

Routinizing Frame Analysis through the Use of CAQDAS

Thomas Koenig

Department of Social Sciences

Loughborough University

t.koenig@lboro.ac.uk

(Draft Version)

Even though frame analysis has become a popular analytical framework in media studies and social movement research, the methodological underpinnings of the empirical identification of frames lack systematization and have consequently remained underdeveloped. This paper consolidates recent advances in the empirical measurement of frames and explores, in how far *computer-assisted qualitative data analysis software* (CAQDAS) can extend on these methodologies. Because framing has become a fairly widely used but ill-defined concept, the paper will start with a brief delineation of framing theory as it is understood for present purposes. Next, current attempts to measure frames empirically in a systematic fashion will be discussed and a methodology, which synthesizes some of these approaches will be proposed. This methodology attempts in a first step to draw on existing knowledge on metanarratives to avoid a purely inductive identification of frames. Additionally, automatic word mapping tools such as *Leximancer*, *Sphinx Survey Lexica* are suggested as interpretative aids. In a second step, the analyst identifies a set of keywords, key phrases, and possibly audial or visual symbols that indicate frames in his data. These indicators are then used in a third step to semi-automatically identify frames in the remainder of the data. Keywords that might acquire different meanings in different contexts are inspected in their contexts by the analyst, who decide on their coding. This method avoids both the rigidities that come with fully automatic keyword clustering, which may lead to the inclusion non-interpretable keywords as well as the exclusion of so-called stop words such as prepositions and articles, which under certain circumstances might indeed be the strongest indicators for certain frames. At the same time it allows for a degree of routinization and systematization in frame analysis, whose quality has notoriously depended on the creativity of the framing researchers.

Five CAQDAS – ATLAS.ti, Kwalitan, MAXqda, NVivo, and Qualrus – are examined with respect to their usability in this type of framing research.

The Use of CAQDAS in Frame Analysis

This paper consolidates recent advances in the empirical measurement of frames and explores, whether CAQDAS can enhance these methodologies. The article is divided into four parts. Because framing has become a fairly widely, but often used but ill-defined, concept, the paper will start with a delineation of framing theory as it is understood for present purposes. Next, a methodology to empirically measure frames will be developed. We will then examine the extent to which *computer-assisted qualitative data analysis software* (CAQDAS) can contribute to the implementation of that methodology. Five CAQDAS – ATLAS.ti, Kwalitan, MAXqda, NVivo, and Qualrus – are examined with respect to their usability in framing research. Finally, a short overview, on how to validate frame models with cluster analysis, factor analysis, and latent class/structure analysis will be made.

Framing Theory

Frame analysis is *en vogue* (Meyer 1999: 85; Reese 2001: 7; Benford and Snow 2000: 611f). The *Social Science Citation Index* counts 1,805 references over the past decade for Goffman's (1974) methodological foundation. That almost doubles Durkheim's ([1894] 1968) classic *Rules of the Sociological Method* (936 references) and also easily surpasses DiMaggio's (1983) re-interpretation of Weber's iron cage, the best known piece of the ubiquitous neo-institutionalism school.

Even so, frame analysis was initially predicted to become a niche method at best. One *Contemporary Sociology* reviewer complained that *Frame Analysis* is cumbersome to read (Davis 1975: 603), the other one wondered, if an adequate systematization of frame analysis would be feasible (Gamson 1975: 605).¹ How can this skeptical approach to framing at its onset be reconciled with its current success?

Probably the single most important factor for the success of Goffman's frame analysis is its unorthodox application. Frame analysis is no longer *Goffman's* frame analysis, but is frequently only loosely connected to the original formulation. Notwithstanding the recurrent symbolic nods to Goffman, today's "frame analysis" spans a number of disparate approaches (D'Angelo 2002; Fisher 1997; Maher 2001: 81f; Scheufele 1999: 103, 118), some of which are even incompatible with each other (Scheufele 1999: 118), While not excluding the possibility of fruitful interaction between the heterogeneous frame analyses (D'Angelo 2002:

¹ Needless to say, the latter reviewer became the probably most eloquent popularizer of frame analysis.

883), conceptual parsimony necessitates the clarification of the framing concept for present purposes.

This is not the place to overview the wide range of approaches that have been subsumed under the heading of frame analysis, a task that others (Benford and Snow 2000; D'Angelo 2002; Scheufele 1999) have already accomplished. Instead, I would like to merge at this juncture certain brands of framing approaches from media analysis and social movement theory to a more specific theoretical framework, whose potential methodological underpinnings will then be discussed.

Frame Concept

In his initial and widely quoted definition, Goffman characterized frames as follows:

"I assume that definitions of a situation are built up in accordance with principles of organization which govern events [...] and our subjective involvement in them; frame is the word I use to refer to such of these basic elements as I am able to identify" (Goffman 1974: 10f)

In other words, frames are basic cognitive structures which guide the perception and representation of reality. On the whole, frames are not consciously manufactured but are unconsciously adopted in the course of communicative processes. On a very banal level, frames structure, which parts of reality become noticed.

For example, a group of persons lined up in an orderly fashion at the side of a road might evoke the frame "bus queue" in a passer-by. This particular frame structures perception in the way that attention is paid to the orderly arrangement of people in a line, which is one indicator of the "bus queue frame" and might have actually triggered it. The frame also directs attention to other latent frame elements, such as a bus stop sign. At the same time, it deflects attention from clothing style, body shape, or communications among the presumed prospective bus passengers.

The adoption of frames is not immune to real world events. If a cab stops at the curb side in front of the line, chances are, the bus queue frame will become rejected and replaced by the "waiting for a taxi" frame.

Todd Gitlin has summarized these frame elements most eloquently in his widely quoted (e.g., Miller 1997: 367; Miller and Riechert 2001b: 115) elaboration of the frame concept:

"Frames are principles of selection, emphasis and presentation composed of little tacit theories about what exists, what happens, and what matters." (Gitlin 1980: 6)

While it is hard to improve theoretically on this definition, the trouble starts, when it comes to the identification and measurement of frames. Precisely because frames consist of tacit rather

than overt conjectures, notorious difficulties to empirically identify frames arise (Maher 2001: 84).

The difficulty of measuring latent frames could partially explain the gradual theoretical shift towards a conceptualization of frames as being more actively adopted and manufactured. Particularly in media studies, it has become commonplace to treat the choice of frames as a more or less deliberate process. Entman's famous definition of frames led the way. For Entman,

“[t]o frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation.” (Entman 1993: 52) Notice the shift towards active *selection* of

frames, a conception that has become dominant in media studies. While indeed not agreeing with Entman on much else, D'Angelo (2002: 873) likewise treats frames as consciously pitched subtle, but powerful discursive devices, as do a number of other eminent communication scholars with a pedigree in the agenda-setting approach ({Iyengar 1991 #540 /ft ": 11"}; {Rogers, Hart, et al. 1997 #530 /ft ": 235"}). Tankard (2001: 97) moves even beyond the mere conscious *selection* of frames, suggesting that journalists at times circulate frames to *deceive* their audiences. Reese (2001: 7) goes furthest in the direction of conscious framing suggesting that framing always *implies* an active process. Consequently, he demands that the analysts "should ask how much 'framing' is going on" (*ibid.*, 13). In a Goffmanian framework, such a question would have been non-sensical, since framing is an innate property of all social processes, not only those most consciously manufactured.

As journalists are professional symbol handlers with a high degree of self-reflexivity, it is probably no accident that media studies treat framing as a more conscious process. Indeed, social movement theory considers "conceptual scaffolding" (Snow and Benford 1988: 213), a metaphor, which in fact is more appropriate to Goffman's. However, social movement theory also elaborates "framing tasks" for successful movement mobilizations (Snow et al. 1986; Snow and Benford 1988), it thus also wanders into the direction of conscious framing.

In conjunction with this bias towards active framing, the ambiguity of the framing concept has led some framing researchers to suggest the conceptualization of frames as a metaphor, alluding to a picture frame (Tankard 2001: 98f; Tankard et al. 1991). In this reading, journalists select certain frames and transform them into "airtight compartments" that make complete social consciousness impossible (Durham 2001: 128). While I doubt that *any* metaphors are suitable for inclusion in sociological theories, *picture* frames are definitely not a metaphor in Goffman's spirit. In fact, it seems that Goffman himself tried to avoid the picture

frame metaphor, sticking to *framework* over the first few pages of *Frame Analysis*. For Goffman, Gitlin, and Gans frames are indispensable for communications, they are the scaffolds for *any* credible stories.

While the approach to consider frames delimiting might also have its merits, e.g., for use in agenda-setting approaches, frames will here be considered as both consciously adopted, but more frequently unconsciously used conceptual scaffolds. That still leaves open the question, which substantive frames can be detected.

Frame Typology

Since framing became a popular approach in the late 1980s, an extensive laundry list of frames has emerged in the literature (Benford 1997: 414). A random selection of only five articles within the framing approach yields 14 frames, many of which are singular to the article in question:

- conflict frame (Neuman et al.1992: 61f; Price et al. 1997: 484; Semetko and Valkenburg 2000: 95),²
- human interest frame (Price et al. 1997: 484; Semetko and Valkenburg 2000: 95f),
- economic interests frame (Triandafyllidou and Fotiou 1998: 4.19; Price et al. 1997: 484; Schwenken 2003: 5),
- moralization frame (Triandafyllidou and Fotiou 1998: 4.1; Ferree et al.2002: 107f; Neuman et al.1992: 75; Semetko and Valkenburg 2000: 96)
- common/collective interest frame (Triandafyllidou and Fotiou 1998: 4.12),
- responsibility frame (Ferree et al.2002: 108,³ Semetko and Valkenburg 2000: 96),
- scientification/rationalization frame (Triandafyllidou and Fotiou 1998: 4.4),
- individual and state frame (Ferree et al.2002: 107),
- troubled industry frame (Gamson1992: 216ff)
- affirmative action frame (Gamson1992: 223ff),
- fetal life frame (Ferree et al.2002: 107),
- women's rights frame (Ferree et al.2002: 107),
- migration autonomy frame (Schwenken 2003: 6f)),
- refugee protection frame (Schwenken 2003: 7f).

While all these frames appear to be reasonable reflections of social reality, it is difficult to tell, how they have been obtained, i.e. which methodology was used to arrive at particular frames and how they have been measured empirically. This disparity of frames leaves one

² Implicitly, the "balancing frame" (Ferree et al. 2002: 107) is part of the conflict frame.

³ Substantively identical with the social justice frame in Ferree et al. (2002: 108).

wonder, whether anything can be framed as a frame. Unfortunately, many studies leave the reader in the dark about the actual process of empirical frame detection. Even otherwise exceptionally well argued studies laconically describe the frame identification process in a footnote with "[f]rames were analyzed from the actual language of the reported claim (direct and reported speech)" (Statham and Mynott 2002: 10, Fn. 6). Another article describes the frame detection process as follows: "After a first reading of the material, five thematic dimensions have been identified, [... this process was then] complemented by an in-depth qualitative analysis of the data " (Triandafyllidou 2002: 3.7 & Fn. 10). A third paper flatly proclaims "during the data collection process, we coded all evidence of acknowledgment of or reactions to the ideas" (McCaffrey and Kayes 2000: 49) without mentioning how the coding was done.

In some cases, at least the measurement model for frames is clarified. In these cases the reader is presented with a list of more or less parsimoniously identifiable frame terms, "attributes" or "devices," which were used as manifest indicators for the identification of frames (e.g., Ferree et al. 2002; Koella [2003] 2003; Semetko and Valkenburg 2000; {Ullrich 1998 #600}). By making their entire coding scheme online available, Ferree *et al.* (2002) are in this respect the trailblazers for a new kind of transparency that has been made possible by the new digital technologies.⁴ While increased transparency and accountability certainly render framing research more credible, they still do not solve the problem of the missing systematization of frame construction. Notwithstanding some efforts to automate frame construction – which will be described below – we thus remain dependent on the creativity of individual scholars (Maher 2001: 84). This large discretion in turn makes framing research susceptible to the criticism of subjectivity, which can easily be construed as frames being merely constructed through "researcher fiat" (Tankard et al. 1991: 5; Tankard 2001: 98).

To counter these objections, the frame identification process should be made more visible and systematic. A first step towards the latter direction is the construction of a frame taxonomy. Notice that the abstraction levels of frames listed above is very different. Almost any social situation could be framed in a conflict or human interest frame, but frames such a "fetal life" are undoubtedly suitable for only a very limited scope of social phenomena. Indeed, a review of the list of frames quoted above shows that the first six frames in above list identify some *structural* scheme, while the other frames focus more on *content* (Benford 1997: 413).

⁴ <http://www.ssc.wisc.edu/abortionstudy/>, accessed: October 6, 2003.

The most frequently mentioned structural frames are the four frames at the top of above list, conflict, human interest, moralization, and economic interests. Additionally, there is ample literature from the agenda-setting approach, which distinguishes *episodic* and *thematic* frames (Iyengar 1991 #540). As structural frames can not be tied to specific topics, the methodology proposed below, which draws on thematic keywords, is not suitable to identify this type of frame.

Within the list of content frames, we can further distinguish between so-called "master frames" or "metanarratives" like "individual and state" that can be keyed in almost any situation and other content frames like

A master frame is not distinguished by its innate substantive contents, but by

- (1) the fact that it is so pervasive that it can be used in almost any situation, and
- (2) its superior credibility, in that it has moved beyond empirical scrutiny.

Empirically, the literature refers repeatedly to three master frames, i.e. the *ethno-nationalist* frame (Billig 1995; Brubaker and Laitin 1998: 428; Eder 1995: 4; Eder and Schmidtke 1998; Greenfeld 1999: 39; Statham and Mynott 2002: 13), the *liberal-individualist citizenship* frame (Berger 1971: 97f; Eder 1995: 4; McAdam 1996: 347; Somers 1995; Statham and Mynott 2002: 13) and the *harmony with nature* frame (Eder 1996: 191; Gamson 1992: 136).

With these clarifications and distinctions in hand, I will now propose a fairly systematic approach to identify *content* frames in *textual* data. Since the methodology rests on the selection of keywords and key phrases, it is less suited to identify structural frames such as the conflict frame, as these frames usually become manifest in the structure, and less in the wording of a speech. Likewise, this methodology is not suited to interpret visual and audial data, unless it is amended by significant researcher input. As much communication outside the field of computer-mediated communication relies on visual and audial cues, it is strongly recommended to supplement the methodology with other research techniques that cannot be described here.

Identifying Frames in Textual Data

Framing in the sense outlined above is a theoretically demanding concept, but – or, rather, as a result – it has proven elusive to measure (Maher 2001: 84). Even though on a conceptual level, frames, more often than not, are latent, read: not spelled out in their entirety, it seems reasonable to assume that *parts of* frames become manifest in speech. If, say, a speaker has adopted or keyed an ethno-nationalist frame, i.e., the conception that quasi-primordial culturally fairly homogenous groups of people can be delineated (and probably should be

granted some degree of self rule), we would expect this speaker to refer some components of that frame in speech. She or he might, for instance speak about peoples, might allude to some historical continuities, or refer to specific (ethnic) nations, such as "the Dutch," etc. These speech figures in turn can be identified by keywords (Entman 1993: 53; Triandafyllidou and Fotiou 1998: 3.7; Miller and Riechert 2001a: 61ff). With the help of these keywords, frames can be empirically identified in large corpora of data.

The first task in the empirical investigation of frames thus becomes the detection of these keywords. As keywords are manifest, this is a much simpler task than the identification of frames themselves. It has even been suggested to generate these keywords automatically, simply by mapping the most frequently words or strings within the data (Koella [2003] 2003: 7; Miller and Riechert 2001a: 70; Miller and Riechert [2003] 1994). In this methodology, the most frequent strings in a set of data, for example newspaper articles or press bulletins, are calculated. From the resulting list, stop words such as prepositions are expunged. In a second step, those terms that according to with the highest X^2 -Square rank are most unevenly distributed between the source documents are chosen as keywords that are presumed to identify frames. As this is a purely mathematical method for the choice of keywords, it avoids a key allegation leveled against more traditional interpretative procedures, namely that frames are established through mere researcher *fiat* (Tankard et al. 1991: 12).

While avoiding researcher bias, this method unfortunately creates three new problems. To begin with, it starts out with exactly a researcher *fiat*, that is by deciding *by convention* on the optimal number of eigenvectors (Miller and Riechert 2001b: 116). This decision might *sound* more "objective," as a number can be pegged onto, but that number is just as arbitrary as the decision on frames. Moreover, the procedure is deeply positivist (in a Carnapian sense), assuming that concepts should arise unmediated from the data. Yet, for reasons that cannot be discussed here, both epistemological and methodological positivism have even been rejected by its former champions (Popper[1934] 1966). However, even within a positivist logic, most statistical tests are based on *a priori* probabilities. By basing the decision in the choice of keywords on *ex post* covariances, these tests become meaningless. While this problem could be circumvented through a split sample, an even more severe problem is that empirically identified keywords clearly cannot be interpreted as an indicator of meaningful frames. Miller & Riechert (1994), for instance, found besides "environmental," "any," and "major" to be identifiers of the "environmental protection" frame. It seems obvious that these are not meaningful framing terms. Indeed, Koella (2002: 8), who most closely follows Miller and

Riechert, deviates in this point, wryly noting that "each set of frame terms was reviewed in context." This proceeding, of course, in turn reintroduces research fiat through the back door.

Frequency counts might thus hint at possible keywords, but in the end an interpretative identification of relevant keywords seems to be the more appropriate and more common route (Andsager et al. 2001: 129; Tankard 2001: 103; Tedesco 2001: 2053, more technically centered: Miller 1997: 369). Reading or listening over a reasonable amount of data, framing researchers should hermeneutically uncover frames and their corresponding keywords. The three master frames mentioned above could help the interpretation of data in this respect, as these frames are likely to surface in any communicative processes in modernity.

Once keywords have been obtained, they can then be used in conjunction with common CAQDAS and word maps such as *WordNet*⁵ or *Wortschatz*⁶ to code large amounts of data in a fairly short time. Initially, all keywords should become *lemmatized*, that is all their inflections forms are to be found. Next, their *listemes*, that is those linguistic representations⁷ which correspond to the *mental lexemes* held by persons involved in the communicative practices that are researched, should be identified. Listemes are the actual conceptual categories in the minds of individuals, regardless of their linguistic representation. Typically, true synonyms represent different linguistic representations of the same listeme, so that any keyword synonyms should be retrieved from the relevant thesauri. Linguistic research has shown that the mind orders listemes in a network structure (Gallmann 1991: 274). It might thus be advisable to also group keywords with their listeme neighbors. I would term the set of a listeme and its most closely related neighbors a *fuzzy listeme*. Figure 1 visualizes a fuzzy listeme for "car", highlighting all associated lemmata in green. If "car" is considered a keyword for a certain frame, then the fuzzy listeme might include the lemmata of "car", its synonyms "auto", "automobile" (as found in *WordNet*) and its significant⁸ collocations "Ford," "GM," "Chrysler," "Honda," "Nissan," "Toyota", "Saturn" as found in *Wortschatz*.⁹ The question, if it is prudent to include these brand names in a fuzzy listeme for research purposes will depend on the context of your data. When researching the discourses on

⁵ <http://www.cogsci.princeton.edu/~wn/>, accessed: November 27, 2003.

⁶ <http://wortschatz.uni-leipzig.de/>, accessed: November 27, 2003; for a selection of more electronic word maps, cf. <http://www.lboro.ac.uk/research/mmethods/resources/analysis/linguistic.html>, accessed: December 7, 2003.

⁷ In written text, these are words, but audio and video data they also refer to visual and audial discursive cues.

⁸ Universal frequency and collocation data are still not available in desirable quality (Quasthoff and Wolf [2003] 2003: 1), but *Wortschatz* currently is the most reliable database in this respect.

⁹ It is apparent that the corpus of *Wortschatz* contains an US-American bias.

economic topics, chances are, aforementioned brand names refer to the respective companies, not their product. If, on the other hand a bulletin board discussion on the use of public transport is researched, it is more likely that a reference to a brand name refers to a private car used by a specific participant.

Both mentioned word maps also tell you that "car" has "cable" as a significant left neighbor. A "cable car" however belongs hardly to the same fuzzy listeme as "car", but rather to "tram." You would, thus, have to deduct "cable car" from your keywords. Likewise, in later keyword searches homonyms might pose a problem (Bolden and Moscarola 2000: 453; Miller 1997: 369; Miller and Riechert 2001a: 65). In the current example, if "Saturn" is chosen to be included in the fuzzy listeme, the homonym planet "Saturn" would be required to be eliminated from analysis.

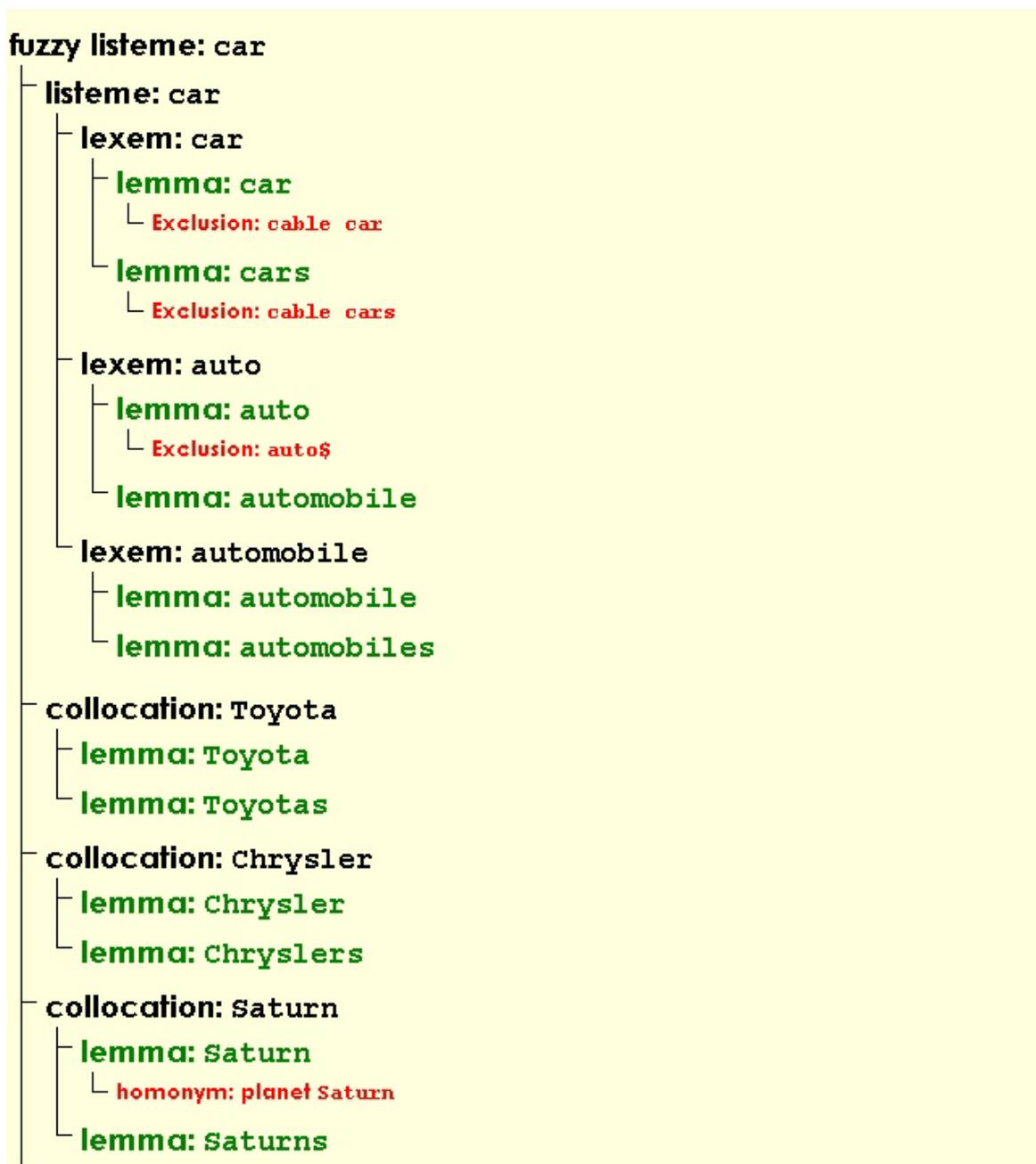


Figure 1 Fuzzy Listeme "Car"

Framing and CAQDAS

The approach to identify a fuzzy listeme through keywords sounds, as if it would be ideally suited to CAQDAS with its GREP search and coding functions. Initially, hermeneutic coding of frames might detect relevant lexeme. Codes could then be automatically generated through searching for the strings that identify lemmata. The instances, in which keywords take on specific meanings to be excluded, such as "cable car" (see above) could be excluded through Boolean search operations. Homonyms could be eliminated from analysis by visually

inspecting the textual environment of the keywords in question and an according interpretative decision on the meaning of the homonym in question. These seem sufficiently circumscribed procedures to be performed by a computer algorithm with the odd human input decision. Alas, the grounded theory bias of CAQDAS (Carvajal [2003] 2002: 3; Coffey et al. [2003] 1996: 7.3; Lonkila 1995; {Welsh 2002 #610 /ft ": 3"}) was quickly apparent and in the end only with a great deal of persistence and software tinkering it was feasible to obtain the desired analysis.

Grounded Theory bias of CAQDAS

By name, CAQDAS are designed for a wide variety of qualitative approaches, in practice many developers and users say "qualitative methodology" and think "Grounded Theory." Admittedly, there is neither a strict adherence among those who claim allegiance to Grounded Theory (Glaser and Strauss, Anselm L1967; Glaser [2003] 2002; Strauss and Corbin, Juliet M1990) to the procedures suggested by Glaser and Strauss (Bong [2003] 2002: 3f Lee and Fielding [2003] 1996: 3.1), nor is Grounded Theory an unequivocal paradigm in itself (Strübing 2002). But to jump from that observation to the conclusion that the centrality of Grounded Theory for CAQDAS is a myth (Gibbs et al. [2003] 2002: 6; Bong [2003] 2002: 6), itself creates a myth.

In fact, even Lee & Fielding's (1996: 3.2f) much more cautious thesis "that grounded theory is an important, but by no means ubiquitous, influence on studies" using CAQDAS seems an understatement. Their observation that "only" 31% of all published work accomplished with THE ETHNOGRAPH cited the major methodological works by Glaser and Strauss is at first glance a convincing indicator for the wide methodological variety of work done with CAQDAS. And the percentages have not changed in favor of Glaser and Strauss over the last 8 years. According to the *Social Science Citation Index*, only 30 of 104 (29%) of those articles that cite QSR's manuals, 25 of 85 (29%) of those that cite ATLAS.ti's manuals, 4 of 27 (15%) that cite winMAX or MAXqda manuals, and 3 of 8 (38%) that cite Kwalitan's manuals also refer to the major Grounded Theory work. But for two reasons these appearances are deceiving. First, as will be seen below, even if indeed only roughly 30% of the work done with CAQDAS consciously adopted a Grounded Theory framework, that proportion would still hugely inflate Grounded Theory's position with qualitative research. Second, much of the remaining work accomplished with CAQDAS does not quote *any* methodological work. Many of the articles based on CAQDAS research are published in the fields of applied research, particularly health sciences (Fielding and Lee 1995: 1f) – an

observation that a quick glance over more contemporary articles confirms. It is therefore not improbable, that the bulk of the remaining work does not follow *any* elaborate methodological path, but relies on a crude idea of Grounded Theory methodology, as Strübing (2002: 319) suspects.

To investigate the alleged Grounded Theory bias further, we therefore examined the archives of QUAL-software, a mailing list that is dedicated to the discussion of CAQDAS.¹⁰ All archived postings, as well as all web pages that were referenced in these postings were downloaded, yielding a total of 9,284 documents. These files were subsequently searched for the occurrence of key words associated with certain qualitative methods (discourse analysis, ethnography, frame analysis, and grounded theory); an approach at the intersection of qualitative and quantitative methodology (network analysis); and a decidedly quantitative method (regression analysis). To see whether a certain methodology is under or over represented in a CAQDAS research, we compared frequency counts to the same keywords search results for all post-1989 journal articles in the Sociological Abstracts database (see Table 1).

Keyword	QUAL-Software (9,284 files)		Sociological Abstracts (167,757 records)		Ratio
	n	%	n	%	
Methodologies					
Grounded Theory	380	4.09	228	0.14	29.94
Network Analysis	95	1.02	770	0.46	2.22
Discourse Analysis	79	0.85	1,812	1.09	0.78
Ethnograph\$ ¹¹	186	2.00	4,566	2.74	0.73
Frame Analysis	1	0.01	172	0.10	0.10
Regression Analysis	3	0.03	1,378	0.83	0.04
Sources					
Focus Group(s)	550	5.92	409	0.25	24.15
Newspapers	61	0.66	1,149	0.69	0.95
Survey	672	7.24	16,908	10.14	0.71
Participant Observation	25	0.76	1,273	0.76	0.35

Table 1 Ground Theory and Focus Group Bias of CAQDAS¹²

¹⁰ <http://www.jiscmail.ac.uk/lists/qual-software.html>, accessed: November 4, 2003.

¹¹ The "\$" sign is a placeholder for any letter character, such as in "ethnography" or "ethnographic." We added this condition to filter for hits on the *Ethnograph* software.

The findings unanimously show that Grounded Theory is the dominant methodology for CAQDAS users — who mention it on average 30 times more frequently than sociologists as a whole. Discourse Analysis and Frame Analysis are less frequent in CAQDAS research than in sociological research in general, with Grounded Theory mentioned 300 times more frequently than Frame Analysis.

Furthermore, as with methodological bias, we can also detect a strong bias towards specific data sources in CAQDAS research. Table 1 also shows that focus groups are 24 times more popular on the CAQDAS listserv than in qualitative researcher as a whole.

Does the bias on Grounded Theory and focus groups that mean that CAQDAS are entirely unsuited for Frame Analysis? In the following section we will investigate the suitability of CAQDAS for the methodology proposed above.

Frame Analysis: CAQDAS in Practice

To assess the usability of CAQDAS for the methodology proposed here, we collected postings from an internet forum. The forum in question is hosted by the website of the Christian Democratic Union (CDU), the main conservative party in Germany. We downloaded the forum thread, in which CDU supporters debated the legitimacy of the dismissal of a CDU MP, Martin Hohmann, from the parliamentary ranks of the party.¹³ The dismissal had been triggered by a speech by Hohmann, which contained elements that were widely regarded as anti-Semitic. Many rank and file members considered the dismissal unjustified and hence a lively debate ensued in the forum. The debate continues to date, but we chose to follow the thread only over the first three months, collecting 2,626 postings from the participants.

Importing the Files

As with almost all internet data, postings from the forum came in HTML format. In this particular case, we obtained a single HTML page incorporating altogether 2626 postings. We split this page using the *csplit*¹⁴ program into 2626 separate files. Since none of the five CAQDAS we examined can actually directly process HTML, the files were stripped of their HTML tags using *NoteTab*. As a result we obtained 2626 plain text files, with each file

¹² For purists — $\chi^2_5 = 2384$ (significant at the $p < .0001$ level).

¹³ <http://www.cdu.de/cgi-bin/forum.cgi?zeigealle=/forum/thema4/ilelEtrLM.ovr>, accessed: December 10, 2003.

¹⁴ http://sourceforge.net/project/shownotes.php?release_id=151105, accessed: December 1, 2003.

representing one posting. Of the five programs, *MAXqda* is the only software unable to process plain text files, requiring instead rich text format. As it is very cumbersome to open such a large amount of files in a text processor, such as *Microsoft Word* or *Corel WordPerfect*, and subsequently convert them into rich text, it was decided to use *ABC Amber Text Converter*¹⁵ to batch convert all plain text into rich text format for usage in *MAXqda*. While the conversion to plain text required only seconds, a Pentium 4 computer with 512MB RAM needed three and a half hours to do the former conversion. While this may not seem long, it would have been fairly simple to collect similar data from other web forums, easily amassing as many as one million postings. At the current rate, a conversion of these postings would have taken more than 50 days.

After doing the necessary conversions, the set of files was imported into the programs. As none of the examined CAQDAS allowed for the import of an entire folder – a standard feature in numerous freeware programs – each set of files had to be imported highlighting the files in question in the Windows 2000 "open file" dialog window. As this dialog window apparently only allows for a limited number of files or characters, at most 390 files could be imported with each batch, holding down the "shift" and "ctrl" keys while highlighting the files, a technique many less experienced Windows users are unaware of, as postings on support fora for CAQDAS indicate.¹⁶ An attempt to import the entire set of files led to various error messages in the different programs, neither of which directed to a maximum number files, so it took quite some tinkering before the mistake was found. NVivo did not even display an error message and simply returned us to the "create documents" dialog without any comment. In none of the programs the *import documents* dialog window would automatically set the working directory to the document folder identified in the first input, meaning that the directory needed to be "re-found" each time. NVivo further surprised us by being unable to process more than 260 files at a time. This in turn required 11 document import processes, as opposed to 7 in all other programs. To make things even worse, NVivo tied up almost all CPU resources during the import process making it infeasible to work in other programs while importing files. If NVivo's *Document Explorer* was open during the import process, importing 200 files would take up to 15 minutes. To avoid crashing, NVivo required

¹⁵ <http://www.thebeatlesforever.com/processtext/>, accessed: November 30, 2003.

¹⁶ <http://www.maxqda.de/~upload/maxqdaforumdeutsch/forum/start.asp?forumid=3&select=498>, accessed: December 9, 2003

<http://www.listserv.dfn.de/cgi-bin/wa.exe?A2=ind0002C&L=atlas-ti&P=R279&I=-3>, accessed: December 9, 2003.

furthermore to save the project after each import and to quit and restart the program after every other import of 200 files. It altogether took 54 minutes to import all files into NVivo, while all other programs required about 8 minutes. Most users, however, would have imported the files one by one, which could easily have taken an entire working day. Again, the fictitious but reasonable one million files would have required at least one working day to import, with NVivo requiring easily more than an entire week.

Free Coding

Once all documents had been imported, simple hermeneutical keyword coding was performed. With the three master frames in mind, and an initial skimming of the documents, we started highlighting and coding those sections of the documents, which we deemed indicative for the frames we saw emerging from the discourse. As inductive coding is standard praxis in Grounded Theory, this type of coding unsurprisingly worked well with most programs. However, we still found some drawbacks in the process.

Kwalitan offers an intuitive coding through highlighting keywords or phrases and a pop-up menu on right-click, which, unfortunately does not automatically show all available codes. Unlike in the other programs, codes can not be order hierarchically. They are also not shown when working on a document, which hinders the coding process considerably, as double or even triple codings likely occur.

NVivo permits quick coding of keywords with two mouse clicks; codes can be created at will and are neatly organized in a handy code menu. Unlike *Kwalitan*, *NVivo* offers a margin in the document window, where code stripes can be displayed. Alas, the display of the code stripes consistently froze the computer, a shortcoming is well known to the developers. QSR laconically suggests the following remedy: The "closing [of] the [coding] stripes display should speed up the editing/coding process."¹⁷ While this is true, we are then back to "blind" coding and face the same problems as in *Kwalitan*. At this point, a remark on *NVivo*'s speed seems in order. Several operations that we performed in *NVivo* drew so many system resources that concurrent work was virtually impossible. Figure 2 displays a rather typical Windows Task Manager, when *NVivo* is running: The program uses almost all CPU time and by far the most memory of all programs, requiring 30 to 40 times more memory than its competitors. *NVivo*'s developers concur that such behavior causes speed problems with larger projects:

¹⁷ <http://www.qsrinternational.com/support/faq/FAQ/answer.asp?ID=137>, accessed: December 9, 2003.

"NVivo is not designed for humungous projects - it was designed for [...] those who wanted fine analysis of rich data, and that's usually not gross amounts of rich data! For bigger projects, we kept the NUD*IST line, with command files and the limitations of plain text and fixed units. N6 projects with very many hundreds of documents report in happily to us. [...] S]plitting of projects is a good strategy if you really want to take more than a few hundred docs into NVivo - but likely you don't ..."¹⁸

That seems a fair enough limitation of NVivo, but at the same time the usage of N6 with its Windows 3.11 style interface upon fairly cryptic DOS commands is a severe impediment to most users, which by now are accustomed to graphical interfaces of more recent provenance. More to the point, however, our 800 typewriter pages of data material hardly amount to a "humongous" project. After all, reading through all postings would take only several hours.

¹⁸ Richards, Lyn (2003, September 15). RE: upper limits in N6. Message posted to QSR-Forum@qsr.com.au.

Image Name	PID	CPU	CPU Time	Mem Usage
NVivo.exe	3232	99	1:05:18	94,688 K
opera.exe	1008	00	0:27:26	23,828 K
WINWORD.EXE	1308	00	7:10:55	14,436 K
Netscp.exe	3208	00	0:02:57	13,524 K
webscanx.exe	208	00	0:00:01	10,916 K
csrss.exe	164	00	0:02:29	9,808 K
svchost.exe	512	00	0:00:01	7,720 K
vshwin32.exe	940	00	0:00:16	7,596 K
mcsshield.exe	928	00	0:21:30	7,568 K
i_view32.exe	3360	00	0:00:01	6,980 K
qttask.exe	1312	00	0:00:00	5,964 K
SPOOLSV.EXE	452	00	0:00:01	5,548 K
services.exe	212	00	0:00:02	5,452 K
Eudora.exe	3224	00	0:01:16	4,968 K
NotePro.exe	3192	00	0:00:36	4,956 K
svchost.exe	796	00	0:00:00	4,828 K
explorer.exe	1028	00	0:05:33	4,724 K
CTFMON.EXE	1336	00	0:00:06	4,368 K
svchost.exe	420	00	0:00:00	4,164 K
NMSSvc.Exe	564	00	0:00:02	4,116 K
taskmgr.exe	3004	00	0:00:01	4,112 K
Imgicon.exe	1324	00	0:00:00	4,064 K
avconsol.exe	1348	00	0:00:13	3,504 K
hpztsb05.exe	1060	00	0:00:00	3,444 K
atiptaxx.exe	1156	00	0:00:00	3,276 K
Atlasti.exe	3180	00	0:01:20	3,200 K
mstask.exe	556	00	0:00:00	3,168 K
MDM.EXE	1092	00	0:00:00	3,064 K
avsynmgr.exe	492	00	0:00:02	2,912 K
max2k.exe	3324	00	0:06:57	2,764 K
vstat.exe	876	00	0:00:01	2,756 K
SCARDS32.EXE	744	00	0:00:33	2,436 K
point32.exe	1296	00	0:00:00	2,376 K
SOUNDMAN.EXE	1276	00	0:00:00	1,656 K
ZipToA.exe	808	00	0:00:00	1,628 K
jusched.exe	1328	00	0:00:00	1,616 K
stisvc.exe	680	00	0:00:00	1,588 K
scardsvr.exe	348	00	0:00:00	1,288 K
AcroTray.exe	1344	00	0:00:00	1,284 K
lsass.exe	224	00	0:00:02	1,228 K
ActivityDisk.ex	532	00	0:00:00	1,096 K
winlogon.exe	184	00	0:00:40	1,052 K

Processes: 48 CPU Usage: 100% Mem Usage: 524908K / 1278080K

Figure 2 Nvivo Resource Drainage

MAXqda does not share NVivo's and Kwalitan's blind coding problems, but free coding is still slightly slower than in other programs, since a code needs to be first created in the – still Windows 3.11 style – *Codes* Window. Only then can be used for coding. Even though all codes are conveniently organized in the *Codes* Window, the drop down menu for codings is

disorganized, which makes it at times difficult to find the desired code. There is a little bug in the coding procedure, as at times not all codes are available for coding. Double clicking the desired code in the coding window, solves this problem.

Both *Qualrus* and *ATLAS.ti* allow for a more intuitive and comfortable free coding procedures. Both open a well organized coding window after right-clicking a highlighted portion of the text and there display of codes in the document margins is impeccable. Qualrus offers additional help, suggesting codes based on correlations between existing codes (Brent and Slusarz 2003: 189), a procedure irrelevant for this study, though.¹⁹ *ATLAS.ti* offers additional "quick" and "in vivo" coding procedures, which allows for the most rapid coding of all programs tested.

Altogether, free coding was easy enough in all five programs, with the missing coding views in Kwalitan and NVivo being the biggest, yet still minor annoyances. Nevertheless, traditional highlighting of printouts with a text marker was an even more efficient way of doing the initial codings, perhaps because it is just more ergonomic to work with pencil and paper than on a screen. We thus read through most of the 1,000 messages we inspected visually on paper and subsequently developed the coding scheme within the software. From the reading, we distilled five hypothesized frames based on two master frames and identified corresponding 71 lemmata, many of which can be found in a single search. Table 2 displays the search terms with the actual search strings set in boldface.

Automatic Coding

Four types of searches were to be performed. Unanimous lemmata such as "Gutmensch" ("do-gooders") require only simple string searches. Lemmata such as "Freiheit" ("freedom") are fairly unanimous, but acquire in specific contexts a different meaning. For instance, "Freiheit" could also be part of the newspaper title *Junge Freiheit*, a neo-right propaganda paper, in which case it would no longer belong to the lexeme "freedom." Boolean searches could automatically eliminate such double meanings. Then there are lemmata that only become the desired frame indicator, if they refer to specific lemmata. For instance, one hypothesized frame in the debate evoked a "normalization" of German ethnicity, claiming a *Sonderweg* in Germanness because of the atrocities during the Third Reich. In this frame, a calls for, or – in case of its "countertheme" (Gamson1992: 135) – against, a normalization Germans'

¹⁹ Furthermore suggestion algorithm, based on a positivist-inductive logic, might in many cases do more harm than help the analysis.

relationship to "their ethnicity." Two lexeme, "normal" and "pride", seemed to be related to this frame, but only if they referred to Germanness. Therefore, they were only coded in the normalization frame, if they were found close to the "German" lexeme. For this procedure, proximity searches were needed. Finally, there are searches that require the interpretative input of the coder, as their multiple meanings cannot be distinguished automatically. For instance, "Spiegel" could refer among others to the popular newsmagazine *Der Spiegel*, to the head of the main German Jewish association, Paul Spiegel, or could simply mean "mirror." These searches do not lend themselves to automatic coding, but require a case by case interpretation by the researcher.

Ideally, simple, Boolean, and proximity searches would thus be coded automatically, while those searches that required user input (highlighted in orange in Table 2) would display the context, in which the word is found to facilitate a swift identification of the proper code. None of the programs fulfilled all our requirements, but there were substantial differences in the adequacy of the different programs.

Master Frame	Frame	Fuzzy Lexemes	Lemmata	Exclusions
Liberal Individualist Citizenship Rights	freedom of speech	freedom of speech	Andersdenkende, -n Freiheit Freiheiten Meinung, -en Meinungsfreiheit, -en Meinungsäußerung, -en Meinungsäußerung, -en	Junge Freiheit
		repression of	repressiv, -e, -en, -er Repressivität unterdrücken, -t, -te, -ten Unterdrückung Maulkorb mündig Sanktionen	
		ensorship	Zensur Zensor, -en, -s zensiert, -e, -er, -en zensieren	
		taboo	Tabu, -isierung, isierungen tabuisieren, -iert, -ierte	
		totalitarian	totalitär, -e, -en, -er Totalitarismus Berufsverbot	
		political correctness	Berufsbetroffene, -r, -m, -n berufsbetroffen, -e, -en, -em Gutmensch, -en berufsmäßigen Vergangenheitsbewältigern Berufsdreckschleudern PC peeeße politisch, -e, -en, -er korrekt, -e, -er, -en Korrektheit political, -ly correct correctness	Berufsverbot
		Constitution (Basic Law)	Grundgesetz Grundrecht Grundrechte Verfassung Bürgerrecht, -e, -en Artikel	
	rebuke of elitism	second chance	2. Chance 2.Chance zweite Chance 2 Chancen zwei Chancen	
		witch hunt	Hexenjagd, -en Hexenverfolgung, -en Hexenverbrennung, -en Hexe, -n Inquisition	Hexentanz
		inciting	Hetzen Hetzeri, -en hetzerisch, -e, -er, -en Hatz	
		metaphor "to keep cooking"	Kochen kocht, -en hochkochen	
		Basis	Basis Parteibasis	
		Christian Democrat leadership	CDU-Führung Partei-Führung Parteiführung Merkel, -s Bosbach, -s Stoiber, -s Koch, -s	
		Media	Medien Presse Spiegel Stern	

			Journalist , -en	
		Social Sciences	Soziologie Soziologen Benz	Mercedes-Benz Daimler-Benz
Ethno-Nationalism	undue Jewish influence	<i>Zentralrat der Juden</i>	Zentralrat ZdJ Friedman , -s Spiegel , -s Berufsjuden	Der Spiegel Im Spiegel Spiegel Artikel
		American Jews	amerikanische Jüdin	
		"Holocaust Industry"	Holocaust-Industrie Holocaustindustrie Finkelstein , -s	
	normalization of German ethnicity	Germanness	Deutschland deutsch , -e, -er, -es, -en Deutsche , -er, -es, -en andere Länder anderen Ländern	
		patriotism	Patriot , -en Patriotismus patriotisch , -er, -em, -en	
		collective guilt	Kollektivschuld kollektiv schuld , -ig	
		repentance	Büßer	
		constant reminder	permanent , -e, -en, -er ewig , -e, -er, -en Erinnerung erinnern	
		German	Deutschland deutsch , -e, -er, -es, -en Deutsche , -er, -es, -en	
		guilt	Schuld schuldig , -e, -er, -en Schuldige , -er, -en	
		normal	normal , -e, -er Normales Normalisierung	
	Anti-Semitism	pride	stolz , -e, -er Stolz , -es Nationalstolz , -es	
		anti-Semitism	Antisemit Anti-Semit Antisemit , -en Anti-Semiten Möller , -s Karsli , -s Walser , -s Jenninger , -s	
		Jews/Jewish/Jewry	Jude , -n jüdisch , -e, -er	
		religion	Religion	
		Nazis	braun Nazi , -s NSDAP Nationalsozialist , -en nationalsozialistische , -r, -n, -m	

Table 2 Framing Devices (search terms set in boldface, homonyms in orange, conditional searches in olive)

On first sight, *Qualrus* appeared to be a prime candidate for our tasks. It is the most recently developed program, and boasts automatic coding functions, which is the reason why some have considered it most suitable to alleviate the Grounded Theory bias (Gibbs et al. [2003] 2002: 10). Yet, *Qualrus* turned out to be the least suited for our purposes. Its search functions are not comprehensive and efficient, if fairly speedy. Using the *Q-Tools* search menu in *Qualrus*, simple string searches were completed within 23 seconds over all

documents. When searching across paragraphs, the same search would take more than an hour. Since we were only interested in instances of frames within documents, the latter problem did not concern us. Boolean search is implemented, but it is only possible to combine "and" and "or," but not "not" operators. Proximity searches are not implemented, thus both more sophisticated search strategies we required were not available. Interactive coding turned out to be fairly cumbersome: From the search window a link for each document needs to be followed, after which the search window disappears, and cannot be retrieved through the familiar options ("ALT-TAB" or "CTRL-TAB" keystrokes or *Windows* menu), but only by reopening *Q-Tools*. In the document window, the search term needed to be found manually. The most important problem however was that *Qualrus* does not allow for automatic coding of keywords. It requires first a manual definition of analytic "segments," which cannot be generated automatically. The program is thus unsuitable for efficient automatic codings of large document samples.

So is NVivo, but for different reasons. NVivo's search functions, which owe much to earlier NUD*IST releases, but no longer beat the competition "hands down" (Weitzman and Miles, Matthew B1995: 248). It is, however, still the only program that allows for *fuzzy searches*, string searches in which the finds differ in one or more characters from the search string. That function is of course of particular importance for Usenet, internet fora, listserv, and chat room research, where users all too frequently misspell words. In our case, for instance, NVivo found 851 cases of lemmata containing the "antisemit" string, while all other programs found only 848 instances. Yet, in 29 seconds a simple fuzzy string search was accomplished still 7 times faster than a regular search in ATLAS.ti, the slowest competitor. When interactive coding is required, the procedure become slightly cumbersome. Keywords cannot be display in their context, it is therefore necessary to open each document that contains a homonym going through three successive windows. Boolean searches of text strings also require somewhat lengthy procedures; the strings in question first need to be transformed into codes, which subsequently can be searched with all Boolean and operators. As in the other CAQDAS, but unlike in the freeware *Inforapid Search & Replace*,²⁰ which we used as a benchmark program, slightly more complex combinations of the type "A AND B AND NOT C" are not permitted and thus must be run successively. While these limitations might be regarded as mere nuisances, it means that any Boolean or proximity search across more than 900 documents will take more than three hours. As the searching time rises

²⁰ <http://www.inforapid.de/html/searchreplace.htm>, cessed: December 11, 2003.

exponentially with the number of documents (600 documents can be searched in about 12 minutes, 300 documents in 30 seconds), these searches became infeasible. Even though the automatic coding functions were working smoothly, if at times somewhat serendipitously, NVivo was thus not suitable for our tasks. This assessment that flies in the face of claims that "unique and innovative developments in QSR software [...] have contributed significantly to [...] advances" in integration of qualitative and quantitative data and methodologies (Bazeley 2002: 230).

ATLAS.ti offers the widest range of autocoding options. As with NVivo, Boolean and proximity searches, require prior coding of single strings. In the somewhat opaque search window, all Boolean operators can be combined, even though AND and OR are not available in a single search. In theory this is an almost ideal autocoding environment. In practice each search and coding procedure took between 6 and 15 minutes for all 2626 files. However, unlike NVivo, *ATLAS.ti* does not tie all system resources, however, so you can work in other programs at the same time. However, *ATLAS.ti* appears less stable than NVivo. After approximately every other autocoding, the program would crash by exiting with the resulting loss of all previous work. Interactive searches require both *Code Manager* and document windows to be open, so at times some juggling of windows is required, but altogether this constitutes the simplest interactive coding of all programs in question.

MAXqda features the most archaic user interface, clearly still grounded in the Windows 3.11 ergonomics. Its search functions are not as powerful as those offered by its competitors. Its Boolean search, for instance, does not allow for the NOT operator (even though via "logical activation" of text can partially be circumvented), proximity searches can only be limited to paragraphs, not to word distances as in the other programs. Yet, *MAXqda* is *more* suitable for the type of research proposed here. To begin with, its interface, however outdated, is quite intuitive. Boolean and proximity searches are performed in a fraction of the time that *ATLAS.ti* or NVivo require and interactive coding is as simple (or as difficult) as in *ATLAS.ti*, while the program is much more stable. Unfortunately, the display keywords in context is not provided for. *MAXqda* may have shortcomings pertinent to other methodologies, but we were able to code above coding scheme within four hours, while in *ATLAS.ti* we needed a full working day to code only the first fuzzy lexeme²¹ and in NVivo and Qualrus we were not able to accomplish our work at all. It might there fore be no accident

²¹ Partially, this slowness was due to us losing work because of program crashes.

that MAXqda's predecessor winMAX was the only CAQDAS we know of that has been used for framing analysis (Van de Steeg 2003).

Unfortunately, we only evaluated the demonstration version of *Kwalitan*, which is restricted to four documents at a time. Therefore we cannot tell, how stable and fast the full version would have been. Its search function is somewhat counterintuitive, as Boolean searches can only be made using the *Filter* window (from which a few translations from the original Dutch are missing). Unlike all other CAQDAS, however, it allows for complex combinations of Boolean searches. Proximity searches are limited to segments, i.e., paragraphs. Interactive coding is somewhat tedious, because of the lack of coding stripes. In summary, *Kwalitan* seems very well suited for our tasks.

Export of Data Matrices

The export of the coding matrices for work in statistical packages or spreadsheets was unproblematic in all programs. Most programs allow for both export of ASCII text and *drag and drop* into windows programs. The only minor problem arose with MAXqda, whose code names mirror the code position within the coding tree, including a backslash separator to separate tree levels. These names cannot be processed by several programs, notably SPSS and lem, and therefore must be shortened in a text editor.

Task	ATLAS.ti	Kwalitan	MAXqda	NVivo	Qualrus	Inforapid
academic license	\$395 €390 £250	€315	\$370 €340 £255	\$445 £270	\$399	freeware
use of system resources	moderate	low	low	very high	low	low
file format required	plain text	plain text	rich text format	plain text	plain text	HTML
batch converter from HTML	several freeware options	several freeware options	ABC Amber Text Converter (US \$24.95)	several freeware options	several freeware options	not required
conversion time	<1s	<1s	3h20min	<1s	<1s	n/a
source import	8min	8min	8min	54min	8min	<1s
manual coding	efficient and intuitive	intuitive, but "blind"	efficient (codes are required to be created first)	intuitive, but codings cannot be displayed while coding	efficient and intuitive	not available
simple search	placeholders available	yes	yes	fuzzy search and placeholders available	no automatic coding	placeholders available, no coding
timing	6min20s	...	10s	19s	23s	20s
Boolean search	all operators, but multiple combinations not possible	all operators, any combination	AND <i>or</i> OR operators, no NOT operator	all operators, but multiple combinations not possible	AND <i>or</i> OR operators, no NOT operator	all operators, any combination
timing for one search ²²	12min20	...	5min10	>>3h	5min20	26s
proximity search	yes	yes	yes, but only with respect to paragraphs	yes, but only with respect to paragraphs	not available	yes, combinable with Boolean search
timing	6min20	...	19s	>>3h	n/a	26s
auto coding	simple, but frequent system crashes	all searches	simple and Boolean	simple searches only	not available	not available
Interactive Coding	easy	serendipitous	unhandy windows	unhandy windows	very cumbersome	not available, context shown, search term highlighted
Export of Coding Matrix	ASCII and <i>drag and drop</i>	<i>drag and drop</i>	efficient, but variable names not suited for direct import into SPSS	ASCII and <i>drag and drop</i>	ASCII and <i>drag and drop</i>	not available

Table 3 Suitability of CAQDAS for Methodology Proposed Here (serious problems set in orange)

²² Includes precoding for NVivo and ATLAS.ti and manual recoding for MAXqda.

Table 3 summarizes the results of our attempt to use CAQDAS for the analysis technique proposed here. All programs contain strength and weaknesses. Qualrus excels in interactive coding, but lacks an automatic coding function, which could only be emulated through an elaborated script. NVivo offers the widest variety of searches, but limits analysis to a couple of hundreds of documents. ATLAS.ti is the most versatile program, but its instability and time consumption pose serious problems. MAXqda is easy to use, much more stable than the previous three programs, but contains some limitations in input, output, and search versatility. Kwalitan has similar limitations as MAXqda, but it is more versatile in the import and export of files. Because we only tested its demonstration version, an assessment of its speed and stability cannot be made.

Surprisingly, the freeware InfoRapid offers the best search functions, does not require any file conversions, as it accepts a wide variety of file types, and is more stable than the newer CAQDAS. Together with its companion programs *InfoRapid Cardfile*,²³ which organizes files in a database, and *InfoRapid Knowledgemap*,²⁴ which visualizes ideational relationships, it might be a viable alternative to genuine CAQDAS.

In the end, two of the five CAQDAS examined cannot be recommended for use with the methodology proposed here. Since Qualrus does not allow for automatic coding, it cannot be used efficiently, and must be dismissed as a candidate, particularly as the freeware InfoRapid would be an efficient helper in coding, if one were to code everything by hand. The two most popular CAQDAS, NVivo and ATLAS.ti, were in principle suited for the analysis, but they become instable when used with larger amounts of files. In fact, NVivo is inherently incapable of handling more than 700 documents and must therefore be excluded from consideration. ATLAS.ti, is somewhat more erratically instable. At times the program works fine, performing ten to twenty autocodings without a problem. Then, there are instances, where a single autocoding is sufficient to crash the program. As we tested the freshly released version 5, these problems might disappear with the maturing of the program. What will not disappear, is the length of the time taken for each coding procedure. Each autosave operation and most searches took several minutes. That may not sound much, but the coding we performed within one working day with ATLAS.ti took 20 minutes in MAXqda. ATLAS.ti can therefore only be recommended with reservation. Particularly, for multi-method approaches that involve data other than simple texts, ATLAS.ti's versatility with multimedia

²³ Shareware (€/\$39), <http://www.inforapid.de/html/cardfile.htm>, accessed: December 13, 2003.

²⁴ Limited freeware, <http://www.inforapid.de/html/knowledgemap.htm>, accessed: December 14, 2003.

data might nevertheless make it a viable choice. That leaves us with MAXqda and the dark horse Kwalitan, both of which can be recommended with some reservations. They are fairly intuitive to use (which sets them apart from the rather idiosyncratic interface of NVivo and the shrouded terminology of ATLAS.ti) and do the job fairly efficiently. Of the two, Kwalitan seems at first glance to be more versatile. The fact that the two lesser known programs turned out to be more suitable for the methodology proposed here in any case dispels the myth that "the most successful qualitative software packages are likely to be constructed in ways that meet a range of methodological approaches" (Jackson 2003: 100).

Still, *InfoRapid Search and Replace*, which outperforms all CAQDAS in those CAQDAS functions it performs, may be a viable low budget alternative. Its only drawback is that it does not allow for coding, but a manual coding of documents in a regular relational database such as Microsoft Access might be a viable option for small projects.

Outlook: Validating Frames

As has been mentioned above, some scholars are criticizing framing approaches, because they feel that all too many frames are made up by researchers at will. To rebut that criticism, it seems useful to develop a technique that tests the empirical adequacy of frame models. Basically, three statistical techniques have been suggested to measure adequacy of frames quantitatively, namely cluster analysis, factor analysis, and latent class analysis. We will consider each of them in turn.

Cluster Analysis

Currently, hierarchical cluster analysis seems to be the most popular method for statistical validation of frames. That is, if you can speak of "popular", when merely a handful of references exist (Dyer 1994; Koella [2003] 2003; Miller 1997; Miller and Riechert [2003] 1994; Miller and Riechert 2001b; Miller and Riechert 2001a). The reason for its relative popularity is probably the existence of a computer program – VBPro²⁵ – that is specifically written for this type of analysis. The reason for its relative unpopularity might be the very same program, that is its command line DOS interface. There are a few other problems with this methodology, though. To begin with, it requires specific chunks of data – documents with around 1,000 words – to perform best (Miller 1997: 369). While this problem could be alleviated by slicing or aggregating data appropriately, the a general problem of all cluster

²⁵ <http://mmmiller.com/vbpro/vbpro.html>, accessed: December 15, 2003.

analyses – be it k-means or hierarchical – cannot be circumvented, namely that it does not offer any real goodness of fit tests (Aldenderfer and Blashfield, Roger K1984), which in turn makes it impossible to choose an optimum number of clusters on an empirical basis (Miller and Riechert 2001b: 116; Trochim and Hover 2003). That means that *any* number of frames could be posited throughout the texts, without any possibility to falsify any frame model, which, once again would return us to researcher *fiat*. On top, hierarchical cluster analysis assumes texts to belong to *either* one *or* the other frame. But it is entirely reasonable, and even likely, that speakers use any number of frames in a given text. In fact, many speakers actively engage in frame alignment processes such as frame bridging (Snow et al. 1986), which presuppose the existence of more than one frame in a text. Moreover, cluster analysis assumes a direct measuring model, but as has been discussed in the theoretical part of this paper, keywords are only indicators of latent frames. Altogether, hierarchical cluster analysis, thus, seems only ill suited for frame model validations.

Factor Analysis

Factor Analysis seems to avoid all the shortcomings of cluster analysis. It knows well-established goodness of fit criteria, it assumes a measurement model that does justice to the latency of frames, and it can decide on an empirical basis, which frame model is more adequate. Yet, to date we know only of one nascent attempt to use frame analysis in framing studies (Risse and Van de Steeg 2003). Interestingly enough, this appears to be the only frame analytic study, in which CAQDAS were used (*ibid*, p.5). While the headway made compared to cluster analysis is considerable, it seems puzzling that the authors do not even discuss the violation of the scale level assumptions of factor analysis, even though it has been shown that this violation can seriously affect the substantial results (Magidson and Vermunt 2004).

Latent Class Analysis

In contrast, latent class analysis exhibits all required features factor analysis offers, but at the same time does not contain the same shortcomings. It should seem therefore straightforward to introduce it into frame analysis studies.

Although the methodological principles of latent class analysis have been already developed in the fifties (Lazarsfeld 1950; Hagenaars1993: 20), it has remained an esoteric statistical method for many social scientists (Reunanen and Suikkanen, Risto1999: 3). Until the early eighties, the absence of quantitative studies using latent class analysis could be explained by the frequently cumbersome estimation of latent class models. Since then,

powerful computational equipment that easily performs these estimations has become widely available. The current draught in studies using this methodology seems instead to be rooted in the fact that none of the major statistical software packages (SPSS, SAS, and STATA) so far include procedures for latent class analysis. Freely available stand-alone programs, such as LCAG and IEM, on the other hand have probably garnished little interest because of their user-interface is not very intuitive.

Basically, latent class analysis can be considered the equivalent of factor analysis for ordinal and nominally scaled variables (McCutcheon 1987: 7). It examines, if a set of observable indicators can meaningfully be projected onto a smaller set of latent, that is, unobservable classes. Most important theoretical concepts, among them frames, do not translate straightforwardly into easily empirically observable, that is: measurable, indicators. Latent class analysis that expressly works with latent, read: unobservable, variables (Lazarsfeld 1950: 363) is therefore in the analysis of frames superior to other log-linear models that operate exclusively with observable data. In comparison to cluster analysis, latent class analysis delivers more unequivocal results, as it allows for a number of well-developed goodness of fit measures. And while it shares with factor analysis the virtue of operating with latent variables, it does not contain the caveat of requiring hard to come by interval scaled data.

References

- Aldenderfer, Mark S and Roger K Blashfield. 1984. *Cluster analysis*. Beverly Hills, CA: Sage.
- Andsager, Julie L., Erica Weintraub Austin, and Bruce E. Pinkleton. Questioning the Value of Realism: Young Adults' Processing of Messages in Alcohol-Related Public Service Announcements and Advertising. *Journal of Communication* 51[1], 121-42. 2001.
- Bazeley, Pat. 2002. "The evolution of a project involving an integrated analysis of structured qualitative and quantitative data: from N3 to NVivo." *International Journal of Social Research Methodology* 5 (3): 229-43.
- Benford, Robert D. 1997. " An Insider's Critique of the Social Movement Framing Perspective." *Sociological Inquiry* 67: 409-30.
- Benford, Robert D. and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology* 26: 11-39.
- Berger, Bennett. 1971. *Looking for America: Essay on Youth, Suburbia, and Other American Obsessions*. Englewood Cliffs, NJ: Prentice Hall.
- Billig, Michael. 1995. *Banal Nationalism*. London, U.K., Thousand Oaks, CA & New Delhi, India: Sage.

- Bolden, Richard and Jean Moscarola. Bridging the Quantitative-Qualitative Divide: the Lexical Approach to Textual Data Analysis. *Social science computer review* 18[4], 450-60. 2000.
- Bong, Sharon A. 2002. "Debunking Myths in Qualitative Data Analysis." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 3 (2).
- Brent, Edward and Pawel Slusarz. 2003. "'Feeling the Beat': Intelligent Coding Advice from Metaknowledge in Qualitative Research." *Social Science Computer Review* 21 (3): 281-303.
- Brubaker, Rogers and David Laitin. 1998. "Ethnic and Nationalist Violence." *Annual Review of Sociology* 24: 423-52.
- Carvajal, Diógenes. 2002. "The Artisan's Tools. Critical Issues When Teaching and Learning CAQDAS." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 3 (2).
- Coffey, Amnda, Beverley Holbrook, and Paul Atkinson. 1996. "Qualitative Data Analysis: Technologies and Representations." *Sociological Research Online* 1 (1): 4.
- D'Angelo, Paul. 2002. "News framing as a multi-paradigmatic research program: A response to Entman." *Journal of Communication* 52 (4): 870-88.
- Davis, Murray S. 1975. "Review of 'Frame Analysis: An Essay on the Organization of Experience' by Erving Goffman ." *Contemporary Sociology* 4 (6): 509-603.
- Durham, Frank D. 2001. Breaching Powerful Boundaries: A Postmodern Critique of Framing. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*, Ed. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dyer, Sam C. 1994. "A Newsflow Analysis of the Sale of a Public Asset." *Australian Journal of Political Science* 29: 172-8.
- Eder, Klaus. 1995. The Institutionalization of Environmentalism: Ecological Discourse and the Second Transformation of the Public Sphere.
- Eder, Klaus. 1996. *The social construction of nature : a sociology of ecological enlightenment*. London ;, Thousand Oaks, Calif. : Sage Publications.
- Eder, Klaus and Oliver Schmidtke. 1998. "Ethnische Mobilisierung und die Logik von Identitätskämpfen." *Zeitschrift für Soziologie* 27: 418-37.
- Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51-8.
- Ferree, Myra Marx, William Anthony Gamson, Jürgen Gerhards, and Dieter Rucht. 2002. *Shaping abortion discourse democracy and the public sphere in Germany and the United States*. Cambridge, New York: Cambridge University Press.
- Fielding, Nigel G and Raymond M Lee. Confronting Caqdas: Choice and Contingency. *Studies in qualitative methodology* 5, 1-23. 1995.
- Fisher, Kim. 1997. "Locating Frames in the Discursive Universe." *Sociological Research Online* 2 (3).
- Gallmann, Peter. 1991. Wort, Lexem und Lemma. In *Rechtschreibwörterbücher in der Diskussion: Geschichte – Analyse – Perspektiven*, Ed. Gerhard Augst and Burkhard

- Schaeder. Frankfurt a/M, FR Germany, Bern, Switzerland, New York, NY & Paris, France: Peter Lang.
- Gamson, William A. 1975. "Review of 'Frame Analysis' by Erving Goffman." *Contemporary Sociology* 4: 603-7.
- Gamson, William A. 1992. *Talking Politics*. Cambridge, U.K.: Cambridge University Press.
- Gibbs, Graham R., Susanne Friese, and Wilma C. Mangabeira . 2002. "The Use of New Technology in Qualitative Research." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 3 (2).
- Gitlin, Todd. 1980. *The Whole World Is Watching: Mass Media in the Making and Unmaking of the New Left*. Berkeley, CA, Los Angeles, CA & London, U.K.: University of California Press.
- Glaser, Barney. 2002. "Conceptualization: On Theory and Theorizing Using Grounded Theory." *International Journal of Qualitative Methods* 1 (2).
- Glaser, Barney G and Anselm L Strauss. 1967. *The discovery of grounded theory; strategies for qualitative research*. Chicago: Aldine Pub. Co.
- Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience*. New York, NY et al.: Harper & Row.
- Greenfeld, Liah. 1999. Is Nation Unavoidable? Is Nation Unavoidable Today? In *Nation and National Identity: The European Experience in Perspective*, Ed. Hanspeter Kriesi, Klaus Armigeon, Hannes Siegrist, and Andreas Wimmer. Chur, Switzerland & Zürich, Switzerland: Rüegger.
- Hagenaars, Jacques A. 1993. *Loglinear Models with Latent Variables*. Newbury Park, CA & London, U.K.: Sage.
- Jackson, Kristi. 2003. "Blending Technology and Methodology: A Shift toward Creative Instruction of Qualitative Methods with NVivo." *Qualitative Research Journal*: 96-110.
- Koella, Jennifer Campell. The Impact of Interest Groups and New Media Framing on Public Opinion: The Rise and Fall of the Clinton Health Care Plan. Annual Meeting of the Midwest Association for Public Opinion Research. 2003. October 21, 2003.
- Lazarsfeld, Paul F. 1950 . The Logical and Mathematical Foundations of Latent Structure Analysis. In *Measurement and Prediction*, Ed. Samuel A Stouffer. Princeton, N.J. : Princeton University Press.
- Lee, Raymond M and Nigel Fielding. 1996. "Qualitative Data Analysis: Representations of a Technology: A Comment on Coffey, Holbrook and Atkinson." *Sociological Research Online* 1 (4).
- Lonkila, Markku. 1995. In *Computer-aided qualitative data analysis : theory, methods and practice*, Ed. Udo Kelle. London : Sage.
- Magidson, Jay and Jereon K. Vermunt. 2004. Comparing Latent Class Factor Analysis with Traditional Factor Analysis for Datamining. In *Statistical Data Mining and Knowledge Discovery*, Ed. Hampersum Bozdogan. Boca Raton, FL & London, U.K.: Chapman & Hall/CRC.
- Maher, T. Michael. 2001. Framing: An Emerging Paradigm or a Phase of Agenda Setting. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*,

- Ed. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.
- McAdam, Doug. 1996. The Framing Function of Movement Tactics: Strategic Dramaturgy in the American Civil Rights Movement. In *Comparative Perspectives on Social Movements: Political Opportunities, Mobilizing Structures and Cultural Framings*, Ed. Doug McAdam, John D. McCarthy, and Mayer N. Zald. Cambridge, U.K.: Cambridge University Press.
- McCaffrey, Dawn and Jennifer Kayes. 2000. "Competitive Framing Processes in the Abortion Debate: Polarization-vilification." *The Sociological Quarterly* 41 (1): 41-61.
- McCutcheon, Alan. 1987. *Latent Class Analysis*. Newbury Park, CA & Beverly Hills, CA: Sage.
- Meyer, David S. 1999. "Tending the Vineyard: Cultivating Political Process Research." *Sociological Forum* 14: 79-92.
- Miller, M. Mark. 1997. "Frame Mapping and Analysis of News Coverage of Contentious Issues." *Social Science Computer Review* 15 (4): 367-78.
- Miller, M. Mark and Bonnie Parnell Riechert. Identifying Themes via Concept Mapping: A New Method of Content Analysis. Association for Education in Journalism and Mass Communication Annual Meeting. 1994. October 20, 2003.
- Miller, M. Mark and Bonnie Parnell Riechert. 2001a. Frame Mapping: A Quantitative Method for Investigating Issues in the Public Sphere. In *Applications of computer content analysis*, Ed. Mark D West. Westport, CT & London, England: Ablex Pub.
- Miller, M. Mark and Bonnie Parnell Riechert. 2001b. The Spiral of Opportunity and Frame Resonance: Mapping the Issue Cycle in News and Public Discourse. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*, Ed. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.
- Neuman, W. Russell, Marion R Just, and Ann N Crigler. 1992. *Common Knowledge: News and the Construction of Political Meaning*. Chicago : University of Chicago Press.
- Popper, Karl R. [1934]1966. *Logik der Forschung*. Tübingen, FR Germany: J.C.B. Mohr.
- Price, Vincent, David Tewksbury, and Elisabeth Powers. 1997. "Switching Trains of Thought: The Impact of News Frames on Readers' Cognitive Responses." *Communication Research* 24 (5): 481-506.
- Quasthoff, Uwe and Christiaan Wolf. Effizientes Dokumentenclustering durch niederfrequente Terme. 2003. November 30, 2003.
- Reese, Stephen D. 2001. Prologue - Framing Public Life: A Bridging Model for Media Research. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*, Ed. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.
- Reunanen, Esa and Risto Suikkanen. 1999. *Latent Class Analysis.: Wandering in Latent Space*. Konstanz, Germany: Universität Konstanz.
- Risse, Thomas and Marianne Van de Steeg. An Emerging European Public Sphere? Empirical Evidence and Theoretical Clarifications . Europeanisation of Public Spheres, Political Mobilisation, Public Communication and the European Union. 2003. Berlin, Germany, Wissenschaftszentrum Berlin.

- Scheufele, Dietram A. 1999. "Framing as a Theory of Media Effects." *Journal of Communication* 49 (4): 103-22.
- Schwenken, Helen. 'Sangatte' - A case study about the political self-organisation of refugees and migrants in the European Union. 2003.
- Semetko, Holli A. and Patti M. Valkenburg. 2000. "Framing European Politics: A Content Analysis of Press and Television News." *Journal of Communication* 50 (2): 93-109.
- Snow, David A. and Robert D. Benford. 1988. "Ideology, Frame Resonance and Participant Mobilization." *International Social Movement Research* 1: 197-219.
- Snow, David A., E. Burke Rochford, Steven K. Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization and Movement Participation." *American Sociological Review* 51: 464-81.
- Somers, Margaret R. 1995. "Narrating and Naturalizing Anglo-American Citizenship Theory: The place of Political Culture and the Public Sphere." *Sociological Theory* 13: 229-74.
- Statham, Paul and Ed Mynott. The Dilemma of Anti-Racist and Pro-Migrant Mobilisation in Britain: Visibility or Political Power? 2002.
- Strauss, Anselm L and Juliet M Corbin. 1990. *Basics of qualitative research : grounded theory procedures and techniques*. Newbury Park, Calif. : Sage Publications.
- Strübing, Jörg. 2002. "Just do it? Zum Konzept der Herstellung und Sicherung von Qualität in grounded theory-basierten Forschungsarbeiten." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (2): 318-42.
- Tankard, James W., Laura Handerson, Jackie Sillberman, Kriss Bliss, and Salma Ghanem. *Media Frames: Approaches to Conceptualization and Measurement*. Association for Education in Journalism and Mass Communication. 1991.
- Tankard, James W., Jr. 2001. The Empirical Approach to the Study of Media Framing. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*, Ed. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tedesco, John C. Issue and Strategy Agenda-Setting in the 2000 Presidential Primaries. *American behavioral scientist* 44[12], 2048-67. 2001.
- Triandafyllidou, Anna. 2002. "Religious Diversity and Multiculturalism in Southern Europe: The Italian Mosque Debate." *Sociological Research Online* 7 (1).
- Triandafyllidou, Anna and Anastasios Fotiou. 1998. "Sustainability and Modernity in the European Union: A Frame Theory Approach to Policy-Making." *Sociological Research Online* 3 (1).
- Trochim, William and Debra Hover. Mapping Student Views of the Benefits of a Course Website. 2003.
- Ullrich, Kerstin. *Soziale Bewegung und kollektive Identität: Der Diskurs über Abtreibung und Reproduktionstechnologien am Beispiel feministischer Identitätskonstruktion*. 1998. Florence, Italy, European University Institute.
- Weitzman, Eben A and Matthew B Miles. 1995. *Computer Programs for Qualitative Data Analysis: A Software Sourcebook*. Thousand Oaks, Calif. : Sage Publications.