

# Transcription Aesthetics

**Keith M. Murphy**  
kmmurphy@uci.edu

---

**Abstract:** I start with a premise that you may disagree with: a lot of the transcripts that discourse analysts publish are bad. I explain that what I mean by this, and then propose a solution: let's make published transcripts good, in the hopes of devising new possibilities for growing discourse analytic knowledge. I offer four provisional and interrelated principles that I argue are critical for developing a new kind of transcription aesthetics, along with some ideas and examples for how to employ them. The first principle sets the stage for those that follow: treat transcription not as a "word process" that produces written-text artifacts, but a graphic process that produces images. The second is to work compositionally, approaching each aspect of the transcript as worthy of its own specific kinds of attention. The third principle is to think openly about order when designing a transcript; and the fourth principle is to prioritize intelligibility. In the end and along the way, I make some arguments for why reformulating transcribing as a design process can help advance discourse analytic practice and theory.

**Keywords:** transcription; graphic design; aesthetics; multimodality

---

## Technically Bad

Let me start with a controversial statement: most of the published transcripts that discourse analysts produce are bad.<sup>1</sup> I know it's not common to use explicitly moralizing terms like this, but I mean it in a technical sense. Many published transcripts are bad because as graphic objects designed to communicate information they often fail at fulfilling their purpose. They repel readers rather than invite them in. They're inconsistent and hard to grasp (see O'Connell and Kowal 2000), requiring a studied deciphering before a reader can even understand the scenes the transcripts represent. They often take up too

much space on a page, or too little space, and the placement of various components seems motivated by convenience instead of intent. When visual images are included in a transcript, they're often sized disproportionately, or so blurry that nothing meaningful can be discerned in them.

There are lots of reasons why this is the case. First, while there's a general consensus for how to transcribe "obvious" linguistic (i.e., semantico-grammatical anchored) material—since the denotational organization of linguistic structure "affords formal description and manipulation," as Alessandro Duranti (1994:4) put it—there's nothing close to a consensus about how to transcribe non-linguistic phenomena (however "non-linguistic" is defined). What this means is that across almost every contemporary instance of transcription, speech-as-text is given conventionalized and comprehensible treatment, but almost every other transcript component—like visual images and annotations, for example—is subject to a mixture of emergent conventions, improvisations, and unorganized ad-hockery.

Second, most of the basic transcription practices used today were developed in the 1960s and 1970s, and to a large extent they've stayed there. This stasis is partly due to professional habits, and partly due to the technologies typically used for transcribing, which in turn support those habits. Once equipment like video cameras became cheaper for discourse analysts to use, features like hand gestures and eye gaze became easier to capture and transcribe, a pattern that's persisted as newly available technologies continue to provide access to previously uncapturable aspects of human communication. Transcripts, however, have remained relatively stable in their basic form, as representations of each "new" feature are simply added to older ones, with the oldest of all—horizontal lines of typed text—positioned firmly at the figurative (and often literal) center. If a contemporary transcript includes things like visual images, they're almost always arranged in ways that nonetheless privilege that familiar grayscale representation of speech—even when, say, the transcript promotes a "multimodal" perspective that theoretically demotes the centrality of verbal language.

Third, transcribing generally requires managing transductions between events, recordings, and transcripts (cf. S. Black 2017). Yet most reflexive attention to how events, recordings, and transcripts relate has looked backward, focusing on how the transcript corresponds with a previously recorded event (see Mondada 2007), rather than forward toward how the transcript will circulate and be used as an analytic instrument. Both perspectives are critical, of course, but priority has been given to the technicalities of truing representations with reality over those that best do the work that constitutes discourse analysis itself: dissecting complex phenomena, making arguments, and building knowledge.

The point I'm making here is very much an aesthetic one. When I say transcripts are bad I'm certainly expressing a normative judgment that the traditional aesthetics of transcription—a dominance of grayscale lines of horizontal written text that are sometimes surrounded by barely organized and often ambiguous visual images—are worn, tired, and unconsidered.<sup>2</sup> But I'm simultaneously stipulating a specific perspective, influenced in part by a graphic design logic, that “bad” graphics are ineffective at doing what their creators intend and, conversely, that “good” graphics work because they are precisely made to do so. There are plenty of ways to describe what “good” design means, but as Lucienne Roberts (2006:13) astutely observed, in all senses of the phrase, “the content of our work and the form it is given have repercussions,” and as such, caring about and caring for those repercussions is a critical aspect of creating things that others consume.

Here's why all of this matters. The capacity for knowledge production that's particular to the perception of complex multimodal visual objects is what Johanna Drucker (2014) calls “graphesis,” and it's been analyzed in all kinds of visual representations—illustrations, photographs, diagrams, models, charts—across many areas of scientific inquiry (Latour 1986). Michael Lynch's (1988) study of the circulation of visual images in biology, for example, detailed how graphic representations, as reduced versions of observed phenomena, afford a kind of understanding, comparison, and communication that other forms of representation typically don't. Similar effects of graphesis have been demonstrated forcefully in fields like epidemiology (Lynteris 2017), physics (Ochs et al. 1996), neuroscience (Beaulieu 2001), software design (Bødker 1998), and particularly with regard to the “Roentgen semiotics” (Cantor 2000) that underpin imaging sciences (Dumit 2004; Prasad 2005; cf. Burri 2008)—but also fields adjacent to STEM, like anthropology (Grimshaw 2001), linguistics (Stewart 1976), and even graphic design itself (Harland 2011). Graphesis also helps shape professional vision (Goodwin 1994)—and professional hearing (Bucholtz 2009)—and contributes to the formation of emergent research communities, as well as the maintenance of existing ones. And while there are many benefits to visual representations, there are, of course, potential downsides, too—even disastrous ones (Tufte 2006)—if care isn't taken to reflect consistently on how information is habitually represented (Suchman 1995). An important takeaway here is that graphic objects do a different kind of knowledge-building work than ones that are more strictly based in written text, particularly within specific professional channels of circulation.

Given what is known about how graphesis works in other scientific fields, it's reasonable to assume that how exactly a transcript looks and feels, how it visually expresses ideas, and the kinds of experiences it invokes in readers—which is to say, its *transcription aesthetics*—are directly consequential not only to discourse analytic knowledge production in general, but also to how specific concepts and relations emerge and come to be seen as relevant to discourse analytic projects. Aarsand and Sparrman (2019:15)

recently called for a “visual ontological politics [...] about how we single out images, how we choose to present images, how we present them, how we look at and scrutinize them, how the visual is merged with the verbal and, finally, about the reflexive approach to this process.” This is a necessary and absolutely welcome call. But this approach doesn’t go far enough, I think, because it still implicitly relies on a traditional view of transcription aesthetics derived from grayscale technologies and their inherent limitations—limitations that likely constrain how theorizing from transcripts plays out.

So, having passed judgment that our default transcription aesthetics are bad, and explained why I think that it matters, it’s time to offer a counterpoint: *let’s make published transcripts good*—or at least less based in old analog ways of working, in the hopes of devising new possibilities for growing discourse analytic knowledge. What I have to say here is a modest first step. In what follows I offer four provisional and interrelated principles that I think are critical for developing a new kind of transcription aesthetics, along with some ideas and examples for how to employ them. The first principle is critical, and sets the stage for those that follow: treat transcription not as a “word process” that produces written-text artifacts, but a graphic process that produces images— notwithstanding the ambiguity of what counts as an image or a text (Mitchell 1984; see Introduction to this issue)—of which written text is often only one component. The second is to work compositionally, approaching each aspect of the transcript as worthy of its own specific kinds of attention. The third principle is to think openly about order when designing a transcript; and the fourth principle is to prioritize intelligibility. All of these principles borrow from graphic design, on the one hand, and the “figurative semiotics” framework outlined by A. J. Greimas (1989), on the other, as I discuss in more detail below.

Finally, note that nothing here is meant to be definitive or prescriptive, despite my coming in hot with judgment, and the examples I provide are not given in the spirit of *do it like this*, but rather as suggestions in the spirit of *here are some things you could do*. My goal, in the end, is to open up a space for exploring future possibilities for transcription norms and technologies that orient to transcribing as a visual “image-forward” practice rather than a written “text-forward” practice.

## **Principle 1: Treat the Transcript as a Graphic Object**

In the mid-twentieth century there was considerable disagreement as to the utility of “performance data”—that is, actual speech—for scientific analysis, when the “native-speaker intuitions” favored by most linguists had worked so well for so long. Sensibilities about performance data first began to shift after William Labov (1963, 1966) used tape recorders to capture phonetic variation in American speech communities, Harvey Sacks stumbled onto a collection of recorded telephone conversations (Heritage 1984), and a

group of anthropologists and psychologists decided to record child speech in action (e.g., Keenan 1974). In direct refutation of long-held understandings of human communication, this cohort of researchers helped establish a new paradigm for exploring what came to be called “naturally-occurring language.”

One of the most critical contributions of this period was the development of a technical approach to transcription from recorded sources. This entailed not only writing down how speakers actually pronounced sounds and formed sentences—that is, how they actually talked—but also devising techniques, conventions, and sets of best-practices for recording, listening to, and representing the things that people say. This endeavor turned out to be more difficult than it might have seemed at first blush and by the late 1970s researchers began publishing reflective critiques of the practice of transcribing itself, the most influential of which was probably anthropologist Elinor Ochs’s (1979) chapter “Transcription as Theory.”

The central takeaway of “Transcription as Theory” was simple enough: transcribing is not a neutral practice but is instead an engaged and consequential *act of selection* within larger interpretive projects. This insight was subsequently expanded to include transcription as “politics” (Bucholtz 2000; cf. Bucholtz 2007; Roberts 1997) and as a “cultural practice” (Duranti 2006). All of these formulations, however, rely on several shared basic points: a transcript is the outcome of choices made by a transcriber; those choices produce particular representations, whether intentional or not, of both language and people; and these representations inevitably influence how we come to understand the phenomena we study.

Many other reflexive critiques followed in the wake of “Transcription as Theory.” Some focused on outlining fundamental principles and considerations relevant to transcribing speech (Du Bois 1991; O’Connell and Kowal 1999; Edwards 2005; Hepburn and Bolden 2012; Henderson 2018). Others offered critical takes on various transcription systems that emerged in and after the 1970s, and the choices required for using them—focusing on, for example, their affordances (Gibson et al. 2014), reliability (Romero et al. 2002), or inconsistencies within and between systems (O’Connell and Kowal 1994, 2000). A related line of inquiry examined the possible consequences of making what could be considered wrong choices in transcription. For example, there have been periodic attempts in conversation analysis and other fields to use a version of literary “eye dialect” to represent naturally occurring phonetic variation—say, writing an utterance as “I’ve gutta fetch mah core” instead of “I’ve got to fetch my car.” As Dennis Preston (1985:328) argued, while the intent of this choice might serve an analytic purpose, “nearly all respellings share in [a] defamation of character” because non-standard orthography almost always invokes images of deviance that project onto the speakers represented in the transcript (Jaffe and

Walton 2000; Miethaner 2000). In response to this critique, the use of eye dialect has almost entirely disappeared from contemporary transcription.

While such treatments have helped cultivate sound transcribing practices, I think their overall thrust misses a more subtle, and probably more consequential, point in Ochs's chapter. Until quite recently (see, e.g., Bezemer and Mavers 2011; Ayaß 2015; Meredith 2016; Aarsand and Sparrman 2019; cf. Norris 2002), almost every reflexive analysis of transcribing has assumed it to be a "word process" in which considerations about selection and representation are focused on the letters, symbols, and other features that represent verbal language (as we see with eye dialect). These features are extremely important, of course, but they are only one part of a larger *design* argument that Ochs was making in her chapter: that *how a transcript looks and feels overall*, and *how its elements are arranged on the page* will directly influence how a reader understands the phenomena presented.<sup>3</sup>

Ochs makes this point most forcefully when discussing the effect of page layout on representing—and thus theorizing about—the basic mechanics of child speech. Most transcripts, then and now, commonly use a "script" format in which utterances are stacked and read vertically, from top to bottom, similar to how dialogue is structured in plays and novels. As John Du Bois (1991) has argued, borrowing forms that readers already recognize increases the accessibility and ease-of-use of a transcript. But when applied in certain contexts, especially ones involving children, Ochs argues, this format may provide an inaccurate depiction of how some situated speaking actually works.

Here's how Ochs explained it: adults overwhelmingly design their utterances to be relevant to the just-prior turn at talk. In a "script" format, any transcribed utterance would typically be linked to the one placed right above it. Thus:

Speaker A: Hello!

Speaker B: Hi, how are you?

Speaker A: I'm good, thanks.

### **Example 1. Script format of standard transcripts**

In such cases, the form of a transcript with sequentially aligned utterances, read from top to bottom, reproduces the form of the interaction represented. But when it comes to transcribing children's talk, this format presents a problem because children don't talk like adults. In fact, they're far more likely to produce an utterance that's relevant to *their own previous turn at talk*, not their interlocutor's. This means that the script format, when used

for children's speech, ends up misrepresenting that speech by depicting it in a form more typical of adult speech. To solve this problem, Ochs proposed using separate columns for each speaker, placed side-by-side, instead of lines placed on top of one another. In this format, a single child's speech is laid out in one column, and the downward flow of the child's utterances in that column reveals linkages between the child's own utterances, rather than to the previous utterances of an interlocutor. In this way the form of the transcript more accurately matches the form of the interaction itself, which in turn provides a more accurate model of children's speech from which to theorize.

The implicit argument that Ochs was making is that the *transcript as a whole* should be treated as a graphic object, as an image produced by an intentional graphic process with its constituent elements receiving their own consideration. Ochs's argument is that transcripts are impactful not only, or even mostly, through the effect of written text on a reader, but through the more comprehensive graphesis entailed in the integration of multiple visual elements into a single transcript. In other words, it was an argument for transcribing through a graphic design process.

A basic way to describe such a graphic design process is as a *deliberate deployment of visual media to communicate some message*. As Jorge Frascara (1988:20, original emphasis) put it, “[g]raphic design is *the activity that organizes visual communication in society*. It is concerned with the efficiency of communication, the technology used for its implementation, and the social impact it effects.” Frascara gave several examples of these communicative effects, including advertising that gets people to buy things and propaganda that gets people to believe things, to which we could add more contemporary examples, like interface design that helps people navigate a webpage and infographics that get people to understand complexity. Here we might follow Guy Bonsiepe's (1999:59) characterization of the graphic designer as “an information manager” (cited in A. Black et al. 2017).

The underlying principle tying all of this together is that graphic design *exerts a push* of some kind on those who view it (or so designers hope); at a minimum, *informing* viewers about something, or better, *persuading* them to think, feel, believe, or even do something. Given that in a very real sense every transcriber also already works in this way—as an information manager making selections about what to include in a transcript, and thus presenting, arranging, and explaining that information on the page—treating the transcript like a graphic object, and transcribing as a process of graphic design, is an almost intuitive perspective to adopt.

One framework for thinking about how exactly to do this can be found in the figurative semiotics of A. J. Greimas (1989; cf. Polidoro 2019). Best known for his application of

structuralist semiotics to narratology, Greimas did spend some time thinking about “visual semiotics” as well. He was particularly interested in distinguishing between the structural armature of the semiotics of written text—linear, unidimensional, often horizontal—and the “planar semiotics” of visual objects and their capacity to produce “meaning effects.” “Planar” objects, Greimas argued, are composed of different “plastic” components, including what he called “chromatic categories” (stuff related to color), “eidetic categories” (stuff related to detailed visible forms), and other “minimal units” all arranged not linearly but “topologically” according to areas, shapes, and directions.

Interpreting visual objects like printed images and other graphic formats requires a specific process, Greimas argued, but not one necessarily linked to how written language is read. While written text primarily relies on a horizontal interpretive process, a visual “object can be grasped only through its analysis. Put simplistically, it can be grasped only through being decomposed into smaller units and through the reintegration of those units into the totalities that they constitute” (Greimas 1989:638), which is to say, through repeated decomposition and recomposition of the graphic object’s constituent “plastic” elements and their relations. Thus, what Greimas called the “meaning effect” of a visual object and what Drucker calls “graphesis” are two versions of a core idea in graphic design, as well as a critical point in Ochs’s “Transcription as Theory”: the capacity of a graphic object, including a transcript, to communicate particular ideas emerges from the specific and intentional arrangements of its constituent semiotic components.

## **Principle 2: Work Compositionally**

Building from Greimas’s (1989) premise that visual objects produce “meaning effects” through the relations of their plastic components, to work compositionally when creating a transcript is to attend to each transcript feature in its own right, while simultaneously considering how all of the features combine to communicate a message as a whole. The point that transcript components are plastic is critical. Traditionally, transcripts contain a set of common features: written text, visual images, and annotations, among others. Most of such features are treated as formally fixed, or mostly fixed. While it’s common, for example, to change a text style to italics or bold, or crop a screen grab to focus on a particular hand gesture, the default orientation to transcribing is to leave most features “as is” when placing them in a transcript. To work *compositionally*, by contrast, is to approach each component as malleable and their details as adjustable in order to craft, with intention, a plausible visual argument. Consider some aspects of what this means.

### **Written Text and Type**

When transcribing naturally-occurring speech as a technical practice first started in earnest, the most advanced piece of machinery for composing and laying out written text, including transcripts, was the IBM Selectric typewriter. Like any other typewriter, the

Selectric produced horizontal lines of text, and allowed adjustments in line-spacing, justification, and other basic formatting features, but did so with a level of precision that was unavailable with other machines. One of the Selectric's key features was a unique ball-shaped typing element, which could fit between 88 and 96 characters on its surface. Crucially, these balls were interchangeable, allowing a user to switch between different styles, type sizes, and typefaces—one of the most common of which was Courier 12. When Gail Jefferson originally developed the set of transcription conventions that are now standard in conversation analysis (CA) and other areas of discourse analysis (Sacks, Schegloff, and Jefferson 1974), she significantly relied on many of the features built into the popular typewriter (Erickson 2004), including the Courier typeface.

But the choice of that typeface was not made simply because it was the Selectric's default option. Courier is a fixed-width (or monospaced) font, which means each character, regardless of its actual dimensions, takes up the same amount of space on a page (so a lowercase l and an uppercase O are spaced the same in Courier). As it turned out, if every character is the same size, laying out a transcript becomes quite easy; for instance, an overlap point in one line of talk can be aligned with the overlap point in the following line of talk without any fuss.

Speaker A: Hello!  
Speaker B: Hi, how are [you?  
Speaker A: [I'm good, thanks.

### **Example 2. Fixed-width font and transcribing overlapping speech**

Courier thus allowed transcribers to represent instances of overlap quite clearly, which in turn allowed overlap as a regular, recurrent, and context-contingent phenomenon to become visible. In this way, the specific features of the Selectric typewriter—including the large number of characters the typing element could accommodate, the qualities of the fonts it worked with, the ease with which a user could control the text layout, and more—became determining factors in how the form of transcription itself was originally imagined as a technical practice.

Then a funny thing happened. Word processing and personal computers came on the scene; and while things changed, they also largely stayed the same. One reason for this was that as new technologies like these emerged, they tended to significantly improve the efficiency and speed of work, but still carried over the same basic functionality of previous technologies. Word processing software was extremely useful for many tasks besides crafting written text, but because it was designed to mimic older typing and printing technologies, it ultimately ended up reproducing the same basic aesthetic that typewriters created: horizontal lines of black text stacked on top of one another on a white

background, often producing a “grayscale blob” gestalt when viewed as a whole. A second reason was more a matter of professional habit. While personal computers have over time steadily unwound the typewriter’s notable limitations, many transcribers still labor according to a typewriter logic: CA still uses Courier as the default typeface, despite the availability of many other monospaced fonts,<sup>4</sup> and most transcription notation conventions are still predominantly based in the orthographic symbols available to typewriters, rather than in other possible manipulations that are possible with modern machines. What this means is that even though there are currently many more possibilities for creating a transcript as a graphic object, most still look like something made in the 1970s.

Breaking free of the restrictions set by the Selectric’s orthography and embracing the tools now at our disposal offers much more flexibility to work with the plastic features of digital text, to do things that—while perhaps possible before computers—were not encouraged by the technologies and professional habits of the Selectric era. For example, modulating typefaces can communicate particular meanings and affect quite efficiently.

**I’m sorry to tell you that your grandmother has passed away.**

**WILL YOU BE  
MY VALENTINE?**

### **Example 3. Mismatching sentiment and typeface**

The two messages in Example 3 seem inappropriate because the sentiment expressed by the words doesn’t match the sentiment expressed by the type in which they’re set (see Murphy 2017). That same idea—that fonts have expressive power—can be exploited in a more robustly designed visual transcription. For example, different fonts can be used to mark different speakers, or different accents, or languages, or registers, or a shift in speed or tone.

laughter could be expressed with Comic Sans

M I just have one problem with the snowman, Peter.  
P Yeah?  
M What do **snowmen look like?**

different registers could be displayed in different typefaces

everyday speech  
**professional jargon**  
respect

different tones or emotions could be displayed in different typefaces

*sing-song*  
**anger**  
*goofiness*

#### Example 4. Using different fonts to mark different aspects of transcribed discourse

What's perhaps even more critical, though, is that modern computers are not restricted to the 96 characters that appeared on an IBM Selectric typing element. Because of Unicode, the standard for encoding the world's writing systems, most computers have access to over 140,000 characters. The vast majority of these won't be useful for creating new transcription notations, because they're actual glyphs in specific writing systems, but hundreds of others certainly could be.



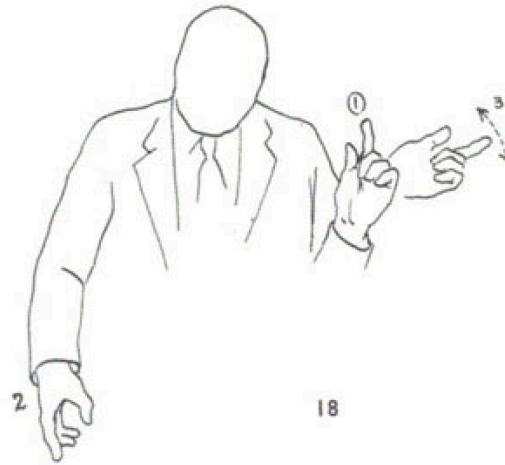
#### Example 5. Unicode symbols usable in transcription

I'm not arguing that these particular symbols should be used to mark any specific phenomenon. I'm simply indicating that there are many more options for notation available to us than are currently being used. A thoughtful approach to updating transcription aesthetics that takes advantage of what Unicode offers could include a much wider range of symbols and integrate them with the other principles discussed here.

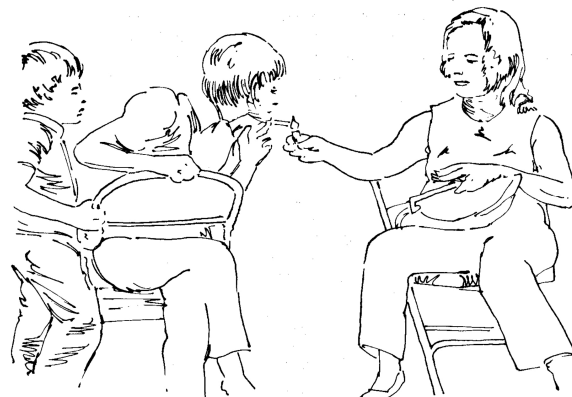
### Visual Images

Other than speech and its representation in written text, the realm of human communication that has received the most transcriptional attention is visual embodiment. Prior to the 1990s it was quite labor intensive, and therefore relatively rare, to include visual images with transcription (but see, e.g., C. Goodwin 1981; Streeck 1983; Heath 1986). Following David Efron's (1941) work in the 1940s (see Example 6), there were periodic attempts to account for quasi-"linguistic" phenomena like gestures, facial

expressions, and body postures. But until quite recently, consistently and systematically incorporating the visual into transcripts has not been granted a high priority (and even today it's not the norm). In as much as visual material is brought into contemporary transcripts, there are inconsistencies in the conventions used and the assumptions motivating the informed selection of the material that seems more or less appropriate for inclusion. Moreover, there are legitimate disagreements as to whether visual semiotic material even matters; as any good conversation analyst would say, it only matters if it's observably relevant to the participants.



**Example 6. A representation of gesture from Efron (1941:143), drawn by Stuyvesant Van Veen**

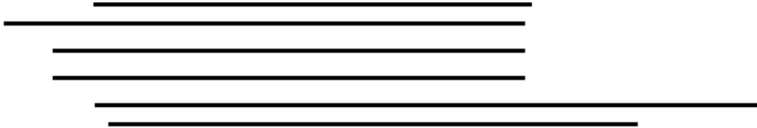


**Example 7. An early use of line drawings used with transcription; from Goodwin 1981:147**

The earliest and still quite common method for integrating human bodies into mostly textual transcripts is the use of simple line drawings (see Example 7). Before the advent of inexpensive video equipment, drawings were pretty much the only option, and indeed, they do offer a number of benefits: they provide participant anonymity, rhetorical clarity, and (as very low-resolution objects) they are easily reproducible both in their original

printing and in subsequent lower-resolution reproduction. But line drawings are at best a partial solution: they are only as good as the skill of the artist—or image filter—allows, there are limits to the level and kind of detail that can be included, and it’s often the case that line drawings enhance the impression of a grayscale blob (see Example 8).

21.                   ..... a ..... b .....
- 118 j; ngalan just yalmba-wi    *guwaar* duuga-athi  
 sun                   sandhill-LOC   *west=R* enter-PAST  
*The sun was just going in west over the sandhills.*  
 L: “B” from front out W at (a), flat, palm down, eyes sighting along  
 arm, held; retracted at (b) flat, back to right elbow.



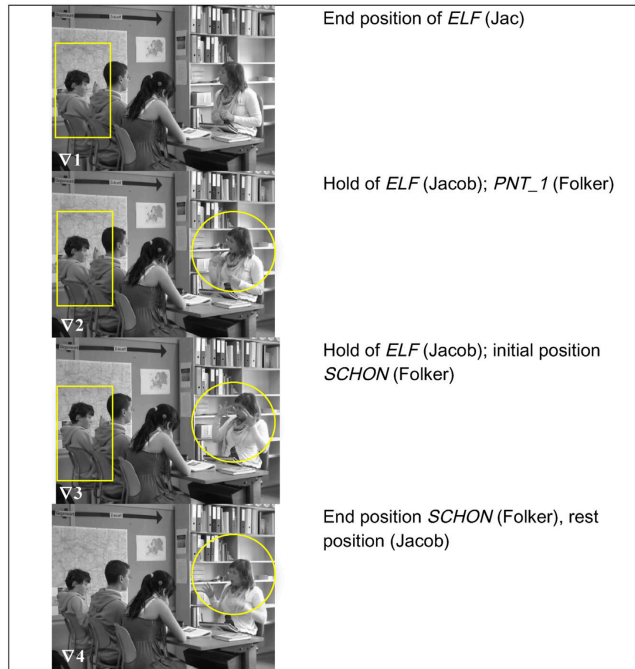
Example 8. A line drawing and its impression; from Haviland 1993:22

Photographic images, either from digital still cameras or (more often) captured from video are now quite easy to incorporate into transcripts, and they offer some benefits over line drawings, including ease of use and a richer sense of “realism.” But using photographic images is not always a simple matter. *Which frame should be captured?* Video cameras record at either 24 or 30ish frames per second (depending on the region of the world the camera was sold in), which means that any given moment of footage contains lots of frames to choose from. To some extent, it doesn’t really matter, but that raises the question: why not? There probably are better or worse frames for representing a phenomenon involving motion, but there’s no clear rule for how to choose.

This leads to a second question: *Does this screen grab clearly represent the phenomenon it’s intended to?* For reasons including the resolution of the original video or software that handles images badly, screen grabs or photographs used in transcripts are often blurry and incoherent, which can result in a transcript that includes visual images with no compelling connection to what’s being described. This can be corrected, but of course there’s a question there, too: *Should adjustments be made in color or contrast, and can images be “marked up” in some way?* It’s quite common for transcribers to simply drop a screen grab into a transcript, either as is or with a filter that hides participants’ identities, because the screen grab represents the most “natural” visual depiction of the event. Unfortunately, without manipulating the image for clarity and coherence, they are often not very useful for much at all.

Finally, we should ask: *How and where should visual images be integrated with transcribed written text?* This includes consideration of what size the image should be, where it should be placed, and how it’s linked to corresponding utterances. None of these issues is standardized, nor is there even a clear shared principle that helps us figure it out.

Let me illustrate what I’m talking about more clearly. The examples below are typical of the range of ways in which some of the points I’ve just made have been handled in publications since about 2010. In Example 9, Groeber and Pochon-Berger (2014) use grayscale images, overlain highlighting, and separate columns for images and descriptions of body movements, which are themselves separated from transcribed verbal utterances (not shown).



**Example 9. Using grayscale image in transcript; from Groeber and Pochon-Berger 2014:126**

In Example 10, Laurier and Brown (2011) use full-color screen grabs integrated in-line with transcribed utterances. In the case below, the visual image, which has no noticeable adjustments or enhancements, is given more space than the written text, and its own line numbers for reference.

1. Oscar: ((speaking quietly to himself))
- 2.
3. Jo: OH! ((taps A's arm and points toward the bins on the right hand screen ))
- 4.
- 5.



- 6.
7. O: What ((still rushing through sequence))
- 8.
9. J: School interior (1.0)

**Example 10. Using full-color screen grab in transcript; from Laurier and Brown 2011:248**

In Example 11, Murphy, Ivarsson, and Lymer (2012) use grayscale images, with some overlain adjustments that correspond with descriptions of non-verbal action, which are themselves written in a different typeface and location than transcribed utterances—locations that are not consistent. The images are not given their own line numbers, but are placed in such a way, and marked with a box, that indicates which line-numbered utterances the embodied actions co-occur with.

1.13 Tom E:h, how they, work with the different corridors, and let them



**thrusts left hand, then right, then left, at angles**

1.14 go to the facade on one level, but not in the next and so on.



**right hand curves away from left, then back**

**Example 11. Using grayscale image, overlain notation, and typeface in transcripts;  
from Murphy, Ivarsson, and Lymer 2012:540**

All of these methods work in their own ways for aligning visual images and written text. But the different choices made by transcribers constructing these transcripts reveal that alongside the emergent consensus that visual images are important components of a transcript, how they are handled and what they are used for is not at all certain.

One of the biggest benefits of digital images, as noted above, is precisely their plastic qualities, that they can be edited in different ways—not to mislead, but to highlight and focus attention on specific things or make a detail clearer. Many of the questions and concerns raised above go away simply by accepting the malleability of digital images, rather than worrying about it, and using that malleability with intent. Backgrounds can be erased or dimmed and details can be enhanced through color-correction or sharpening. Items in a scene can be brought closer together, moved further apart, or arranged differently in order to demonstrate particular relationships. And while the current techniques for integrating visual images and written texts are dominated by square images alongside lines of text, it need not be that way (as Charles and Marjorie Goodwin have demonstrated for years; see, e.g., C. Goodwin 1994; Goodwin and Goodwin 1996; M. Goodwin 1998). For example, written text can be placed around visual images, rather than vice versa, as in Example 12.



### Example 12. Intercalating written text and visual images in transcripts

Edited figures can be placed within and between lines in ways that emphasize actions that co-occur with speech (Example 13).

Peter	Tända det. Light it.
Matti	Huh? Huh?
Peter	Tända. Light.
Matti	Tända alla? Light them all?
Peter	Ja::: Yes!



### Example 13. Emphasizing co-occurrence of speech and action with edited figures

It is even possible to place moving images—small video or GIFs—into transcripts that are hosted on platforms that support them, like websites and some digital books; and if the PDF standard is ever supplanted, perhaps its replacement will support these formats, too.

## Color

For many years, color was a difficult quality to incorporate into transcripts. The typewriters used to make transcripts, the journals that published articles, the photocopiers used to reproduce those articles, and even the first digital files downloaded and printed on home printers, were all grayscale technologies oriented to depicting a monochromatic world. While it's been technically possible to incorporate color into published transcripts for decades, it only became economically feasible once enough readers began moving away from printed pages and looking at transcripts primarily on full-color screens. By the mid-2010s, color finally achieved the status of a full-fledged feature for most published transcripts.

This means that Greimas’s (1989) “chromatic category” can now appear in a number of different places in a transcript. For example, in contrast to black-and-white images, color images tend to provide a richer sense of detail and realism in the scenes they represent. This isn’t to say that they are somehow “true” representations in any strong sense. Nonetheless, they do offer some sense of “being there” in a depicted scene that a black-and-white image simply doesn’t, at least for most readers.

But color is also useful for other transcript components. For example, color annotations can mark different kinds of information, or link different things to each other—say, a yellow box around an image that’s paired with a yellow box around a segment of written text. Color can separate information, signify mood or affect, and nudge a reader to focus on a particular element or area. It can also link elements from different places—maybe across different transcripts—through repetition. Color gradients can indicate progress in space or time, or relative intensities (Example 14).



#### **Example 14. Using color and gradient in transcripts**

And colored text can do more than simply highlight or call attention to particular words. When combined with other plastic features, like typefaces and styles, colored text can furnish otherwise unexpected kinds of representation. The point here is that color isn’t a transcript component that should be left to the defaults provided by the software we use. It’s a feature we can control, manipulate, and strategically deploy in order to communicate specific messages.

#### **Points, Lines, Textures**

Three of the most basic building blocks of graphic design traditionally include point, line, and texture, and these can also be useful for constructing transcripts topologically. *Points* are the smallest units of a composition, and they’re generally organized into some kind of recognizable form that itself offers some sense and feeling. Well-ordered collections of points generally create *lines*, while less-ordered collections of points can form *textures*—

which often fit into Greimas’s chromatic category—though of course these components aren’t always mutually exclusive.

One of the most obvious kinds of “points” in a transcript would be a single letter. I’ve argued that traditional transcription aesthetics rely on a default pattern dominated by stacks of horizontal lines composed of orthographically-marked letters set in a limited number of typefaces. But there’s more that can be done with these points and lines. Example 15 contains the same chain of utterances, all from one speaker, presented as different points and lines with different textures. Version (a) uses Courier and line breaking typical in conversation analysis; version (b) uses the same line breaking, but a different font and slightly tighter line-spacing—note that this version feels heavier and more compact; version (c) uses the same typeface, but puts the utterances into longer lines, and uses looser spacing between letters; version (d) uses a very light typeface, and produces different textures depending on whether it’s justified against the left or right margin.

- a**  
I like that little guy,  
it’s just that is looks a little strange in relation to—  
to this.  
Or a little soft.  
If this were a little (.) I don’t know,  
maybe softer in form.
  
- b**  
I like that little guy,  
it’s just that is looks a little strange in relation to—  
to this.  
Or a little soft.  
If this were a little (.) I don’t know.  
maybe softer in form.
  
- c**  
I like that little guy, it’s just that is looks a little strange in relation to— to this.  
Or a little soft. If this were a little (.) I don’t know, maybe softer in form.
  
- d**  
I like that little guy,  
it’s just that is looks a little strange in relation to—  
to this.  
Or a little soft.  
If this were a little (.) I don’t know,  
maybe softer in form.
  
- e**  
I like that little guy,  
it’s just that is looks a little strange in relation to—  
to this.  
Or a little soft.  
If this were a little (.) I don’t know,  
maybe softer in form.

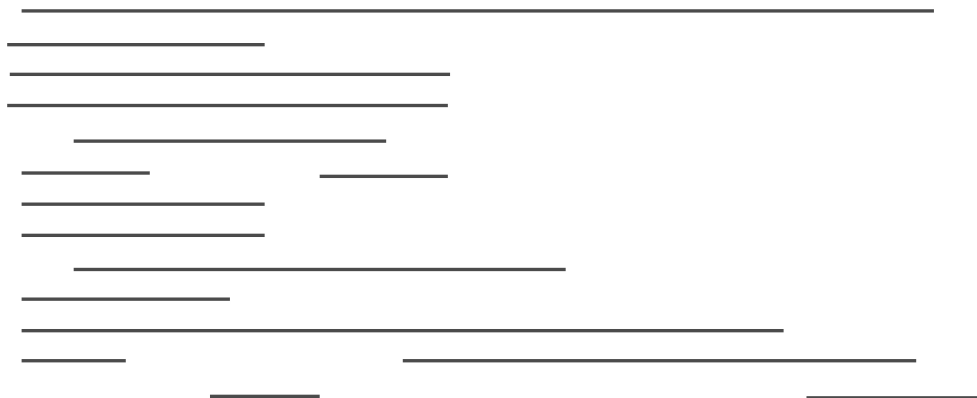
### Example 15: Using points, lines, and textures in transcripts

These different arrangements of points (that is, letters and words in particular fonts) and lines produce very different textures and chromatic expressions in the overall transcript segment, which in turn can stimulate different feelings or interpretations. Segments with darker letters and tighter line-spacing feel heavier than segments with regular text and looser line-spacing, which can impact how a reader interprets the text. From a compositional point of view, these are features that a transcriber can exploit in order to craft a visual argument.

### Principle 3: Think Openly about Order

Above, I argued that the rudimentary transcript structure of a grayscale blob was locked in place when electric typewriters allowed non-professionals to work with complex typesetting and layout (a significant feature at that time); and further, that technological advances now allow transcribers to do so much more with a transcript, though this basic structure persists as the default form.

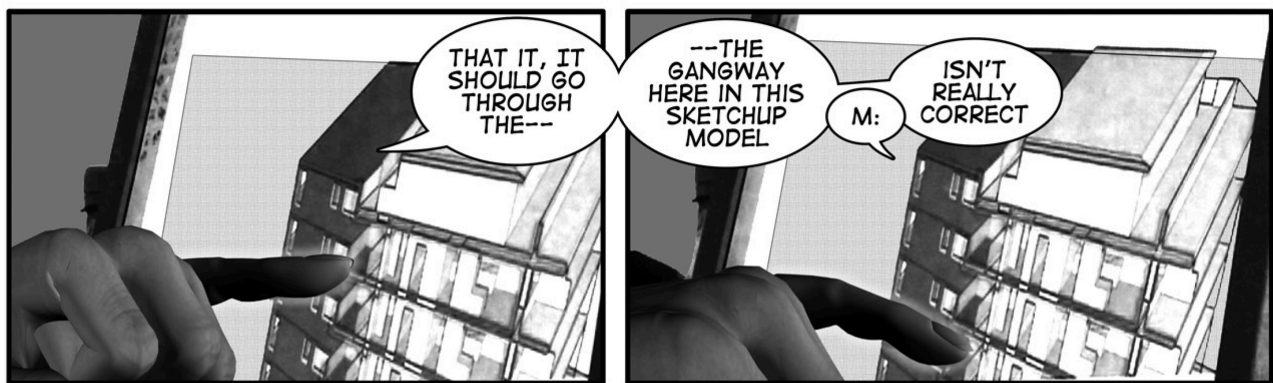
→ Roger: ((To Jim)) Are you just agreeing because you feel you wanna uh  
 → Jim: Hm?  
 → Roger: You just agreeing?  
 → Jim: What the hell's that.  
 Al: It's- Agree[ing?  
 → Roger: [Agreeing.  
 → Jim: Agree::n.  
 → Roger: Yeah.  
 Al: With us. Just going along with us.  
 → Jim: No.  
 → Roger: Saying 'yes, yes' [hehheh hh hehhh hh hehheh hh  
 → Jim: [Well, i-i-it's-it's true. Everything he sai(h)d  
 is true, so [GTS:2:2:70]



Example 16. A transcript and its impression; from Sacks et al. 1974:708

One alternative is to treat transcribing as a planar layout process, as many multimodal and visual transcription techniques have demonstrated (Ayaß 2015; Meredith 2016; Aarsand and Sparrman 2019). Instead of relying on the word processing software’s proclivity to order everything in horizontal lines, the transcription space becomes a topological plane on which every element is placeable, movable, and adjustable, and can be intentionally related to other elements in some way. Written text is instantly demoted from its position as the centerpiece of the transcript, rendering speech (at least in the transcript) one among multiple components. It allows a transcriber to select which transcribed features receive more attention at any given time, and to experiment with ways to represent those features and their relations.



Take, for example, what Eric Laurier (2014) calls a “graphic transcript,” a format inspired by traditional techniques to draw comic books, as a way to organize information. In transcripts like Example 17 (also see, e.g., McIlvenny 2009; Hoffmann-Dilloway in this issue), strips of interaction are broken into “beats” and “panels” and utterances are placed in speech bubbles. As Laurier (2014) makes clear, this format provides a new lens for how we conceive of space and time in relation to social action. It captures speech, bodies, and context in a holistic way that traditional transcript aesthetics simply cannot, and presents “moments” and “scenes” —and transitions between them— in forms that are easily and immediately comprehensible.



**Example 17. A “graphic transcript” using comic book aesthetics to organize information; from Ivarsson 2010:184**

Another emerging format tries to capture the complexities of communication as it unfolds on internet live-streaming. What’s particularly interesting for live-streaming interactions is that platforms like Twitch (Recktenwald 2017) already use a kind of transcription in their chat functions, and when recorded live-stream videos are uploaded to platforms like YouTube (Choe 2019), there’s an additional transcription-like function provided by the comments. One of the challenges, and opportunities, for researchers working in these

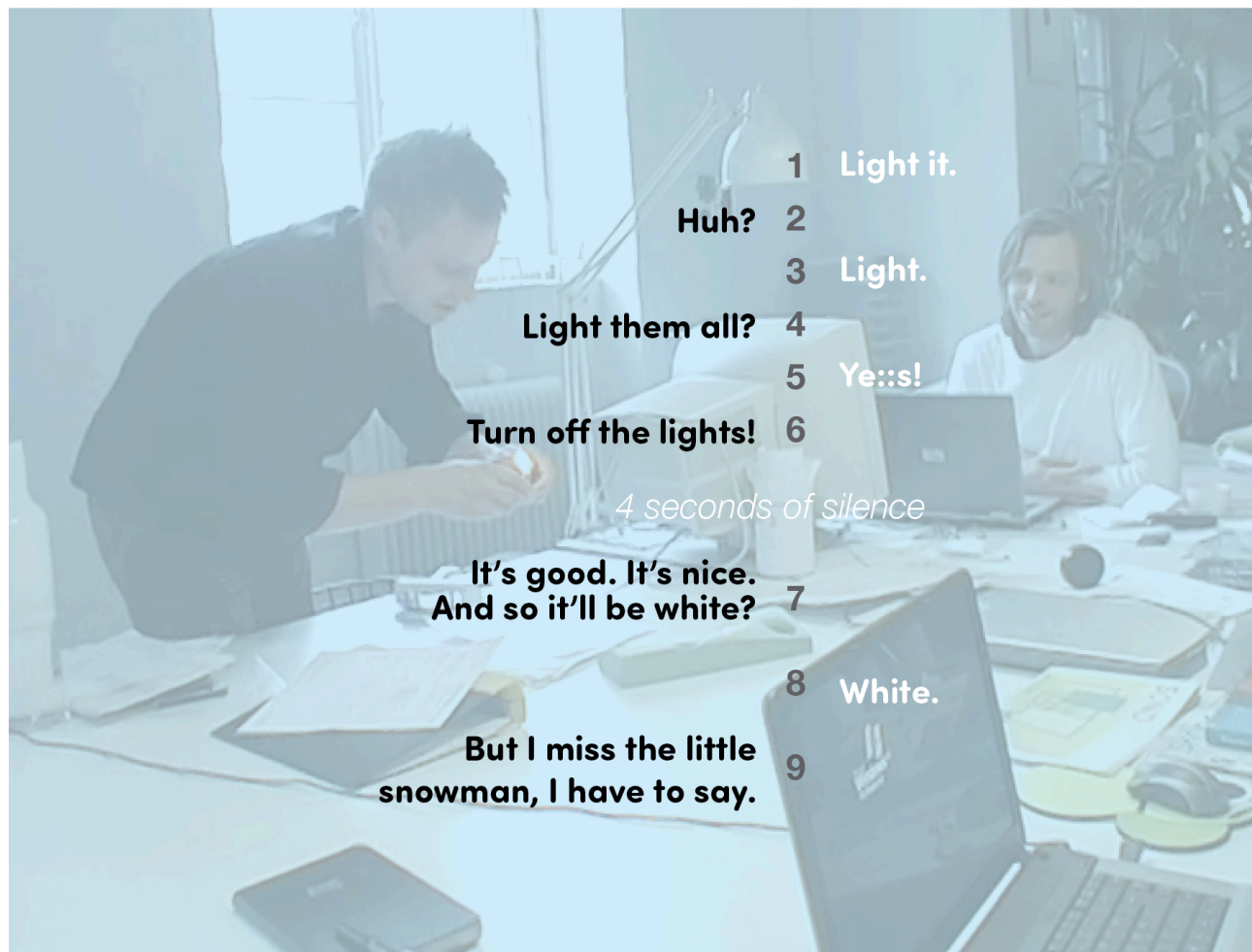
areas is to figure out how to blend the features inherent to these platforms with the transcription conventions used in discourse analysis.

Image	BJ ChangHyun		Viewers			
	Line	Spoken action	Bodily action	Line	Typed action	ID
	1	자 우리 밤부먹자 'Okay, Bamboo, say ah'	Right hand bringing boneless fried chicken closer to the camera			
	2		Waiting for 3 seconds	A	밤부님 남 장전!!!! 'Bamboo is ready to eat!!!!'	21bib21
	3	밤부 남 'Bamboo, try it'		B	순살인가? 'Is it boneless chicken?'	krbfl433
				C	?	hwpgud1
				D	리얼사운드아닌가? 'Isn't it high quality sound?'	djadmlwlo
	4	리얼사운드냐고요? 네 리얼사운드죠 'Did you ask if it's high quality sound? Yes, it is'				
	5	소리 왜요? 'Is there any problem with the sound?'		E	BJ 창현님의 방송 추천 1위 'BJ ChangHyun's broadcast is currently ranked No.1 for the most recommended livestream'	website notification

### Example 18. Transcribing internet live-streaming; from Choe 2019:188


One benefit to treating the transcription space like a planar canvas is the addition of a third dimension through layers. Word processors operate primarily along a horizontal x-axis, as text is typed from left to right (or vice versa, depending on the language), and a vertical y-axis, as lines of text stack on top of one another. Graphics software, however, not only allows free movement of components on a topological plane, but also provides a z-axis through the use of multiple layers. Rather than requiring components to snap to horizontal lines of written text, they can easily be placed in front of or behind each other. Transcripts can be maximally constructed to include several layers of information that are hidden and revealed at different times for different reasons, and components can be sequentially transformed through the addition of layers without having to manipulate their original forms.

In Example 19, a screen grab is placed in the background layer to provide a sense of the scene, while a transparent blue layer on top attempts to capture the jovial spirit of the interaction. The utterances of each speaker are placed in relation to their positions in the image, which is primarily intended to provide context rather than indicate a specific action. The colors of the type used for each speaker are also matched to the colors of their clothing.



### Example 19. Using layers in transcripts

In Example 20, which is another version of the same scene, the visual image is the placed as the centerpiece, while speakers are again divided into different columns that correspond with their positions in the image; each is given a different typeface, and the background is chosen to match the color scheme of the image, mirroring the lighting of the environment.

1		Light it.
2	<b>Huh?</b>	
3		Light.
4	<b>Light them all?</b>	
5		Ye::s!
6	<b>Turn off the lights!</b>	
		
7	<b>It's good. It's nice. And so it'll be white?</b>	
8		White.
9	<b>But I miss the little snowman, I have to say.</b>	

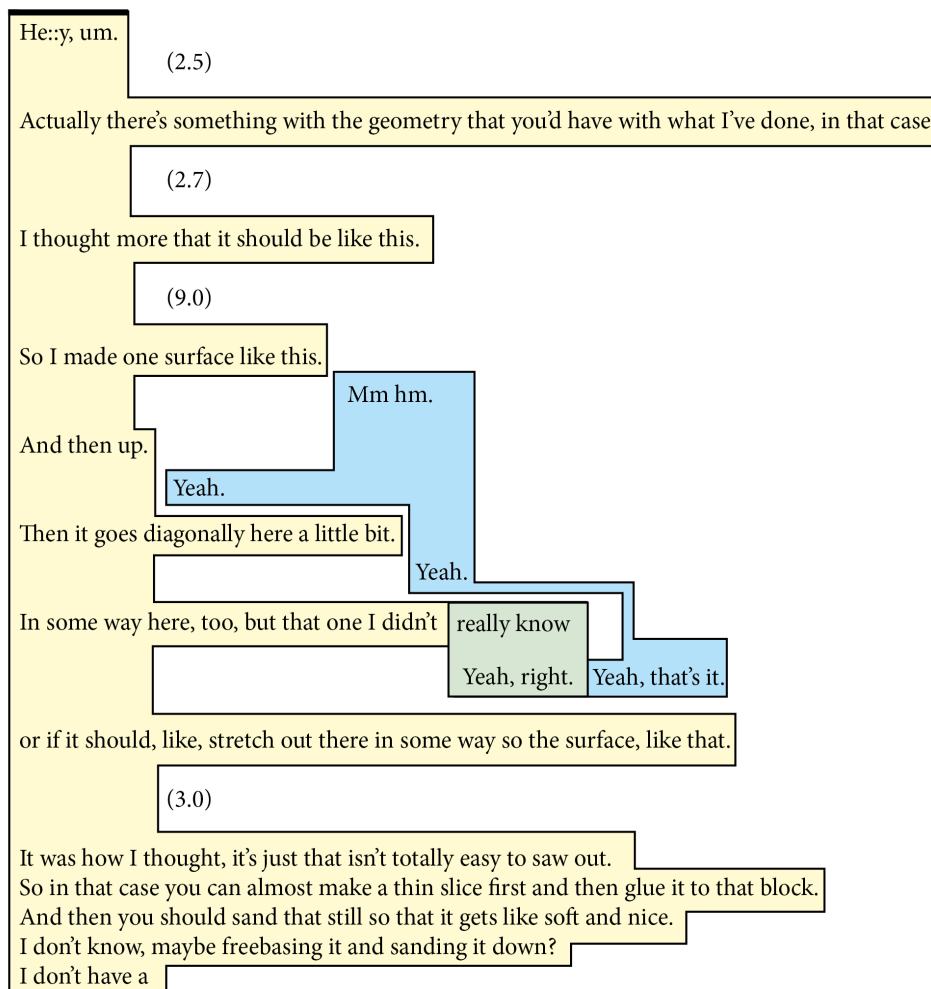
**Example 20. Using background color, typeface, and columns to echo visual image in transcript**

Other ways to use topological layout techniques to visualize communicative phenomena include lines and subdivided planar areas. The following example demonstrates how placing speakers into separate planes of different shapes and sizes can reveal the relative amount of floor time each participant uses in an interaction. In Example 20, the speaker on the left takes up more floor time, but we can see that the dynamic also shifts over several turns at talk.

Hello there my name— I'll be taking care of you today.	Okay.
Now then, could you tell me your name please?	Pat.
I'm sorry could you repeat that I didn't hear.	I said it's Pat.
Thank you I'll write that (.) here now.	How long will this take?
I know you'd rather be elsewhere hheheh	I just have to get home soon

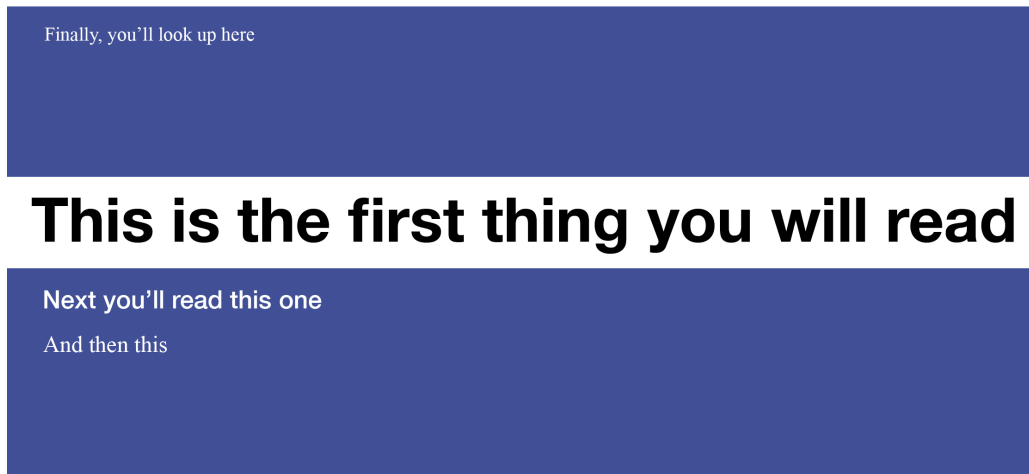
**Example 21. Planar organization and shape of transcript used to reveal dynamics of talk**

In Example 22, the relative amounts of time spent talking during this interaction are revealed through color blocking.



**Example 22. Color blocking used to reveal dynamics of talk**

Finally, an additional principle of graphic design is that information can be effectively ordered through hierarchies. There are different ways to create hierarchies using many of the elements already discussed, like typography, color, size, and placement, but the logic underlying all of them is that information of different kinds and values can be given more or less priority in a document depending on how it's handled. Thus, when looking at Example 23:



**Example 23. Hierarchies in graphic design as priority-giving and attention-directing device**

Using this principle, a transcriber can guide a reader's eye to focus on the most important information in the proper order, without having to provide much explicit instruction.

## Principle 4: Prioritize Intelligibility

Early written-text-only transcripts were relatively easy to understand. Once standard conventions were in place, reading a transcript could be taught or picked up quickly. However, as discourse analysis moved toward incorporating more information, and developing new—and often sui generis—methods for representing various phenomena, many transcripts became overloaded, unbalanced, and difficult to comprehend without careful study. To prioritize intelligibility in designing a transcript is to make it inviting, accessible, and decipherable without much effort, rather than complicated and frustrating. From a graphic design point of view, crafting an intelligible transcript requires attention to the plastic features of transcript components, as well as to the relations that obtain between them, and to a range of functional attributes, including readability (does the transcript make sense?), fidelity (does it at least minimally match the empirical reality it represents?), and cogency (does it contribute to a strong argument?).

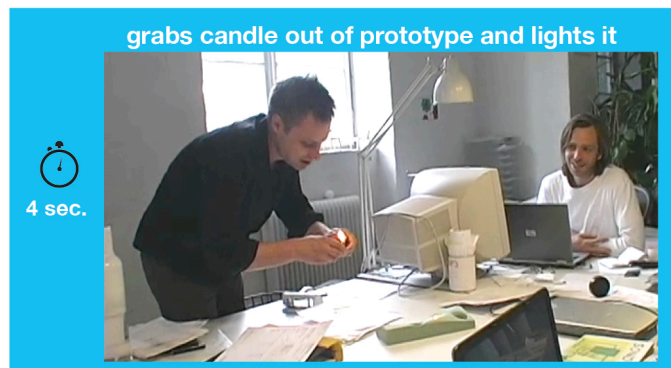
In order to do this, it's important to consider adopting a few basic practices, including expanding the core suite of software that's used to create a transcript, since it's the

software that establishes the basic parameters for how transcript components can be transformed and placed. Word processing software, like Microsoft Word or Apple Pages, seems to be the most commonly used for creating published transcripts, but these applications were not designed to handle anything other than very simple integrations of written text and visual images. In addition, they don't really allow a user to do things easily like adjust the margins or crop the size of a whole transcript, or export the transcript in different formats and resolutions. In a very real sense this software behaves like a glorified typewriter. Various layout applications—big ones, like Adobe InDesign, and similar smaller alternatives, like Affinity Publisher or Scribus, which is open-source—are better because they are specifically designed for creating semiotically complex forms. They may seem daunting at first, but a little practice goes a long way with these applications.

Resolution—that is, the number of “dots” per square inch (dpi) of any graphic—should also be an important consideration for creating transcripts. Relative resolution levels are implicated at practically every stage of the discourse analytic process. The cameras we use to record an event capture images in higher or lower resolution, the software we use to create transcripts export higher or lower resolution files, and the devices on which we read transcripts have higher or lower resolution screens. As recently as the early 2010s, cameras, files, and screens mostly all worked in low resolution, but with high-resolution cameras, screens, and projectors rapidly turning those older technologies obsolete, low-resolution images and transcripts have simply become inadequate in terms of how people now consume digital media, including academic articles. For example, older defaults, like 72 dpi for images displayed on the web, were adopted at a time when low-resolution computer monitors were the norm and images were viewed at only one size. But we now interact with visual images differently. We constantly and dynamically change their sizes by tapping and zooming and scrolling, and if a visual image has a lower resolution, it will look “blocky” when it's expanded, rendering it essentially useless. This means that if we want our published transcripts to look good, and to “work” in ways that are compatible with the devices we use, we need to consider the resolution of every element we're placing in the transcript, as well as the final version of the overall transcript we create.

Another way to prioritize intelligibility is to avoid the use of complex notation conventions if they're not relevant to a specific point that's trying to be conveyed. Conventions work, of course, and they're critical to transcript graphesis, but they are not always required, and in many cases simple and familiar notation might also work well (and not require extra decoding by the reader). Example 24 uses almost no technical conventions, and relies on graphics and layout to mark relevant features of the interaction—English translations are placed just below the original Swedish, and set in a different font; actions are represented with screen grabs placed in blue boxes, which are sized according to how long the action took; and the timing of the action is marked in a common non-technical notation.

- P Tända det.  
Light it.
- M Huh?  
Huh?
- P Tända.  
Light.
- M Tända alla?  
Light them all?
- P Ja:::  
Ye::s!
- M Släkt ljuset!  
Turn off the lights!



- M Det är bra. Det är fint.  
It's good. It's nice.
- Och så vitt på det?  
And so it'll be white?
- P Vitt.  
White.



- M Men jag saknar en liten  
But I miss the little  
snögubbe, måste jag säga.  
snowman, I have to say.

**Example 24. Using graphics and layout to highlight features of interaction**

Not every example of language-in-use can be slotted into this sort of format, of course. But every example of language-in-use can absolutely be put into a transcript that is somehow designed in ways that highlights clearly and with force the specific features of language (and discourse more generally) the transcriber is trying to communicate.

## Conclusion

Transcription is, at its core, a solution—a kludgy, imperfect solution—to the problem of translating rich empirical “lived” phenomena into the comparatively “dead” forms required for sustained scientific analysis. As it’s been traditionally conceived, to be a skillful transcriber is to listen well to a recording, as a professional listens, and to represent what is recorded as accurately as possible using specified conventions that other professionals are expected to understand.

Transcription is also a significant creative process involving multiple sets of compositional choices for building, shaping, and *designing* a transcript. As such, transcribers don’t have to just make a transcript, they can make a transcript compelling, or inviting, or even beautiful, and in so doing encourage deeper and more sustained interaction with it. They can make something that’s easy to read, and that people want to look at and spend time with. They can make graphic objects that through their specific forms marshal arguments about the lived phenomena that the transcripts represent, arguments that would be impossible to make using grayscale technologies alone. Perhaps it seems like making a transcript beautiful requires learning an entirely new set of skills, and well, yes, it does, but it’s certainly no different or more daunting than having to learn professional listening, or how to use a transcription system accurately. It’s just another way of working to develop the overall transcribing process.

Each of these perspectives represents a different responsibility for the transcriber. There’s the responsibility to the recording, the event, and the event’s participants, a requirement that the transcriber *gets the representation right*, whatever that might mean in the instance and to whatever degree it’s possible. There’s also the responsibility to future readers who will “activate” that representation, in one way or another, when reading and using the published transcript. These two responsibilities are in essence always linked, yet for several decades of discourse analytic work we’ve focused almost exclusively on the former while short-shrifting the latter.

As an alternative, acknowledging these two responsibilities simultaneously prompts at least two transformations in how to envision discourse analytic work. First, it invites us to reimagine the role of the transcriber not as a passive, and often reluctant, mediator between representation and reality, but as an active and motivated composer of complex instruments of visual communication. Second, it asks us to embrace, rather than avoid or

ignore, the actual “creative” conditions of our research practice, including that transcripts are aesthetic objects designed for specific purposes. This doesn’t mean transcription is about making things up. Rather, it means making things with care and intent from the very start of a research project such that they fit well into the knowledge circuits within which they travel.

To be sure, there might be some discomfort with a push to treat transcribing as design because design feels almost too *constructivist* for how transcription has been ideologically configured. What’s important to bear in mind, though, is that embracing transcribing as a process of design is mostly just a re-description of what’s already going on, which is to say, *transcripts are already designed and we are their designers*. We just don’t typically conceive of the work that way. This is why they are, thus, often “bad”; first, because we frequently don’t give proper care to their design; and second, by transcribing, we are already committed to a set of values and criteria—to their communicability, usability, to-be-looked-at-ness—that our transcripts often fail to satisfy.

But by re-describing transcription using a register borrowed from graphic design, the consequences of design choices in making transcripts are brought to the fore, and the critical point that Ochs offered early on—that both the constituent elements and overall form of a transcript matter to their efficacy, and thus require reflexive attention—can finally be realized in full. The hope is that in treating transcripts as designed rather than simply as made, and by attending to the details of transcription aesthetics, theory-building by means of transcription can produce new, and hopefully better, analyses of how humans interact with each other and their environments.

## Acknowledgments

This article started life as a presentation at the American Anthropological Association meetings in 2013. It was reborn in workshops at the Conference on Language, Interaction, and Culture (CLIC) at UCLA in 2014, and at the University of Chicago in 2016. I want to thank the participants of those workshops for their feedback, Costas Nakassis, Meghanne Barker, and two anonymous reviewers for their comments on earlier drafts, and three of my mentors—Candy Goodwin, Chuck Goodwin, and Elinor Ochs—for planting the seeds of these ideas a long time ago.

## Endnotes

1. I’m focusing here on published transcripts because of their status as Latourian (1986) immutable mobiles that are made specifically for people other than the transcriber to look at. There are lots of other intermediate stages of transcription, and software like ELAN that produces particular kinds of transcripts, but the focus here is on the versions that tend to (and probably should) get the most attention.↩

2. One could argue that what I'm describing as traditional transcription aesthetics have clearly been shown to work, so why should we change them? And that's a good point! But, I think it's a point with some problems. First, it assumes that beyond orthodox conversation analysis there even is some standard at all, which is certainly not accurate in the strict (and even loose) sense of "standard." Second, it assumes that what we call "conventions" don't change or vary. But any cursory review of discourse analytic publications would reveal that it's common for researchers to add their own symbols when existing systems don't include a notation for some observed phenomenon. Moreover, even within a rigid system there are often multiple ways of marking the same thing, like laughter or emphasis. Third, there is nothing even approaching a standard way for including images and other visual information in a transcript. All of which is to say, the systems we use are already malleable—even if we orient to them as if they're not—so perhaps it's time to start considering a new transcription aesthetics. ↩

3. We can see similar effects in other language-oriented fields. In her book *Graphic Representation of Models in Linguistic Theory*, Ann Harleman Stewart (1976) explored the aesthetic conventions involved in graphically modeling language within the field of linguistics—or at least the conventions dominant in the 1970s. She focused, in particular, on three graphic types: *trees* (used for constituent analysis and displaying taxonomic or genetic relations); *boxes* (also for constituent or componential analysis); and *matrices* (for componential analysis and taxonomy). While these model-types serve different immediate purposes, they all share some common features. First and foremost, they present and structure complex information clearly and legibly. They also facilitate editorial discretion by removing "noise" from the model, which is to say, all the stuff that doesn't fit the paradigm the model is trying to capture (though of course there's always a question as to what counts as "noise"). These models also reveal significant relations among elements and afford comparison between particular examples (for instance, any given sentence can be slotted into a tree diagram). Crucially, as Stewart argued, in doing all of this, these graphic models *produce arguments* and help *generate scientific knowledge*, functioning as critical tools for doing the fundamental work of linguistics, and contributing to—or even constituting—the very epistemological basis of linguistics as a scientific field. Much of how the ontology of language itself was imagined by linguists in the mid-to-late twentieth century can be traced directly back to the aesthetics of these models. ↩

4. At this point Courier does little more than loosely index a "technical" style without providing any actual technical benefit. To use it is to perform a symbolic act, one that signals membership within a specific community of practice. Courier's ability to help with turn alignments has long been superseded by, on the one hand, the availability of many other fixed-width fonts (Inconsolata and Roboto Mono, for instance, both of which look better than Courier, and are available to download for free from Google) and, on the other hand, the ability to use other graphic design (and even word processing) techniques for aligning text. ↩

## References

- Aarsand, Pål and Anna Sparrman. 2019. Visual Transcriptions as Socio-Technical Assemblages. *Visual Communication*. July 2019. doi: <https://doi.org/10.1177/1470357219852134>.
- Ayaß, Ruth. 2015. Doing Data: The Status of Transcripts in Conversation Analysis. *Discourse Studies* 17(5):505–28.
- Beaulieu, Anne. 2001. Voxels in the Brain: Neuroscience, Informatics and Changing Notions of Objectivity. *Social Studies of Science* 31(5):635–80.
- Bezemer, Jeff and Diane Mavers. 2011. Multimodal Transcription as Academic Practice: A Social Semiotic Perspective. *International Journal of Social Research Methodology* 14(3):191–206.
- Black, Alison, Paul Luna, Ole Lund, Sue Walker, Paul Luna, Ole Lund, and Sue Walker. 2017. *Information Design: Research and Practice*. New York: Routledge.
- Black, Steven P. 2017. Anthropological Ethics and the Communicative Affordances of Audio-Video Recorders in Ethnographic Fieldwork: Transduction as Theory. *American Anthropologist* 119(1):46–57.
- Bødker, Susanne. 1998. Understanding Representation in Design. *Human-Computer Interaction* 13(2):107–25.
- Bonsiepe, Guy. 1999. *Interface: An Approach to Design*. Maastricht: Uitgeverij De Balie.
- Bucholtz, Mary. 2000. The Politics of Transcription. *Journal of Pragmatics* 32(10):1439–65.
- Bucholtz, Mary. 2007. Variation in Transcription. *Discourse Studies* 9(6):784–808.
- Bucholtz, Mary. 2009. Captured on Tape: Professional Hearing and Competing Entextualizations in the Criminal Justice System. *Text & Talk – An Interdisciplinary Journal of Language, Discourse & Communication Studies* 29(5):503–23.
- Burri, Regula Valérie. 2008. Doing Distinctions: Boundary Work and Symbolic Capital in Radiology. *Social Studies of Science* 38(1):35–62.
- Cantor, Robert M. 2000. Foundations of Roentgen Semiotics. *Semiotica* 131(1–2):1–18.

Choe, Hanwool. 2019. Eating Together Multimodally: Collaborative Eating in Mukbang, a Korean Livestream of Eating. *Language in Society* 48(2):171–208.

Drucker, Johanna. 2014. *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press.

Du Bois, John W. 1991. Transcription Design Principles for Spoken Discourse Research. *Pragmatics* 1(1):71–106.

Dumit, Joseph. 2004. *Picturing Personhood: Brain Scans and Biomedical Identity*. Princeton, NJ: Princeton University Press.

Duranti, Alessandro. 1994. *From Grammar to Politics: Linguistic Anthropology in a Western Samoan Village*. Berkeley: University of California Press.

Duranti, Alessandro. 2006. Transcripts, Like Shadows on a Wall. *Mind, Culture, and Activity* 13(4):301–10.

Edwards, Jane A. 2005. The Transcription of Discourse. In Deborah Schiffrin, Deborah Tannen, and Heidi E. Hamilton, eds. *The Handbook of Discourse Analysis*, pp. 321–48. Oxford: Wiley.

Efron, David. 1941. *Gesture and Environment: A Tentative Study of Some of the Spatio-Temporal and Linguistic Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living Under Similar as Well as Different Environmental Conditions*. New York: King's Crown Press.

Erickson, Frederick. 2004. Origins: A Brief Intellectual and Technological History of the Emergence of Multimodal Discourse Analysis. In Philip LeVine and Ron Scollon, eds. *Discourse and Technology: Multimedia Discourse Analysis*, pp. 196–207. Washington, DC: Georgetown University Press.

Frascara, Jorge. 1988. Graphic Design: Fine Art or Social Science? *Design Issues* 5(1):18–29.

Gibson, Will, Helena Webb, and Dirk vom Lehn. 2014. Analytic Affordance: Transcripts as Conventionalised Systems in Discourse Studies. *Sociology* 48(4):780–94.

Goodwin, Charles. 1981. *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

- Goodwin, Charles. 1994. Professional Vision. *American Anthropologist* 96(3):606–33.
- Goodwin, Charles and Marjorie Harness Goodwin. 1996. Seeing as Situated Activity: Formulating Planes. In David Middleton and Yrjo Engeström, eds. *Cognition and Communication at Work*, pp. 61–95. Cambridge: Cambridge University Press.
- Goodwin, Marjorie Harness. 1998. Games of Stance: Conflict and Footing in Hopscotch. In Susan Hoyle and Carolyn Temple Adger, eds. *Kids' Talk: Strategic Language Use in Later Childhood*, pp. 23–46. New York: Oxford University Press.
- Greimas, Algirdas Julien. 1989. Figurative Semiotics and the Semiotics of the Plastic Arts. *New Literary History* 20(3):627–49.
- Grimshaw, Anna. 2001. *The Ethnographer's Eye: Ways of Seeing in Anthropology*. Cambridge: Cambridge University Press.
- Groeber, Simone and Evelyne Pochon-Berger. 2014. Turns and Turn-Taking in Sign Language Interaction: A Study of Turn-Final Holds. *Journal of Pragmatics* 65:121–36.
- Harland, Robert. 2011. The Dimensions of Graphic Design and Its Spheres of Influence. *Design Issues* 27(1):21–34.
- Haviland, John B. 1993. Anchoring, Iconicity, and Orientation in Guugu Yimithirr Pointing Gestures. *Journal of Linguistic Anthropology* 3(1):3–45.
- Heath, Christian. 1986. *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press.
- Henderson, Holly. 2018. Difficult Questions of Difficult Questions: The Role of the Researcher and Transcription Styles. *International Journal of Qualitative Studies in Education* 31(2):143–57.
- Hepburn, Alexa and Galina B. Bolden. 2012. The Conversation Analytic Approach to Transcription. In Jack Sidnell and Tanya Stivers, eds. *The Handbook of Conversation Analysis*, pp. 57–76. Oxford: Wiley.
- Heritage, John. 1984. *Garfinkel and Ethnomethodology*. Cambridge: Polity.
- Ivarsson, Jonas. 2010. Developing the Construction Sight: Architectural Education and Technological Change. *Visual Communication* 9(2):171–91.

Jaffe, Alexandra and Shana Walton. 2000. The Voices People Read: Orthography and the Representation of Non-Standard Speech. *Journal of Sociolinguistics* 4(4):561–87.

Keenan (Ochs), Elinor. 1974. Conversational Competence in Children. *Journal of Child Language* 1(2):163–83.

Labov, William. 1963. The Social Motivation of a Sound Change. *Word* 19(3):273–309.

Labov, William. 1966. The Effect of Social Mobility on Linguistic Behavior. *Sociological Inquiry* 36(2):186–203.

Latour, Bruno. 1986. Visualization and Cognition: Thinking with Eyes and Hands. *Knowledge and Society: Studies in The Sociology of Culture Past and Present*, vol. 6, edited by H. Kuklick, pp. 1–40. Greenwich, CT: JAI Press.

Laurier, Eric. 2014. The Graphic Transcript: Poaching Comic Book Grammar for Inscripting the Visual, Spatial and Temporal Aspects of Action: The Graphic Transcript. *Geography Compass* 8(4):235–48.

Laurier, Eric and Barry Brown. 2011. The Reservations of the Editor: The Routine Work of Showing and Knowing the Film in the Edit Suite. *Social Semiotics* 21(2):239–57.

Lynch, Michael. 1988. The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences. *Human Studies* 11(2–3):201–34.

Lynteris, Christos. 2017. Zoonotic Diagrams: Mastering and Unsettling Human-Animal Relations. *Journal of the Royal Anthropological Institute* 23(3):463–85.

McIlvenny, Paul. 2009. Communicating a ‘Time-out’ in Parent–Child Conflict: Embodied Interaction, Domestic Space and Discipline in a Reality TV Parenting Programme. *Journal of Pragmatics* 41(10):2017–32.

Meredith, Joanne. 2016. Transcribing Screen-Capture Data: The Process of Developing a Transcription System for Multi-Modal Text-Based Data. *International Journal of Social Research Methodology* 19(6):663–76.

Miethaner, Ulrich. 2000. Orthographic Transcriptions of Non-Standard Varieties: The Case of Earlier African-American English. *Journal of Sociolinguistics* 4(4):534–60.

Mitchell, W. J. T. 1984. What Is an Image? *New Literary History* 15(3):503–37.

Mondada, Lorenza. 2007. Commentary: Transcript Variations and the Indexicality of Transcribing Practices. *Discourse Studies* 9(6):809–21.

Murphy, Keith M. 2017. Fontroversy! Or, How to Care About the Shape of Language. In Jillian R. Cavanaugh and Shalini Shankar, eds. *Language and Materiality: Ethnographic and Theoretical Explorations*, pp. 63–86. Cambridge: Cambridge University Press.

Murphy, Keith M., Jonas Ivarsson, and Gustav Lymer. 2012. Embodied Reasoning in Architectural Critique. *Design Studies* 33(6):530–56.

Norris, Sigrid. 2002. The Implication of Visual Research for Discourse Analysis: Transcription beyond Language. *Visual Communication* 1(1):97–121.

Ochs, Elinor. 1979. Transcription as Theory. In Elinor Ochs and Bambi B. Schieffelin, eds. *Developmental Pragmatics*, pp. 41–72. New York: Academic Press.

Ochs, Elinor, Patrick Gonzales, and Sally Jacoby. 1996. “When I Come Down, I’m in a Domain State”: Talk, Gesture, and Graphic Representation in the Interpretive Activity of Physicists. In Elinor Ochs, Emanuel A. Schegloff, and Sandra Thompson, eds. *Interaction and Grammar*, pp. 328–69. Cambridge: Cambridge University Press.

O’Connell, Daniel C. and Sabine Kowal. 1994. Some Current Transcription Systems for Spoken Discourse: A Critical Analysis. *Pragmatics* 4(1):81–107.

O’Connell, Daniel C. and Sabine Kowal. 1999. Transcription and the Issue of Standardization. *Journal of Psycholinguistic Research* 28(2):103–20.

O’Connell, Daniel C. and Sabine Kowal. 2000. Are Transcripts Reproducible? *Pragmatics* 10(2):247–69.

Polidoro, Piero. 2019. Image Schemas in Visual Semiotics: Looking for an Origin of Plastic Language. *Cognitive Semiotics* 12(1). <https://doi.org/10.1515/cogsem-2019-2006>.

Prasad, Amit. 2005. Making Images/Making Bodies: Visibilizing and Disciplining through Magnetic Resonance Imaging (MRI). *Science, Technology, & Human Values* 30(2):291–316.

Preston, Dennis R. 1985. The Li’l Abner Syndrome: Written Representations of Speech. *American Speech* 60(4):328–36.

Recktenwald, Daniel. 2017. Toward a Transcription and Analysis of Live Streaming on Twitch. *Journal of Pragmatics* 115(July):68–81.

Roberts, Celia. 1997. Transcribing Talk: Issues of Representation. *TESOL Quarterly* 31(1):167–72.

Roberts, Lucienne. 2006. *GOOD: An Introduction to Ethics in Graphic Design*. First Printing edition. Lausanne Worthing: AVA Publishing.

Romero, Catherine, Sabine Kowal, and Daniel C. O’Connell. 2002. Notation Systems for Transcription: An Empirical Investigation. *Journal of Psycholinguistic Research* 31(6):619–31.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50(4):696–735.

Stewart, Ann Harleman. 1976. *Graphic Representation of Models in Linguistic Theory*. Bloomington: Indiana University Press.

Streeck, Jürgen. 1983. *Social Order in Child Communication: A Study in Microethnography*. Philadelphia: John Benjamins Publishing.

Suchman, Lucy. 1995. Making Work Visible. *Communications of the ACM* 38(9):56–64.

Tufte, Edward R. 2006. *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*. 2nd edition. Cheshire, CT: Graphics Press.

© Copyright 2021 Keith M. Murphy

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.