gait. Furthermore, there is consideration of different types of communicative conflict, specifically active conflict such as arguing versus passive conflict like stonewalling. It is relevant to suggest that we may see evidence of both of these styles of conflict in the sphere of public debate, and it is important for this text to have raised awareness of subtler manifestations of potential arrogance.

In the most recent text considered here, Hwang and Matsumoto (2016) focus more on the physical and physiological cues to anger. Some of these are too subtle to merit consideration in this study, such as secretion from the salivary glands, but eyes bulging and reddening of the skin may be visible depending on the quality of camera recording. Nevertheless, this text will be useful considered in conjunction with other texts in this section, compounding evidence of communication cues to anger, and therefore arrogance.

Richmond et al's (2008) text not only discusses the communication cues to arrogance, but approaches the topic from an interpersonal perspective, finding that anger cannot be reliably perceived from one area of the face alone. Two areas of the face must be perceived for anger to be judged accurately, which indicates that the cameras used to record the public debate sessions must capture each participants face completely. That said, anger is identified as one of the easiest emotions to identify from vocal cues alone. This text also outlines five guidelines on vocal cues, as well as indicating that fast speech tempo, high pitch and loudness among other features are indicative of anger. Richmond et al consider vocal patterns in greater depth than most of the other texts considered here.

Andersen's (2008) text *Nonverbal Communication: Forms and Functions* gives a detailed discussion of the facial components of anger. Much of the information in this text can be seen reflected in the other texts discussed here, demonstrating reliability of data. Andersen's text is more of an overview of nonverbal communication, and does not have a particular focus unlike some of the other texts on the topic. That said, it is still useful in further consolidating data on communication cues to anger.

Although Remland's (2000) text is considerably older than the others here, it merits inclusion for its unique exploration of turn taking. Alongside standard information on facial cues such as brow action and nostril dilation, Remland discusses how turn taking can indicate arrogance. Specifically, Remland engages with how turn requests can be denied by an individual intent on holding the floor, and the idea that this is an arrogant quality. Furthermore, despite Remland's information on communication cues for anger being of little difference to those in other texts, it shows that these cues have remained consistent for the best part of two decades.

The consistency between different texts' communication cues for anger suggests that they are reliable, and most of these texts offer a unique focus, or closer exploration of a specific part of nonverbal communication. Between them, it is possible to ascertain a good idea of what we must look for in public debate. Of course, arrogance and anger are not identical concepts, but it seems logical that communication cues for anger are most appropriately transferrable to arrogance.

5. Conclusion

This literature review has sought to cover the most recent and relevant literature regarding multimodal transcription and coding. Furthermore, it has succeeded in using this literature to identify four general transcription styles, and consequently consider which will be appropriate for the Changing Attitudes in Public Discourse project.

Multimodal transcription remains a developing field, but reviewing texts mostly from the last two decades has shown great amounts of progress within the field. Within the past decade especially, different styles of multimodal transcription, predominantly tabular transcription, have started to become conventionalised. The requirement for multimodal transcription is only set to grow as we develop new ways of processing data, hence there are certainly further avenues for study of multimodal transcriptions. Timeline transcriptions are starting to be recognised in published work, and their usage will likely see rapid increase in future publication. Whether the Changing Attitudes in Public Discourse project chooses to use tabular or timeline transcription style, it will undoubtedly help to shape how these methods will be used in the future.

Similarly, transcription software has developed to the stage that there is a variety available to the researcher, although many of them aim towards similar goals. However, with the rise in utility of specialised software such as Praat, it seems that transcription programs will develop to address specific needs within the multimodal transcription community.

The review of literature relating to nonverbal communication and communication cues has proved fruitful in terms of identifying cues for anger. Arrogance, however, lies among a set of traits that have not seen much specific study thus far. This project will expand the knowledge of communication cues at least to include arrogance, and measure whether anger and arrogance share communication cues on a functional level.

In all, the Changing Attitudes in Public Discourse project will contribute to an expanding canon of literature on multimodal transcription. Not only will it transcribe a great deal of multimodal data during its course and add valuable information to the field; it will further standardise the usage of multimodal transcription and further the boundaries of its capability.

Annotated Bibliography

Andersen, P.A. (2008). *Nonverbal communication: forms and functions*. Long Grove, ILL., Waveland Press, Inc.

Andersen's text regarding nonverbal communication goes slightly further than others in exploring gestural and postural cues to emotion. Thus, it contributes to a collection of information on communication cues which may prove vital for the transcription of arrogance in public debate. Alongside other texts, Andersen's work will be fundamental in creating a comprehensive idea of what communication cues will require transcription.

Ayaß, R. (2015). 'Doing data: The status of transcripts in Conversation Analysis.' *Discourse Studies* 17(5): 505-528.

This text considers the usage of Conversational Analysis techniques in transcribing multimodal texts. Ayaß provides a number of examples of pre-existing audiovisual transcription methods, breaks down their various elements and considers how they are effective. This text is useful as a small collection of multimodal transcription methods, amid less relevant discussion of the liminal status of the transcripts produced. Some of the transcription methods. Ayaß depicts are innovative forms, and her extensive discussion of each one, how it may be effective or limited, is of great use for deconstructing these transcription methods.

Baldry, A. and P.J. Thibault (2006). *Multimodal transcription and text analysis: a multimedia toolkit and coursebook*. London, Equinox.

This foundational text in the field of multimodal transcription analysis provides extensive information on transcription methods for many forms of multimodal text. First looking at printed media, then web pages, the book then turns to its largest chapter on film, or audiovisual texts. Baldry and Thibault provide two major methods for audiovisual transcription, which they describe as macro-analytical and integrated. Exemplar transcriptions are provided for both styles, preceding a comprehensive breakdown of the elements recorded within each transcription. This book is a very useful introduction to the field of multimodal analysis, demonstrating some effective methods for transcribing audiovisual texts.

Berger, A.A. (1998). *Media analysis techniques*. Thousand Oaks, CA, Sage Publications.

This text focuses on conceptual methods of analysis such as psychoanalytical, or Marxist interpretation. Thus, our requirement for technical methods of analysis, in this case transcriptions methods, is not found within this book. There is a small amount of basic information regarding camera techniques and shot types which may partially support an analysis of camera work in an audiovisual text, but there is little relating to multimodal transcription. This information regarding camera techniques could especially be of use if a tabular transcription method were to be selected, wherein a column regarding shot composition could benefit from specific camera technique notation.

Bezemer, J. and D. Mavers (2011). 'Multimodal transcription as academic practice: a social

semiotic perspective.’ *International Journal of Social Research Methodology* 14(3): 191-206.

This text provides an thorough discussion of multimodal transcription, or the creation of 'transvisuals'. Beginning with a brief history of multimodal transcription methods and theories on the topic, Bezemer and Mavers go on to frame how elements within a potential transcript are selected then ranked for salience. The text outlines several different units of analysis, which may prove useful in deciding how a multimodal text is broken down for transcription. In all, this text is of significant value to any individuals looking to engage in multimodal transcription as it provides some unique insights into the composition of a transcription.

Boersma, P. and D. Weenink. (2017). Praat: doing phonetics by computer.

Praat is an advanced audio analysis and manipulation software. Unlike other transcription software considered by this review, Praat does not seek to aid with overall transcription, but to offer extensive tools specifically relating to audio analysis. Praat is likely to be useful in conjunction with one of the other pieces of transcription software when chosen.

Burgoon, J.K., L.K. Guerrero and K. Floyd (2016). *Nonverbal communication*. New York, Routledge.

This text offers some of the widest range of information available on nonverbal communication, and thus it will be instrumental in aiding transcription of arrogance markers. Burgoon et al go further than preceding texts in considering elements such as gait. Moreover, a very useful passage provides insight into less obvious forms of communication arrogance, such as stonewalling within a debate. This text will prove very useful in identifying both active and passive communication cues for transcription.

Cowan, K. (2014). ‘Multimodal transcription of video: examining interaction in Early Years classrooms.’ *Classroom Discourse* 5(1): 6-21.

This text begins with a thorough overview of literature on multimodal transcription methods, predominantly from the 2000's. This collection of literature is useful both in contextualising Cowan's own research within the developing field, but equally in pointing to further reading for a researcher. The most valuable element of the text, however, is Cowan's own transcriptions. Drawing on the past literature, Cowan makes a series of transcriptions in various styles based on her own multimodal data captured within the classroom. Cowan provides two styles of multimodal transcription, which she calls 'multimodal: tabular layout' and 'multimodal: timeline layout', and discusses their usage. These two transcription systems are some of the best developed multimodal transcription methods, and will undoubtedly prove useful.

Edwards, J.A. and M.D. Lampert (1993). *Talking data: transcription and coding in discourse research*. Hillsdale, N.J, Lawrence Erlbaum Associates.

Whilst this text provides some useful methods for transcription within discourse analysis, it only touches very briefly on transcription of multimodal, audiovisual media. No transcription method is given, but a discussion of the advantages and difficulties of using audiovisual data. Due to the age of the text, multimodal transcription was not such a current issue, and the technology referenced in this text is now outdated.

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

Flewitt presents an overview of multimodal transcription, its challenges and what it entails in an introductory style. The text is useful as a broad, but shallow discussion of the topic of multimodal transcription, though it does go on to present some examples of transcription methods. There is not so much discussion of the restrictions of these particular methods, but it is an adequate collection for a researcher to begin collecting examples of multimodal transcription. Some of the transcription methods provided are somewhat outdated, but their presence still allows us to analyse why they were deemed ineffective.

Flewitt, R.S., R. Hampel, M. Hauck and L. Lancaster (2009). 'What are multimodal data and transcription?' *The Routledge Handbook of Multimodal Analysis*. ed. C. Jewitt. London, Routledge: 40-53.

This text provides a methodical discussion of the requirement for multimodal data analysis, dealing both with difficulties the practice entails, transcription methods themselves and how multimodal transcription might affect theorycrafting. Flewitt's text examines these concepts in relation to their use in transcribing classroom interaction, but it is easy to see how the conclusions drawn can translate to other uses. Flewitt does provide some concrete examples of multimodal transcription amidst wider discussion of the requirements of the mode, thus her method of multimodal transcription may be important within a study of available methods. Moreover, a wider discussion of data collection methods allows us to see where multimodal transcription may fit within a wider research project.

Fouse, A., N. Weibel, E. Hutchins and J. Hollan (2011). 'ChronoViz: a system for supporting navigation of time-coded data'. *PART 1: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*. Vancouver, BC, ACM.

ChronoViz is not so much of a transcription software as an aid to visualising multimodal data on screen. Though it does allow for manual text transcription in its editor, is has no resources for complete export of data. Instead, it specialises at displaying data simultaneously.

Lecumberri, G., M. Luisa., and J.A. Maidment (2000). *English transcription course*. London, Arnold.

This text is an accessible guide to transcribing speech in English based on IPA sounds in RP English. Whilst it does not deal with multimodal transcription, the methods described for transcribing speech may be useful for transcribing the speech within a video. Where speech will be transcribed as part of multimodal data, we must decide which conventions are to be used, and this text may provide some insight as to

effective methods.

Guichon, N. and C.R. Wigham (2015). ‘A semiotic perspective on webconferencing-supported language teaching.’ *ReCALL* 28(1): 62-82.

Although Guichon and Wigham's text is based in a semiotic perspective, it can be of some use to us in its discussion of multimodal transcription. Perhaps the most important part of this text is its discussion of the ELAN software used to transcribe multimodal texts by separating their various modalities. Having an exemplified transcription using this software will be beneficial to analysing its effectiveness for further multimodal transcription. Moreover, the text seeks to distinguish different types of action within a transcript, and different camera shot types. These subtypes may be useful for future transcription.

Helm, F. and M. Dooly (2017). ‘Challenges in transcribing multimodal data: A case study.’. *Language Learning and Technology* 21(1): 166-185.

Helm and Dooly focus their paper on multimodal transcription of online interaction, specifically considering an online communication resource called 'Soliya'. The text is useful as the multimodal text they aim to transcribe involves multiple communicative elements such as speech, text speech, gaze, and gesture. Therefore, the methodology the authors discuss is useful as it discusses selection of parts to transcribe, and goes on to present some transcription methods. Whilst it is important to bear in mind that the transcriptions provided are designed specifically for the program Soliya, their presentation may be useful in constructing a transcription method for similar multimodal data.

Hwang, H. and D. Matsumoto (2016). ‘The cultural bases of nonverbal communication’. *APA Handbook of nonverbal communication*. Washington, DC, American Psychological Association.

Hwang and Matsumoto provide a thorough guide to physical and physiological communication cues to the major emotions. Alongside other texts considering communication cues, this text should provide insight as to our transcription requirements.

Kipp, M. (2014). ANVIL: A Universal Video Research Tool. *Handbook of Corpus Phonology*. ed. U.G.J. Durand and G. Kristofferson. Oxford, Oxford University Press: 420-436.

ANVIL is a transcription software with an innovative interface similar to a timeline transcription. Data processed through ANVIL is well placed for transcription, and it can also display audio data as waveform or pitch contour. ANVIL is one of the most modern softwares available.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, Lawrence Erlbaum Associates.

CLAN is a piece of computer software designed to aid researchers in transcription of multimodal data. Featuring extensive resources to link sections of transcript to sections of the referent audio or video media. CLAN also aims to provide easy navigation of the media. Competence in using CLAN software could be useful as a means through which to produce a transcription which is flexible in its use for analysis.

MacWhinney, B. and J. Wagner (2010). 'Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository.' *Gesprachsforschung: Online-Zeitschrift zurverbalen Interaktion* 11: 154-173.

This text is a companion to the CLAN software, moving through an overview of what the software aims to accomplish to detail on the finer aspects of CLAN's functions. If a researcher were to decide upon using CLAN as a transcriptive aid, this text would be invaluable in assisting the user in beginning to use CLAN.

Mondada, L. (2001). Conventions for multimodal transcription. No further information available.
<Available at https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/ Mondada_conv_multimodality.pdf>

In this text, Mondada creates a thorough orthographic transcription method for multimodal media. Focusing especially on how actions could be transcribed with their temporal information in a conventional, vertical transcript. Mondada aims to create new conventions for transcribing multimodal information, and whilst she does succeed in covering the majority of what is required, the result is a rather complex system relying on a multitude of symbols. It would be required, therefore, to learn the transcription system, and is not accessible to a lay audience. Mondada does go on to demonstrate how images might be added to the transcript to provide more precise demonstration of action. In all, this is a useful text in creating a multimodal transcription quite different to others, rooted in traditional Conversational Analysis techniques.

Mondada, L. (2016). 'Challenges of multimodality: Language and the body in social interaction.' *Journal of Sociolinguistics* 20(3): 336-366.

Arguably a spiritual successor to Mondada's 2001 text establishing her conventions for multimodal transcription, this text demonstrates its usage in a number of situations. Mondada takes her transcription technique, and applies it to a series of multimodal texts for transcription. Each is accompanied by a discussion of how her transcription deals with the more challenging areas of each text, such as how we could transcribe sensory information we presume a participant to be experiencing. This text serves as a useful companion to Mondada's initial text outlining her multimodal transcription conventions.

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

This text aims to create a transcription method for the online streaming service Twitch, used mostly for streaming videogames. Although Recktenwald's transcription method is specifically designed for Twitch, some of its merits could translate to translation of other multimodal data. Recktenwald opts for a 'play-script format' vertical layout, and requires transcribed information for in-game events, the streamer pictured on camera and the accompanying text chat box. Twitch is clearly not the only multimodal communicative platform set out in roughly this way, and is distinctly similar to many videoconferencing platforms. Thus, this text and its transcription strategy may be useful in informing more general guidelines on multimodal transcription.

Remland, M.S. (2000). *Nonverbal communication in everyday life*. Boston, Houghton Mifflin.

Remland's text is useful in tandem with other resources on nonverbal communication, as it serves to reinforce the communicational cues associated with various emotions. Furthermore, Remland's consideration of turn taking contributes another element to potential markers for arrogance, as we might measure arrogant behaviour through how speakers manipulate their turns within a debate.

Richmond, V.P., J.C. McCroskey and M. Hickson (2008). *Nonverbal behavior in interpersonal relations*. Boston, MA, Pearson/Allyn and Bacon.

This text is useful for its thorough discussion of nonverbal communication in interpersonal settings. Chapters are dedicated to various nonverbal communication methods, and each chapter contains detailed information regarding how such nonverbal communication manifests, its effects, and salience within communication. This text is also especially useful regarding vocal behaviour, giving a range of cues for various emotions, but also five key findings on vocal cues. This text will be useful in identifying features of potential arrogance and anger which merit transcription.

Satar, H.M. (2013). ‘Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze.’ *ReCALL* 25(1): 122-142.

Satar's paper is a thorough analysis of gaze in the multimodal format of gaze in the setting of online videoconferencing through a webcam. The text develops a set of gaze types which can be employed by participants, and Satar also presents methods of transcribing the gaze actions from the video data. Satar's text focuses closely on one element of multimodal transcription which is often not given so much attention in other literature on multimodal transcription. His paper, therefore, will be invaluable in demonstrating how to transcribe gaze within a multimodal setting.

Schmidt, T. and K. Wörner (2009). ‘EXMARaLDA – creating, analysing and sharing spoken language corpora for pragmatic research.*’ Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 19(4): 565-582.

EXMARaLDA is a transcription software with a number of different export layouts and formats. It uses a useful interface something between a tabular and timeline transcription method, and has a number of integrated transcription conventions. Thus, it is well equipped for multimodal transcription.

Sloetjes, H., and Wittenburg, P. (2008). ‘Annotation by category – ELAN and ISO DCR’. *Proceedings of the 6th International Conference on Language Resources and Evaluation LREC*, Marrakech, Morocco.

ELAN is a transcription software using a timeline transcription layout. Like most of the other software considered, it can express audio data in a number of ways and can replay video data within its interface. ELAN is one of the best programs for creating an complex transcription with many different modalities as its use of the timeline style ensure it maintains clarity.

Taylor, C. (2003). ‘Multimodal Transcription in the Analysis, Translation and Subtitling of Italian Films.’ *The Translator* 9(2): 191-205.

In this text, Taylor takes Thibault and Baldry's multimodal transcription method, and seeks to refine it by slightly restricted the material transcribed. This article is useful, therefore, in the progression of multimodal transcription towards a concise and accessible transcription method. Through comparison between this article and the original transcription method, we can judge whether the restriction of some information entering the transcript benefits the resultant transcript or not. Furthermore, it offers another set of examples of completed multimodal transcription to be considered within an overview of methodologies.

Taylor, C. (2013). Multimodality and audiovisual translation. *Handbook of Translation Studies*. ed. Y. Gambier. Amsterdam, John Benjamins: 98-104.

In this chapter, Taylor gives a brief overview of multimodal transcription combined with two examples. One of these examples is an audiovisual text, using a similar transcription method to Baldry and Thibault's foundational method. Whilst this text is useful in portraying Taylor's refined transcription method, other papers by the author offer a more thorough discussion of the requirements and challenges of multimodal transcription.

Taylor, C. (2016). ‘The multimodal approach in audiovisual translation.’ *Target. International Journal of Translation Studies* 28(2): 222-236.

Taylor's text responds to ongoing discussion about multimodal translation and transcription. The text analyses these concepts largely from a semiotic perspective, moving through the various elements of a multimodal text before considering transcription directly. Thus, a relatively small portion of the text is given over to transcription directly, but an example transcription is given nonetheless. Taylor considers the limitations and difficulties of aligning all the semiotic resources deployed by a text in a multimodal transcription. Thus, this text is useful in demonstrating another means of multimodal transcription combined with further discussion of that method.

Woods, D. and C. Fassnacht (2017). *Transana*. Version 3.10. Wisconsin, Spurgeon Woods LLC. <Available at https://www.transana.com> [accessed 20th June 2017].

Transana is a premium transcription software with a focus towards coding and categorising data. Transana allows multiple users to contribute to the document from different locations in real time. That aside, it has few unique features.

Computer Software (also included above)

Boersma, P. and D. Weenink. (2017). Praat: doing phonetics by computer.

Praat is an advanced audio analysis and manipulation software. Unlike other transcription software considered by this review, Praat does not seek to aid with overall transcription, but to offer extensive tools specifically relating to audio analysis. Praat is likely to be useful in conjunction with one of the other pieces of transcription software when chosen.

Fouse, A., N. Weibel, E. Hutchins and J. Hollan (2011). 'ChronoViz: a system for supporting navigation of time-coded data'. *PART 1: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*. Vancouver, BC, ACM.

ChronoViz is not so much of a transcription software as an aid to visualising multimodal data on screen. Though it does allow for manual text transcription in its editor, is has no resources for complete export of data. Instead, it specialises at displaying data simultaneously.

Kipp, M. (2014). ANVIL: A Universal Video Research Tool. *Handbook of Corpus Phonology*. ed. U. G. J. Durand and G. Kristofferson. Oxford, Oxford University Press: 420-436.

ANVIL is a transcription software with an innovative interface similar to a timeline transcription. Data processed through ANVIL is well placed for transcription, and it can also display audio data as waveform or pitch contour. ANVIL is one of the most modern software pieces available.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, Lawrence Erlbaum Associates.

CLAN is a piece of computer software designed to aid researchers in transcription of multimodal data. Featuring extensive resources to link sections of transcript to sections of the referent audio or video media. CLAN also aims to provide easy navigation of the media. Competence in using CLAN software could be useful as a means through which to produce a transcription which is flexible in its use for analysis.

Schmidt, T. and K. Wörner (2009). 'EXMARaLDA – creating, analysing and sharing spoken language corpora for pragmatic research.*' Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 19(4): 565-582.

EXMARaLDA is a transcription software with a number of different export layouts and formats. It uses a useful interface something between a tabular and timeline transcription method, and has a number of integrated transcription conventions. Thus, it is well equipped for multimodal transcription.

Sloetjes, H., and Wittenburg, P. (2008). 'Annotation by category – ELAN and ISO DCR'. *Proceedings of the 6th International Conference on Language Resources and Evaluation LREC*, Marrakech, Morocco.

ELAN is a transcription software using a timeline transcription layout. Like most of the other software considered, it can express audio data in a number of ways and can replay video data within its interface. ELAN is one of the best programs for creating an complex transcription with many different modalities as its use of the timeline style ensure it maintains clarity.

Woods, D. and C. Fassnacht (2017). *Transana*. Version 3.10. Wisconsin, Spurgeon Woods LLC. <Available at https://www.transana.com> [accessed 20th June 2017].

Transana is a premium transcription software with a focus towards coding and categorising data. Transana allows multiple users to contribute to the document from different locations in real time. That aside, it has few unique features.