Representing Simultaneity in Polychannel Linguistic Events: A Multitrack Method for Transcription

Noah C.G. Johnson
University of Iowa

Abstract

Despite the advances in audio and video recording that have been made in the last century and the implications of the more recent advent of the internet and other information sharing systems, the study of linguistic anthropology requires us from time to time to present reproductions of recorded linguistic events in a textual format – a transcript. In addition, transcripts can also prove invaluable during analysis, making clear what we miss in the auditory or visual reproductions of a recording device. And yet, the transcript is not a neutral tool but, like a lens, it has the ability to distort everything that is seen through its paradigm. A discourse-centered approach to the study of language advocates that we seek to investigate linguistic events that occur in natural, dynamic and often messy situations. Unfortunately, our transcription methods, though now greatly bolstered by sophisticated technology, still lack methodological robustness when applied to polychannel interactions – or conversations involving multiple speakers – which are often typified by interruptions or overlapping. In this article, a method of transcription that relies upon the same organizing logic as multitrack recording is proposed, the process by which this method was developed is explained and additional directions for future research are suggested.

Keywords: transcription, multitrack, overlap, linguistics, methodology

Introduction

Beginning in the 1960s and carrying on to the end of the 1990s, the processes of transcription – up until then thought to be a rather straightforward and neutral affair – came under scrutiny in several landmark publications by linguistic anthropologists, sociologists and communication scholars who greatly problematized the seemingly dependable and innocent transcript (see Ochs 1979; Jefferson 1985; Goodwin and Heritage 1990; Sherzer 1992; and Bucholtz 2000). Though the insightful arguments from this time period greatly advanced our understanding of the strengths, limitations and concerns attendant upon the transcription process, they should not be an end to conversation on the methodological and theoretical problematics of transcribing living speech into text. Yet, even as technological advances in the last decade have made audio and video recording innovations available more than ever before, our
methodological advances have failed to keep pace, and the conversation has largely fallen silent.

Stemming from my own experience with the limitations of transcription methodology I present a model for transcribing conversations with multiple participants that are collected in an informal setting. This model is not dependent upon a particular technological device, such as a multitrack recorder, but is rather an attempt to optimize the translation of audio or visual records (up to and including multitrack recorders) into a textual format. As such, it should be seen as a methodological approach more concerned with how to properly use the space on the page than a recording technology 'how-to' (for a discussion on practical considerations of recording see Lapadat 2000; and Farnell and Graham 1998).

Beginning with the graduate seminar assignment that generated this model I will describe the several ways that I attempted to match conversation to transcription before formulating my approach, and the theoretical concerns that these attempts bring to light, rather than simply presenting the 'finished product.' In the same spirit of an evolving process I will then discuss the limitations of this model as well as some possible ways that it could be expanded or adapted for other research situations. In this paper I will discuss not only the considerations that guided the development of this transcription method, but also the inspirations that I took from my particular life-history which lead to its development. It is my hope that an inclusion of my process will demonstrate the stops and starts, U-turns and dead ends that I encountered in trying to make our current models of transcription work in scenarios that they are not suited for. The need to interrogate and reflect upon our methodology in transcription is far from complete, and advances in recording technology do not promise to solve these problems for us, but more often present different problems that it is our responsibility to consider.

Thinking Like a Student: The Assignment

Though I am a sociocultural anthropologist by inclination, as part of my course work in the PhD program at the University of Iowa, I enrolled in a graduate-level seminar on Linguistic Anthropology. Figure 1 is a twenty second excerpt of my first (rather amateurish) attempt to transcribe a linguistic event I had recorded as part of an assignment for that class. Purposefully, our seminar leader, Dr. Laura Graham, had arranged the syllabus so that this assignment would be completed before we had become well acquainted with existing transcription methods. It was her expressed hope that we would try to approach this assignment with 'fresh eyes', and that such an approach would generate interesting 'attempts.' It was understood that our transcriptions would not be perfect, and that the focus of the assignment was on process more than product.

As such, I came to this assignment, not informed by the limits of methods for
capturing and recording the information that takes place in a linguistic event, but rather from a perspective concerned with discourse-centered methods. Barbara Farnell and Laura Graham (1998) describe discourse-centered methods as a general turn in what is held to be important to current anthropological inquiry into the use of language, rather than a distinct school of thought. Central to this shift, “researchers pay close attention to how language is used in and across social situations, focusing particularly on “naturally occurring discourse” — that is, utterances that occur in the context of social interaction, in contrast to utterances specifically elicited by a linguist or ethnographer” (Farnell and Graham 1998: 411). Guided by this focus on dynamic, interactional discourses, I chose to record a family from the Midwestern United States during an informal get-together, as they played cards around the dining room table. To me, there are few linguistic events that are more interesting or more “naturally occurring” than the patter of conversation that typifies a card game amongst friends or family.

What I found, to my dismay, is that the transcript I first generated, and as shown in Figure 1, does very little to capture the actual atmosphere of that interaction, and even fails to properly represent the interpersonal dynamics of the conversation (or conversations) that took place. As I have learned more about transcription practices I have become aware that we lack the proper tools for systematically representing polychannel interactions — or conversations involving multiple speakers — particularly if speakers are interrupting or talking over each other to any great degree.

In part this discrepancy may be the legacy of what linguistic anthropology considered to be its intellectual territory before the discourse-centered turn. In critiquing this disciplinary history, Charles Goodwin and John Heritage pointed out that “within both linguistics and social theory, the actual behavior that occurs within interaction was viewed as disorderly, and indeed inherently defective... that actual talk was such a degenerate sample of ideal linguistic competence that linguists should ignore it entirely and work instead with idealized sentences constructed by the analyst” (Goodwin and Heritage 1990: 285). With the understanding that research goals can limit or even determine methodological innovation (Lapadat 2000), it should come as little surprise if models and methods were not developed to properly cope with those situations that do not come from an anthropologist’s armchair musings or the early ethnographer’s elicitations of participants’ speech acts. Yet, as Goodwin and Heritage further argue, we must recognize that “the stream of speech actually produced by a speaker in conversation” (1990: 287), though messy, is also endlessly interesting and central to many of the questions that we, as academics, would like to answer. Because of this, “discourse-centered work emphasizes the heterogeneous, multifunctional, and dynamic character of language use and the central place it occupies in the social
construction of reality” (Farnell and Graham 1998: 412), but is hampered in polychannel settings by the inadequate methodological tools that we have at our disposal to transcribe such events. In order to continue delving into such complex and dynamic interactions we must also work to develop tools more suited to the situations that we would seek to investigate. Clearly, we require models and methods that allow us to realize that “the basis of intersubjective knowledge in linguistics must be found in speech language as it is used in everyday life by members of the social order, that vehicle of communication in which they argue with their wives, joke with their friends, and deceive their enemies” (Labov 1972: xiii). This article is an account of the process I went through to create such a method of transcription, as well as a discussion of the related theories and methods that connect to this approach.

Figure 1. A twenty second excerpt from a transcription of six persons with familial ties playing cards. The names of participants have been removed and replaced with designators (e.g. “Prt4”)
Thinking Like an Actor: The Problematics of Simultaneity

The most ubiquitous form of transcription has always appeared to me to look like a script for a play, perhaps due to the fact that my undergraduate degree is not in anthropology, but is in theater. Because of this training and my personal life-history, as I used this transcription format in my first attempt, I became aware that the structure of this transcription falls victim to the same pitfall that is presented by dramatic scripts. A common mistake among beginning actors when they are interpreting a script is that they miss the points of overlap, which are not often written into the script. This leads to dialogues where first one actor speaks and then waits while the other actor speaks their own line - in its entirety - before they then deliver their next line. In a dramatic performance, the lines in the script are meaningful, and should not be trampled over, but without overlap at key points, the reading or performance that results is boring and without dynamic tension – precisely because it does not reproduce the characteristics of real conversations.

Though people may often take turns in an orderly manner, the danger of presenting conversation as something that is regulated and is turn-based is that it recommends models of transcription that are best suited for only that type of conversation. Elinor Ochs has remarked that “many types of sequences... are based on turn units. For the analysis of such sequences as well as other areas of concern, our transcript ought to display turn units in a systematic manner... In the work on adult conversational behavior in middle class Anglo society, turns are considered to be verbal units bounded either by the talk of another speaker or by a significant pause” (Ochs 1979: 69). I come from what I consider to be a “middle class Anglo society,” and the instance that I recorded, which was amongst a family of middle class Anglos, was far less organized than what Ochs describes. In her defense, she wrote this passage over thirty years ago, and perhaps the interactional norms for middle class Anglo discourse have changed. Regardless, the linguistic event that I feel more closely resembles the conversation had over cards is one described by John Haviland in his work with the Tzotzil:

In the rapid flow of conversation, speakers must continually monitor each other's turns, so that they know what will count - in the moment - as agreement or disagreement... the two speakers are in an intricate dance of doubt, agreement, and confirmation, marked by evidential particles that track the state of discursive play at each moment (1996: 56).

Haviland also noticed that the very act of transcription seemed to work in such a way as to effect “a smoothing of interactional edges” (1996: 61) as the story was made more coherent in the text, losing the struggles and missteps that characterized the interaction as he had experienced it in the role of participant and recorder (Haviland 1996: 57). These “interactional edges” are also the qualities that lend a dramatic
performance its immediacy and confer upon it a certain verisimilitude. In polychannel linguistic events, the orderly representation of text can confer a sense of turn-taking, regulated by unspoken but universally understood and perpetually agreed upon rules that may be misleading. This was the major concern that I faced in developing a transcription of the linguistic event that I had chosen to record.

As Goodwin and Heritage point out, “One of the best known, but most misunderstood, contributions of [Conversational Analysis] to the specification of sequential organization can be found in its analysis of turn-taking” (Goodwin and Heritage 1990: 289). It is easier and far preferable to believe that conversations systematically follow a set of rules and will lay themselves open to analysis without a proverbial fight. Unfortunately, as the scholars who developed Conversational Analysis (or CA, as it is commonly referred to) realized early on, the process of actual conversation is not maintained only by repair in the case of overlap, but actually depends upon instances of overlap as “one of the systematic products of the system” (Goodwin and Heritage 1990: 290). Thus, CA is specifically interested in the effect that such interaction, as these elements denote, has upon the participants, as well as the effect the participants’ choices have on the interaction. What CA scholars have found is that there is a great importance in the “sequences of actions that participants perform to create and manage meaning between themselves” (O'Sullivan 2010: 21) which leads us to pay greater attention to the priority and temporality of speech enunciation (Goodwin and Heritage 1990; O’Sullivan 2010).

As such, any model that would seek to put to text the interchange between multiple speakers, who consistently and constantly overlap, would require a way to spatially organize and demonstrate the instances of overlap and of repair in a way that would be legible and useful to analysis.

Thinking Like A Recorder: The Problematics of Entextualization and Formal Structures

Though linguistic anthropology has advanced to the point where it recognizes the importance of studying speech in very contextualized, situated, and “natural” settings, the methodological tools have not kept pace. In addition to issues of epistemology, this problem may also stem from the way that technology-- far from alleviating this problem -- has exacerbated it. The fact remains that we must often find a way to write down what we have recorded despite the growing presence of online journals that can be linked to actual recordings. Likewise, though an audio or visual recording device can offer an unparalleled fidelity, the ability to write out the course of an interaction is still an indispensable component in the deep analysis of a linguistic event, and a necessity for any scholar who would like to publish in most academic settings. As Ochs remarked, “the utilization of mechanical means of recording may appear to eliminate some [methodological] problems... [but, in fact,] the problems of selective observation are not eliminated with the use of recording equipment. They are
simply delayed until the moment at which the researcher sits down to transcribe the material from the audio- or videotape” (Ochs 1979: 44).

From the perspective that “a transcription format should be related to the goals of analysis” (Farnell and Graham 1998: 427), I found myself frustrated by the way that my initial attempt had failed to replicate (or even remotely approximate) the dynamic interactions and multiple layers of conversation that had been present in the actual linguistic event. I did not have a method of transcription that could reliably represent these characteristics, let alone capture the simultaneity of the actual speech utterances of my participants on the page. First and foremost, all forms of transcription are meant to represent the discourse that they report upon. In this case, the method of transcription I first employed did not represent the discourse, but in fact, misrepresented it to the point of being misleading and almost incomprehensible.

Following from Richard Bauman and Charles Briggs (1990) we can see that any form of transcription is a practice in entextualization-- an act whose specific aim is to remove or detach text from its context and render it portable. Greg Urban notes, “If entextualization is understood as the process of rendering a given instance of discourse a text, detachable from its local context, replication is one way, seemingly, of implementing detachment. It tries to portray the textual as opposed to contextual aspects of the original discourse, and to capture, thereby, the decontextualized or polycontextual meanings associated with it” (Urban 1996: 21). In decontextualizing the text from the speech act and rendering it portable to other settings, the transcriber accomplishes an act of legerdemain, insisting that what was said in a specific contextualized linguistic event speaks also to other events, and thus has a polytextual meaning or set of meanings that exist beyond the context in which it was generated.

Through entextualization practices then, all forms of transcription are composed of choices made by the transcriber about what will be included, what will be removed, and what will be added to the text as it is removed from a linguistic event into a textualized form of writing (Ochs 1979; Haviland 1996; Urban 1996). These decisions are acts of power both politically and intellectually (Bucholtz 2000), but are to a certain degree unavoidable if the aim is to textualize a spoken discourse. These practical limitations are not only methodological but also greatly shape theoretical concerns as well as the ability to carry out analysis. Interestingly, these problems appear to be systemic, inherent in the form, rather than simply reflecting the shortcomings of the researcher.

In their work with transcribers drawn from the same communities which they studied, both Urban (1996) and Haviland (1996) noted that these transcription practices of elaboration, removal, and normalizing were generated by interpersonal relations, but also by the very form of representation that transcripts call for. Though working from different considerations, the practices of transcribers from both communities effected the repair of inconsistencies in the transcription, presenting the sense of an orderly and
coherent performance (Haviland 1996, Urban 1996). Urban noted that metadiscursive comments were not reproduced by transcribers (1996: 33), and Haviland remarked that the Tzotzil he worked with were also likely to remove contextual details of the setting – including participants that were not considered central to the discourse by the transcriber (1996: 67) – as well as the give and take between conversation participants, summarily smoothing over and homogenizing the interaction. “There is no longer a negotiable universe of discourse... Any text resulting from writing eliminates the warmth of the sun that the original conversants could point to... the tension between tellers... and the secret animosities between rivals... that were all too obvious to us gossipers” (Haviland 1996: 72).

Though we might expect the creation of these “monologic narratives” (Haviland 1996:69) from researchers raised and trained in the positivist paradigm of Western academe, Haviland clearly demonstrates that “These speakers, whose experience with reading and writing in any language (let alone their own) is next to nil, nonetheless by their practice are able implicitly to indicate what a text should be like” (Haviland 1996: 49).

**Thinking Like a Musician: Multitrack Transcription of Polychannel Linguistic Events**

Upon realizing the shortcomings of my initial attempt to transcribe the conversation I had recorded, the smart thing to do would have been to read up more on the theory of transcription methodology. In short order I would have become aware of the problematic possibilities inherent in transcription that I have presented so far (by no means exhaustive). Fortunately, I wasn’t too smart, and instead of examining the literature (which after all, went against the spirit of the assignment), I fell back on what I know most.

For over two decades I have been a musician and in the last decade or so I have worked almost daily with multitrack recording devices, both physical machines and computer-based software. From this perspective, I realized that it would have been most advantageous to have recorded each of my participants on a separate track, using a multitracking device and dedicated microphones. Unfortunately, before these methodological and theoretical concerns became apparent, I only had my trusty audio recorder with its omnidirectional microphone when I recorded the conversation. Though I had already lost the chance to record each channel of speech separately, I realized that I could at least multitrack the utterances of my participants when I went to encode the data textually.

Though Steven Feld and Aaron Fox characterize the flow of information between music theory and linguistics to be primarily one-sided, with theoretical models from linguistics being applied to music (1994: 26), the contributions that musical theory has made to transcription practices in linguistic anthropology should not be minimized. As Joel Sherzer (1992) notes, some of the most progressive ideas that were developed for
the transcription of Native American studies in linguistic anthropology came from the utilization of poetic and musical genres.

As early as 1965, Dell Hymes was concerned that the way previous scholars had represented transcriptions of Native American performance were not adequate (Hymes 1965). These considerations were picked up by Dennis Tedlock (1978) who paid special attention to the “expressive features of the voice, especially pause and pitch patterns, in his representation of Zuni narratives” (Sherzer 1992: 427). The adoption of line breaks and dots to represent pauses, upper-case letters to mark louder speech, punctuation to denote the rise and fall of pitch, and dashes to represent vowel lengthening (Farnell and Graham 1998: 425-426), to name but a few devices, all served to make the text on the page “look to the eye how it sounds to the ear” (Schenkein, 1978: xi). In turn, authors such as Joel Sherzer and Sammie Wicks adapted methods from music theory (Sherzer and Wicks 1982), and scholars such as Laura Graham (1984, 1995) made use of musical staff to represent Shavante/Xavante wailing and chanting, using regular scoring to great effect. Though the Western musical staff offers interesting ways forward with the model I present here, at this point I would like to focus on its qualities as a model inspired by multitracking, as I developed it initially for the assignment under discussion.

From my background with multitracking it seemed only natural to represent polychannel linguistic events with a multitrack representational model in my transcription process. Some of the features that recommend this process are the ability to clearly and easily demonstrate instances of overlap, slippage and repair. This is important, because even if this does not return the feeling of “the warmth of the sun”, as Haviland (1996) put it, this format allows even a linguistic layman to grasp some of the give and take evoked by the chaotic and messy characteristics of a “naturally occurring” event.

It is the capture or reproduction of these interactional dynamics that I felt most important to the representation of my recorded conversation and the model that I developed was guided by this concern more than the multiplicity of others that could have directed me. First and foremost, the decision of how to transcribe a conversation should recognize what the researcher seeks to investigate, while also honestly observing and remarking upon the implications of these choices (Bucholtz 2000). In effect, this model is but one possibility amongst many others, which is specifically suited to a particular type of linguistic event and particular theoretical concerns. As in anything else, the most expedient way to accomplish a task is to select the proper tool for the job at hand. For polychannel linguistic events, I recommend the model herein described.

Out of consideration for space, I have here included only the same twenty second excerpt from the conversation that I began this article with, and though it is brief, this excerpt sufficiently demonstrates the mechanics of this transcription model.
One who is familiar with Jeffersonian transcription (1985), which is regularly utilized by Conversation Analysis researchers (Niemelä 2010), will see some similarity between my method and that of CA studies. The differences that exist are important ones and are predicated by my own research concerns. Noticeably, I use a much truncated system of representational punctuation and other devices. The great advantage of Jeffersonian Transcription, devised by Gail Jefferson in the 1970’s, is that “the symbols and format used enable the user to access details essential to understanding what is going on in an interaction, in ways which stress the social, embodied nature of language” (O’Sullivan 2010: 26). This advantage, which allows for the depiction of stress, hesitations, and other verbal anomalies, is balanced by the necessity that one be familiar with an expansive suite of symbols and stylistic markers.

I also pay a greater attention to the specificity of spatial calibration than is conventional in CA and thus I have also made use of the Courier New font rather than Times New Roman or other “reader-friendly” proportional-width fonts because the fixed-width font of Courier New allows for the precise alignment of characters to show “overlaps and latching” (Bucholtz 2000: 1453).

In addition, I have opted to focus on the morphological level of accuracy (Preston 1982), rather than representing any derivation from the “ideal” phonetic production of utterances. Although such fine-grained considerations could be instructive given the right research aims, here they would only distract from the goals of my analysis, which was primarily concerned with representing the interrelation and simultaneity of utterances. Given this primary concern, I have taken temporality as my guiding principle in the arrangement of space in this model. “Ideally, we want our transcript to meet practical as well as theoretical considerations... We want it to encode not only prior and subsequent behaviors, but cooccurrent and interoccurrent behaviors as well” (Ochs 1979: 59).

Such a consideration is advocated by Ochs (1979) in her discussion of how transcriptions display priority through the acculturation of Western literary ideals:

As members of a culture, we, the transcribers, bring into the transcription process a biased spatial organization... Across many cultures, there is a convention whereby written language is decoded from the top to the bottom of each inscription. The reading of conversational transcripts takes no exception to this norm, and, generally, the history of a discourse is unfolded in a downward direction. Utterances that appear below other utterances are treated as occurring later in time (1979: 45-46).

Though it is a politicized decision – which I readily admit (Bucholtz 2000), I feel that it is most expedient to conform to this bias of spatial organization in the arrangement of this model so that it is easily comprehensible to those most likely to come in contact with it or to make use of this model. As Ochs further notes, “The
European culture of literacy socializes its members to encode ideas not only from top to bottom, but from left to right of the writing surface... Leftness is linked with priority and also with inception of a statement or entire discourse” (1979: 49).

**Figure 2.** A twenty second excerpt from a transcription of six persons with familial ties playing cards, now in the suggested format, with the spatial arrangement of words on the page used to represent temporality. The names of participants have been removed and replaced with designators (e.g. “Prt4”).
Thus, when reading this model, the temporality of utterances can be discerned by reading from left to right within each ‘frame’ (aided by the time code in the top bar that corresponds to the beginning of each utterance), noting that the frames closer to the top of the page represent sections of the discourse that occurred before the frames that are placed below it. Further, the implications of the top-to-bottom bias and my concerns with temporality also organize the order in which speakers and their assigned tracks are represented within the transcript. Beginning with the start of the record, the person who speaks first (temporally speaking) is assigned the first track, which is the track that is positioned highest within the frame. As the record proceeds, other participants are assigned tracks in descending order, predicated by the relative temporality of their first utterances. These assignments are fixed and in each frame the track assignments are represented even if the speaker does not contribute vocally to the conversation within the duration represented by that frame. In this respect, my model differs from other models of transcription, such as those conventionally used in Conversation Analysis (Niemelä 2010), sacrificing a more strident space management, in order to provide consistent representation of the interlocutors involved in the linguistic event (Haviland 1996).

Understandably, these are not unproblematic decisions. Due to my research aims, I have decided the left should represent primacy in terms of temporality, and that the ranking of participants from top to bottom should also be dictated by temporality (in the order of utterances as they are performed), rather than any other extralinguistic concerns such as age, relationship, etc. In large part this is a model reflecting the research concerns to which it is being applied: the exploration of speech in overlapping and simultaneous generations; time and temporality being the elements of most practical interest, and from which all other inferences may be built.

These decisions are from my own biases, though I have reasoned through these choices in a deliberate manner with my eyes upon the intended uses I perceive this model to be best suited for. A limited, but important panacea to the problems that arise from the unavoidably subjective construction of a transcription method is a reflexive engagement with the transcription. “A reflexive transcription practice, then, is one in which the researcher is conscious of her or his effect on the unfolding transcript, and the effect of the transcript on the representation of speakers whose discourse is transcribed” (Bucholtz 2000: 1462).

**Drawbacks and Limitations of the Model**

The first drawback to this method that one may notice is that it is very space intensive and that much of the space within a frame is blank. I have found it expedient to maintain these open tracks in each frame for the sake of consistency, and to make the record more easily readable. As Ochs says, “one of the important features of a transcript is that it should not have too much information. A transcript that is too
detailed is difficult to follow and assess. A more useful transcript is a more selective one... the transcript should reflect the particular interests- the hypotheses to be examined- of the researcher” (Ochs 1979: 44).

With this in mind, it should be apparent that this method of multitrack transcription is only suitable when other forms of transcription are not adequate to the job. The more common transcription style that I have previously referred to as dramatic script (and which I may have seemed to earlier malign), is actually quite a useful approach when one is dealing with a single speaker, or multiple speakers that are politely and consistently respecting the turn-taking sequence. Also, this method does not recommend itself, as it has thus far been presented, for the translation of one language into another. Rather, a transcription with either a side-by-side layout, or a linear interspersing (Sherzer 1992; Haviland 1996) of both the record in the original language and the target language would be preferred.

If one were interested in describing and relating the tonality of a sing-song performance of speech I would not recommend this model as presented thus far. Rather I would suggest that they consider their intended audience and choose from some of the fine methods already existing, to which I referred earlier. Specifically, the model of representation that was developed by Wicks in the article she co-authored with Sherzer (Sherzer and Wicks 1982) may be one of the more ingenious models that I have seen for representing pitch and temporality - chiefly because it can be so easily apprehended.

This list of shortcomings is far from exhaustive, but I leave the greater enumeration to be revealed by those who may make use of this model. Any further faults, will invariably be found in the field as the model fails to live up to specific tasks. For the purposes of my use, I have found the model to be well-suited for transcribing polychannel linguistic events in a way that is easily readable and open to analysis.

Additional Direction for Future Studies and Further Adaptation of the Model

Even in the first presentation of this model to my seminar group in Linguistic Anthropology some adaptations that could be incorporated into my model became apparent. One of my colleagues, Elias Gonzalez, used a very similar model to transcribe polychannel communication with the useful addition of continuous lines where participants were not speaking (2012, in class discussion). Though I had concerns about cluttering the page unnecessarily (Ochs 1979), I found that this simple device assisted in orienting the reader to the presence of momentarily silent participants (Figure 3).

This is a useful device if one is concerned with the same shortcomings of transcription that also bothered Haviland (1996) regarding the effacement of participants who are not actively speaking but nonetheless add to the contextual situation of the linguistic event. With the addition of these “scoring lines” by Gonzalez,
the model begins to resemble a musical staff more than a multitrack, calling to mind
the work of scholars such as Graham (1984, 1995) who have made use of the Western
musical staff. If one had an interest to incorporate an analysis of tonality and pitch
into a transcript using the multitracking method (and assuming that one is dealing with
speech that is not performed in a sing-song affectation or a true musical performance),
I would suggest that one adopt the pared-down three-bar staff that T.M. Scruggs

![Figure 3. Dashes are added to the model to allow easier orientation for the reader
in a method developed by Elias Gonzalez and reproduced here with his permission.
The use of this model could reasonably be referred to as the Johnson-Gonzalez
method (in contrast to the Johnson multitrack method of transcription discussed in
this article). The names of participants have been removed and replaced with
designators (e.g. “Prt4”).

By necessity then, this would create a multimodal representation of the
discourse for each participant. Within each track, every participant would have
multiple “channels” (in this case only two), relating different types of data that would
still be rendered in a legible format, constrained and regulated by the calibration of the
information to the space on the page so that the analysis of this information can be
examined in terms of temporality.

Pushing the ideas of this configuration even further, one could incorporate even
more analytical information or modal representation. Sherzer made use of multilineality
to represent grammatical elements of Kuna performance (1992: 431-432). Alternatively
(or additionally!) if one were concerned with representing movement, positioning or
gestures that occur within the discourse, they could make use of a separate
multimodal channel to record such behavior. As Ochs notes “nonverbal behavior may
be an alternative rather than an accompaniment to verbal behavior” (Ochs 1979: 52;
see also Farnell and Graham 1998: 433). Though nonverbal behavior may prove
difficult to write within the same space on the page as verbal behavior (Ochs 1979) one
could make use of a modified Labanotation (Farnell 1999) on its own multimodal
channel. Because Labanotation is written from the bottom-up, incorporation into this
model of multitrack transcription could prove highly problematic and would likely
require a diligent and inspired methodologist’s hand.
Such a wealth of information for each speaker in the conversation would be labor intensive to create, and perhaps more difficult to read, but the analytical benefits of such a representation may still persuade a scholar with very exacting research aims to adopt such an approach. The advantage of a multitrack transcription with multimodal channels would be to allow for all the information to be viewed and analyzed within the same model – upon the same page – held together by considerations of temporality.

Much like the list of the presented model’s shortcomings, this list of possible adaptations and elaborations is incomplete in a way that can only be filled-in by real world application and field testing that will be carried out upon the model. I recognize that this model, like all models is not “a foolproof transcription system that forecloses ideological positioning” (Bucholtz 2000: 1461). So it is my hope, that if anyone should find this model desirable for their use and modification, that they share their experiences, so that we may better refine and task the model to a multiplicity of situations.

Conclusion

Though transcription is often presented as a transparent rendering of a events – a faithful record of the facts (Ochs 1979) – it is seen upon closer inspection that the practice of transcription and the choices that a transcriber must make are politically charged (Bucholtz 2000), socially situated and generative of influences upon the way we view data (Farnell and Graham 1998: 424; Ochs 1979), and can greatly impact and shape our analytical and theoretical endeavors (Sherzer 1992: 426). As I have aimed to show, this model of representing polychannel discourse is no less biased than any other model, and is potentially a politically, as well as a theoretically, loaded instrument (Bucholtz 2000).

In recognition that all models of transcription will err as well, though in different ways, we are left with little alternative, but to make sure that our tools are suited to our tasks. It is my hope that I have conducted the description of my choices in developing this transcription model in a way that makes my guiding principles and their considerations transparent to the reader. Not only in the name of honesty, but also because I believe that the ability to see the “inner-workings” of the theoretical concerns that directed the development of this model will also offer easily grasped handles for its further development and adaptation by those who would make use of it in other situations.

In recognizing which issues are worth our investigation, I take my cues from such authors as Sherzer, when he says, “Discourse is at the heart of language-culture-society relations. For this reason we must take seriously the study of discourse, in all of its aspects” (Sherzer 1992: 439). Often these aspects can be messy or difficult to grapple with if our analytical and methodological tools are found wanting. As I had chosen a conversation not governed by turn-taking conventions and with multiple
overlapping, I found it necessary to develop an adapted approach for studying the kind of discourse I feel is worth investigating: raucous, dynamic and informal linguistic events, with multiple participants.

It is my desire that this model might find a home in the methodological toolboxes of up-and-coming anthropologists, alongside the more traditional and well-established models of transcription. It has a very specific set of uses that it can be turned to in its present form, but I would urge anyone who chooses to use this model to also think beyond my limited horizons and let their inquiry shape this model to their uses, rather than shaping their research to this tool (or any other).

As students who are just now pushing out into the field, or hope to shortly, I feel that this model of multitrack transcription and the process I used to develop it, may have more to offer to the readers of this journal than other audiences. My journey to develop this model also points out the importance of a liberal arts education to a certain degree. Were it not for my background in theater and my continuing familiarity and work with music and recording technologies, I may never have thought of this model. In fact, it seems to me impossible to have conceived of it without such a grounding. If there is any lesson to be taken from my example, I would hope it is the importance of keeping our minds open to the possibility that inspiration, even along analytical or methodological lines, may come from all quarters.

Acknowledgments: I would like to thank Elias Gonzales for his contribution to the further refinement of the proposed model, and Laura Graham for putting the assignment in my way that generated this model and for pointing me in the direction of key materials that guided my theory in this piece. As well, this article has greatly benefited from the attention of the reviewers at Student Anthropologist and I would like to thank them for making this a stronger article.

Author Biography: Noah C.G. Johnson is a graduate student of sociocultural anthropology at the University of Iowa in Iowa City, IA. He has conducted research in Okinawa, Japan, studying karate as a localized cultural practice that has become an international phenomenon by working with transnational karate practitioners. Noah’s larger research interests include the issues of representation, politico-cultural discourse and struggle, the place of ritual and the process of self-identification, culture-loss and cultural survival, assimilation practices “from below”, and practical applications of anthropology.
References

Bauman, Richard, and Charles Briggs

Bucholtz, Mary

Farnell, Barbara

Farnell, Brenda, and Laura R. Graham

Feld, Steven, and Aaron A. Fox

Goodwin, Charles and John Heritage

Graham, Laura R.

Graham, Laura R.

Haviland, John

Hymes, Dell

Jefferson, Gail

Labov, William

Lapadat, Judith

Niemelä, Maarit
Ochs, Elinor

O'Sullivan, Terry

Preston, Dennis R.

Schenkein, Jim, ed.

Sherzer, Joel

Sherzer, Joel, and Sammie Wicks

Tedlock, Dennis

Tedlock, Dennis

Urban, Greg