# DOING NEWS FRAMING ANALYSIS II

## Empirical and Theoretical Perspectives

*Edited by Paul D'Angelo*

Second edition published 2018

ISBN: 978-1-138-18854-9 (hbk)
ISBN: 978-1-138-18855-6 (pbk)
ISBN: 978-1-315-64223-9 (ebk)

9

### A METHODOLOGICAL APPROACH FOR INTEGRATIVE FRAMING ANALYSIS OF TELEVISION NEWS

*Viorela Dan*



DOI: 10.4324/9781315642239-13

The funder for this chapter is Ludwig-Maximilians-Universität München

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

# 9

# A METHODOLOGICAL APPROACH FOR INTEGRATIVE FRAMING ANALYSIS OF TELEVISION NEWS

*Viorela Dan*

Frames in news are essentially coherent interpretations of issues and people by actors—including sources, advocates, and journalists themselves—who rely on various kinds of reasoning devices and framing devices (D'Angelo, 2002; Entman, 1993; Matthes & Kohring, 2008; Reese, 2001; Van Gorp, 2007). If we understand frames in this way, then we must acknowledge the possibility that frames are articulated not just in printed and spoken news texts, but also via still and moving visuals. In *Framing Public Life*, and later, in the first edition of *Doing News Framing Analysis*, both Messaris and Abraham (2001) and Coleman (2010), respectively, clarified the call to analyze news visuals for the frames they entail (see also Grabe & Bucy, 2009). During the last two decades, the number of visual framing studies has increased considerably, as the obstacles related to researching visuals have gradually been ironed out. According to Coleman (2010), therefore, the way to advance theory building and methodology is to move toward "integrative work," by which she meant framing analyses combining words and visuals (p. 235).

Nowhere else is this situation as pressing as in analyses of television news. At least since several articles in a seminal issue of *American Behavioral Scientist* intimated that framing analysis would be useful in understanding "picture messages" (Graber, 1989; see also Gamson, 1989), researchers have tried to "cop[e] with the duality of television news" (Barkin, 1989, p. 153) both from a content perspective and from the perspective of audience comprehension and learning. When getting news from television, an audiovisual medium, people are exposed to words *and* visuals, and not to words *or* visuals. Yet, as commonsensical as the argument is that two integrated modalities cannot be fully understood when one of them is neglected (Dan, 2018), it is not reflected in empirical analyses: Most framing analyses of TV news do not pay attention to the interwoven duality of the verbal and the visual stream of information.

DOI: 10.4324/9781315642239-13

While our field has moved beyond archaic views that visuals do not matter, one preconception persists: Visuals are often assumed to play a supporting role to the words, deemed the undisputed protagonist (see Dan, 2018; Fahmy, Bock, & Wanta, 2014; Grabe & Bucy, 2009). Often enough, however, visuals go beyond the words. Think of the TV news coverage of ISIS/ISIL. How often have you seen excerpts of positively toned ISIS propaganda videos shown in TV news? I, for one, have seen plenty. They showed young men standing on top of tanks proudly waving the ISIS flag and appearing to be enjoying themselves greatly. The voice-over was typically the exact opposite: ISIS was described as a very dangerous militant group, and it was stated that the video shown was excerpted from a propaganda video.

Grimes (1990) must have had in mind similar examples when he deemed poor audiovisual correspondence to be one of the "loose ends [that] mar many newscasts" (p. 16). Of course, one could counterargue that there was a verbally delivered disclaimer and that this should be able to wipe out or tone down the positive visuals. However, a substantial body of experimental research suggests that this is unlikely, as mismatches lead audiences to absorb and recall the visuals over the words, leading scholars to speak of a *picture superiority effect* (PSE) (Aust & Zillmann, 1996; Gibson & Zillmann, 2000; Graber, 1990). This is precisely the point of conducting integrative framing analyses. Recently, Geise and Baden (2015) made a strong argument in favor of including both the verbal and the visual component of news in studies of framing effects.

Integrative framing analyses go beyond classical audiovisual redundancy literature, as they are not interested in finding out whether the verbal component of TV news was illustrated with redundant visuals. Rather, they are designed to find out whether the same interpretation of the issue/people at hand was articulated both verbally and visually. This informs us about how news contributes to the social construction of that issue/those people. By advancing verbal and visual frames that are incongruent, journalists might construct reality in ways that they did not intend and impact audiences accordingly. This is not to say that audience members who viewed news stories such as the one described above will become radicalized; however, they might be less inclined to view ISIS as a problem in need of a solution and reject governmental spending on counterterror strategies. Judging by the verbal component of such news stories, this is the opposite of what journalists wanted to convey.

In the following literature review, I present some recent and classical content analyses of TV news and experimental studies to substantiate this anecdotal evidence. It will become apparent that news words and news visuals in the same news story often convey conflicting meanings (Walma van der Molen, 2001) and even frames (Dan, 2018; Powell, Boomgaarden, De Swert, & de Vreese, 2015). My main argument in reviewing the literature is this: The relative lack of framing analyses on both words and visuals in TV news is likely due to the paucity of methodological advice on how to conduct such analyses. Thus, in the second part of this

DOI: 10.4324/9781315642239-13

chapter, I offer in-depth methodological advice, which will hopefully demystify integrative framing analyses of TV news (see Van Hoof, Takens, & Oegema, 2010; Walma van der Molen, 2001) and show that they are not just worthwhile but also doable. Before I go through the literature in those two sections, however, it will be useful to consider the verbal and visual components within the overall composition of a TV newscast, as any methodological approach to integrative framing analysis must take all of them into account.

## The Structure of a TV Newscast

Figure 9.1 represents a summary of current understandings on how TV newscasts are typically made up (Clausen, 2003; Liebler & Bendix, 1996; Machin & Polzer, 2015; Schaefer & Martinez, 2009; Silcock, 2007). Many TV newscasts begin with a headline-type overview of the day's events. Then, each story is presented individually using one of the following techniques (see Figure 9.1). One possibility is to present the story using anchor narration, which may be paired with still or moving images shown full-screen or in the background. When images are shown in the background, news segments of this type are referred to as "studio anchor scenes." When images are shown full-screen, the segments are called VO (voice-over) or VO/SOT (voice-over sound-on-tape). The difference between VO and VO/SOT is the following: In VO, the anchor reads out the entire script live while a news video is played. In VO/SOT, the anchor reads out the script live while video is shown until a sound bite is played (video or audio). The anchor is quiet while the SOT is played and potentially continues reading the news story after the SOT is over.

The second technique, which is far more popular (Cushion, Rodger, & Lewis, 2014), involves the use of a lead-in together with a news package. The lead-in can take the form of classical anchor narration, but anchor/reporter exchanges are also conceivable (e.g., in the studio or using a split-screen). The news package is a pre-produced news item that comes after the lead-in. To the extent that the newscast is interrupted by commercials, upcoming stories are teased before each break. Some newscasts end with a recap of the stories covered (see Figure 9.1).

Given the popularity of news packages, it is worthwhile to consider their components. According to previous literature, and as shown in the upper part of Figure 9.1, the verbal component of a news package consists of an overview of the issue and reactions (Clausen, 2003; Liebler & Bendix, 1996; Machin & Polzer, 2015; Schaefer & Martinez, 2009; Silcock, 2007). In the overview of the issue, facts, journalistic assessments of recent developments, and background information are offered (e.g., context, history, views on the issue). The reporter voice-over is the most common format for this segment, but it is also possible that the reporter is shown talking on camera, as denoted by the first dotted arrow in Figure 9.1. The second verbal component of news packages consists of reactions to the story, which are "short sections in which people [other than the reporter]

DOI: 10.4324/9781315642239-13

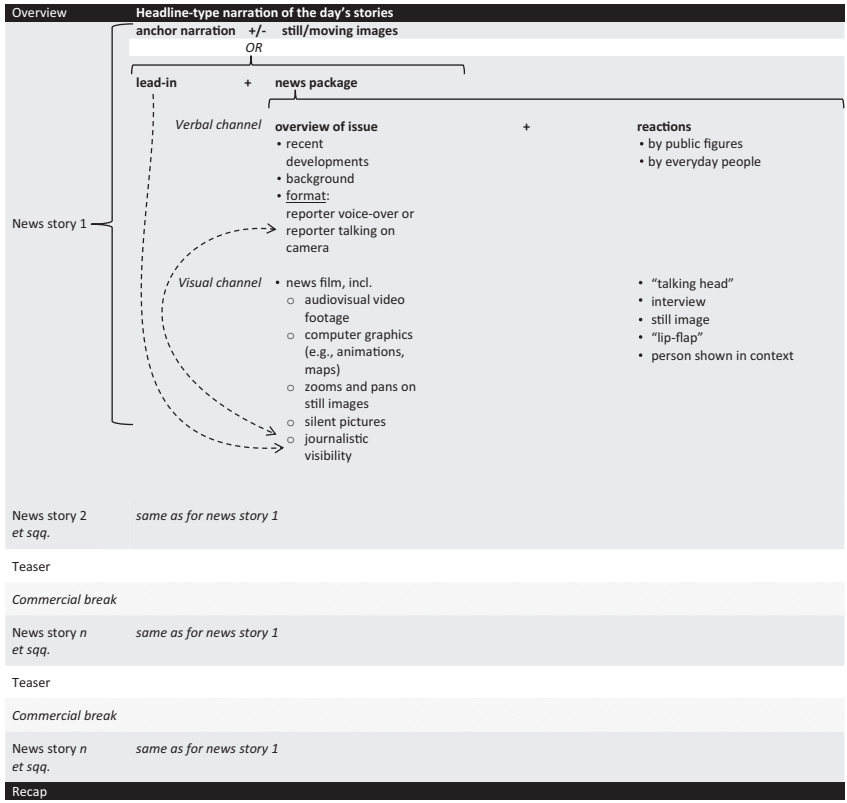


**FIGURE 9.1** The structure of a TV newscast

talk on camera" (Stephens, 2005, p. 316). These people can be public figures like politicians or everyday people ("man-on-the-street").

The visual components of a news package differ along these lines. When an overview of the issue is offered, still/moving images are paired with reporter voice-over. Alternatively, the reporter is shown talking on camera ("journalistic visibility"). The images combined with voice-over take various formats, including news film (i.e., video footage of people and organizations involved in the issue), zooms and pans on still images, and computer graphics (e.g., animations and maps). These images can be either topical or taken from the archive; they can also be standard news pictures (Brosius, Donsbach, & Birk, 1996; J. Robinson & Levy, 1986; Walma van der Molen, 2001). In journalistic visibility segments, for instance during a sign-off/wrap-up, several formats are conceivable. The reporter may be shown standing still and talking on camera ("stand-up"), walking toward the camera while talking and gesticulating ("on-location reporting"), or on a split-screen when interacting with the anchor ("anchor/reporter exchange"). Alternatively, when no video footage of the reporter is available, he or she may be

DOI: 10.4324/9781315642239-13

shown merely in a still photograph, e.g., when they are connected via telephone. Camera people use shorter camera distances for the stand-up segment than for the on-location reporting (close-up vs. medium/wide shot). The visuals used when reactions to the story are presented include audiovisual excerpts of an interview or a public appearance ("sound-bites" or "talking heads") or merely visual material paired with reporter voice-over (i.e., individuals shown but not heard talking, aka "lip-flap"). Camera distances vary from close-ups for interviews and statements to medium/wide shots for public appearances.

It is important to understand that the components of a news package are not necessarily presented in the order depicted in Figure 9.1, nor are they offered in full, one after another. Rather, they tend to be presented as "installments" throughout the news package (Clausen, 2003, p. 68). This means that bits of each component will alternate several times during one news package. For example, some background information could be offered before the presentation of some reactions by public officials, followed by more information on recent developments, a man-on-the-street segment, and so on.

## What We Know So Far about the Duality of TV News

### *Content Analytical and Framing Studies*

The proliferation of research into the possibility that words and visuals in TV news are mismatched—and thus, that both should be analyzed—owes much to the devotedness of a few scholars. Doris Graber was surely the most productive of these, as her numerous studies were able to clarify the relevance of this endeavor to communication scholars. A case in point is a now classical study in which she examined how, if at all, TV news visual went beyond the verbal story line (Graber, 1990). She found that almost half (49%) of the TV news stories in her sample used visuals that illustrated aspects left unspoken. About a decade later, Walma van der Molen (2001) analyzed news items directed at children and adults. She found that adult-oriented news was more likely to use visuals indirectly related to the voice-over (43.3%) than was children's news (37%), which preferred directly related/redundant visuals (42.8%) that were far less common in adults' news (22.3%). More recently, Horvat (2010) investigated the TV news coverage of the 2005 riots in France using a qualitative approach. She reported various instances in which the interpretations advanced verbally conflicted with those articulated visually. For example, inconspicuous words were paired with visuals presenting Muslims as dangerous intruders, and vice versa. Similar examples can be found in other studies on a variety of topics (Bahador, 2008; Kelly, 2010; Liebes & Kampf, 2009; C. Robinson & Powell, 1996).

This brief review is representative of how scholars' understanding of the rapport between the verbal and the visual component of TV news has evolved over time. Initially, scholars were interested in whether or not the audio and the visual track

DOI: 10.4324/9781315642239-13

of TV news conveyed the same meaning. They bundled their research under the keyword "audio-visual redundancy" and looked for a full overlap or perfect match between the modalities, as in visuals synchronized with their word labels (Reese, 1984). Over the years, scholars' use of the term became less literal (see Zhou, 2005). It became increasingly common to think of the degree to which the verbal matches the visual as a continuum, with the endpoints being full overlap and complete contradiction. In this way, researchers considered it more appropriate to think of the interplay between words and visuals in terms of various degrees of congruence as opposed to dichotomous yes/no (Brosius et al., 1996; Walma van der Molen, 2001).

It was a long time before scholars used framing as a theoretical guide in work designed to determine the extent to which the verbal and the visual component of TV news transported congruent interpretations.[1] In fact, to my knowledge, Van Hoof et al.'s (2010) study is the only one that explicitly does this. Of course, other framing studies of TV news investigate its verbal and visual components; however, they are generally vague about the degree of congruence between these modalities (Entman, 1991; Liebler & Bendix, 1996). In this regard, the study by Reynolds and Barnett (2003) is less explicit than Van Hoof et al.'s (2010), but more explicit than the other two (Entman, 1991; Liebler & Bendix, 1996). I now review each of them and highlight similarities and differences in their approach.

Liebler and Bendix (1996) investigated the coverage of ABC, CBS, and NBC on a controversy regarding whether or not harvest limits on old-growth forests should be issued. The authors identified two frames: prosave and procut. The "prosave" frame presented these forests as irreplaceable and clear-cutting as non-esthetical. Also, it emphasized the beneficial aspects of the harvest ban on the environment (including saving a threatened bird species), with an argument that such limits would not cause job losses in the timber industry. By contrast, the "procut" frame reduced the debate to a narrow conflict between owls (old-growth forests were the habitat of spotted owls) and people, arguing that "the owls couldn't possibly need so much land, and weren't worth the cost if they did" (Liebler & Bendix, 1996, p. 55). Though the authors did not explicitly compare the verbal and the visual frames in each newscast, one finding engenders the assumption that they must have been mismatched. Specifically, the procut frame was clearly preferred in the verbal component of the news stories, whereas the visuals were almost evenly distributed among the two frames. Such observations are only possible by reading between the lines.

Entman (1991) explored CBS evening newscasts on two separate yet very similar military operations. In both cases, civilian planes wrongfully identified as hostile targets were subsequently shot down, killing almost 300 people each. Despite these similarities, Entman found considerable differences in the way the two incidents were framed. Specifically, a moral frame was applied for the 1983 incident, in which the perpetrator was a Soviet fighter plane (i.e., the incident was presented as a deliberate act, typical of the Soviets, who were to blame).

DOI: 10.4324/9781315642239-13

By contrast, a technical frame was used for the 1988 shooting down, in which the perpetrator was a US Navy ship (i.e., tragic accident, technical information, no blame attributions). But judging by the way Entman (1991) presented results of the TV news analysis,[2] it is hard to tell whether there were any mismatches between the verbal frames and the visual frames.

Drawing conclusions about the extent to which verbal frames match visual frames in TV news is nearly impossible based on the two studies I just discussed. This is because neither Liebler and Bendix (1996) nor Entman (1991) identified visual *frames*. Rather, they acknowledged visual content in the analysis of verbal frames. Thus, these authors could not compare verbal frames with visual frames. Notably, too, these two studies reached different results (partial congruence vs. redundancy).

Research making more explicit claims about verbal–visual frame congruence was conducted by Reynolds and Barnett (2003) and Van Hoof et al. (2010). The former examined the verbal and visual framing of the 9/11 attacks during the first 12 hours of "breaking news" coverage on CNN. The latter were interested in the verbal and visual framing of the 2006 Dutch national election. Reynolds and Barnett's (2003) qualitative study found that one single dominant frame was conveyed, namely, "that a U.S. military-led international war would be the only meaningful solution to prevent more terrorist attacks" (Reynolds & Barnett, 2003, p. 91). While the video track did not match the audio in a literal sense, the war-on-terror frame was employed consistently across the modalities investigated. Van Hoof et al. (2010) went further and quantified the congruence between verbal and visual content in TV news. Generally, they found low degrees of congruence, which prompted them to attest to "poor framing … in political news" (p. 1). They noticed differences by frame and by network. Specifically, the verbal and visual tracks were most congruent for the human interest frame (59.5%) and least congruent for the development frame (19%). As for the differences by network, they discovered that one particular network, SBS6, paid more attention to congruence between the verbal and the visual modalities (48.9%) than the other two networks investigated, i.e., NOS (24.5%) and RTL4 (24.8%) (Van Hoof et al., 2010).

The studies I just reviewed paint a mixed picture. While the findings of Liebler and Bendix (1996) and Van Hoof et al. (2010) suggest that low congruence is the norm, Entman's (1991) and Reynolds and Barnett's (2003) studies make the opposite seem plausible. It appears that the respective topics chosen for analysis may serve as an explanation for these differences. Both Entman (1991) and Reynolds and Barnett (2003) investigated extreme situations with immediate consequences and considerable loss of human life. The "rally 'round the flag effect" may thus serve as an explanation for congruent (verbal and visual) media coverage during wars and international crises (see Goldstein & Pevehouse, 2013). Environmental matters like the one analyzed by Liebler and Bendix (1996) and elections (Van Hoof et al., 2010) also have consequences, but unlike military operations, their impact is not immediately visible. Thus, controversial news framing (manifested

DOI: 10.4324/9781315642239-13

also through verbal and visual mismatches) may be inherent in topics such as these and counterintuitive for wars/international crises. Other explanations for this variation in research results may be found in the methodological choices made by these authors. This is something I will come back to in the next main section (see "Methodological Questions Arising from Existing Work").

## *Experimental Studies*

The review of content and framing analyses presented earlier suggested that there are at least three ways in which the verbal and the visual channel of communication in TV news interact with each other: redundancy, congruence, and conflict. Accordingly, I now review experimental studies that analyzed the effects of each of these ways of interaction.

Experimental research in this area was conducted through one of the following theoretical lenses: (a) the *Belongingness Hypothesis* (Grimes, 1991; Kahneman, 1973; Treisman & Davies, 1973), (b) the *Cue-Summation Theory* (Hsia & Jester, 1968; Reese, 1984; Son, Reese, & Davie, 1987), or (c) the *Semantic Overlap Hypothesis* (Walma van der Molen & Klijn, 2004; Walma van der Molen & Van der Voort, 2000a). As their names already suggest, these frameworks are quite similar. They start from the same premise: that people's information processing capacities are necessarily limited and that this prevents them from processing TV news stories in their entirety under certain conditions.[3]

These conditions refer to the interplay between the verbal and the visual channel. When the two channels convey conflicting information, attentional capacity is overwhelmed. In this situation, the properties of visuals, in particular their capacity to convey meaning more quickly and easily (see Messaris & Abraham, 2001), make them more likely to be acknowledged and remembered than words. As already mentioned in this chapter's introduction, this can be understood with reference to the PSE (Graber, 1990). In the words of Drew and Grimes (1987), "Because television presents information in two channels, there is potential for overloading the information-processing capabilities of viewers. TV news has an even greater potential for this because many of its stories are voice-over, that is, the voice track and the video track are not necessarily isomorphic" (p. 452). By contrast, presenting congruent or even redundant verbal and visual information in TV news stories can assist audiences in processing the information conveyed.

The *Belongingness Hypothesis* posits that "two distinct perceptual stimuli will be attended to as if they were a single stimulus when they appear to belong together" (Grimes, 1991, p. 268). It is expected that less effort is necessary to process both modalities in TV news when they are "perceived as a semantic unit … because attention does not have to be distributed among different stimuli" (Grimes, 1991, p. 270). However, when the two modalities "are discordant enough, they will be regarded as separate units, each demanding attention" (Grimes, 1991, p. 271). Similarly, the *Cue-Summation Theory* (Hsia & Jester, 1968) hypothesizes that when

DOI: 10.4324/9781315642239-13

words and visuals are redundant/congruent, each modality provides cues for the other, thus facilitating the processing and the (accuracy of) recall for the information conveyed; by contrast, irrelevant cues would decrease learning. Finally, the *Semantic Overlap Hypothesis* (Walma van der Molen & Klijn, 2004) theorizes that verbal-visual congruent TV stories are more easily processed and recalled than print stories, but that this is not the case for TV stories where the two modalities are mismatched. As Walma van der Molen and Klijn (2004) stated, "If television stories contain a greater amount of semantically related audiovisual information, television can be more effective [i.e., recalled more] than print" (p. 90), adding:

> [W]hen verbal and visual information do not correspond, viewers' attentional capacity is exceeded and … priority will be given to processing of visual information. This implies that when text and pictures are not semantically related, the main message of a television news story, which is usually presented through the audio channel, will be lost to most viewers. (p. 90)

The experimental studies in this tradition employ a similar research design. Study participants in studies of the *Belongingness Hypothesis* are exposed to three versions of the same TV news stories. While story narration is kept constant across conditions, the verbal and the visual channel are either redundant, congruent, or conflicting[4] (Drew & Grimes, 1987; Grimes, 1990, 1991). The operationalization of redundancy has been consistent across studies (i.e., full overlap), but that of congruence and conflicting information has been progressively adjusted. Initially, congruence—understood as thematic relatedness—was operationalized by 50% of the shots that were redundant and 50% that were conflicting (Drew & Grimes, 1987). This was later criticized as invalid, because it seemed "less a condition than a combination of conditions" (Grimes, 1990, p. 16). In his later studies, Grimes operationalized congruence as "a true medium match between the audio and the video, with the two channels neither unrelated nor tightly bound together semantically" (Grimes, 1990, p. 16; see also, Grimes, 1991). Similarly, the operationalization of conflicting modalities evolved from mismatched audio and moving images (i.e., no semantic relationship) (Drew & Grimes, 1987) to incoherent video footage, "shots of events that were mixed together in a kind of visual potpourri" (Grimes, 1990, p. 18) in order "to maximize channel conflict" (Grimes, 1991, p. 275). In studies through the perspective of *Cue-Summation Theory*, researchers used only two conditions for the interplay of the verbal and the visual channel in TV news: redundant or conflicting (Reese, 1984; Son et al., 1987). This is because these authors were also interested in the combined effects of these conditions with recaps or verbal captions.

In studies guided by the *Semantic Overlap Hypothesis*, researchers contrasted TV news stories that varied in the degree of verbal and visual congruence not with each other, but with printed transcripts of the same news stories (Walma van der Molen & Klijn, 2004; Walma van der Molen & Van der Voort, 2000b). The way

DOI: 10.4324/9781315642239-13

"congruence" was operationalized is worth noting. Like work in the *Belongingness Hypothesis* tradition, these studies distinguished between redundant, congruent, and conflicting modalities. But they were the first to acknowledge in their experiments that news stories do not merely consist of voice-over paired with moving images, but rather, also of journalists and news sources talking on camera. Accordingly, Walma van der Molen and Klijn (2004) decided to put "talking heads" in an extra category. They offered the following rationale: Even though talking heads are not illustrating what is said (as in the redundant or congruent conditions), they cannot be categorized as conflicting with the verbal message either. Walma van der Molen (2001) used this categorization consistently, going back to earlier content analytical studies.

The studies conducted through each of these perspectives obtained consistent results. Studies of the *Belongingness Hypothesis* revealed that moving images supporting the voice-over in both the redundant and the congruent conditions led to improved recall, understanding, memory, and attention for the audio (Drew & Grimes, 1987; Grimes, 1990, 1991). However, when modalities disagreed, visuals were recalled best, and the information conveyed verbally was poorly understood (Drew & Grimes, 1987; Grimes, 1990, 1991). Similarly, studies of the *Cue-Summation Theory* discovered that congruent TV news stories increased viewers' cued recall of story details (Son et al., 1987). Reese (1984) found that redundant audio and moving images enhanced learning. Brosius (1989) analyzed the effects of TV news stories delivered by studio anchors with still images in the background. He found that redundant visuals improved memory for person-focused news, but lessened memory for event-focused stories. Finally, studies of the *Semantic Overlap Hypothesis* showed that congruent and redundant news stories were recalled more than printed news transcripts (Walma van der Molen & Klijn, 2004; Walma van der Molen & Van der Voort, 2000b). Recall of TV news stories in which the verbal and the visual modality conveyed conflicting meanings was even lower than for story transcripts (Walma van der Molen & Klijn, 2004).

As for "talking heads," when they were part of news stories, redundant or congruent, they did not contribute to superior recall (Walma van der Molen & Klijn, 2004). However, when they were used in news stories otherwise divergent in terms of modalities, they were partly responsible for the inferior recall: they sidetracked participants' attention from the verbal message. Support for this is found in previous literature, too. Several studies suggested that information gain from TV audio is lower for talking heads than when matching or otherwise interesting visuals are used (Edwardson, Grooms, & Proudlove, 1981; Gunter, 1979, 1980).

While none of the studies reviewed in this section consulted framing theory, important lessons for framing research can be derived from this literature. Overall, we need a more nuanced discussion of the interplay of the verbal and visual channels of TV news when we conduct framing studies. Even when the words employed do not literally describe or refer to the visuals, the two modalities can still convey the same frame. In classical redundancy literature, scholars would have

DOI: 10.4324/9781315642239-13

categorized audiovisual material such as this as conflicting. But the frame may well be clearly conveyed to audiences. Using framing would make the delineation of stimulus material for experiments more straightforward. In the redundant condition, the same frame would be verbally and visually conveyed; in the congruent condition, related but different frames would be used; in the divergent condition, a verbal frame and a visual frame that are contradicting each other would be used. From the perspective of framing theory, Walma van der Molen et al.'s results build anticipation for effects studies pairing a verbal frame with various talking head information. It is unclear whether the talking head impacts the verbal frame picked up by audiences. Future studies should consider the situation of partial overlap (several news frames per news story), where some of the verbal frames are reinforced visually and others are not.

## Methodological Questions Arising from Existing Work

Scholars investigating verbal and visual framing in TV news differ in their approach to a number of methodological questions.[5] Fully reflective of the plurality of understandings circulating in academia about what frames are and how they can be identified (see D'Angelo, 2002; Entman, 1993), these studies differ in their preference for qualitative, quantitative, or critical approaches. They also vary regarding the strategies used to identify frames, either splitting them into their elements or coding them holistically (see David, Atun, Fille, & Monterola, 2011; Matthes & Kohring, 2008). Issues such as these are evergreens in theory building efforts surrounding framing. Elsewhere, I have attempted to provide a comprehensive review of the existing approaches and their respective advantages and disadvantages (Dan, 2018). Here, I focus on the benefits and shortcomings connected to three aspects that are particularly relevant to framing studies of TV news and which have been dealt with differently in previous studies. These aspects are: (a) transcription, (b) the unit of analysis, and (c) the strategy used to determine verbal–visual frame congruence.

### *Transcription*

Transcribing a video file means converting it into a text file to be used as the primary material during data collection.[6] The resulting text file typically consists of a transcription of the spoken words, a transcription of the text shown on screen, and a description of the images shown (Reynolds & Barnett, 2003; Walma van der Molen, 2001). These components are then matched to one another so that the transcript reflects how they co-occurred in the video—that is, which words were heard and/or seen on screen when certain images were shown.

For a long time, studies of TV news framing relied exclusively on transcripts of the audio track, as offered by large databases such as Vanderbilt Television News Archive or LexisNexis (e.g., Coe, 2011; Kim, Carvalho, Davis, & Mullins, 2011;

DOI: 10.4324/9781315642239-13

Rowling, Sheets, & Jones, 2015). This means that the visuals and sometimes the text on screen were entirely neglected. As for those studies acknowledging both the verbal and the visual component of TV news, the following observations can be made.

Just like other researchers of communication, framing scholars can analyze TV news straight from the tape or can choose to start off their analysis with a transcription of the original material. It seems that scholars conducting quantitative analyses are more inclined to extract their data straight from the video, without transcription (Entman, 1991; Liebler & Bendix, 1996; Van Hoof et al., 2010), while qualitative studies rely on transcription (Reynolds & Barnett, 2003). For reasons I will present later, I argue that researchers should consider using the original audio and visual track instead of transcripts, while pairing the text shown on screen with the audio track. Matching the text on screen with the audio track simplifies the coding process, as it makes the identification of people and places more straightforward.

Using transcripts instead of the original audio/video files seems inadvisable for two reasons. First, that approach tends to artificially reduce meaning and hinder the accurate collection of data—all while increasing the cost and duration of data collection. Arguing that these drawbacks can be condoned in the light of an increased comfort when coding from text is no longer feasible, as digital data formats and user-friendly video players allow researchers to play, pause, and replay files without having to physically rewind tapes. While the material must still be experienced in real time—unlike texts, which can be skimmed, speed-read, and searched using key words—the overall duration of data collection is unlikely to exceed that of creating transcripts and coding them.

The second reason why transcripts should not replace the original files is that the assumption according to which extracting information from written words is bound to be more accurate than coding audio or moving images relies on a widely contested belief in the primacy of written words over orality and visuals (see Frisch, 2013; Grabe & Bucy, 2009). Provided that researchers deal with small portions of material at a time, coder performance, understood as accuracy, speed, and ability to reach high intercoder reliability, should not be an issue.

## *Unit of Analysis*

The question regarding the appropriate unit of analysis is one of the first decisions scholars confront when attempting to analyze words and visuals in TV news (Graber, 1989). Two aspects need clarification in this context: (a) Should the news item be coded at once, or should it be truncated into shots or scenes? and (b) Should data from the verbal and the visual component of each unit be collected at the same time or separately?

Previous investigations into issue framing mostly used the entire news story as the unit of analysis (Entman, 1991; Liebler & Bendix, 1996; Reynolds & Barnett,

DOI: 10.4324/9781315642239-13

2003). This is reminiscent of the gestalt coding approach (Graber, 1986, 1987), which also involved coding whole news items. A few studies used the scene (e.g., Choi & Lee, 2006) or the shot (Grabe & Bucy, 2009) instead, with the latter study investigating character frames, or frames about people. Contrary to this apparent preference for the entire news story, the following review suggests that it is necessary to parse news stories in the sample into shots or scenes. Using the scene as the unit of analysis may be most suitable in studies of issue framing. Investigations into character framing are best advised to maintain their use of the shot as the unit of analysis.

The three possible units of analysis can be placed on a continuum based on their duration, with the shot at one end and the entire news story at the other. The scene would be located somewhere in the middle. A news story lasts several minutes and can consist of several scenes and shots. A shot is a fragment of video material filmed by one camera without interruption (Gianetti, 1982; Mascelli, 1965). In a shot, the camera movement is unedited, which means that there are no cuts (Van Leeuwen & Jewitt, 2001). A shot is between four and 14 seconds long (Banning & Coleman, 2009; Walma van der Molen, 2001). In Walma van der Molen's (2001) study, there were about 24 shots per news story. By contrast, a scene is "a composition of one or more shots with a unifying character, place, theme, idea, topic, or perspective portrayed in the news" (Choi & Lee, 2006, pp. 706–707). In Choi and Lee's (2006) study, a scene averaged 18 seconds, and there were about eight scenes per news story.

There are certain disadvantages to using the entire news story as the unit of analysis. Using the entire news story connects with concerns regarding coder accuracy and ability to extract in-depth information, as both may be jeopardized when analysts are exposed to full-length stories. Also, those using the entire news story may have to limit themselves to identifying the most dominant frame in that story (Entman, 1991; Reynolds & Barnett, 2003), which is unlikely to be an accurate representation of news on contested topics (Choi & Lee, 2006).

Thus, using scenes or shots as the unit of analysis seems more advisable. In order to decide which one should be used instead of the entire news story, scholars must consider study manageability and how much detail is needed to answer one's research questions and hypotheses. One way to think of study manageability is to estimate how many scenes or shots one would have to cope with if either were used in a study of 150 TV news stories. Assuming the number of shots per news story in Walma van der Molen's (2001) study as typical, the analysis of 150 TV news stories would require coping with 3,600 shots. Similarly, assuming the number of scenes per news story in Choi and Lee's (2006) study as typical, analysts would have to cope with 1,200 units of analysis in a study of 150 news stories. While the number of units of analysis is still high, it represents just one-third of those necessary in a shot-based study. Thus, using the scene as the unit of analysis may be easier to manage. Deciding on the amount of detail to observe using either the shot or the scene as the unit of analysis is less straightforward. Using the shot

DOI: 10.4324/9781315642239-13

as the unit of analysis proved to be a good strategy in studies of character framing, where visual material could be coded in the minutest detail, such as minor changes in nonverbal behavior, for example (Grabe & Bucy, 2009). By contrast, when it comes to issue framing, using the scene as the unit of analysis is likely to be sufficient. As stated by Choi and Lee (2006), the scene represents "the smallest unit that contains a meaningful narration, argument, or perspective of a news story" (p. 704).

Once scholars have decided on using shots or scenes as the unit of analysis, they must proceed with truncating the videos in the sample accordingly. For shot-based studies, this is very straightforward, as a cut denotes the beginning of a new shot. The division of news stories into scenes may require more training, but very helpful guidelines exist. According to Choi and Lee (2006), a new scene begins whenever considerable changes in format or content occur. In keeping with the terminology used in Figure 9.1, changes in format mean transitions from one of the following formats to any other one listed here: (a) studio anchor scene; (b) video footage (VO; VO/SOT; pre-recorded audio); (c) computer graphics; (d) still images; (e) silent pictures; (f) journalistic visibility; and (g) reactions, including establishing shots. Changes in content refer to changes in (h) the who, what, when, where, why, and how (based on Choi & Lee, 2006). Choi and Lee (2006) clarify that "[a] format change always marks a scene change, regardless of content changes, but when there is an obvious shift in content, it is also considered a scene change" (p. 707).

The second aspect that needs clarification and pertains to the unit of analysis regards the mix of modalities during coding. Judging by the methodological information offered in previous studies, analysts collected data from the verbal and visual channels of information at the same time (Entman, 1991; Liebler & Bendix, 1996; Van Hoof et al., 2010). Currently, this is the state of the art, though scholars have suggested that conducting two separate analyses would be more appropriate in quantitative studies (Coleman, 2010; Reynolds & Barnett, 2003). This is because, although exposing coders to multimodal units of analysis (e.g., an audiovisual scene)—as opposed to the audio component and the visual component independently—may seem like a good idea (because it replicates the experience of the members of the audience), it is not.

At first glance, it may seem reasonable to assume that coders should be able to identify possible mismatches between the verbal and the visual channel right off the bat (Graber, 1986, 1987; Walma van der Molen, 2001). However, like everybody else, coders cannot help but see the visuals through the lens suggested by the accompanying words, and vice versa (see the "Experimental Studies" section). Thus, in the case of a mismatch, the data collected would reflect which channel "won" from the perspective of that individual coder, as opposed to recording the meaning embedded in the words and visuals, respectively. In other words, this strategy almost incorporates an experiment in content analysis, where the coders act as study participants. Researchers then run the risk that the data will say more

DOI: 10.4324/9781315642239-13

about the coders than about the material. To borrow other researchers' terminology, one risks identifying coder frames instead of news frames[7] (see Matthes & Kohring, 2008; Wirth, 2001). For these reasons, scholars are advised to expose coders to each modality separately.

### *Verbal and Visual Frame Congruence*

Researchers who analyze both the verbal and the visual framing in TV news must, at some point, say something about the extent to which the verbal frames matched the visual frames—i.e., about verbal–visual frame congruence. This is not an easy task, as no publication to date deals with this question explicitly. Several studies attend to this task using overall estimations or side notes (Entman, 1991; Liebler & Bendix, 1996; Reynolds & Barnett, 2003). It seems that more explicit and more objective ways to determine congruence are needed. In search of such ways, the following two studies seem particularly instructive: a conference paper on framing in TV news by Van Hoof et al. (2010) and a journal article by Walma van der Molen (2001) on text-picture correspondence in TV news.

Van Hoof et al. (2010) were interested in what they referred to as "redundancy of television news frames" (p. 1). This term would lead one to expect that the authors compared the verbal frames with the visual frames in each unit of analysis and determined when the same frame was conveyed across modalities (i.e., was redundant). But, in fact, Van Hoof et al.'s (2010) analysis went beyond redundancy (yes/no), in that it investigated various degrees of congruence between verbal frames and visual *content* using a formula: They calculated the percentage of sentences in a news story where one element of the verbal frame—e.g., an actor involved in the issue—was shown (Van Hoof et al., 2010).

Thus, in search of ways to estimate frame congruence across modalities, the take-away from Van Hoof et al.'s (2010) study is that formulas can be used instead of overall estimations. They are likely to produce more accurate assessments of the way TV news are made up, as the interplay between words and visuals is unlikely to be dichotomist (fully redundant vs. totally divergent). Moving forward, framing scholars should attempt to identify visual *frames* instead of visual *content*. This is important, as visuals are a modality in their own right, which can convey frames in and for themselves, and these frames can be more or less congruent with the verbal frames. I do realize that this means "essentially [conducting] two studies in one," and that this may seem like a "daunting task" (Coleman, 2010, p. 235). Also, things do not necessarily get easier once the analysis is completed, as the two datasets, one for each modality, must then be "puzzle[d] together in a meaningful and coherent way," which may be "difficult" (Reynolds & Barnett, 2003, p. 91). Solutions to these problems are suggested in the next sections.

Walma van der Molen's (2001) study reveals other valuable lessons for the assessment of verbal–visual frame congruence, even though this author did not use framing theory. She developed a coding scheme for assessing various degrees

DOI: 10.4324/9781315642239-13

of verbal-visual congruence in TV news. For each shot, she asked coders to estimate the level of text-picture correspondence using one of four coding categories: direct, indirect, divergent, or talking head. Thus, unlike Van Hoof et al. (2010), Walma van der Molen (2001) used coders' overall estimations. I already explained that, to me, this technique appears less desirable than using a formula—at least in quantitative studies.

For those interested in finding ways to assess frame congruence between modalities in TV news, the added value of this study lies in the acknowledgment of the "talking head" as an extra category, for which the interplay between words and visuals is not easily determined (Walma van der Molen, 2001). I very much agree with this, and would even go further, arguing that talking heads are not the only types of visuals to which special attention must be paid when assessing verbal-visual congruence. It seems to me that this is the case for all TV visuals that are "little more than visualized radio" (see Gunter, 2015, p. 96), provided that the cropping is tight and/or no meaningful visuals are shown in the background, specifically lead-ins, anchor narrations, reactions, and journalistic visibility (see Figure 9.1). Also, this is the case with silent pictures.

When confronted with this type of video footage, other scholars chose the same path. For example, Liebler and Bendix (1996) did not code visual framing for journalistic visibility segments, as they deemed them not to be engaging. Similarly, Van Hoof et al. (2010) skipped the sections with the anchor on screen and assumed that redundancy is by definition zero for this part of the news item. Yet, as already described in the "Experimental Studies" section, newer effects studies suggest that we must find a way to factor in this type of video footage when determining verbal–visual frame congruence. This is because such talking heads distract attention from the verbal message in stories that are incongruent anyway, but don't have a negative effect on memory of congruent stories.

## Analytical Steps in Integrative Framing Analyses of TV News

In the preceding sections, I presented the different ways in which scholars interested in both the verbal and the visual component of TV news conducted framing analyses. I offered a discussion of the advantages and disadvantages of various methodological decisions surrounding transcription, the unit of analysis, and the way to assess the interplay between words and visuals. Scholars interested in conducting such analyses can use this review to inform their decision about how to set up an integrative framing study of TV news. I used this review of literature to make up my own mind in this regard. This section presents the result of this decision process in the form of a sequence of analytical steps to be taken when conducting integrative framing analyses of TV news.

Figure 9.2 shows the seven-step process of integrative framing analysis.[8] Some of these steps are comparable to those known from other types of research (e.g.,

DOI: 10.4324/9781315642239-13

codebook development and data collection). What stands out is the proposal to carry out four of the seven steps separately for each modality, as indicated by labels such as "Step 3a" and "Step 3b," where "a" stands for carrying out the step in the verbal sample and "b" for the visual sample. The block arrows at the bottom of the figure symbolize the three stages of the analysis: preparation, data collection, and data analysis. In the remainder of this section, I account for each of the steps in detail, referring back to this visualization.

As shown in Figure 9.2, Step 1 involves the *segmentation of each news story in the sample into shots or scenes.*[9] The segmentation can begin after an ID number has been assigned to each news item in sample and the duration of each news item has been recorded. For the data analysis, it is important to record here the verbal and visual format of each scene/shot, as shown in Figure 9.1 (e.g., reactions, silent images, journalistic visibility). To increase the accuracy of segmentation in shots, researchers can use Coleman and Banning's (2006) rule of thumb and set the minimum duration of each shot at 4 seconds. For dividing the scenes, I suggest using the guidelines drafted by Choi and Lee (2006) briefly reviewed earlier. After training, research assistants can assume this task. Here, intercoder reliability has to be measured both for the number of shots/scenes per news story and for the division points within a story. The duration of each shot/scene should be recorded in seconds.

Next, we can proceed to Step 2, namely, the *separation of audio and visual tracks*. Here, we must separate each of the shot/scenes into the verbal and the visual component, respectively. Freeware able to do this is available by searching for "batch demuxing." To fully understand the verbal component of the news story, researchers must also account for the text shown on screen, which will go into the verbal analysis. This can be accomplished by simply typing the words into the document used for data collection (see Figure 9.3, column "Text Shown on Screen"). A more sophisticated way would involve the use of audio-annotation software, which has the advantage that this information appears just at the right time in the audio file (i.e., we know who the person talking is when they begin talking).

At the end of this step, we have four IDs in both datasets ("1," "2," "3," "4"); they are interconnected, as denoted by the dashed arrows in Figure 9.3 (which assumes the use of the scene as the unit of analysis). The descriptors reveal that, in fact, we now have eight units of analysis: four for the verbal scenes and four for the visual scenes (e.g., there are two descriptors for unit of analysis 1: "1_verbal_scene_1" in the verbal dataset and "1_visual_scene_1" in the visual dataset).

In preparation for the coding, I recommend preparing one file for each modality, where the IDs and their descriptors are already typed in and where these columns are locked in place. This should prevent coders from deleting a case by mistake[10] and ensure that the researchers will be able to automatically merge the datasets in a meaningful and coherent way for the data analysis. Also, if the software used for coding allows it, I recommend using hyperlinks to connect the descriptor of each unit of analysis to the original file (as denoted by the

DOI: 10.4324/9781315642239-13

DOI: 10.4324/9781315642239-13

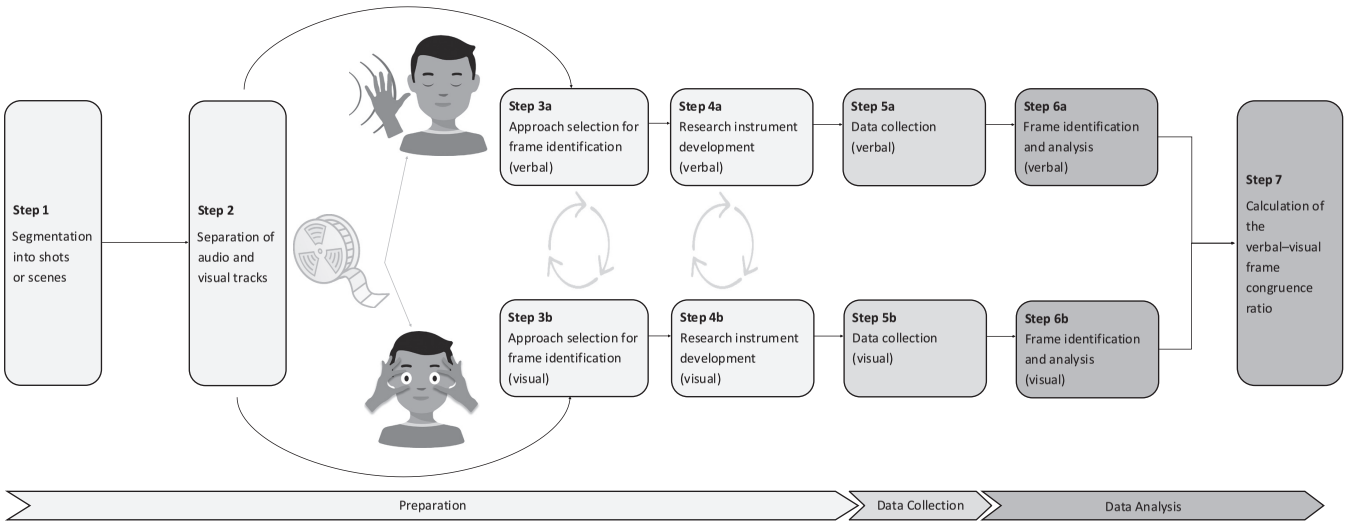


**FIGURE 9.2** Step-by-step process for integrative framing analysis of TV news

*Note.* Adapted from Dan (2018), reproduced with permission. Artwork by Robér Rollin.

Dataset for the Analysis of the Verbal Component of TV news

| ID | News Story No. | Number of Scenes per News Story | Descriptor Unit of Analysis (ID) | Duration in Seconds | Text Shown on Screen |
|---|---|---|---|---|---|
| 1 | 1 | 4 | 1_verbal_scene_1 | 18 | Text text |
| 2 | | | 1_verbal_scene_2 | 32 | N/A |
| 3 | | | 1_verbal_scene_3 | 41 | Text text |
| 4 | | | 1_verbal_scene_4 | 37 | Text text |
| 5 | 2 | 5 | 2_verbal_scene_1 | 19 | N/A |
| 6 | | | 2_verbal_scene_2 | 27 | Text text |
| 7 | | | 2_verbal_scene_3 | 10 | N/A |
| 8 | | | 2_verbal_scene_4 | 43 | N/A |
| 9 | | | 2_verbal_scene_5 | 21 | N/A |

Dataset for the Analysis of the Visual Component of TV news

| ID | News Story No. | Number of Scenes per News Story | Descriptor Unit of Analysis (ID) | Duration in Seconds |
|---|---|---|---|---|
| 1 | 1 | 4 | 1_visual_scene_1 | 18 |
| 2 | | | 1_visual_scene_2 | 32 |
| 3 | | | 1_visual_scene_3 | 41 |
| 4 | | | 1_visual_scene_4 | 37 |
| 5 | 2 | 5 | 2_visual_scene_1 | 19 |
| 6 | | | 2_visual_scene_2 | 27 |
| 7 | | | 2_visual_scene_3 | 10 |
| 8 | | | 2_visual_scene_4 | 43 |
| 9 | | | 2_visual_scene_5 | 21 |

**FIGURE 9.3** Structure of the verbal and the visual datasets for integrative framing analyses

DOI: 10.4324/9781315642239-13

hand mouse pointer next to "1_verbal_scene_1"). Creating these hyperlinks may require some time, but this will likely speed up the coding process and prevent mistakes, as coders will be able to view/hear the original material in a pop-up player window. This step is complete when the verbal component of the sample is archived separately from the visual component, and when it is clear which visuals correspond to which words. Looking at the two datasets should produce an image as shown in Figure 9.3.

Step 3 consists of the *approach selection for frame identification*. Specifically, Step 3a requires that researchers select an approach for the identification of verbal frames, whereas Step 3b refers to the selection of an approach for the identification of visual frames. For each modality, framing scholars have a relatively wide range of approaches from which to choose. Elsewhere, I have described all procedures I encountered in an in-depth review of literature, as well as the strengths and weaknesses of each of them (Dan, 2018). Suffice it to say that there are six approaches to verbal and five to visual framing analysis. They differ in a number of ways, including (a) the preference for deductive or inductive methods, (b) the use of human coders or automated content analysis, and (c) their focus on content, frame function (problem definition, cause, treatment, and evaluation), selection, or salience. As a general observation, it seems that the acts of selection and emphasis get greater emphasis in visual framing studies, whereas verbal framing studies focus more on the content and function of frames (see for a review Dan, 2018). Because none of these approaches was developed specifically for the analysis of audiovisual TV news, they may require some tweaking before they can be implemented. For instance, researchers might want to acknowledge techniques specific to moving images (such as camera movement) on top of the more typical categories for still images, such as camera angle and camera distance.

In Dan (2018), I also explained that researchers are free to choose the approaches they deem meaningful and high performance, that correspond to their understanding of frames, and that align best with their needs, methodological preferences, and resources. Yet, because of the relatively large number of approaches available, scholars might want to justify their choice. The approaches chosen must allow the collection of comparable data from the two modalities, as suggested by the circular arrows between Steps 3a and 3b in Figure 9.2. For example, scholars who include verbal metaphors in their analysis of verbal frames should also acknowledge visual metaphors, symbols, and so on when coding the visuals. Researchers can move to the next step as soon as they have chosen at least one approach for the identification of verbal frames and at least one for the identification of visual frames. They might want to combine several approaches for each modality to reduce the drawbacks of each approach.

Next, the researcher can proceed with the *development of the research instruments* in Step 4. Two codebooks are needed: one for words and one for visuals. As denoted by the circular arrows between Steps 4a and 4b, the development of

DOI: 10.4324/9781315642239-13

these two codebooks requires some iteration loops and comparisons between the modalities (see Figure 9.2). This is needed to determine whether a verbal and a visual expression of each aspect of interest exist and, if so, how their components can be captured in each modality. The goal is to define which features of the verbal and visual components of TV news in the sample are of interest and how they can be translated into variables. This step is a sensitive one, as those new to visual research may try to jam visual meaning into categories derived verbally; it is essential that this does not happen. For this reason, this step is possibly the most demanding one in integrative framing analyses.

As already mentioned in the introduction to this chapter, frames involve acts of selection and emphasis. The results of selection processes can be captured by looking at the very nature of the frames: What aspects were selected and which were left out? Giving specific advice about which categories should be included in the codebooks is a Herculean task. Because there are so many factors involved in this decision, the task is difficult. Still, I can give some generic advice:

(1) Develop categories based on a thorough survey of literature—re: what did previous framing and content analyses find out regarding this topic or related topics?
(2) Complement this with an inductive approach.[11]
(3) Think of ways in which variations in content may organize reality in different ways—e.g., who are the good/bad guys, according to this interpretation?
(4) Ascertain how content performs each of the four functions of framing delineated by Entman (i.e., problem definition, causal interpretation, moral evaluation, and treatment recommendation).
(5) Include categories that allow the measurement of dominance and prominence—e.g., at which position in the news package was the frame presented? Was it the last comment/last image? How much space/time was awarded to each frame in the news package?
(6) Record actors/sources sponsoring each frame, providing sound bites and image bites, in an attempt to contribute to literature on frame building and framing contests.

In all, Step 4 is complete when two separate but connected research instruments exist and after the pre-tests and reliability tests have revealed acceptable intercoder reliability.

Step 5 consists of *data collection*. Similarly to Steps 3 and 4, this one is performed in the verbal and visual samples, respectively. However, contrary to the procedure for Steps 3 and 4, this step is performed in the verbal sample (5a) independently of its execution in the visual sample (5b), as denoted by the absence of circular arrows in Figure 9.2. To prevent the information processing biases reviewed at the beginning of this chapter from kicking in, researchers are advised to make sure that the coding of the verbal component of the TV news story at hand is

DOI: 10.4324/9781315642239-13

not influenced by the visual component, and vice versa. To this end, scholars can do the following:

(1) Complete the data collection of one modality before beginning the other.
(2) Hire different coders for each modality; it may even be advisable to hire coders other than the research assistants who truncated the material/helped prepare the dataset for coding, as they might already be too familiar with the material.
(3) Assign the verbal counterpart of the visuals in one coder's subsample to the subsample given to the other coder, and vice versa.

This step is complete when two separate datasets emerge: one is relevant to visual framing and the other to verbal framing.

In the penultimate Step 6, scholars engage in *frame identification and analysis*. Again, this is something done independently in each modality's sample (notice again the missing circular arrows between Steps 6a and 6b in Figure 9.2). In order to identify frames,[12] scholars can use cluster analysis, factor analysis, latent class analysis, and index building upon the calculation of Cronbach's alpha coefficients, among others. Upon frame identification, scholars must convincingly explain that the constructs identified are frames in the way they were defined in the literature review. One aspect that is often neglected concerns an explanation of how the constructs identified organize social reality.

In my experience, the variables recording news frames often use nominal measures. In this case, researchers can use tests of significance such as Chi-square (or, if necessary, Fisher's exact test or Monte Carlo test). Information about the effect size is required, for example, Cramér's V. For crosstabs larger than 2 × 2, column proportions tests are necessary. Other levels of measurement will make different statistical procedures necessary, such as *F*-tests in one-, two-, or three-way analyses of variance (including post hoc tests chosen based on the result of the homogeneity test, e.g., Levene and Tukey's HSD or Welch's & Dunnett's C). Effect sizes can be reported using Eta squared ($\eta^2$).

Finally, when scholars reach Step 7, the task consists of *calculating the verbal–visual frame congruence ratio*, that is, the extent to which the verbal frames conveyed in a news story matched the visual frames in that story. I have already justified my preference for using a formula to determine congruence instead of overall estimations (see "Verbal and Visual Frame Congruence"). Here, I present a formula for the calculation of a verbal–visual frame congruence ratio ($CR_{Frames}$) in TV news; it draws on the one I presented in an earlier publication for frames in print materials, that is, still images and written words (Dan, 2018). I propose that the congruence ratio can be calculated in TV news by dividing the total duration of scenes/shots in which the same frame was conveyed both verbally and visually ($\sum t_{verbal\ =\ visual}$) by the total duration of the news story ($\sum t_{News\ Story}$):

DOI: 10.4324/9781315642239-13

$$CR_{Frames} = \frac{\sum t_{verbal\ =\ visual}}{\sum t_{News\ Story}}$$

Already in Step 1, we have recorded the duration of the news story and that of every scene within it. This means that a variable recoding $\sum t_{News\ Story}$ is available. To determine the total duration of the scenes in which the same frame was conveyed both verbally and visually ($\sum t_{verbal\ =\ visual}$), some preparation is required. To this end, researchers are advised to start by merging the verbal and visual datasets using the corresponding IDs assigned to each unit of analysis prior to coding. This new dataset includes the following variables that interest us to compute a value for $\sum t_{verbal\ =\ visual}$. They record the duration of each scene ($t_{Scene}$) and the verbal ($Frame_{verbal}$) and the visual frame ($Frame_{visual}$) conveyed in that scene, respectively. Thus, to obtain $\sum t_{verbal\ =\ visual}$, researchers have to add the duration of all scenes ($t_{Scene}$) in which the values recorded in the variables $Frame_{verbal}$ and $Frame_{visual}$ were identical.

Figure 9.4 contains some fictitious data for four news stories that consist of various numbers of scenes (in this example, the scene is used as the unit of analysis). Let's assume that we have identified five frames in our sample and that each of them had both a verbal expression and a visual one. Accordingly, the variables $Frame_{verbal}$ and $Frame_{visual}$ can adopt values ranging from 1 to 5, where "1" stands for the first frame, "2" for the second one, and so on. The value "0" means that no frame was conveyed in the respective scene through that modality.[13] We can now create a new variable, which records whether the same frame was identified in both the verbal and the visual samples ($Same_{frame}$). $Same_{frame}$ can assume either the value "1" (yes) or "0" (no). For each news story, the durations of the scenes in which $Same_{frame}$ takes on the value "1" have to be added to determine $\sum t_{verbal\ =\ visual}$ (recorded in seconds).

Applying the formula results in a new interval-scaled variable with values ranging from 0 to 1. The closer the value is to 1.00, the higher is the congruence between the frames in the two modalities in that news story. Several calculation examples are offered in Figure 9.4. The congruence ratio ranges from 1.00 for story 2 to 0.00 for story 3. To understand how the formula works, let's look at the way the congruence ratio was calculated for story 1. There, two scenes conveyed the same frame verbally and visually; they lasted 40 seconds and 50 seconds, respectively. Overall, the news story lasted 142 seconds. The congruence ratio for this story was thus calculated as follows:

$$CR_{Frames} = \frac{\sum t_{verbal\ =\ visual}}{\sum t_{News\ Story}} = \frac{90}{142} = 0.63$$

It is worthwhile to take a look at the way the congruence ratio was calculated for story 3. There, frame "1" was conveyed in two verbal scenes and also in two visual scenes. However, applying the congruence ratio formula yields the value

DOI: 10.4324/9781315642239-13

DOI: 10.4324/9781315642239-13

| ID | News Story No. | News Story Duration ($\sum t_{News\ Story}$) | No. of Scenes | Scene Duration ($t_{Scene}$) | Descriptor Verbal Unit of Analysis | Descriptor Visual Unit of Analysis | Format | Verbal Frame ($Frame_{verbal}$) | Visual Frame ($Frame_{visual}$) | Same frame ($Same_{frame}$) | Duration of Scene/s where $Same_{frame} = 1$ ($\sum t_{verbal = visual}$) | CR Frames |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 142 | 4 | 15 | 1_verbal_scene_1 | 1_visual_scene_1 | 1 | 1 | 0 | 0 | 90 | 0.63 |
| 2 | | | | 20 | 1_verbal_scene_2 | 1_visual_scene_2 | 2 | 1 | 0 | 0 | | |
| 3 | | | | 40 | 1_verbal_scene_3 | 1_visual_scene_3 | 3 | 3 | 3 | 1 | | |
| 4 | | | | 50 | 1_verbal_scene_4 | 1_visual_scene_4 | 3 | 2 | 2 | 1 | | |
| 5 | | | | 17 | 1_verbal_scene_5 | 1_visual_scene_5 | 2 | 2 | 0 | 0 | | |
| 6 | 2 | 90 | 3 | 30 | 2_verbal_scene_1 | 2_visual_scene_1 | 3 | 1 | 1 | 1 | 90 | 1.00 |
| 7 | | | | 25 | 2_verbal_scene_2 | 2_visual_scene_2 | 3 | 1 | 1 | 1 | | |
| 8 | | | | 35 | 2_verbal_scene_3 | 2_visual_scene_3 | 3 | 1 | 1 | 1 | | |
| 9 | 3 | 197 | 5 | 40 | 3_verbal_scene_1 | 3_visual_scene_1 | 3 | 4 | 1 | 0 | 0 | 0.00 |
| 10 | | | | 26 | 3_verbal_scene_2 | 3_visual_scene_2 | 3 | 2 | 1 | 0 | | |
| 11 | | | | 54 | 3_verbal_scene_3 | 3_visual_scene_3 | 3 | 5 | 3 | 0 | | |
| 12 | | | | 63 | 3_verbal_scene_4 | 3_visual_scene_4 | 3 | 1 | 3 | 0 | | |
| 13 | | | | 14 | 3_verbal_scene_5 | 3_visual_scene_5 | 3 | 1 | 3 | 0 | | |
| 14 | 4 | 120 | 4 | 17 | 4_verbal_scene_1 | 4_visual_scene_1 | 1 | 2 | 0 | 0 | 49 | 0.40 |
| 15 | | | | 22 | 4_verbal_scene_2 | 4_visual_scene_2 | 2 | 2 | 0 | 0 | | |
| 16 | | | | 32 | 4_verbal_scene_3 | 4_visual_scene_3 | 4 | 0 | 3 | 0 | | |
| 17 | | | | 49 | 4_verbal_scene_4 | 4_visual_scene_4 | 3 | 5 | 5 | 1 | | |
| 18 | 5 | 139 | 5 | 28 | 5_verbal_scene_1 | 5_visual_scene_1 | 1 | 3 | 0 | 1 | 78 | 0.56 |
| 19 | | | | 19 | 5_verbal_scene_2 | 5_visual_scene_2 | 2 | 3 | 0 | 1 | | |
| 20 | | | | 33 | 5_verbal_scene_3 | 5_visual_scene_3 | 3 | 4 | 4 | 1 | | |
| 21 | | | | 45 | 5_verbal_scene_4 | 5_visual_scene_4 | 3 | 5 | 5 | 1 | | |
| 22 | | | | 14 | 5_verbal_scene_5 | 5_visual_scene_5 | 2 | 1 | 0 | 0 | | |

**FIGURE 9.4** Calculation examples of the verbal–visual frame congruence ratio

0.00 for the congruence ratio, which might strike readers as an inaccurate expression of this news item. I argue that 0.00 is the correct value, because the respective frame is not presented at the same time verbally and visually. I rely here on Mayer and Moreno's (1998) dual-processing experiment, which suggested that congruent audio and visuals only improve memory when the congruent information is presented at *the same time* in the two modalities (cited in Walma van der Molen, 2001). This is not the case in story 3, where the same frame was conveyed in the two modalities at *different times*, in different scenes.

## Conclusion

By building on Renita Coleman's (2010) contribution to the first edition of this volume, this chapter continues an ongoing discussion of news framing analyses. Here, I intended to show the merit of integrative framing analyses of TV news. This was done from the perspective that studies investigating frames just in the verbal or just in the visual component of audiovisual material are unable to reveal the interpretations presented to audiences. After an extensive review of the literature, I proposed a seven-step methodological approach to integrative framing analyses of TV news. The goal was to find a way in which such analyses can be conducted in a meaningful and manageable way, both practically and economically.

Grappling with methodological questions is not an end in itself. Rather, answering methodological questions is a prerequisite for conducting meaningful empirical studies. I hope that the step-by-step delineation of a methodological approach to integrative framing analysis proposed here will increase the number of integrative studies in our discipline and, through this, our understanding of how journalists and their editors frame issues and people.

The proposal of this approach was informed by a large and sprawling body of literature. Yet, while I am confident that no major aspects were missed, I acknowledge that this process was much like a dry-run surfing lesson, in which standing up straight is a lot easier when the water is calm. Empirical studies attempting to use my approach (and especially the formula for the calculation of the congruence ratio) might experience problems that I did not anticipate. Moving forward, studies implementing this approach might consider publishing video tutorials, sharing their SPSS syntax, or developing an SPSS macro.

## Notes

1 Recently, more studies acknowledging still images and written words in print and online news have been published (Dan, 2018; Wessler, Wozniak, Hofer, & Lück, 2016; Wozniak, Lück, & Wessler, 2015).

2 Entman (1991) included a wider range of news outlets in his analysis. Given the focus of this chapter, my review is limited to his findings with regard to TV news.

DOI: 10.4324/9781315642239-13

3 One is reminded of the Limited Capacity Model (Lang, 2000), which posits that news in a video format—that is, consisting of many modalities, such as words, visuals, and sound—can overtax the processing system. Lang argued that the complexity of visuals, especially their capacity to be emotionally arousing, impaired the processing of words and other modalities. For example, information processing and recall were dominated by emotional visuals in a study by Brosius (1993). By focusing on message complexity as opposed to variations in content, the Limited Capacity Model does not speak directly to the interplay between words and visuals, and is not as relevant for this chapter.
4 I use these terms for the purpose of consistency with the literature reviewed earlier. Note that scholars use a variety of other labels, too, including high/medium/no correspondence and high/medium/no redundancy.
5 Editor's Note: The chapter by Schwalbe, Keith, and Silcock (this volume) also addresses methodological issues involved in doing visual news framing research.
6 For a more detailed discussion of transcription, see Ferguson (2000).
7 The following account clarifies this point—and illustrates how I learned this lesson the hard way. In a previous study (Dan, 2018), I had initially attempted to collect data from multimodal materials at the same time. I was interested in verbal and visual frames conveyed for people living with HIV/AIDS (PLWHA) in news, special interest publications, and public service announcements. Most of the materials consisted of still images and written text. During the pre-test, I noticed that intercoder reliability was difficult to reach because the text was coded regularly through the perspective suggested by the photos. For example, verbal frame elements suggestive of the survivor frame (i.e., strong personality, excellent health, activism) were overlooked because the image used in that article aligned with the carrier frame instead (i.e., PLWHA as deviant and dangerous). To be more specific, one such article told the story of an eloquent and vocal HIV/AIDS activist who—for a very brief period of time in the past—had worked in the sex industry, where she had contracted HIV. An older photo was paired with this verbal account; it showed her in a crude pose wearing transparent clothes and flamboyant makeup.
8 This step-by-step process represents an enhancement of the methodological approach to integrative framing analysis of written text and still images (e.g., in newspaper articles) that I presented elsewhere (Dan, 2018). It has been revised and adapted for the analysis of audiovisuals.
9 As already explained, the scene should be used as a unit of analysis in studies of issue framing, whereas the shot may be more suitable for analyses of character framing.
10 Should the addition of a new row be necessary for whatever reason, researchers must ensure that the row is added in both files. Researchers might want to use a file storage and synchronization service, which should allow them to collaboratively edit the document in which coders enter their codes.
11 Scholars can use a stratified subsample of the material to add coding examples to each variable. This process might lead researchers to eliminate variables for which no coding examples could be found from the codebooks and to add new variables that were not described in previous studies.
12 Frame identification is not applicable in studies using holistic measures, i.e. those coding the entire frame at once (see David et al., 2011).
13 This can be a direct function of the format of that scene ("Format"). This variable can take on values such as 1 = anchor narration; 2 = journalistic visibility; 3 = audiovisual news film; 4 = silent pictures; 5 = reactions (see "The Structure of a TV Newscast" section for details). These values were recorded in Step 1; "4" is not informative of the

DOI: 10.4324/9781315642239-13

verbal framing of an issue; "1," "2," and "5" are not informative of the visual framing of that issue. We cannot determine whether the same frame was conveyed verbally and visually in the same scene if that scene consists solely of visuals (format = 4) or if it is not informative for the visual framing of the issue at hand—for example, because it only shows people talking on camera with tight cropping (format = 1, 2, or 5). However, scholars interested in character framing (as opposed to issue framing) are advised to look for frames in these segments, too. For instance, the non-verbal behavior of the person talking on camera can be informative, as can the camera angles and camera distances used for that person (e.g., in terms of symbolic power of the protagonist or credibility). Also, when background information is visible, this may well contribute to issue framing. For instance, this would be the case when a journalist reporting on labor conditions in the textile industry stands before a run-down sweatshop with underage workers as opposed to standing in front of a modern building with happy-looking workers.

## References


Aust, C. F., & Zillmann, D. (1996). Effects of victim exemplification in television news on viewer perception of social issues. *Journalism & Mass Communication Quarterly*, *73*, 787–803.

Bahador, B. (2008). Framing the 2006 Israel-Hezbollah war. Paper presented at the *Annual Meeting of the International Studies Association* (ICA), Chicago, IL.

Banning, S., & Coleman, R. (2009). Louder than words: A content analysis of presidential candidates' televised nonverbal communication. *Visual Communication Quarterly, 16*(1), 4–17.

Barkin, S. M. (1989). Coping with the duality of television news: Comments on Graber. *American Behavioral Scientist, 33*(2), 153–156.

Brosius, H.-B. (1989). Influence of presentation features and news content on learning from television news. *Journal of Broadcasting & Electronic Media, 33*(1), 1–14. doi:10.1080/08838158909364058

Brosius, H.-B. (1993). The effects of emotional pictures in television news. *Communication Research, 20*(1), 105–124.

Brosius, H.-B., Donsbach, W., & Birk, M. (1996). How do text-picture relations affect the informational effectiveness of television newscasts? *Journal of Broadcasting & Electronic Media, 40*(2), 180–195. doi:10.1080/08838159609364343

Choi, Y. J., & Lee, J. H. (2006). The role of a scene in framing a story: An analysis of a scene's position, length, and proportion. *Journal of Broadcasting & Electronic Media, 50*(4), 703–722. doi:10.1207/s15506878jobem5004_8

Clausen, L. (2003). *Global news production*. Copenhagen, Denmark: Copenhagen Business School Press.

Coe, K. (2011). George W. Bush, television news, and rationales for the Iraq war. *Journal of Broadcasting & Electronic Media, 55*(3), 307–324. doi:10.1080/08838151.2011.597467

Coleman, R. (2010). Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. In P. D'Angelo & J. A. Kuypers (Eds.), *Doing news framing analysis: Empirical and theoretical perspectives* (pp. 233–261). New York: Routledge.

Coleman, R., & Banning, S. (2006). Network TV news' affective framing of the presidential candidates: Evidence for a second-level agenda-setting effect through visual framing. *Journalism & Mass Communication Quarterly, 83*(2), 313–328.

Cushion, S., Rodger, H., & Lewis, R. (2014). Comparing levels of mediatization in television journalism: An analysis of political reporting on US and UK evening news bulletins. *International Communication Gazette, 76*(6), 443–463. doi:10.1177/1748048514533860


DOI: 10.4324/9781315642239-13

Dan, V. (2018). *Integrative framing analysis: Framing health through words and visuals*. New York: Routledge.

D'Angelo, P. (2002). News framing as a multiparadigmatic research program: A response to Entman. *Journal of Communication, 52*(4), 870–888.

David, C. C., Atun, J. M., Fille, E., & Monterola, C. (2011). Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures, 5*(4), 329–351. doi:10.1080/19312458.2011.624873

Drew, D. G., & Grimes, T. (1987). Audio-visual redundancy and TV news recall. *Communication Research, 14*(4), 452–461.

Edwardson, M., Grooms, D., & Proudlove, S. (1981). Television news information gain from interesting video vs. talking heads. *Journal of Broadcasting, 25*(1), 15–24. doi:10.1080/08838158109386425

Entman, R. M. (1991). Framing U.S. coverage of international news: Contrasts in narratives of the KAL and Iran Air incidents. *Journal of Communication, 41*(4), 6–27.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.

Fahmy, S., Bock, M. A., & Wanta, W. (2014). *Visual communication theory and research. A mass communication perspective*. New York: Palgrave Macmillan.

Ferguson, S. D. (2000). *Researching the public opinion environment: Theories and methods*. Thousand Oaks, CA: Sage.

Frisch, M. (2013). Three dimensions and more: Oral history beyond the paradoxes of method. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 221–240). New York: Guilford.

Gamson, W. A. (1989). News as framing: Comments on Graber. *American Behavioral Scientist, 33*(2), 157–161.

Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory, 25*(1), 46–69. doi:10.1111/comt.12048

Gianetti, L. D. (1982). *Understanding movies*. Englewood Cliffs, NJ: Prentice Hall.

Gibson, R., & Zillmann, D. (2000). Reading between the photographs. The influence of incidental pictorial information on issue perception. *Journalism & Mass Communication Quarterly, 77*(2), 355–366.

Goldstein, J. S., & Pevehouse, J. C. (2013). *International relations* (10th ed.). New York: Pearson Longman.

Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. Oxford, UK: Oxford University Press.

Graber, D. A. (1986). Portraying presidential candidates on television: An audiovisual analysis. *Campaigns and Elections, 7*, 14–21.

Graber, D. A. (1987). Television news without pictures? *Critical Studies in Mass Communication, 4*(1), 74–78. doi:10.1080/15295038709360115

Graber, D. A. (1989). Content and meaning: "What's it all about?" *American Behavioral Scientist, 33*(2), 144–152.

Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication, 40*(3), 134–155.

Grimes, T. (1990). Audio-video correspondence and its role in attention and memory. *Educational Technology Research and Development, 38*(3), 15–25. doi:10.1007/bf02298178

Grimes, T. (1991). Mild auditory-visual dissonance in television news may exceed viewer attentional capacity. *Human Communication Research, 18*(2), 268–298.

DOI: 10.4324/9781315642239-13

Gunter, B. (1979). Recall of brief televised news items: Effects of presentation mode, picture content and serial order. *Journal of Educational Television, 5*, 57–61.

Gunter, B. (1980). Remembering television news: Effects of picture content. *Journal of General Psychology, 102*(1), 127–133. doi:10.1080/00221309.1980.9920970

Gunter, B. (2015). *The cognitive impact of television news: Production attributes and information reception*. London, UK: Palgrave Macmillan.

Horvat, K. V. (2010). Multiculturalism in time of terrorism. *Cultural Studies, 24*(5), 747–766. doi:10.1080/09502380903549855

Hsia, H., & Jester, R. E. (1968). Output, error, equivocation and recalled information in auditory, visual and audiovisual information processing with constraint and noise. *Journal of Communication, 13*(December), 325–353.

Kahneman, D. (1973). *Attention and effort*. New York: Holt, Rinehart & Winston.

Kelly, M. (2010). Regulating the reproduction and mothering of poor women: The controlling image of the welfare mother in television news coverage of welfare reform. *Journal of Poverty, 14*(1), 76–96. doi:10.1080/10875540903489447

Kim, S.-H., Carvalho, J. P., Davis, A. G., & Mullins, A. M. (2011). The view of the border: News framing of the definition, causes, and solutions to illegal immigration. *Mass Communication and Society, 14*(3), 292–314. doi:10.1080/15205431003743679

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication, 50*(1), 46–70.

Liebes, T., & Kampf, Z. (2009). Black and white and shades of gray: Palestinians in the Israeli media during the 2nd Intifada. *International Journal of Press/Politics, 14*(4), 434–453.

Liebler, C., & Bendix, J. (1996). Old-growth forests on network news: News sources and the framing of an environmental controversy. *Journalism & Mass Communication Quarterly, 73*(1), 53–65.

Machin, D., & Polzer, L. (2015). *Visual journalism*. Basingstoke, UK: Palgrave Macmillan.

Mascelli, J. V. (1965). *The five C's of cinematography*. Hollywood, CA: Cine/Graphic.

Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication, 58*(2), 258–279.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.

Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226). Mahwah, NJ: Lawrence Erlbaum.

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication, 65*(6), 997–1017. doi:10.1111/jcom.12184

Reese, S. D. (1984). Visual-verbal redundancy effects on television news learning. *Journal of Broadcasting, 28*(1), 79–87. doi:10.1080/08838158409386516

Reese, S. D. (2001). Prologue—Framing public life: A bridging model for media research. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 7–30). Mahwah, NJ: Erlbaum.

Reynolds, A., & Barnett, B. (2003). "America under attack": CNN's verbal and visual framing of September 11. In S. M. Chermak, F. Y. Bailey, & M. Brown (Eds.), *Media representations of September 11* (pp. 85–102). Santa Barbara, CA: Greenwood Publishing Group.

Robinson, C., & Powell, L. A. (1996). The postmodern politics of context definition: Competing reality frames in the Hill-Thomas spectacle. *Sociological Quarterly, 37*(2), 279–305.

DOI: 10.4324/9781315642239-13

Robinson, J. P., & Levy, M. R. (1986). *The main source: Learning from television news*. Thousand Oaks, CA: Sage.

Rowling, C. M., Sheets, P., & Jones, T. M. (2015). American atrocity revisited: National identity, cascading frames, and the My Lai massacre. *Political Communication, 32*(2), 310–330. doi:10.1080/10584609.2014.944323

Schaefer, R. J., & Martinez, T. J. (2009). Trends in network news editing strategies from 1969 through 2005. *Journal of Broadcasting & Electronic Media, 53*(3), 347–364. doi:10.1080/08838150903102600

Silcock, B. W. (2007). Every edit tells a story: Sound and framing routines of videotape editors in global news cultures. *Visual Communication Quarterly, 14*(1), 3–15.

Son, J., Reese, S. D., & Davie, W. R. (1987). Effects of visual-verbal redundancy and recaps on television news learning. *Journal of Broadcasting & Electronic Media, 31*, 207–216.

Stephens, M. (2005). *Broadcast news*. New York: Holt, Rinehart & Winston.

Treisman, A., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic.

Van Gorp, B. (2007). The constructionist approach to framing: Bringing culture back in. *Journal of Communication, 57*(1), 60–78.

Van Hoof, A., Takens, J., & Oegema, D. (2010). Poor framing in television news: Redundancy between audio and visual modalities in political news. Paper presented at the *Annual Meeting of the International Communication Association* (ICA), Suntec International Convention Centre, Singapore.

Van Leeuwen, T., & Jewitt, C. (2001). *Handbook of visual analysis*. Thousand Oaks, CA: Sage.

Walma van der Molen, J. H. (2001). Assessing text-picture correspondence in television news: The development of a new coding scheme. *Journal of Broadcasting & Electronic Media, 45*(3), 483–498. doi:10.1207/s15506878jobem4503_7

Walma van der Molen, J. H., & Klijn, M. E. (2004). Recall of television versus print news: Retesting the semantic overlap hypothesis. *Journal of Broadcasting & Electronic Media, 48*(1), 89–107. doi:10.1207/s15506878jobem4801_5

Walma van der Molen, J. H., & Van der Voort, T. H. A. (2000a). Children's and adults' recall of television and print news in children's and adult news formats. *Communication Research, 27*(2), 132–160.

Walma van der Molen, J. H., & Van der Voort, T. H. A. (2000b). The impact of television, print, and audio on children's recall of the news. *Human Communication Research, 26*(1), 3–26. doi:10.1111/j.1468–2958.2000.tb00747.x

Wessler, H., Wozniak, A., Hofer, L., & Lück, J. (2016). Global multimodal news frames on climate change: A comparison of five democracies around the world. *The International Journal of Press/Politics, 21*(4), 423–445. doi:10.1177/1940161216661848

Wirth, W. (2001). Der Codierprozess als gelenkte Rezeption. Bausteine für eine Theorie des Codierens [The coding process as a controlled reception. Toward a theory of coding]. In W. Wirth & E. Lauf (Eds.), *Inhaltsanalyse: Perspektiven, Probleme, Potentiale* (pp. 157–182). Cologne, Germany: Halem.

Wozniak, A., Lück, J., & Wessler, H. (2015). Frames, stories, and images: The advantages of a multimodal approach in comparative media content research on climate change. *Environmental Communication, 9*(4), 469–490. doi:10.1080/17524032.2014.981559

Zhou, S. (2005). Effects of arousing visuals and redundancy on cognitive assessment of television news. *Journal of Broadcasting & Electronic Media, 49*(1), 23–42. doi:10.1207/s15506878jobem4901_3

DOI: 10.4324/9781315642239-13